CLINICAL PREDICTION RULES

FOR LOW BACK PAIN

Robin Haskins

B.Phty (Hons), Cred.MDT

Thesis submitted for the degree of

Doctor of Philosophy

The University of Newcastle, Australia

February 2015

STATEMENT OF ORGINALITY

The thesis contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. I give consent to the final version of my thesis being made available worldwide when deposited in the University's Digital Repository, subject to the provisions of the Copyright Act 1968.

……

Robin Haskins Date: 20/02/15

ACKNOWLEDGEMENT OF AUTHORSHIP

I hereby certify that this thesis is submitted in the form of a series of published papers of which I am a joint author. I have included as part of the thesis a written statement from each co-author; and endorsed by the Faculty Assistant Dean (Research Training), attesting to my contribution to the joint publications (Appendix 1).

……

Robin Haskins **Date: 20/02/15**

ACKNOWLEDGEMENTS

I would first like to acknowledge my supervisors Professor Darren Rivett and Dr Peter Osmotherly. Their support, expertise and generosity over these years has been sincerely appreciated. I am very proud of the work we have achieved together and acknowledge that none of it would have been possible without their flexibility, enthusiasm and commitment. Darren and Peter are highly regarded researchers, clinical experts and professional role models who will continue to inspire me.

I wish to also give my appreciation to Dr Erica Southgate and Dr Frank Tuyl for their respective expertise and mentoring in the fields of qualitative research methods and statistics. Erica and Frank were integral to the success of several aspects of this work and I sincerely thank them for their generosity, kindness and wisdom.

Thank you also to Dane Fehlberg for his assistance with screening and appraising the literature for the systematic reviews that form part of this thesis. His valued and tireless commitment to the long hours involved in study selection and quality appraisal is sincerely appreciated.

I also give thanks to my colleagues in Hunter New England Local Health District. Throughout my journey as a part-time research student, the flexibility and support I have received from Hunter England Health has enabled me to

remain working as a clinician. I genuinely appreciate the opportunities, assistance and encouragement they have provided to me.

Finally and above all, I wish to give my sincere appreciation to my beautiful wife Lauren. I thank her for her unwavering support and for maintaining my sanity throughout this journey. She is my best friend and I owe her everything.

PUBLICATIONS AND PRESENTATIONS

The following publications and presentations were a direct result of the work completed in this thesis:

Published papers

1. Haskins, R., Rivett, D. A., & Osmotherly, P. G. (2012). Clinical prediction rules in the physiotherapy management of low back pain: a systematic review. *Manual Therapy*, *17*(1), 9-21.
2. Haskins, R., Osmotherly, P. G., Southgate, E., & Rivett, D. A. (2014). Physiotherapists' knowledge, attitudes and practices regarding clinical prediction rules for low back pain. *Manual Therapy*, *19*(2), 142-151.
3. Haskins, R., Osmotherly, P. G., Southgate, E., & Rivett, D. A. (2015). Australian physiotherapists' priorities for the development of clinical prediction rules for low back pain: a qualitative study. *Physiotherapy 101*(1), 44-49.
4. Haskins, R., Osmotherly, P. G., Tuyl, F., & Rivett, D. A. (2014). Uncertainty in Clinical Prediction Rules: The Value of Credible Intervals. *Journal of Orthopaedic & Sports Physical Therapy*, *44*(2), 85-91.
5. Haskins, R., Osmotherly, P. G., & Rivett, D. A. (2015). Diagnostic clinical prediction rules for specific subtypes of low back pain: A systematic review. *Journal of Orthopaedic & Sports Physical Therapy, 45*(2), 61-76.

6. Haskins, R., Osmotherly, P. G., & Rivett, D. A. (2015). Validation and impact analysis of prognostic clinical prediction rules for low back pain is needed: a systematic review. *Journal of Clinical Epidemiology, 68*(7), 821-832.

Editorial

1. Rivett, D.A., & Haskins, R. (2009). Weighing up CPRs. *In Touch* (pp. 4). Victoria, Australia: Musculoskeletal Physiotherapy Australia, Australian Physiotherapy Association.

Published abstracts

1. Haskins, R., Rivett, D. A., & Osmotherly, P. G. (2011). Clinical prediction rules in the physiotherapy management of low back pain: a systematic review. In *APA Physiotherapy Conference 2011 Abstract Presentations,* Brisbane, Australia. October 2011.
2. Haskins, R., Osmotherly, P. G., Southgate, E., & Rivett, D. A. (2013). Physiotherapists' knowledge, attitudes and practices regarding clinical prediction rules for low back pain. In *APA Physiotherapy Conference 2013 'New Moves' Conference Abstracts*, Melbourne, Australia. October 2013.

Poster presentations

1. Haskins, R., Osmotherly, P. G., Southgate, E., & Rivett, D. A. (2013). Physiotherapists' knowledge, attitudes and practices regarding clinical

prediction rules for low back pain. e-Poster. *APA Physiotherapy Conference 2013 'New Moves',* Melbourne, Australia. October 2013.

2. Haskins, R., Osmotherly, P. G., & Rivett, D. A. (2015). Prognostic and prescriptive clinical prediction rules for low back pain: a systematic review. *World Confederation for Physical Therapy Congress 2015.* Singapore. May 2015.

Conference and other invited presentations

1. Haskins, R., Rivett, D. A., & Osmotherly, P. G. (2011). Clinical prediction rules in the physiotherapy management of low back pain: a systematic review. Oral presentation. *APA Physiotherapy Conference 2011.*Brisbane, Australia. October 2011.
2. Haskins, R., Rivett, D. A., & Osmotherly, P. G. (2011). Clinical prediction rules in the physiotherapy management of low back pain. Oral presentation. *Hunter New England Health Allied Health Research Forum*. Newcastle, Australia. October 2011.
3. Haskins, R. (2012). Clinical prediction rules for low back pain. Oral presentation. *Rural Spinal Seminar.* Tamworth, Australia. June 2012.
4. Haskins, R. (2012). Clinical prediction rules. Are we building better mousetraps? Oral presentation. *University of Newcastle Physiotherapy Seminar.* Newcastle, Australia. August 2012.
5. Haskins, R. (2013). Probabilistic reasoning and clinical prediction rules. Oral presentation. *Cessnock District Hospital.* Cessnock, Australia. June 2013.

6. Haskins, R (2014): Clinical prediction rules for low back pain. Readiness for clinical application. Oral presentation. *University of Newcastle Physiotherapy Seminar.* Newcastle, Australia. November 2014.
7. Haskins, R., Rivett, D. A., & Osmotherly, P. G. (2015). A systematic review of diagnostic clinical prediction rules for low back pain. Oral presentation. *World Confederation for Physical Therapy Congress 2015*. Singapore. May 2015.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

ABSTRACT

Low back pain (LBP) is prevalent, costly and a significant contributor to societal burden. It is not a single specific condition, but instead generally considered to be comprised of smaller more homogenous subgroups of patient presentations that meaningfully differ with regard to their symptomology, prognosis and responsiveness to different interventions. The identification of subgroups of patients with LBP is a research priority and several classification mechanisms have been proposed. Traditionally, such classification approaches have been predominantly based upon expert opinion and biologic plausibility, with little concordance among them. More recently, there has been a greater focus upon empirically derived subgrouping methods, notably including the development of clinical prediction rules (CPRs). A CPR is a clinical tool designed to be used with an individual patient, and is based on the statistical identification of a parsimonious composite of variables that facilitate the prediction of a given diagnosis or outcome.

The main objective of this thesis is to facilitate the development of CPRs with the greatest potential to positively influence the physiotherapy management of LBP. This was achieved through a series of five published studies and a published Clinical Commentary that together sought to address three primary research aims:

1. Identify and assess the degree to which CPRs for LBP may be confidently applied in clinical practice using a hierarchical framework for CPR development and an appraisal and synthesis of the existing evidence base.
2. Explore the range of factors that may influence the implementation of CPRs for LBP within Australian physiotherapy practice.
3. Examine the areas of perceived need for LBP CPRs and the range of characteristics such tools need to encompass to be considered clinically meaningful and useful within Australian physiotherapy practice.

Three systematic reviews were conducted which sought to synthesise the available body of evidence to; (1) identify CPRs relevant to the assessment and management of LBP; (2) assess the degree to which such tools may be confidently applied in clinical practice; and (3) identify opportunities to improve the methodological quality and reporting of LBP CPR development studies. The evidence considered within these reviews identified that a large number of diagnostic, prognostic and prescriptive LBP CPRs are under development, however the majority of these tools have not undergone validation and therefore cannot be recommended for direct use in clinical practice at this time. The current lack of impact analysis studies also prevents the assessment of whether the application of LBP CPRs in clinical practice results in beneficial effects on patient outcomes or resource efficiencies. A small number of LBP CPRs have undergone validation, such that clinicians may have some confidence in the predictive accuracy of these tools when applied in similar patient populations and settings. Further, several

opportunities to improve the methodological rigour of future CPR development studies have been identified.

Two qualitative studies using focus groups and involving a sample of Australian physiotherapists who manage patients with LBP were undertaken concurrently to address research aims 2 and 3. The findings of the first of these studies highlighted that physiotherapists' knowledge of LBP CPRs may be quite varied and few participants in that study reported having ever used them to inform their clinical decision-making. Barriers to the use of LBP CPRs included a negative connotation associated with the term 'rule', a perception that CPRs are overly-complex and infrequently applicable, clinical experience obviating the need for such tools, and the potential threat to clinical autonomy and for misuse by third-party payers. Physiotherapy participants felt that LBP CPRs were best used within the suite of clinical reasoning processes physiotherapists typically employ and considered as second opinions or safety nets that were able to be overruled by the clinician.

The findings of the second qualitative study indicated that prognostic forms of CPRs for LBP that function to predict future meaningful outcomes may be welcomed by practising physiotherapists. CPRs that identify likely responders to interventions are likely to be considered useful, as well as diagnostic forms of CPRs that function to identify serious causes of LBP such as fracture and cancer. CPRs that identify which patients are more likely to experience an adverse outcome or to not require physiotherapy intervention may also be welcomed by clinicians. Participants thought that LBP CPRs should be

uncomplicated, easy to remember, easy to apply, accurate and precise, and well-supported by research evidence. It was believed that LBP CPRs should not contain an excessive number of variables, use complicated statistics, or contain variables that have no clear logical relationship to the dependent outcome. It was further considered by participants that LBP CPRs need to be compatible with traditional clinical reasoning and decision-making processes, and sufficiently inclusive of a broad range of management approaches and common clinical assessment techniques.

A published Clinical Commentary was produced as a resource for clinicians and researchers based on findings arising from this research indicating the potential importance of the predictive precision of CPRs. The Clinical Commentary highlights the importance of considering uncertainty in clinical prediction, and provides a technical guide to the calculation and approximation of posterior probability uncertainty intervals. This and other study findings presented in this thesis have direct immediate implications for clinicians contemplating the application of LBP CPRs in clinical practice, and for researchers involved in the development of these tools. Opportunities for further research in this area have also been identified and are presented in the final chapter of this thesis. It is anticipated that consideration of the study findings in this program of research may support the development of CPRs with the greatest capacity to benefit the physiotherapy management of patients with LBP, and also strategies and future research projects designed to facilitate the successful translation of CPR research evidence into clinical practice.

CHAPTER 1

INTRODUCTION

This chapter provides an overview of the thesis including a summary of the relevant background knowledge. An outline of subsequent chapters is also provided.

1.1 Background and context

Low back pain (LBP) is common, costly and a cause of significant societal burden. Approximately one in 10 people are affected by LBP at any point in time and the majority of the population will experience at least one episode of LBP within their lifetime (Balague, Mannion, Pellise, & Cedraschi, 2012; Hoy et al., 2014; Walker, 1999; Walker, Muller, & Grant, 2004b). The high prevalence of LBP, coupled with its high negative impact on health (Salomon et al., 2013), results in LBP being the single largest cause of disability in the world (Hoy et al., 2014). The economic impact of LBP is also significant with costs estimated to be more than $9 billion in Australia (Walker, Muller, & Grant, 2003), with the majority of direct costs attributable to non-surgical management (Dagenais, Caro, & Haldeman, 2008).

Significant investment has consequently been devoted in recent decades in reducing the individual and societal burden imposed by LBP. Since the first clinical practice guideline for LBP was published in 1987 (Spitzer, 1987),

there have been a plethora of guidelines developed (Childs, Flynn, & Wainner, 2012; Koes et al., 2010) synthesising the exponential output of LBP research that has arisen in recent years. A near uniform recommendation arising from practice guidelines is the use of a diagnostic triage, whereby patients with LBP are grouped as either having either (1) serious spinal pathology (e.g. fracture, cancer), (2) radicular syndrome, or (3) non-specific LBP (Koes et al., 2010). The latter group accounts for the majority of patients with LBP encompassing up to 94% of those presenting to primary care (Maher, Williams, Lin, & Latimer, 2011). The term 'non-specific' reflects the inability to identify a single specific pathoanatomic cause of the complaint.

Despite significant investment into the optimised management of LBP, it is apparent that little, if any, positive impact has been made. Reported prevalence rates have not declined (Briggs & Buchbinder, 2009; Deyo, Mirza, & Martin, 2006; Freburger et al., 2009; Harkness, Macfarlane, Silman, & McBeth, 2005) and indeed, costs have actually escalated without notable improvements in health outcomes (Deyo, Mirza, Turner, & Martin, 2009; Martin et al., 2008; Waddell, 1996). When compared to a placebo, the mean effect size of most treatments for non-specific LBP is small (Machado, Kamper, Herbert, Maher, & McAuley, 2009) and there is often negligible difference in the treatment effect observed between different interventions (Foster, Hill, & Hay, 2011).

Such findings have led many to speculate that non-specific LBP is not a homogeneous clinical population, but instead comprised of smaller subgroups that meaningfully differ in regard to their symptomology, prognosis and response to various treatments (Foster, Hill, O'Sullivan, & Hancock, 2013). While some patients with non-specific LBP respond well to a given intervention, others may experience negligible benefit or perhaps may even worsen from that same intervention. Thus, the average effect size observed in clinical trials may in fact be diluted and may be much greater if only provided to those who are most likely to benefit from that approach (Foster et al., 2011).

The identification of empirically derived, valid and meaningful subgroups of patients with LBP has been a priority area of research (Borkan & Cherkin, 1996; Bouter, Pennick, Bombardier, & Editorial Board of the Back Review Group, 2003; Costa et al., 2013; Foster, Dziedzic, Windt, Fritz, & Hay, 2009; Henschke, Maher, Refshauge, Das, & McAuley, 2007). Clinicians who manage patients with LBP generally believe that subgroups of patients are identifiable (Kent & Keating, 2004), however there is little consensus amongst clinicians regarding each subgroup's defining characteristics (Kent & Keating, 2005). Numerous classification approaches have been proposed in the literature (Karayannis, Jull, & Hodges, 2012; Kent & Keating, 2005; C. McCarthy, Arnall, Strimpakos, Freemont, & Oldham, 2004; Riddle, 1998) but many are entirely reliant on expert opinion or biological plausibility, with little concordance between them.

A more recent approach to the sub-classification of LBP involves clinical prediction rules (CPRs). A CPR is defined as “a clinical tool that quantifies the individual contributions that various components of the history, physical examination and basic laboratory results make towards the diagnosis, prognosis, or likely response to treatment in an individual patient” (McGinn et al., 2008, p. 493). CPRs are tools designed to help inform clinical decision-making and comprise a small number of variables that have been statistically identified to be independently predictive of a particular diagnosis or outcome. There are three major types of CPRs; (1) diagnostic CPRs which function to facilitate diagnostic decision-making, (2) prognostic CPRs which assist in the prediction of future outcomes, and (3) prescriptive CPRs which help inform treatment decision-making by identifying those patients with a higher likelihood of treatment response to a given intervention (C. Cook, 2008). LBP has been identified as an ideal target for the development of CPRs due to its heterogeneity and the large number of treatment alternatives (Fritz, 2009).

A key factor in determining the readiness of a CPR to be applied in clinical practice is its stage of development, which broadly occurs across three phases – derivation, validation and impact analysis (Childs & Cleland, 2006; McGinn et al., 2000). A CPR is initially derived by the identification and selection of variables that are empirically demonstrated to be predictive of a particular diagnosis or outcome. Next, a CPR undergoes a process of validation whereby its accuracy is assessed in new patient cohorts and across different clinical settings. Validation of a CPR is important as predictors identified in the initial derivation phase may reflect chance

statistical associations, or be specific to the population or setting in which it was derived (McGinn et al., 2008). Once a CPR has been successfully validated, it may be considered for application in clinical practice with confidence in its predictive accuracy (McGinn et al., 2000). Impact analysis is the final phase of a CPR's development and involves the investigation of whether the clinical application of a rule changes clinicians' behaviours and results in improved patient outcomes or efficiencies in resource consumption (Reilly & Evans, 2006). Such evidence is required before a CPR can be confidently recommended to be used in clinical practice with knowledge that its application has beneficial clinical consequences (McGinn et al., 2000). Knowledge regarding the stage of a CPR's development therefore underpins the appropriateness of its clinical application and is also crucial to guiding future research efforts.

CPRs that have been appropriately developed may not necessarily be successfully adopted in clinical practice. Similar to other evidence-based innovations such as practice guidelines (Côté, Durand, Tousignant, & Poitras, 2009), many individual and system level barriers may impede the implementation of CPRs (Beutel, Trehan, Shalvoy, & Mello, 2012; Brehaut et al., 2006; Brehaut et al., 2005; Eagles, Stiell, Clement, Brehaut, Taljaard, et al., 2008; Graham et al., 2001; Graham, Stiell, Laupacis, O'Connor, & Wells, 1998; Stiell et al., 2006). Understanding and addressing such barriers, in addition to recognising the factors that facilitate their adoption is integral to the successful translation of CPR research evidence into practice. A parallel consideration is ensuring that the function and modifiable characteristics of

LBP CPRs match the needs and preferences of the target users – that is, clinicians who manage patients with LBP (Eagles, Stiell, Clement, Brehaut, Kelly, et al., 2008). Substantial time and resources are required to develop a CPR, and it is therefore crucial that they will be accepted by clinicians and viewed as helpful in addressing a clinically meaningful problem.

The main objective of the program of research detailed in this thesis was to facilitate the development of CPRs with the greatest potential to positively influence the physiotherapy management of LBP.

1.2 Structure of the thesis

This is a thesis by publication and details a program of research on the topic of CPRs for LBP. The five studies and one commentary that comprise the thesis have all been published in peer-reviewed scientific journals. The Word (Microsoft Corporation, Washington, USA) version of each manuscript is presented within this thesis as a separate chapter. With the following exceptions, each corresponding chapter is an exact reproduction of the material that has been published.

1. A concise overview is provided at the beginning of each chapter describing the study within the context of the broader research program.
2. Heading, Table, Figure, Equation and Appendix numbering have been amended and reformatted for consistency throughout the thesis.
3. Appendices for each study appear within the relevant chapter.

4. Within-text referencing has been modified throughout the thesis in the style of American Psychological Association (APA) 6th edition.
5. All references appear at the end of the thesis and have been reformatted in the style of APA 6th edition.

1.3 Overview of the thesis

Chapter 2

This chapter describes the relevant background literature concerning LBP that is pertinent to understanding the broader context of the studies reported in this thesis. The considerable problem of LBP is described and the evolution of approaches to its assessment and management are detailed. Chapter 2 forms the platform from which the conscientious use of evidence, probabilistic reasoning, and the development of statistically derived sub-classification approaches to LBP have arisen.

Chapter 3

Chapter 3 presents the relevant background literature on the topic of CPRs. An overview of the conscientious use of statistics in healthcare decision-making is provided. The long-standing and still relevant discussion points regarding the integration of statistical and clinical decision-making strategies are also discussed. The process of deriving, validating and assessing the impact of a CPR is discussed in the context of the readiness of a CPR to be applied in clinical practice. Important methodological considerations in a CPR's development are also detailed. The integration of CPRs within clinical

decision-making is discussed and potential barriers to the adoption of LBP CPRs are outlined with consideration of the body of evidence in Emergency Medicine and for other similar evidence-based innovations. The chapter concludes with a discussion addressing clinician priorities for CPR development including the types of problems they aim to address and the modifiable characteristics that may enhance their perceived utility.

Chapter 4 – Study 1

This chapter is the first of five studies that comprise the program of research detailed in this thesis. Chapter 4 is a systematic literature review of CPRs in the physiotherapy management of LBP and has been published in a peer-reviewed scientific journal (Haskins, Rivett, & Osmotherly, 2012). At the time this study commenced, only one review had been previously published on the topic of CPRs relevant to musculoskeletal physiotherapy practice, although this was limited to CPRs for physical therapy interventions in the derivation phase of development (Beneciuk, Bishop, & George, 2009). The present systematic review was undertaken to identify all forms of physiotherapy CPRs for LBP at any stage of their development and to assess their readiness for clinical application. Twenty-three publications were included in the review describing the development of 25 LBP CPRs including 15 diagnostic, seven prescriptive, and three prognostic tools. Most of the tools were found to be in their initial stage of development and no tools had been demonstrated to positively impact clinical practice. A number of opportunities to improve the methodological rigour of future CPR

development studies were identified. The major finding from this study was that although several tools had been derived, the current body of evidence did not enable direct clinical application of the LBP CPRs included in the review.

Chapter 5 – Study 2

The study reported in Chapter 5 has been published in a peer-reviewed scientific journal (Haskins, Osmotherly, Southgate, & Rivett, 2014) and is the second of five studies in this research program. This study was conducted to gain a greater understanding regarding the range of factors that may influence the implementation of LBP CPRs within physiotherapy clinical practice. It is postulated that such knowledge is integral to the translation of CPR research into clinical practice. The recognition of the barriers and facilitators to the use of CPRs for LBP by practising physiotherapists is anticipated to guide policy development and future research efforts that may optimize the integration of CPRs into the best practice management of LBP. Qualitative research methodology was used to explore Australian physiotherapists' knowledge, attitudes and practices regarding LBP CPRs. The findings from the systematic review presented in Chapter 4 were used to help develop a LBP case scenario that was in turn used to inform discussions across four focus groups. Focus group participants were practising physiotherapists who manage patients with LBP in public and private settings across metropolitan and regional areas. Several potential barriers and

facilitators were identified and the implications of these findings are discussed further in the chapter.

Chapter 6 – Study 3

Chapter 6 details an investigation into Australian physiotherapists' priorities for the development of CPRs for LBP and is the third study of this thesis. Identifying and addressing the needs and preferences of the target clinical consumers is postulated to be an important aspect in the successful translation of CPR research into clinical practice. This study was conducted simultaneously with the study reported in Chapter 5, and has been published in a peer-reviewed scientific journal (Haskins, Osmotherly, Southgate, & Rivett, 2015). Several areas of perceived need for CPRs in the assessment and management of LBP were identified by study participants. Modifiable characteristics of CPRs, including the precision of predictions, were found to be influential to the perceived utility of such tools. The chapter details these findings and their clinical and research implications.

Chapter 7 – Clinical Commentary

Chapter 7 discusses the precision of posterior probability estimates. This commentary has been published in a peer-reviewed scientific journal (Haskins, Osmotherly, Tuyl, & Rivett, 2014) and overviews appropriate methods to calculate or approximate uncertainty intervals for posterior probabilities. The rationale for this commentary evolved from the findings reported in Chapter 4 and Chapter 6. The systematic review (Chapter 4)

identified that the precision of posterior probability estimates in CPR development studies is rarely reported. However, qualitative research findings from the third study (Chapter 6) highlighted that such information may be an important consideration by physiotherapists in the clinical application of such tools. Consequently, this commentary aimed to address this gap by providing a resource for clinicians and researchers that would facilitate the incorporation of uncertainty intervals in the calculation of posterior probabilities.

Chapter 8 – Study 4

The study reported in Chapter 8 has been published in a peer-reviewed scientific journal (Haskins, Osmotherly, & Rivett, 2015a) and is the fourth of the five studies that comprise this thesis. This study is a comprehensive systematic review of diagnostic CPRs for LBP and is intended to serve as a resource for clinicians and researchers. The diagnostic LBP CPRs that are currently under development and their appropriateness for clinical application are detailed. The review identified 14 publications not previously reported in earlier reviews on this topic. As detailed in this chapter, three LBP CPRs were found to have undergone validation, but none had been assessed for their ability to produce beneficial clinical consequences. The clinical and research implications arising from this study are discussed in detail within this chapter.

Chapter 9 – Study 5

Chapter 9 is the final study in this program of research and is a systematic review of prognostic forms of LBP CPRs. This study has been published in a scientific journal (Haskins, Osmotherly, et al., 2015b) and was conducted in parallel to the study reported in Chapter 8 using similar methodology. Thirty prognostic LBP CPRs were identified with three tools known to have undergone validation. No impact analysis studies were identified. The readiness of each tool to be used in clinical practice to inform decision-making is discussed and the identified opportunities to reduce the risk of bias in future prognostic CPR development studies are detailed.

The systematic reviews reported in Chapter 8 and Chapter 9 differ from and extend upon the systematic review reported in Chapter 4 in the following ways;

1. In contrast to Chapter 4, the systematic reviews in Chapters 8 and 9 are entirely focused upon either diagnostic (Chapter 8) or prognostic (Chapter 9) forms of LBP CPRs. The findings of these studies are discussed with greater specificity and depth in regard to the type of tool under investigation.
2. Using the methodology reported in Chapter 7, uncertainty intervals have been calculated or approximated for posterior probability estimates in instances where reported data permits.
3. CPRs for LBP were included irrespective of the health discipline(s) involved in their development. In contrast, Chapter 4 was limited to tools developed by physiotherapists. It is anticipated that CPRs

developed external to physiotherapy may be of relevance to the profession and that opportunities to assess the generalizability of such tools to physiotherapy practice may be identifiable.

4. The updated reviews include the large volume of more recent research in their respective fields and in contrast to the earlier review, were not restricted to CPRs developed after a given date.
5. A more sensitive search strategy was employed based upon the incorporation of a newly developed sensitive search string designed to identify prediction model studies. Further, a greater number of electronic databases were searched allowing the identification of a greater number of potentially eligible studies.
6. The quality appraisal of included studies is more comprehensive and incorporates recent methodological considerations pertinent to the development of CPRs. In contrast to the earlier review reported in Chapter 4, the two updated reviews additionally appraise the methodological quality of included studies based upon their underlying study design.

Chapter 10

The final chapter of the thesis summarises the key findings arising from this research within the context of the existing body of evidence. The key conclusions and limitations are discussed. The research and clinical implications are detailed and recommendations for future research are outlined.

1.4 Delimitations

The scope of this program of research concerns CPRs for LBP. CPRs are just one of several approaches that facilitate the identification of meaningful subgroups of patients with LBP who differ in regard to their symptomology, prognosis and response to given interventions. Alternative approaches to the sub-classification of LBP are beyond the scope of this research.

CPRs have been developed for a range of clinical problems that are relevant to the practice of physiotherapy. This program of research, however, is focused upon those tools that specifically aim to assist in the non-pharmacological conservative assessment and management of patients presenting with LBP. This clinical condition has been selected due to its large societal burden and the hypothesised capacity of such tools to improve patient outcomes. CPRs designed to assist in decision-making for conditions other than LBP are outside the scope of this thesis.

1.5 Significance and research aims

The purpose of this program of research is to facilitate the development of CPRs with the greatest potential to positively influence the physiotherapy management of LBP. This was achieved through a series of five studies and a Clinical Commentary that together sought to address the three primary research aims (Table 1.1):

Table 1.1 Primary research aims

Research aim 1	Identify and assess the degree to which CPRs for LBP may be confidently applied in clinical practice using a hierarchical framework for CPR development and an appraisal and synthesis of the existing evidence base.	Study 1 Study 4 Study 5 Commentary
Research aim 2	Explore the range of factors that may influence the implementation of CPRs for LBP within Australian physiotherapy practice.	Study 2
Research aim 3	Examine the areas of perceived need for LBP CPRs and the range of characteristics such tools need to encompass to be considered clinically meaningful and useful within Australian physiotherapy practice.	Study 3 Commentary

Research aim 1 is addressed via the systematic and critical evaluation and synthesis of the current body of evidence on LBP CPRs (Studies 1, 4 and 5). The clinical significance of this research is in providing clinical practice recommendations regarding the evidence-based implementation of CPRs in the management of patients with LBP. Concurrently, the Clinical Commentary provides a resource for clinicians to assist with the interpretation of CPR research findings concerning the precision of posterior probability estimates to inform clinical decisions. The research significance of Studies 1, 4 and 5 relate to identifying opportunities to progress the development of LBP CPRs and opportunities to improve the methodological

rigour by which they are derived, validated and evaluated in regards to their clinical impact. The Clinical Commentary serves as a resource for researchers to inform the calculation and reporting of uncertainty intervals for posterior probabilities reported in CPR development studies.

Research aim 2 is addressed in Study 2, in which qualitative research methodology is employed to explore the range of factors that may facilitate or impede the clinical implementation of LBP CPRs within an Australian physiotherapy setting. The significance of this research relates to identifying opportunities to reduce the potential barriers to CPR implementation throughout a tool's development, and in informing strategies that may facilitate the translation of CPR research findings into practice.

Research aim 3 is addressed in Study 3, in which the areas of perceived need for LBP CPRs amongst a sample of physiotherapists who manage patients with LBP are explored using qualitative methodology. Within this study, the range of characteristics which LBP CPRs need to encompass in order to be considered clinically meaningful by physiotherapists are also investigated. The significance of this research relates to identifying opportunities to develop CPRs that specifically align with the identified needs and preferences of the target clinical consumers. It is anticipated that explicitly addressing clinician needs and preferences throughout the development of a CPR, may beneficially impact the effective translation of CPR research evidence into clinical practice. The Clinical Commentary (Chapter 7) was developed as a resource for clinicians and researchers

based upon the finding from Study 3 which highlighted the potential clinical importance of the precision of posterior probability estimates arising from the application of some forms of CPRs.

CHAPTER 2

LITERATURE REVIEW: LOW BACK PAIN

2.1 Introduction

LBP is not a specific disease, but rather a condition that is characterised by an unpleasant sensory experience in the posterior bodily region between the lower margin of the twelfth ribs and the inferior gluteal folds (Cieza et al., 2004; Dionne et al., 2008; Hoy, March, et al., 2010; Krismer & van Tulder, 2007; Loeser & Treede, 2008). The condition is prevalent, universal and a major source of global socioeconomic burden (Hoy et al., 2012; Hoy, Brooks, Blyth, & Buchbinder, 2010).

This chapter will examine the multi-faceted problem of low back pain in modern society, focussing on its prevalence, burden and economic consequences. The heterogeneity of LBP will be discussed and contemporary approaches to its primary classification will be outlined. Finally, the sub-classification of non-specific LBP into meaningful clinical subgroups will be examined with regard to its identified need, differing methodologies and approaches, and the current body of evidence.

2.2 A brief historical perspective

Our understanding of LBP has changed substantially over time. Such change however has not always occurred in a linear or progressive fashion (Allan &

Waddell, 1989). The earliest known written account of LBP appears in the Edwin Smith Papyrus, circa 17th century BC (van Middendorp, Sanchez, & Burridge, 2010). This ancient Egyptian medical text of the Second Intermediate Period (Sullivan, 1995) details the assessment and management of a range of predominantly orthopaedic presentations including joint dislocations, fractures and sprains (Breasted, 1930). The final case presented within this text concerns the management of a patient suffering with “a sprain in a vertebra of his spinal column”. Although the text of this final case is disappointingly incomplete, it and other cases presented in the papyrus, illustrate a rational and non-mystical understanding and management approach to conditions affecting the spine (van Middendorp et al., 2010).

The conceptual understanding of LBP from classical antiquity to the early modern era was largely influenced by the works of Hippocrates (460BC – 370BC) and Galen (129 – ca.199) (Coxe, 1846; Gruber & Boeni, 2008; Waddell & Allan, 2004). The prevailing view throughout this period was that medical ailments were attributable to an imbalance in the four fluids thought to regulate the human body known as ‘humors’ – yellow bile, black bile, blood and phlegm (Coxe, 1846). An English translation of the work of Hippocrates suggests a belief that LBP/sciatica was commonly caused by long exposure to the sun causing the hip joints to become heated and thereby drying out the humors (Coxe, 1846). Understandably, many treatment approaches for LBP advocated during this time may seem irrational to the modern observer. One such treatment for LBP, attributed to Galen, was to powder the wings of a

swallow and to bleed it to death via a cut to its leg. The deceased bird was to then be cooked and eaten by the patient. Oil from the swallow's carcass was also to be rubbed into the patient for three days (Deming, 2010). Given the favourable natural history of some acute presentations of LBP, it would seem probable that confirmation bias helped to reinforce some of these practices (Nickerson, 1998).

Seemingly little progress with regard the understanding of LBP had been made even as late as the eighteenth century, during which time LBP was generally considered to be a form of rheumatism caused by exposure to cold and damp (Allan & Waddell, 1989; Sydenham, 1848). A build-up of 'rheumatic phlegm' flowing from the brain to the lumbar region was considered the cause of pain (Dembe, 1996; Waddell & Allan, 2004) and treatment commonly involved bloodletting and in at least one documented case, whipping the buttocks with nettles (Boonen & van der Linden, 2002).

The hypothesis that LBP may be caused by local irritation of spinal structures only gained popular acceptance less than 200 years ago. In 1821, in a letter to the Editor of the Quarterly Journal of Science, the English surgeon Mr Richard Player explicitly hypothesised that the spinal structures themselves may be the cause of back pain and associated referred symptoms (Player, 1821). This theory was expanded by Scottish physician, Dr Thomas Brown, and the term 'spinal irritation' was coined seven years later (Brown, 1828).

During the 19th century, the invention of the steam locomotive caused rail transportation to become increasingly utilised throughout most of the industrialised world (Wolmar, 2012). Back pain attributed to railway transportation and accidents became known as 'railway spine' and was believed in some instances to be a result of a concussion of the spine (Dembe, 1996). Importantly, it was postulated for the first time that disabling LBP could be caused by trivial and/or accumulative trauma, that symptoms could take some time to develop, and that patients may present without any overt signs of physical injury (Erichsen, 1867). The diagnosis of railway spine eventually feel into disrepute over the subsequent decades and became primarily considered a psychologically-based condition, in part due to a lack of evidence of a local identifiable pathoanatomic lesion (Dembe, 1996; T. Keller & Chappell, 1996; Waddell & Allan, 2004). Nevertheless, it is considered that the legacy of railway spine was the precursor to modern perceptions that back pain results from injury to the local anatomical structures, and that disability may persist from back pain in the absence of an identifiable lesion (Waddell & Allan, 2004).

A prominent orthopaedic surgeon and researcher summarised much of 20th century research effort regarding our understanding of LBP as the 'dynasty of the disc' (Waddell & Allan, 2004). Since the 1934 discovery that a prolapsed intervertebral disc may cause sciatica (Mixter & Barr, 1934), other theories proposing the disc as a cause of LBP emerged (Mixter & Ayre, 1935) and ushered in a period of intense research and interventions focused upon this structure (Parisien & Ball, 1998). Consequently, other potential causes of

LBP received significantly less research interest during this period. Lutz, Butzlaff, and Schultz-Venrath (2003) examined 5,185 issues of an established German primary care journal (Deutsche Medizinische Wochenschrift) to identify studies published on the topic of LBP from 1900 to 1999. Each of the 464 identified studies were examined to determine the aetiologies purported to be causing LBP and these were broadly categorised as relating to the bone, muscle, nerve, or disc. Figure 2.1 below illustrates the key findings of this study and highlights the dominance of disc-related research throughout much of this century.

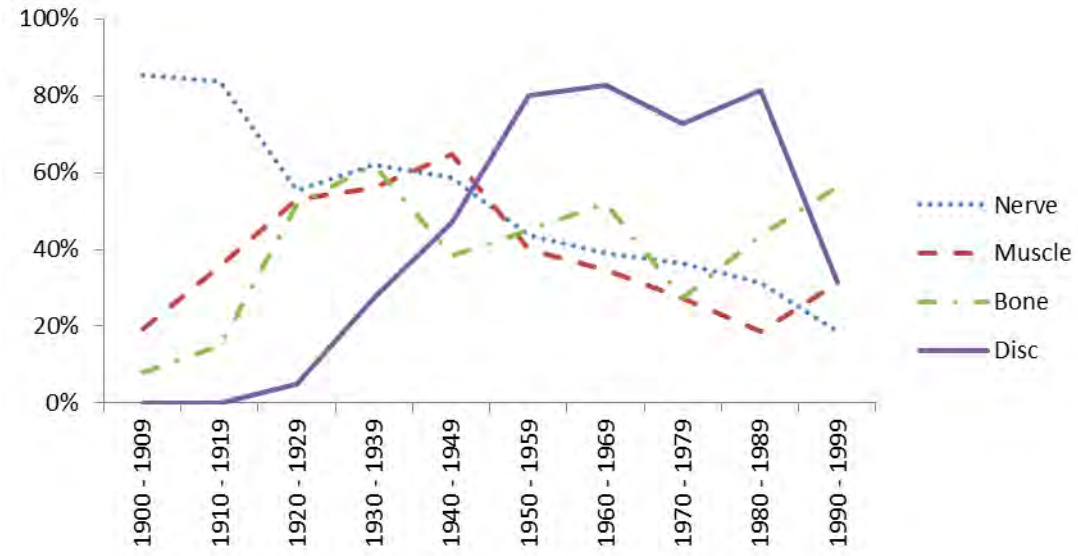


Figure 2.1 **Purported pathoanatomic causes of LBP described in 464 studies published in Deutsche Medizinische Wochenschrift from 1900 – 1999 (data from Lutz et al. (2003))**

Despite significant gains in our understanding of LBP during the 20th century and the development of new treatments and technologies, it has been proposed that many such 'advances' may have inadvertently exacerbated the societal burden of LBP (Waddell, 1996). Costs in the management of LBP have escalated in the absence of evidence of improved health outcomes (Deyo et al., 2009; Martin et al., 2008). Further, there is no evidence to suggest that the prevalence of LBP has declined (Briggs & Buchbinder, 2009; Deyo et al., 2006; Freburger et al., 2009; Harkness et al., 2005), however occupational disability attributable to the condition exponentially grew in the latter half of the 20th century (Waddell, 1996).

So far in the 21st century, LBP is generally considered within a broader biopsychosocial context, in which the social and psychological factors that influence pain perceptions are incorporated into the assessment and management of an individual patient (Gatchel, Peng, Peters, Fuchs, & Turk, 2007; C. McCarthy et al., 2004). There is greater confidence in the acceptance of what is not known, and rather than a restricted focus on understanding the aetiology of the condition, contemporary research examines the best ways that the problem of LBP may be holistically managed (Lutz et al., 2003).

2.3 The problem of low back pain

LBP is an escalating global problem. The condition is highly prevalent, costly and a significant contributor to societal burden. In response to growing

disability trends associated with LBP in recent decades, the World Health Organisation (WHO) introduced the 'Low Back Pain Initiative' (Ehrlich & Khaltaev, 1999). This multinational and multidisciplinary program has helped profile and raise awareness of the global epidemic of LBP and its substantial economic and social consequences.

Within the Australian context, the significant problem of LBP is reflected in the condition being identified as a National Health Priority Area (NHPA) within the 'Arthritis and musculoskeletal disorders' category (Australian Institute of Health and Welfare, 2013c). These are the diseases and illnesses that have been identified to significantly contribute to burden within the community and require the focused attention and resources of the various levels of government (Australian Institute of Health and Welfare, 1997).

However until relatively recently, LBP has not been an explicit priority for government or health policy investment. In earlier studies of global disease burden, LBP has ranked relatively low (Lopez & Murray, 1998; Mathers, Fat, & Boerma, 2008) compared to other conditions. However, there are notable limitations in these earlier estimates, including a paucity of suitable data and issues concerning the classification of LBP (Hoy, March, et al., 2010). More recent work has demonstrated that previous reports have substantially under-represented the burden attributable to LBP (Hoy et al., 2014), and this will be explored in greater detail in later subsections of this chapter.

The following subsections provide an overview of the epidemiology, disease burden, and economic consequences of LBP.

2.3.1 Epidemiology

As with many other conditions, the reported incidence and prevalence of LBP is notably sensitive to how the condition is operationally defined. Variables including the anatomic location of pain, minimum episode duration, prevalence period (e.g. point, 1 year, lifetime), and extent of activity limitation are known to influence epidemiological estimates (Hoy et al., 2012).

In response to the variability in the definition of an episode of LBP, Dionne et al. (2008) used a modified Delphi research design to help establish consensus-based standardised definitions that could be used in future prevalence studies. A minimum definition of low back pain that incorporated the anatomic area, observed symptoms, and time-frame of the measures was derived and formulated into two questions for research purposes (see Table 2.1).

Table 2.1 Minimal definition of an episode of LBP for prevalence research as proposed by Dionne et al. (2008)

Question 1	In the past 4 weeks, have you had pain in your low back?
Question 2	If yes, was this pain bad enough to limit your usual activities or change your daily routine for more than one day?

One of the most recent high quality investigations on the global prevalence of LBP was conducted by Hoy et al. (2012). This systematic review included 165 population-based studies published between 1980 to 2009 and was conducted to explicitly inform the analysis of the 2010 Global Burden of Disease study (Hoy et al., 2014; Murray, Ezzati, et al., 2012). The number of countries represented within the review was 54. For the purposes of this review, LBP was defined as “activity limiting low back pain (+/- pain referred into 1 or both lower limbs) that lasts for at least 1 day” (Hoy et al., 2012, p. 2028). This definition was selected for consistency with the consensus-based recommendation of Dionne et al. (2008) (Table 2.1 above) and its prior use in the 2005 Global Burden of Disease study (Hoy, March, et al., 2010).

Multivariable analysis in the review of Hoy et al. (2012) identified that reported LBP prevalence estimates were significantly related to the gender and age of study participants, prevalence period, anatomic and episode duration definitions of LBP, study coverage (community, regional or national) and urbanicity (rural or urban). In the adjusted analysis, the mean reported point-prevalence of LBP, as operationally defined above, was estimated to be 11.9% (SD 2.0%), and the mean reported 1-month prevalence was estimated to be 23.2% (SD 2.9%). Females and persons aged between 40 and 80 years were identified as having higher prevalence rates of LBP. The distribution of point, 1 month, 1 year, and lifetime prevalence estimates reported in the studies included in this review are depicted graphically in Figure 2.2.

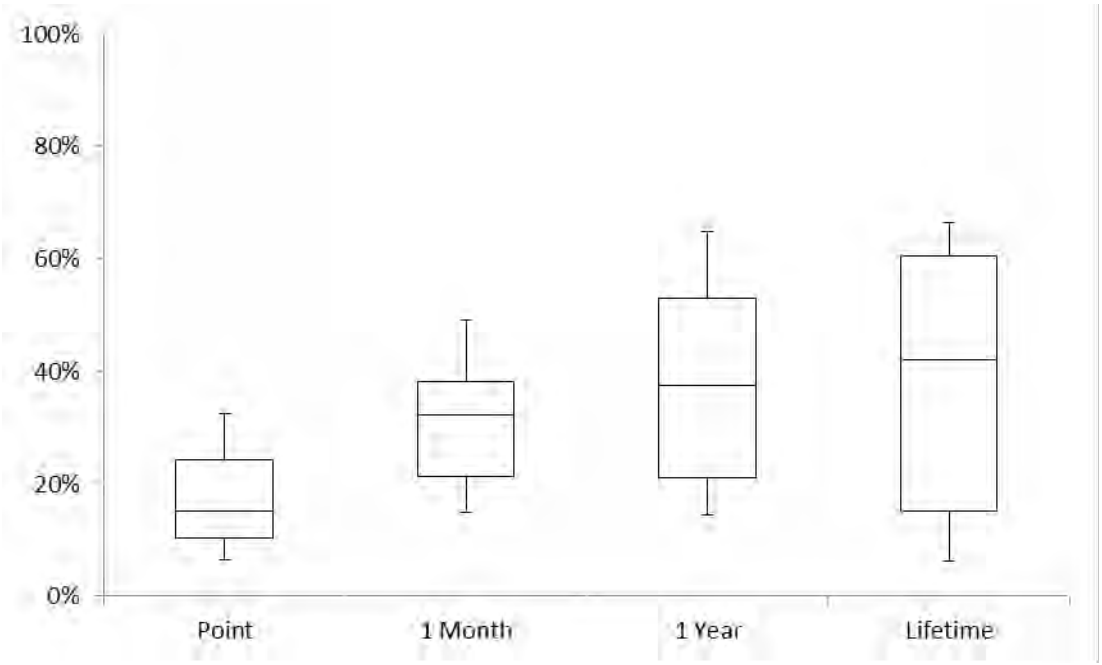


Figure 2.2 **Box plots detailing the median, interquartile range and 10% and 90% percentiles of the point, 1 month, 1 year and lifetime prevalence estimates of LBP from Hoy et al. (2012)**

Within Australia, data from the 2011-2012 Australian Bureau of Statistics National Health Survey provides an estimated population point prevalence of self-reported long-term back pain of 12.7% (Australian Bureau of Statistics, 2012). When combined with self-reported long-term sciatica and curvatures of the spine, the point prevalence increases to 13.3% (Australian Institute of Health and Welfare, 2013g). These estimates approximate the central tendency of the point prevalence figures reported in other countries and studies (Andersson, 1999; Hoy et al., 2012). However, it is notable that the question item within the Australian National Health Survey that informs the

reported prevalence figure specifically enquires about health conditions that have lasted, or were expected to last, for 6 months or more (Australian Bureau of Statistics, 2013). It is therefore probable that the reported prevalence figure may be much higher if expanded to include all other episodes of back pain that interfered with usual activities for more than one day, as per the recommended definition given in Table 2.1 above. Conversely, given that the Australian National Health Survey does not discriminate between problems in the region of the lower back and back problems located elsewhere, the reported prevalence figures specific to LBP may be substantially lower.

Another methodological issue that impacts upon the epidemiological measurement of LBP concerns the recurrent nature of the condition, and how recurrence and recovery from LBP are operationally defined (Wasiak, Pransky, & Webster, 2003). The clinical course of LBP varies considerably between individuals (Cassidy, Côté, Carroll, & Kristman, 2005; Hestbaek, Leboeuf-Yde, & Manniche, 2003), and episodic flare-ups of a persistent LBP presentation require delineation from instances where an episode of LBP has finished and a new episode has commenced. In a systematic review of the literature, Stanton, Latimer, Maher, and Hancock (2009) found very little concordance in the definitions of recovery and recurrence across studies that had attempted to measure the effectiveness of various treatments in reducing recurrent LBP episodes. Further, explicit definitions were often lacking and when provided were generally developed independently by each research team without consideration of the definitions used in previous research. An

output from this review was the proposal of evidence-informed recommendations concerning the definitions of recovery from an episode of LBP and recurrence of LBP, adapted from a definition originally proposed by de Vet et al. (2002). These definitions are summarised in Table 2.2 below.

Table 2.2 Recommended definition of recovery from low back pain (LBP) and recurrence of LBP, as proposed by Stanton, Latimer, et al. (2009)

Recovery from LBP	Pain-free for at least 1 month
Recurrence of LBP	Preceded by a period of recovery from LBP as defined above, and a minimum duration of LBP of at least 24 hours, and intensity of pain or degree of functional limitation greater than or equal to the minimal important change for the chosen scale.

The one year recurrence rate of LBP is commonly reported to be high. A 2003 systematic review identified two studies reporting on the cumulative risk of LBP recurrence at one year and provided a pooled risk estimate of 73% (95%CI 59% - 88%) (Pengel, Herbert, Maher, & Refshauge, 2003). Other studies have provided similarly high estimates (Hestbaek et al., 2003; Marras, Ferguson, Burr, Schabo, & Maronitis, 2007). It has been proposed however, that many reported LBP recurrence rates are unreliable due to study designs that have included patients who have not recovered, but

whose episode has instead persisted (Stanton et al., 2008). To address this limitation, Stanton et al. (2008) undertook a prospective cohort study that was limited to those patients whose episode of acute LBP had resolved and were followed up at 12 months. The proportion of patients who reported to have experienced a recurrent episode of LBP in this study at 12 months was 24% (95%CI 20% - 28%). When data analysis was supplemented using self-reported pain measurements at 3 and 12 months follow-up, the proportion defined as having experienced a recurrence increased to 33% (95%CI 28% - 38%). The findings of this study thus highlight the limitations in using twelve month recall as a stand-alone measurement tool, and also challenge the very high recurrence rates reported in earlier studies that do not restrict study participants to those who have recovered from their initial episode of LBP.

2.3.2 Burden of disease

The metric by which the burden attributable to a health condition is measured is the Disability-Adjusted Life Year (DALY) (Murray, 1994). Conceptually, one DALY is the equivalent of one lost year of healthy life, and represents the gap between current and ideal health statuses (World Health Organization, 2014b). The DALY incorporates the morbidity and premature mortality that may result from a health condition, and can be expressed as the sum of Years of Life Lost (YLL) and Years Lived with Disability (YLD) (Prüss-Ustün, Mathers, Corvalán, & Woodward, 2003) (Equation 2.1).

Equation 2.1 Disability-Adjusted Life Year (Prüss-Ustün et al., 2003)

Disability-Adjusted Life Year (DALY) =	Years of Life Lost (YLL) + Years Lived with Disability (YLD)

The YLL for a health condition is equivalent to the product of the number of premature deaths it causes, and the standard life expectancy at the age at which death occurs (World Health Organization, 2014b) (Equation 2.2).

Equation 2.2 Years of Life Lost (World Health Organization, 2014b)

Years of Life Lost (YLL) =	Number of deaths X Standard life expectancy at age of death in years

For health conditions such as LBP that do not directly cause premature death, the YLL is zero and therefore the DALY is equivalent to the YLD (Hoy et al., 2014). It could, however, be argued that some treatments for LBP (e.g. surgery) may unintentionally result in premature death, although such 'adverse effects of medical treatments' are considered as a separate category within studies of disease burden (Naghavi et al., 2015). The YLD has been calculated in studies of global disease burden using different methods and is often expressed as the product of a health condition's incidence (I), average duration of disability (L), and it's 'disability weight' (DW) (Prüss-Ustün et al., 2003) (Equation 2.3).

Equation 2.3 Years Lived with Disability (Prüss-Ustün et al., 2003)

$$\text{Years Lived with Disability (YLD)} = \text{Incidence (I)} \times \text{Duration of disability (L)} \times \text{Disability weight (DW)}$$

The DW is a metric designed to reflect the severity of a health condition on a scale from 0 (perfect health) to 1 (equivalent to death) (World Health Organization, 2014a). A disability weight of 0.05 implies that 1 YLL is equivalent to 20 years lived with a particular health condition. The calculation of DWs is complex and the most recent iterations for 220 specified health conditions involved surveying more than 30,000 people across five countries and contrasting randomly selected pairs of health statuses to identify which condition was regarded as healthier (Salomon et al., 2013). The DWs for acute and chronic LBP, with and without leg pain, are detailed in Table 2.3.

Table 2.3 Disability Weights for Low Back Pain (LBP) (with 95% confidence intervals) from Salomon et al. (2013)

	Without leg pain	With leg pain
Acute LBP	0.269 (0.184 – 0.373)	0.322 (0.219 – 0.447)
Chronic LBP	0.366 (0.248 – 0.499)	0.374 (0.252 – 0.506)

These data would suggest that 1 YLL due to premature mortality is equivalent to 3.7 years (95%CI 2.7 – 5.4) of living with acute LBP without leg pain, and equivalent to 2.7 years (95%CI 2.0 – 4.0) of living with chronic LBP with concomitant leg pain.

The 2010 Global Burden of Disease study utilised these DWs and the best available evidence concerning the prevalence of LBP to provide estimates of the global health burden attributable to LBP, and its ranking amongst other health conditions (Hoy et al., 2014; MurrayVos, et al., 2012). LBP has subsequently been identified as the single largest contributor to disability globally, and in 12 of the 21 world regions, including Australia (Hoy et al., 2014). LBP is estimated to account for 10.7% of global YLD (Vos et al., 2012). Table 2.4 lists the largest age-standardised contributors to disability globally and within Australia (Institute for Health Metrics and Evaluation, 2013).

When considered within the context of overall disease burden, including health conditions that cause premature mortality, LBP is the 6th overall cause of disease burden globally, and the number one cause of disease burden within Australia (Hoy et al., 2014). Table 2.5 lists the top 10 age-standardised contributors to disease burden globally and within Australia (Institute for Health Metrics and Evaluation, 2013).

Table 2.4 Top 10 health conditions contributing to Years Lived with Disability (YLD) globally and within Australia (Institute for Health Metrics and Evaluation, 2013)

	Globally	Australia
1	Low back pain	Low back pain
2	Major depressive disorder	Major depressive disorder
3	Iron-deficiency anaemia	Neck pain
4	Neck pain	Other musculoskeletal
5	Chronic obstructive pulmonary disease	Anxiety disorders
6	Other musculoskeletal	Asthma
7	Anxiety disorders	Migraine
8	Diabetes	Drug use disorders
9	Migraine	Falls
10	Falls	Chronic obstructive pulmonary disease

In Australia, data from the 2011-2012 National Health Survey indicates that the self-perceived overall health of people with back problems is also poorer than those without back problems (Australian Institute of Health and Welfare, 2013b). Further, this population is more likely to have relatively higher levels of psychological distress (Australian Institute of Health and Welfare, 2013b). A large retrospective analysis of insurance claims in the US involving 101,294 patients with chronic LBP and age/sex/region matched controls confirmed a similar association (Gore, Sadosky, Stacey, Tai, & Leslie, 2012).

In this study, patients with chronic LBP were identified to be more likely to suffer from other comorbidities, including depression (OR = 2.3), anxiety (OR = 2.5), sleep disorders (OR = 3.2) and a range of other musculoskeletal disorders (including osteoarthritis) (OR = 4.5). Similar observations have been made in other regions of the world (Bener et al., 2013).

Table 2.5 Top 10 health conditions contributing to Disability-Adjusted Life Years globally and within Australia (Institute for Health Metrics and Evaluation, 2013)

	Globally	Australia
1	Ischaemic heart disease	Low back pain
2	Lower respiratory infections	Ischaemic heart disease
3	Stroke	Road injury
4	Diarrheal diseases	Major depressive disorder
5	HIV / AIDS	Drug use disorders
6	Low back pain	Neck pain
7	Chronic obstructive pulmonary disease	Other musculoskeletal
8	Malaria	Asthma
9	Road injury	Chronic obstructive pulmonary disease
10	Preterm birth complications	Anxiety disorders

Such associations are important considerations for the holistic management of patients with LBP, however to date there is limited evidence to suggest that such relationships are causal. Nonetheless, a recent prospective study provides limited evidence of temporality with the observation of worsening scores on scales for depression, anxiety and stress following the onset of chronic LBP (Mathew, Singh, Garis, & Diwan, 2013). A significant limitation of this study, however, was that the pre-morbid scores were measured after the onset of LBP via patient recall and may plausibly not provide an accurate baseline measurement.

2.3.3 Economic considerations

The direct and indirect economic costs associated with LBP are large and have significant individual and society-level consequences. It has been estimated that the total costs attributable to LBP in Australia are more than AU$9.1 billion annually (Walker et al., 2003). Notably, this estimate was based upon data for the year 2001, and is thus likely to be significantly lower than the true current costs. All else being equal, adjusting this figure by the average annual inflation rate in Australia from 2001 to 2013 (2.8% pa) would give an updated estimate of AU$12.6 billion annually (Reserve Bank of Australia, 2014).

Direct costs associated with the management of LBP include hospital admissions, non-hospital services (e.g. general practice, medical specialists, physiotherapy, chiropractic), imaging, prescription medications, over-the-counter drugs and ancillary services. Approximately AU$1.2 billion dollars

was spent in Australia during the 2008/09 financial year on direct costs relating to management of back problems (Australian Institute of Health and Welfare, 2014). Taking into consideration the population of Australia at that time (Australian Bureau of Statistics, 2009), this equates to approximately AU$54 per capita. The majority of these costs related to hospital inpatient admissions (47%, AU$560m) and out-of-hospital services (40%, AU$465m). Prescribed pharmaceuticals consumed a relatively smaller but nevertheless non-trivial proportion of the overall direct costs (13%, AU$153m) (Australian Institute of Health and Welfare, 2014). Studies examining the direct costs associated with LBP in other countries (Boonen et al., 2005; Luo, Pietrobon, X Sun, Liu, & Hey, 2004) have indicated that the cost per capita may in fact be much higher in those other nations (Dagenais et al., 2008).

In the 2011-12 financial year, there were 28,700 hospital admissions in Australia for LBP (ICD-10-AM code M54.5) (Australian Institute of Health and Welfare, 2013e). This represents just 0.31% of all Australian hospital separations for the same time period (Australian Institute of Health and Welfare, 2013a). The number of hospital separations for the management of LBP has grown in recent years, however as illustrated in Figure 2.3 below, the average length of stay for this presentation has gradually declined from 3.5 days in 1998-99 to 2.4 days in 2011-12 (Australian Institute of Health and Welfare, 2013d). Accordingly, the change in direct costs of care resulting from hospital admissions for LBP may not necessarily follow the trend of increasing separations.

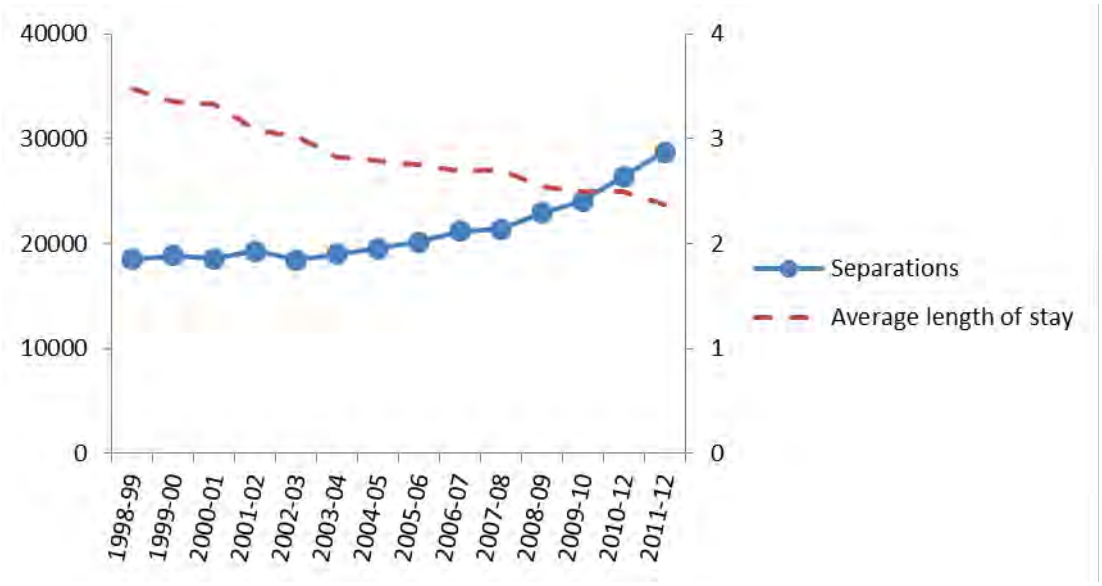


Figure 2.3 **Australian hospital separations and average length of stay (days) for low back pain (ICD-10-AM code M54.5) from 1998-99 to 2011-12 (Australian Institute of Health and Welfare, 2013d)**

Approximately half of Australians with an episode of LBP do not seek care from a health professional. However when care is sought, the majority of patients with LBP seek care from more than one healthcare provider (Walker, Muller, & Grant, 2004a). The greatest out-of-hospital costs are attributable to chiropractic, physiotherapy, general practice and massage therapy (Walker et al., 2003). Back problems constitute the majority of Australian chiropractic patient encounters, representing approximately 62% of all consultations (French et al., 2013). Back complaints are the 9th most frequently managed condition in Australian general practice, and are the primary reason for presentation in approximately 2.6% of GP-patient encounters (Cooke, Valenti, Glasziou, & Britt, 2013). A population-based mail survey of 1913

randomly selected Australian adults suggests that 15% and 13% of those who suffered with an episode of LBP in the past six months consulted a massage therapist and a physiotherapist respectively (Walker et al., 2004a).

Several studies across different countries have highlighted that the indirect economic costs attributable to LBP greatly outweigh the direct costs (Dagenais et al., 2008). One Australian study estimated the indirect costs to be approximately eight times larger than the direct costs (Walker et al., 2003). Indirect costs are often more difficult to accurately measure as they arise from LBP causing work absenteeism, reduced productivity, early retirement and the reduced ability to perform usual non-paid activities, such as housework (Dagenais et al., 2008).

The disability created by LBP has significant ramifications for workforce participation and the capacity to generate income. Within Australia, persons aged 15-64 years with self-reported long-term back problems are 1.2 times less likely (relative risk) to be in the labour force compared to those without back problems (Australian Institute of Health and Welfare, 2013f). This problem may be magnified in older Australians, with a retrospective study of persons aged between 45 and 64 years finding that those with back problems had 3.6 times greater odds of being out of the labour force (Schofield, Shrestha, Passey, Earnest, & Fletcher, 2008). It has been estimated that up to AUD$4.8 billion is lost in annual individual earnings each year in Australia as a consequence of early retirement caused by LBP (Schofield, Shrestha, et al., 2012). Independent of workforce participation,

and after adjusting for age, sex and education level, Australians with back problems are more likely to be living in poverty (OR = 2) compared to those without back problems (Schofield, Callander, et al., 2012).

The longer term consequences of workforce participation restrictions at an individual-level caused by LBP importantly include reduced retirement wealth. Australians aged between 45 and 64 years who have retired early due to back problems are less likely (absolute risk reduction = 18.2%) to own income producing assets (e.g. superannuation, shares, property) compared to those still able to participate in gainful employment (Schofield et al., 2011). Further, the median wealth at age 65 of those people with back pain forced to retire early from the workforce may be up to 80-97% less compared to those with back problems who are able to continue to work even in a part-time capacity (Schofield, Kelly, et al., 2012). From a societal perspective, workforce participation restrictions caused by LBP also result in reductions in government wealth. An Australian study identified that early retirement caused by LBP in people aged 45 to 64 years resulted in a AUD$497 million loss in taxation revenue, and an additional AUD$622 million in welfare payments (Schofield, Shrestha, et al., 2012).

Government and business also incur costs associated with compensable work-related back injuries. Towards the end of the 20th century the costs associated with workers' compensation claims for LBP dramatically escalated throughout much of the industrialised world, with at least one Australian state experiencing a three-fold increase within a decade (Buchbinder, 2008;

Waddell, 1996). Across Australia, 20,060 worker's compensation claims were made for a lower back injury resulting in one or more week's absence from work in the 2010-11 financial year. These injuries account for 16% of all claims (Safe Work Australia, 2013). In New South Wales, Australia's most populous state, almost one quarter of all major workplace injuries are back injuries (WorkCover New South Wales, 2010). In the 2008/09 financial year, there were 7,214 back injuries in NSW, of which 8% resulted in permanent disability. These injuries cost AUD$138 million and represent approximately 31% of the total costs incurred by WorkCover NSW for that financial year. The observation that the costs of workers' compensation claims for LBP are proportionally higher than the number of claims made does not appear to be confined to an Australian context (Webster & Snook, 1994).

Favourably however, both the total number (Figure 2.4) and incidence rate (Figure 2.5) of work-related LBP injuries have gradually reduced in Australia over the past decade (Safe Work Australia, 2013; WorkCover New South Wales, 2010). This may in part be the result of successful population-based mass media campaigns designed to change public attitudes, beliefs and behaviours regarding back pain (Buchbinder & Jolley, 2004; Buchbinder, Jolley, & Wyatt, 2001).

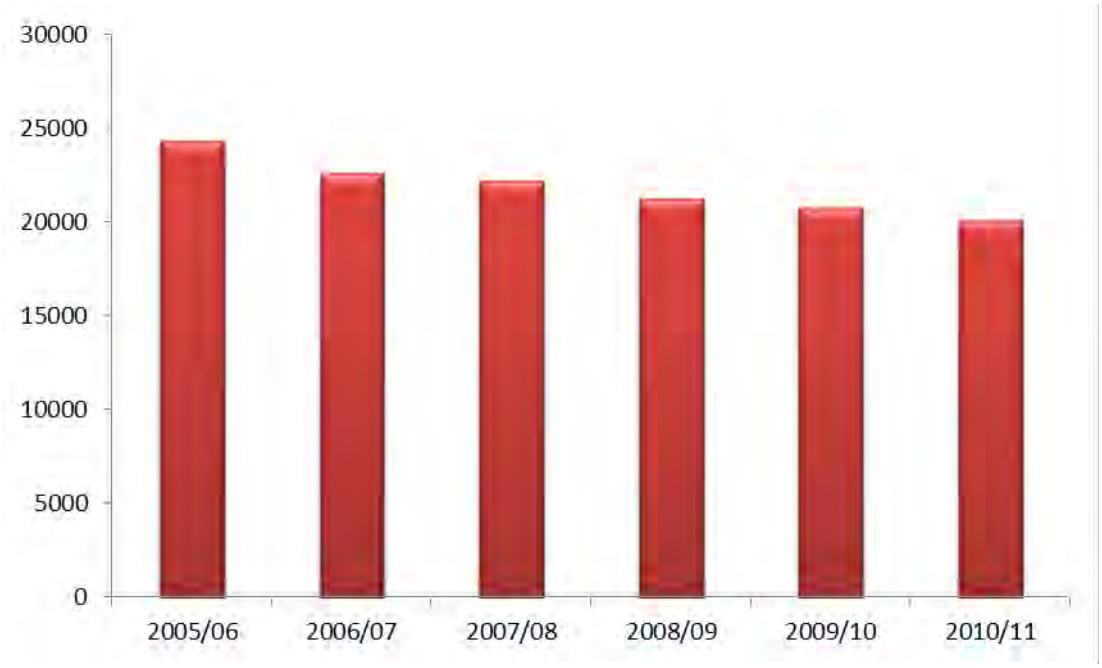


Figure 2.4 **Lower back injury workers' compensation claims resulting in absence from work ≥ 1 week in Australia from 2005/06 to 2010/11 (Safe Work Australia, 2013)**

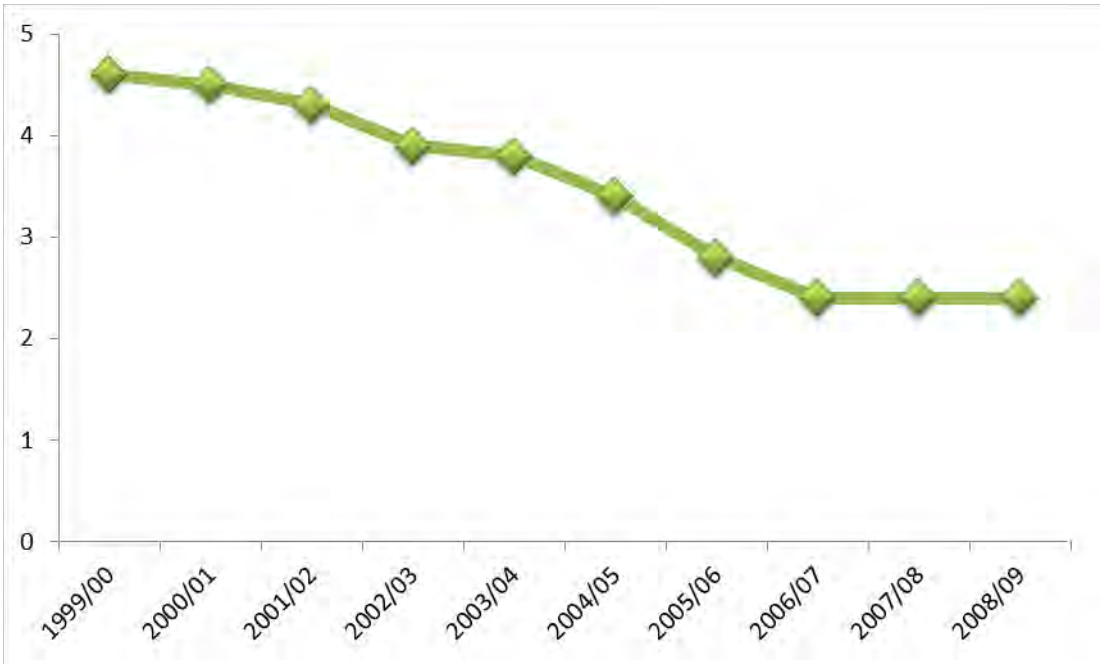


Figure 2.5 **Rate of workplace back injuries per 1,000 employees in New South Wales, Australia from 1999/00 to 2008/09 (WorkCover New South Wales, 2010)**

2.4 Heterogeneity and diagnostic triage of low back pain

LBP is a broad term that encapsulates a variety of clinical presentations that share, at least in part, an anatomically similar distribution of symptoms. Accordingly, LBP is not a single specific disease, but is rather comprised of numerous differing presentations. Such presentations range from the rarer but more serious conditions such as cancer, infection and fractures, to less serious presentations that may involve more simple sprains and strains of the local soft tissues (Bogduk, 1999; Delitto et al., 2012). The LBP patient population is therefore highly heterogeneous and the symptomology,

prognosis and treatment response to given interventions is correspondingly variable.

The heterogeneity of LBP was addressed in the first clinical practice guideline for spinal disorders published by the Quebec Task Force in 1987 (Spitzer, 1987). This guideline highlighted that LBP is not a homogenous presentation, but instead comprised of many different clinical presentations, only some of which have an identifiable pathoanatomic origin. The Task Force highlighted that wide-ranging nomenclature was often used to describe LBP presentations encompassing symptomatic descriptors, radiological findings and/or unsubstantiated pathoanatomic hypotheses. Such unwarranted variation was considered a major barrier to clinical decision-making, healthcare evaluation and scientific research. Consequently, an 11-category classification system was proposed for activity-related spinal disorders (summarised in Table 2.6). Central to this classification system was the delineation of presentations with and without a specific identifiable origin. In the absence of serious or specific causes of LBP, further categorisation of LBP was based upon clinical variables including the nature, distribution and duration of presenting symptoms, and the response to treatment.

Table 2.6 Quebec Task Force's classification of activity-related spinal disorders as it relates to LBP (Spitzer, 1987)

Classification	Description
1	Pain without radiation (no pain below gluteal fold).
2	Pain with radiation to proximal extremity (above knee).
3	Pain with radiation to distal extremity (below knee).
4	Pain with radiation to limb with neurologic signs (focal muscle weakness; reflex asymmetry; dermatomal sensory loss; or intestinal, bladder or sexual dysfunction).
5	Presumptive compression of a spinal nerve on simple X-ray (eg. spinal instability, fracture).
6	Compression of a spinal nerve confirmed by specific imaging (eg. CT, MRI) and/or specific investigations (eg. nerve blocks).
7	Spinal stenosis confirmed by CT or myelography.
8	Post-surgery within 6 months.
9	Post-surgery greater than 6 months ago. This category is further sub-classified as those who are asymptomatic (9.1) or symptomatic (9.2).
10	Chronic pain syndrome, characterised by the persistence of preoccupying pain in the absence of a treatable active disease.
11	All other diagnoses, including cancer, fracture, spondylitis, visceral disease.

Since the proposal of the Quebec Task Force's classification system, numerous clinical practice guidelines for LBP have been published with progressively improved quality (Bouwmeester, van Enst, & van Tulder, 2009). A near universal recommendation across guidelines for LBP is the use of a

diagnostic triage whereby presentations are categorised as having either; (1) confirmed or suspected serious or specific pathology; (2) radicular syndrome; or (3) non-specific LBP (Koes et al., 2010; Koes, van Tulder, Ostelo, Burton, & Waddell, 2001).

Serious or specific conditions that cause LBP are considered rare and include spinal fracture, cancer, infection, inflammatory disorders (e.g. ankylosing spondylitis), cauda equina syndrome and visceral disease (e.g. abdominal aneurysm) (Delitto et al., 2012; Deyo & Weinstein, 2001). Studies reporting on the prevalence of serious causes of LBP have provided variable estimates, which may relate to differences in the clinical setting, patient population, study design and choice of reference standard (Henschke, Maher, & Refshauge, 2007, 2008). A prospective study of 1,172 consecutive patients with acute LBP presenting in Australian primary care provides estimates regarding the prevalence of such conditions in this setting (Henschke et al., 2009). In this study, the prevalence of spinal fracture was 0.7% (95%CI 0.4% - 1.3%), cancer was 0% (95%CI 0% - 0.3%), infection 0% (0% - 0.3%), cauda equina syndrome was 0.1% (95%CI 0% - 0.5%) and inflammatory disorders was 0.2% (0.1% - 0.6%). The reference standard used in this study was patient self-report of a serious spinal pathology at one year following the initial assessment, in addition to specialist follow-up of a random sample of those not reporting serious spinal pathology. Ideally all patients would receive the appropriate reference tests (e.g. blood tests, imaging, specialist review etc.) for each serious spinal pathology (Whiting et al., 2011), however as acknowledged by the study authors, this would be

entirely unfeasible, expensive and has the potential to cause harm. It is also important to note that as this study used an inception cohort of patients presenting with acute LBP (<6 weeks duration) it may underestimate the total prevalence of serious causes of LBP presenting in primary care as those patients presenting with longer durations of symptoms were excluded.

Identifying patients with LBP caused by serious spinal disorders has important clinical implications. Patients with suspected serious spinal pathologies are contraindicated to many physiotherapy interventions (Hancock, Maher, & Latimer, 2008; Houghton, Nussbaum, & Hoens, 2010), and require further investigation and/or referral to medical specialist services for appropriate assessment and management (Australian Acute Musculoskeletal Pain Guidelines Group, 2003; Delitto et al., 2012). Practice guidelines universally recommend screening for serious causes of LBP via the identification of ‘red flags’ in a patient’s clinical history and physical examination (Koes et al., 2010). Table 2.7 below details the ‘red flags’ listed in the most recent practice guidelines for LBP produced by the American Physical Therapy Association.

Table 2.7 Red flag variables from the history and physical examination suggestive of serious spinal pathology (Delitto et al., 2012)

Serious spinal pathology	Red flag
Back-related tumour	Constant pain not affected by position or activity; worse with weight bearing, worse at night Age over 50 History of cancer Failure of conservative intervention (failure to improve within 30 days) Unexplained weight loss No relief with bed-rest
Cauda equina syndrome	Urine retention Faecal incontinence Saddle anaesthesia Sensory or motor deficits in the feet (L4, L5, S1 areas)
Back-related infection	Recent infection (e.g., urinary tract or skin), intravenous drug user/abuser Concurrent immunosuppressive disorder Deep constant pain, increases with weight bearing Fever, malaise, and swelling Spine rigidity; accessory mobility may be limited
Spinal compression fracture	History of major trauma, such as vehicular accident, fall from a height or direct blow to the spine Age > 50 Prolonged use of corticosteroids Point tenderness over site of fracture Increased pain with weight bearing
Abdominal aneurysm (≥ 4 cm)	Back, abdominal or groin pain Presence of peripheral vascular disease or coronary artery disease and associated risk factors (age over 50, smoker, hypertension, diabetes mellitus) Smoking history Family history Non-Caucasian Female Symptoms not related to movement stresses associated with somatic low back pain Abdominal girth < 100 cm

Serious spinal pathology	Red flag
	Presence of a bruit in the central epigastric area upon auscultation Palpation of abnormal aortic pulse Aortic pulse ≥ 4cm

Interestingly, clinical practice guidelines for LBP produced in different countries and by different professional groups are somewhat inconsistent with the 'red flags' recommended (Downie et al., 2013). This may be because there is little evidence to support the diagnostic utility of most of the recommended 'red flag' variables, with many remaining untested or found to have very high false positive rates (Downie et al., 2013; Henschke et al., 2013; Williams et al., 2013). For example, in the study of Henschke et al. (2009), it is reported that 80.4% of patients with LBP seeking treatment in primary care had one or more 'red flags', however just 0.9% of patients were identified as having a serious spinal pathology. Such findings have highlighted the need for clinical tools to be developed that are able to accurately identify individuals with LBP from a suspected serious spinal pathology (Downie et al., 2013).

Radicular syndrome is the second category within the diagnostic triage endorsed by LBP clinical practice guidelines (Koes et al., 2010; Koes et al., 2001). These patients have concomitant signs and symptoms suggestive of possible neurological involvement (Rossignol et al., 2007) and represent approximately 5% of patients with LBP encountered in primary care (Maher et al., 2011). A herniated intervertebral disc with associated nerve root

compression is the most common cause of radicular syndrome in young and middle-aged adults, however with advancing age symptomatic lumbar spinal stenosis is an increasingly probable cause (Atlas & Deyo, 2001; Chou et al., 2007; Koes, van Tulder, & Peul, 2007).

The diagnosis of lumbar radicular syndrome is predominantly based upon the clinical history and physical examination (Koes et al., 2007). A dermatomal distribution of radiating symptoms is considered to be a sensitive diagnostic variable for disc herniation with nerve root compression (Vroomen, de Krom, & Knottnerus, 1999). The straight leg raise test has also been identified to be highly sensitive (91%, 95%CI 82% - 74%) for this presentation, but is limited by poor specificity (26%, 95%CI 16% - 38%) (Deville, van der Windt, Dzaferagic, Bezemer, & Bouter, 2000). Conversely, the crossed straight leg raise test has relatively high specificity (88%, 95%CI 86% - 90%), but low sensitivity (29%, 95%CI 24% - 34%) (Deville et al., 2000). Many practice guidelines recommend a neurological clinical examination to investigate motor, sensory and/or reflex deficits, however these tests, at least in isolation, may only have very limited diagnostic utility (Al Nezari, Schneiders, & Hendrick, 2013; Hancock, Koes, Ostelo, & Peul, 2011; van der Windt et al., 2010). Diagnostic imaging is reserved for those presentations where it becomes necessary for further management decision-making, such as the consideration of the need for surgical intervention (Chou et al., 2007; Koes et al., 2007).

The identification of patients with radicular syndrome has meaningful prognostic and management implications. Although a large proportion of patients with disc herniation causing nerve root compression resolve over time with conservative management (Vroomen, De Krom, & Knottnerus, 2002), the extent and timeframe of improvements are generally less favourable when compared to non-specific LBP (J. C. Hill, Konstantinou, et al., 2011; Rossignol et al., 2007). In an inception cohort study of 123 patients with acute LBP presenting in primary care, the presence of at least two concomitant neurological signs at baseline significantly increased the odds of non-recovery at three months (OR = 4.6, 95%CI 1.4 – 14.9) (Grotle et al., 2005). A similar observation was made in a different study of 1,247 consecutive patients with LBP presenting in primary care in which the six month pain and disability outcomes were identified to differ in association with baseline symptom distribution (J. C. Hill, Konstantinou, et al., 2011). Self-reported radiating pain distal to the knee is not specific to radicular syndrome, as indicated by the relatively high prevalence (38%) of this symptom in this sample, however it is commonly used as a proxy for sciatica in large epidemiological studies (Ashworth, Konstantinou, & Dunn, 2011). This study identified that patients presenting with LBP and radiating pain distal to the knee were on average 4.4 points (95%CI 3.7 – 5.2) worse on the 24-point Roland Morris Disability Questionnaire (Roland & Morris, 1983), and 1.7 points (95%CI 1.3 – 2.0) worse on an 11-point composite pain severity scale (Dunn, Jordan, & Croft, 2010) at six months compared to those with localised LBP. Further, those with pain below the knee were less likely (49%

vs 61%, OR = 0.60, 95%CI 0.46 – 0.78) to report feeling improved at six months compared to those with localised LBP symptoms.

Failure to sufficiently improve with conservative management in patients with suspected radicular syndrome is an indicator for further investigation and referral to medical specialist services (Koes et al., 2007). Surgical intervention is only indicated for a limited proportion of patients with radicular syndrome, however the available evidence suggests that in selected patients with disc herniation or spinal stenosis, surgery may provide faster or greater clinical improvements compared to conservative management (Gibson & Waddell, 2007; Kovacs, Urrútia, & Alarcón, 2011; Lurie et al., 2014; Peul et al., 2007; Weinstein et al., 2010). A recent systematic review aimed to identify predictors of subsequent surgery in patients with sciatica receiving conservative management (Verwoerd et al., 2013). Among the 33 candidate predictor variables investigated, only greater leg pain intensity at the baseline assessment was found to be a strong predictor of progression to surgery. Within the non-surgical context, the available research evidence suggests that prognostic variables in radicular syndrome may in fact differ to those identified for non-specific LBP (Ashworth et al., 2011).

Non-specific LBP is the third and remaining category within the diagnostic triage advocated by practice guidelines (Koes et al., 2010; Koes et al., 2001) and encompasses the overwhelming majority (>85%) of all LBP presentations (Deyo, Rainville, & Kent, 1992; Maher et al., 2011). The term ‘non-specific’ relates to the lack of an attributable specific source of

symptoms, thereby making it a diagnosis of exclusion (Balague et al., 2012; Koes, van Tulder, & Thomas, 2006). Current practice guidelines for the assessment and management of LBP recommend that presentations are classified as ‘non-specific’ in the absence of suspected serious spinal pathology or radicular syndrome (Koes et al., 2010).

The low back muscles, fascia, ligaments, bone, intervertebral discs, zygapophyseal joints and sacroiliac joints are all potential pathoanatomic sources of non-specific LBP (Bogduk, 1999; Bogduk & Karasek, 2005). These structures have nociceptive innervation and therefore have the capacity to give rise to pain in this bodily region (Bogduk, 2005). Studies in which asymptomatic participants have received provocation injections into the sacroiliac joints (Fortin, Dwyer, West, & Pier, 1994) or zygapophyseal joints (McCall, Park, & O'Brien, 1979) have demonstrated that pain distributions, similar to that seen in patients presenting with LBP, are able to be produced when these structures are noxiously stimulated. Provocation of the lumbar intervertebral discs via discography (Schwarzer et al., 1995) or intradiscal heating (O’Neill, Kurgansky, Derby, & Ryan, 2002) may also reproduce concordant back +/- leg pain in many patients with LBP, but only in a small proportion of asymptomatic participants (Carragee et al., 2000; Derby et al., 2005; Walsh et al., 1990). Such findings, coupled with biologic plausibility, suggest that local somatic structures are potentially identifiable sources of the nociceptive origin of a non-trivial proportion of non-specific LBP presentations (Bogduk, 2005).

Indeed, it has been proposed that approximately 70% of patients with persistent LBP would be able to be diagnosed with a specific pathoanatomic source for their symptoms if invasive diagnostic procedures were to be applied (Bogduk & Karasek, 2005; Laslett, McDonald, Tropp, Aprill, & Oberg, 2005).

The use of lumbar provocation discography as a reference standard for the identification of intervertebral discs as the nociceptive source of LBP has been controversial, particularly due to the high rate of false positives reported in earlier studies in asymptomatic populations (Cohen et al., 2005). The International Association for the Study of Pain (IASP) has provided guidelines for the diagnosis of lumbar discogenic pain using standardised procedures for provocation discography to optimise its validity (Table 2.8) (Merskey & Bogduk, 1994). Using these recommendations, the prevalence of pain resulting from the noxious stimulation of the intervertebral disc in patients with persistent LBP has been estimated to be in the range of 26% to 42% (DePalma, Ketchum, & Saullo, 2011; Laslett, Aprill, McDonald, & Oberg, 2006; Manchikanti et al., 2001; Schwarzer et al., 1995). The prevalence rate may also vary with age, with reported findings of higher rates of pain during lumbar provocation discography in younger patients with persistent LBP (DePalma et al., 2011).

Table 2.8 Diagnostic criteria for lumbar discogenic pain and internal disc disruption using provocation discography (Merskey & Bogduk, 1994)

• Provocation discography of the putatively symptomatic intervertebral disc reproduces the patient's accustomed pain, and • the provocation of at least 2 adjacent intervertebral discs do not reproduce the patient's pain, and • the pain cannot be ascribed to some other source innervated by the same segments that innervate the putatively symptomatic intervertebral disc.
• For the diagnosis of internal disc disruption, discography must also demonstrate a grade 3 or greater annular disruption as defined by the Dallas discogram scale (Sachs et al., 1987).

The identification of lumbar zygapophyseal joints as the nociceptive origin of LBP has been conducted predominantly using controlled nerve blocks of the medial branches of the dorsal rami (Bogduk, 2004b). Prevalence estimates are notably sensitive to the percent reduction in symptoms required to define a positive test response (Bogduk & Karasek, 2005; Laslett, McDonald, Aprill, Tropp, & Oberg, 2006; Manchikanti, Pampati, & Cash, 2010). Based on data from studies that have used a reference criterion of at least 80% reduction in pain, the prevalence of 'positive' controlled nerve blocks of the lumbar zygapophyseal joints has been estimated to be within the range of 21% to 40% of persistent LBP presentations (Manchikanti, Datta, et al., 2010). Several studies have identified that blockade of the lumbar zygapophyseal

joints is more commonly positive in older patients with persistent LBP (DePalma et al., 2011; Laslett, McDonald, et al., 2006; Revel et al., 1992).

In contrast to zygapophyseal joints, identification of the sacroiliac joint as the nociceptive source of LBP cannot be diagnosed using nerve blocks due to its much broader innervation (Laslett, 2008). Instead, fluoroscopically guided controlled joint blocks are recommended (Bogduk, 2004b). In a review conducted by Manchikanti, Datta, et al. (2010), four studies were identified that provided prevalence estimates for the sacroiliac joint as the nociceptive source of LBP using a criterion of at least 70% reduction in pain following invasive diagnostic blocks. These data gave an estimated prevalence of between 10% to 27% in patients with persistent LBP (Irwin, Watson, Minick, & Ambrosius, 2007; Laslett, Young, Aprill, & McDonald, 2003; Maigne, Aivaliklis, & Pfefer, 1996; Manchikanti et al., 2001).

Importantly, there are notable criticisms regarding the interpretation and significance of diagnostic provocation/blocking procedures. Pain is not analogous to nociception, and each can occur in the absence of the other (Moseley & Vlaeyen, 2015). Consequently, identifying a putative nociceptive source may not necessarily have implications regarding a patient's pain experience. The validity of axial diagnostic blocks has also been challenged with the absence of a true 'gold standard' negating the ability to validate these procedures using conventional methods (Carragee, Haldeman, & Hurwitz, 2007). Furthermore, inter-study variability in the standard by which a diagnostic anaesthetic block is considered 'positive' has plausibly contributed

to discrepancies in reported prevalence and false positive rates (Manchikanti et al., 2013). Engel, MacVicar, and Bogduk (2014) have subsequently proposed eight criteria, based largely on the Bradford Hill criteria for determining causality (A. B. Hill, 1965), to enable the empirical assessment of a given diagnostic anaesthetic block's validity. However, these criteria have not yet been applied to diagnostic block procedures relevant to LBP.

A further critical question concerns the value of attempting to establish a pathoanatomic diagnosis using invasive diagnostic procedures in patients who would otherwise be classified as having 'non-specific' LBP. Diagnostic procedures are invasive, difficult to access (Maher et al., 2011) and have the potential to cause serious complications (Bogduk et al., 2008). For example, discography may cause discitis, nerve root injury, bleeding, epidural abscess, subarachnoid puncture, disc herniation or chemical meningitis (Cohen et al., 2005; Junila, Niinimäki, & Tervonen, 1997; Poynton, Hinman, Lutz, & Farmer, 2005; Rathmell, Saal, & Saal, 2008; Walker III, El Abd, Isaac, & Muzin, 2008). A ten year follow-up study comparing 50 patients with LBP who received discography to 52 matched control subjects, demonstrated that those who received discography experienced accelerated degenerative changes in the lumbar disc over this time (Carragee et al., 2009). Substantial consideration of the relative benefits and risks of such diagnostic procedures is therefore arguably merited.

Given that the prognosis of acute LBP is generally favourable (Costa et al., 2012), invasive diagnostic procedures to identify a non-sinister somatic

source of symptoms in this population is unlikely to be considered justifiable in routine clinical care. However, it has been argued that the pursuit of a pathoanatomical diagnosis in those with persistent LBP that has been unresponsive to therapy may in some instances be meaningful to informing patient management (Bogduk, 2004a). Nonetheless, there is currently a lack of high quality evidence to suggest that knowledge of the pathoanatomic source of symptoms meaningfully changes management or leads to improved patient outcomes (Chou et al., 2007; Maher et al., 2011). Consequently, clinical practice guidelines do not presently provide a recommendation for the use of invasive diagnostic techniques to identify a specific pathoanatomic cause of symptoms in those who would otherwise be classified as having non-specific LBP (Koes et al., 2010).

The current body of evidence also indicates that less invasive procedures, including clinical tests and imaging are unlikely to provide a suitable alternative as most have been found to be of limited diagnostic utility in informing pathoanatomic diagnoses in patients with LBP (Chou, Qaseem, Owens, & Shekelle, 2011; Endean, Palmer, & Coggon, 2011; Hancock et al., 2007; Wassenaar et al., 2012).

Hancock et al. (2007) conducted a systematic review of studies published up until February 2006 investigating the diagnostic accuracy of clinical tests in identifying a pathoanatomic source of symptoms in patients with LBP. A notable finding of this study was the absence of any evidence concerning the diagnostic utility of a large number of clinical tests traditionally considered to

be suggestive of specific pathoanatomic diagnoses. Of the few clinical tests that have been investigated, the evidence considered within this review found their magnitude of diagnostic accuracy to be of limited clinical value. For example, the centralisation of symptoms with repeated lumbar movements was found to be the only clinical test that increases the probability of a patient having a positive discography. However, the pooled positive likelihood ratio (+LR) of 2.8 (95%CI 1.4 – 5.3) suggests that this test is unlikely to meaningfully change the probability of this diagnosis in most contexts.

Abnormalities on lumbar MRI are commonly found in asymptomatic individuals, which leads to questions about the value of such findings in patients presenting with LBP (Boden, Davis, Dina, Patronas, & Wiesel, 1990; Jensen et al., 1994). In a large population-based study involving 1,043 volunteers, findings on MRI were compared between people with and without a self-reported history of LBP, which was operationally defined as ever having pain in the low back for more than two weeks requiring physician consultation or treatment (Cheung et al., 2009). The results demonstrated a weak relationship between a history of LBP and degenerative disc disease (OR 2.2, 95%CI 1.4 – 3.4) and disc herniation (OR = 2.1, 95%CI 1.4 – 3.1), but no relationship between LBP and annular tears (OR = 0.98, 95%CI 0.63 – 1.5) or Schmorl's nodes (OR = 1.3, 95%CI 0.7 – 2.5). Hancock et al. (2012) however demonstrated, that the value of certain MRI findings in delineating symptomatic and asymptomatic populations significantly improves when restricting the symptomatic patient cohort to those with clinical features that

raise the diagnostic probability of discogenic pain (such as the presence of centralisation of symptoms).

A further consideration is the risk of harm. Routine imaging in patients with non-specific LBP has been demonstrated to not only *not* improve clinical outcomes (Chou, Fu, Carrino, & Deyo, 2009), but to instead have a deleterious influence on patient well-being (Ash et al., 2008) and lead to future avoidable cancers as a result of unnecessary radiation exposure (Flynn, Smith, & Chou, 2011). The costs of care in the management of patients with LBP are also notably amplified as a result of routine imaging (Chou, Deyo, & Jarvik, 2012).

2.5 Sub-classification of low back pain

Non-specific LBP is not considered to be a homogenous population, and in the absence of meaningful valid pathoanatomic classification in routine clinical care, various sub-classification approaches have been proposed and implemented to inform patient management. The identification of valid LBP subgroups, whose identification leads to improved patient outcomes, has been a priority area of research for at least the past two decades (Borkan & Cherkin, 1996; Borkan, Koes, Reis, & Cherkin, 1998; Costa et al., 2013; Foster et al., 2009; Henschke, Maher, Refshauge, et al., 2007).

Serial cross-sectional surveys and group-based workshops involving attendees of the International Forum for Primary Care Research on Low

Back Pain have provided insights into the changing research priorities in the field of LBP. Most recently, Costa et al. (2013) surveyed attendees of the 10th International forum for Primary Care Research on Low Back Pain Forum in Boston in 2009. Data provided from 145 participants (37% physiotherapists) enabled the ranking of the current perceived LBP research priorities and also enables comparison to previous findings in 1995 (Borkan & Cherkin, 1996) and 1997 (Borkan et al., 1998) (Table 2.9). Many research priorities have changed over time, or split into more discrete research questions, such as the separation of investigating subgroups of LBP into the identification of subgroups and the identification of specific treatments for those subgroups. Nevertheless, it is evident that the identification of meaningful subgroups of patients with LBP remains a high research priority.

Table 2.9 Primary care research agenda from studies on attendees of the International Forum for Primary Care Research on Low Back Pain presented as rankings of priority (Borkan & Cherkin, 1996; Borkan et al., 1998; Costa et al., 2013)

	10th 2009 USA n = 145	**2nd 1997 The Netherlands n = 45**	**1st 1995 USA n = 41**
Identification of clinically relevant subgroups of LBP	1	1	1
Specific treatment for specific subgroups of LBP	2	-	-
Translation of research into clinical practice	3	7	4
Understanding causes and	4	-	-

	10th 2009 USA n = 145	2nd 1997 The Netherlands n = 45	1st 1995 USA n = 41
mechanisms of LBP			
Understanding non-specific treatment effects	5	-	9
Identifying the most clinically and cost-effective treatments	6	-	5
Organisation of primary care services to improve efficiency	7	-	-
Understanding how to improve self-care strategies	8	4	8
Prevention of an episode of LBP or recurrence	9	2	13
Identifying effective diagnostic tests	10	-	-

Evidence concerning the presence of subgroups of patients with non-specific LBP has been provided through multiple avenues including; the examination of clinician beliefs and behaviours, observed variance in symptomatic manifestations and clinical prognoses, and observed variance in treatment effects.

Research evidence indicates that clinicians who manage patients with non-specific LBP do not believe that it is a single clinical condition. Kent and Keating (2004) conducted a postal survey of 651 Australian primary clinicians (39% physiotherapists) who manage patients with LBP. The overwhelming majority (93%) of practitioners indicated that they believed non-specific LBP

to be comprised of more than one condition. Further, 74% believed that at the time of the postal survey it was possible to identify the differing subgroup presentations, however this latter view varied significantly between professional disciplines. Physiotherapists (90%), osteopaths (85%), and chiropractors (72%) were more likely to indicate they could identify subgroups compared to general practitioners (40%) and musculoskeletal medicine practitioners (55%). Independent of identifying subgroups, almost all surveyed clinicians (93%) indicated that they varied their management based on an individual patient's pattern of signs and symptoms.

From the same study, Kent and Keating (2005) further identified minimal concordance amongst primary care clinicians in both the sub-classification labels attributed to differing presentations, and the signs and symptoms that delineate them. Sub-classification labels used by clinicians most commonly (84%) related to a putative pathoanatomic diagnosis with the five most common relating to the zygapophyseal joint, intervertebral disc, sacroiliac joint, instability and postural syndrome. However, there was less than 10% agreement concerning the three most common signs or symptoms that could be used to characterise these presentations. Agreement was identified to be greater within professional disciplines than across disciplines, which lead the authors to hypothesise that discipline-specific views, or 'clinical cultures', may in part account for some of the observed variance.

Observed variation in the prognosis of inception cohorts of patients with non-specific LBP also provides evidence for the existence of meaningful

subgroups. In a recent meta-analysis of 11 prospective cohort studies of patients presenting with acute (< 3 months duration) LBP in primary care, Itz, Geurts, Kleef, and Nelemans (2013) identified a large variation in the recovery from pain over 12 months. Their analysis demonstrated that while a proportion of patients experienced recovery from pain within three months, many patients continued to experience pain for at least 12 months following their initial presentation. The results of the pooled analysis are detailed in Figure 2.6 below.

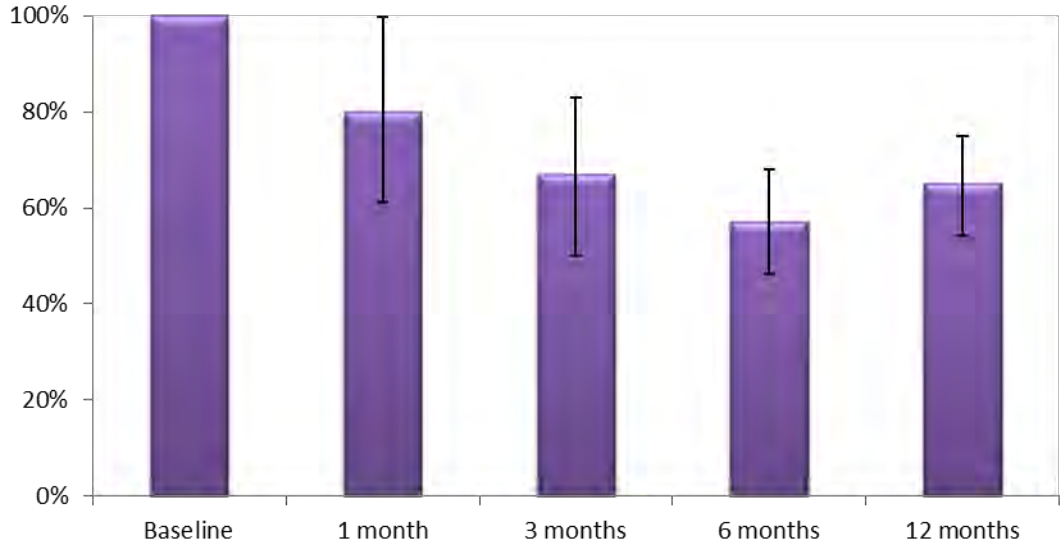


Figure 2.6 Proportion (95% CI) of patients with acute non-specific LBP presenting in primary care with pain at 1, 3, 6 and 12 months. Data from pooled results of Itz et al. (2013)

Costa et al. (2012) found similar results in their meta-analysis of prospective cohort studies of patients with acute (< 3 months duration) or persistent (3 –

12 months) non-specific LBP. This review investigated the degree of change in continuous pain and disability outcomes across 12 months. The pooled estimates demonstrate that the largest degree of improvements in pain and disability occurred within the first six weeks, in both patients with acute and persistent symptoms, with subsequent improvements being of a low magnitude after this time. Patients with an acute duration of symptoms at the baseline assessment improved to a greater extent compared to those with persistent symptoms, as illustrated in Figure 2.7 (p. 66) and Figure 2.8 (p. 68). Importantly, the within-study standard deviations of these outcomes were generally around 20 / 100. Given that the data are normally distributed, this indicates that approximately one third of patients experienced outcomes more than 20 standardised points different from the study mean, highlighting a moderate degree of between-patient variability in these outcomes.

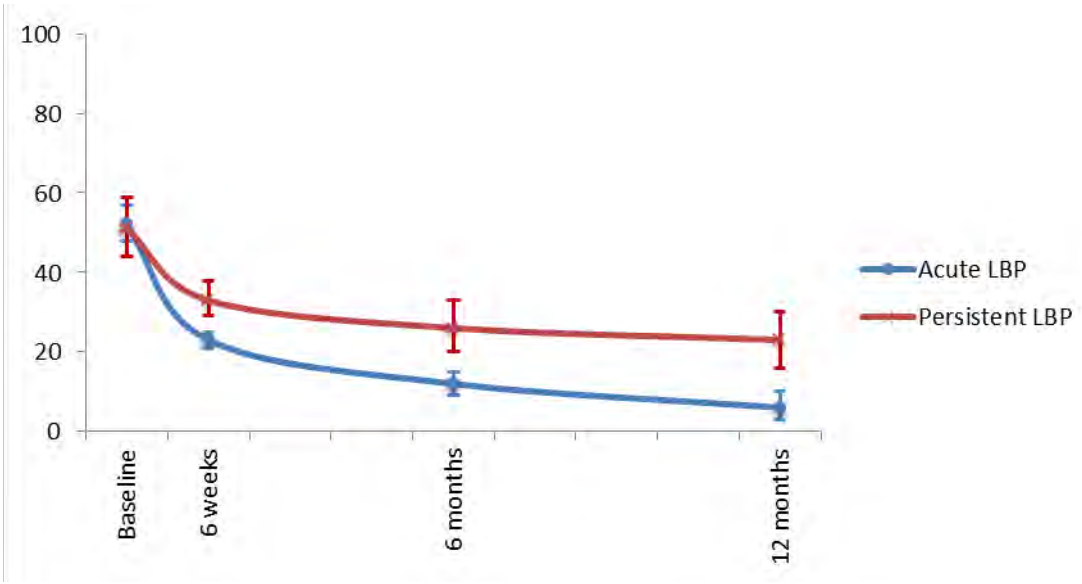


Figure 2.7 **Standardised (0-100) mean pain outcomes (95% CI) in patients with acute and persistent non-specific LBP across 1 year. Data from pooled results of Costa et al. (2012)**

Another argument for the existence of meaningful subgroups of patients with non-specific LBP has arisen from the observation of low effect sizes in most clinical trials of patients with non-specific low back pain (Assendelft, Morton, Emily, Suttorp, & Shekelle, 2003; Hayden, van Tulder, Malmivaara, & Koes, 2005; Machado, De Souza, Ferreira, & Ferreira, 2006; Machado et al., 2009). A meta-analysis of LBP clinical trials included in Cochrane reviews published up until 2005 was conducted to investigate the effect size for treatments for non-specific LBP compared to placebo, sham or no intervention (A. Keller, Hayden, Bombardier, & van Tulder, 2007). A key finding arising from this study was that the pooled effect sizes for pain at short and long-term follow-

up for a range of interventions for non-specific LBP were only low to moderate.

Similar results were identified by Machado et al. (2009) in their meta-analysis of randomised placebo-controlled clinical trials across 34 different treatments for non-specific low back pain. They found that almost half of the 76 included studies reported an average effect size for pain of less than 10 points on a standardised 0-100 scale, and just 15% reported effect sizes of greater than 20 points. The meta-analysis also found that several commonly-employed treatments, including acupuncture, spinal manipulative therapy, behavioural treatments, exercise and traction, were no more effective than a placebo, at least when considering the pooled average effect-size for pain.

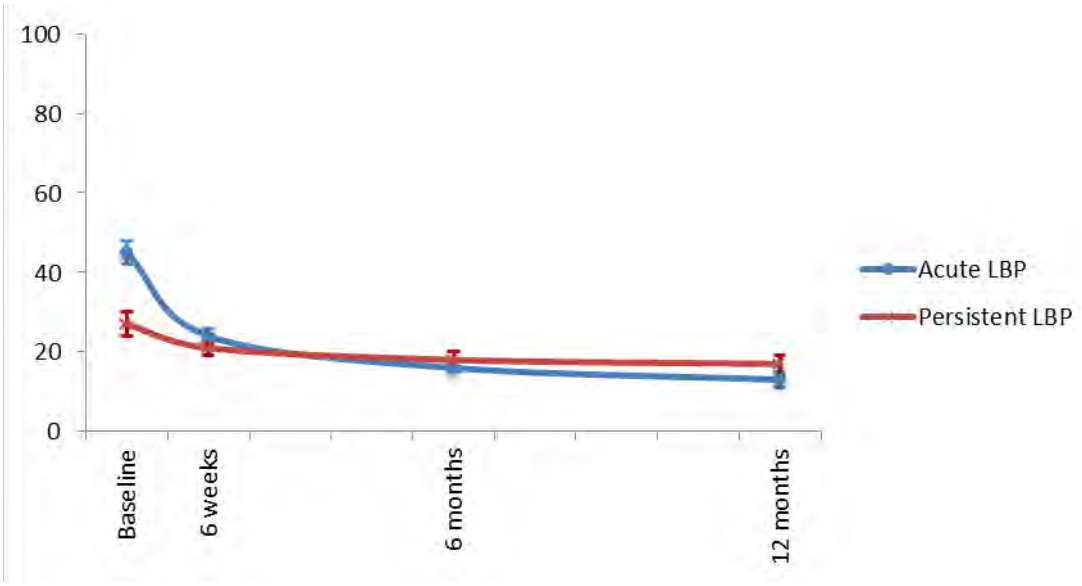


Figure 2.8 **Standardised (0-100) mean disability outcomes (95% CI) in patients with acute and persistent non-specific LBP across 1 year. Data from pooled results of Costa et al. (2012)**

The observation of low mean effect-sizes in clinical trials of patients with non-specific LBP has led many to hypothesise that treatment effects in such trials may be diluted as a result of differing degrees of treatment response across separate subgroups of patients (Foster et al., 2011). That is, while some patients within a clinical trial may meaningful benefit from a given intervention, others may not or may in fact worsen, thus giving rise to an overall average effect size of close to zero. Figure 2.9 (p. 69) adapted from Foster et al. (2011), illustrates the hypothetical distribution of treatment effect from an intervention evaluated in a controlled clinical trial, and the subgroup that experiences a meaningful degree of improvement.

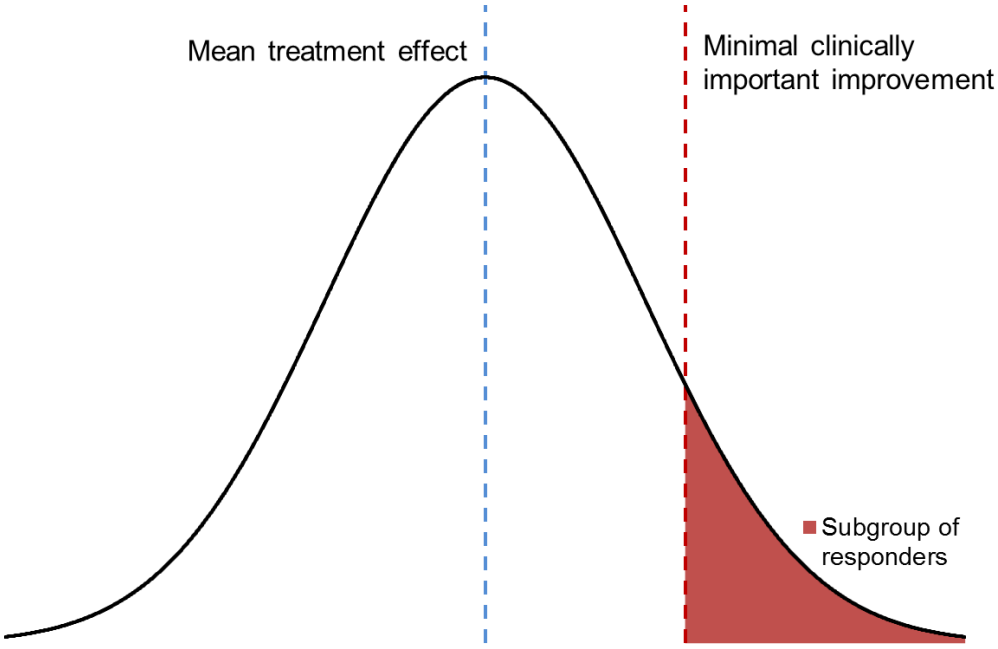


Figure 2.9 Hypothesised distribution of treatment effect and subgroup of responders (adapted from Foster et al. (2011))

Numerous approaches have been developed to sub-classify LBP, however, no existing system has universal acceptance. Fairbank et al. (2011) conducted a systematic review of the literature published up until January 2011 that sought to describe a clinical classification system for chronic LBP. Twenty-eight unique classification systems were identified in this review, which the authors separated into three broad types of classification approaches: diagnostic (Table 2.10, n=16) – those based on patterns of signs/symptoms or putative pathoanatomic labels without intending to predict outcomes or treatment effects: prognostic (Table 2.11, n=7) – approaches

based on prognostic outcomes irrespective of treatment; and treatment-based (Table 2.12, n=5) – approaches linked to treatment effects. Six additional publications describing classification systems were excluded from this review as they related only to acute low back pain. Most notably, the findings of this review concur with that of previous literature reviews that have sought to identify and contrast the various sub-classification approaches in that it highlighted the notably few similarities amongst them (Billis, McCarthy, & Oldham, 2007; C. McCarthy et al., 2004; Petersen, Thorsen, Manniche, & Ekdahl, 1999; Riddle, 1998). Further, it is apparent that most systems have been derived sorely on the basis of expert opinion and biologic plausibility, and despite the large number of classification systems that have been developed, very few have been formally evaluated for the their ability to positively impact clinical practice (Fairbank et al., 2011).

Table 2.10 Diagnostic classification systems for chronic low back pain (adapted from Fairbank et al. (2011))

1.	Back pain classification scale (Leavitt, Garron, Whisler, & Sheinkop, 1978)	Based on the use of back pain descriptors (e.g. 'cruel') to classify
2.	Pathology-based classification (Kirkaldy-Willis & Hill, 1979)	Based upon pathoanatomy (eg. segmental instability)
3.	Patient description of symptoms (Nachemson & Andersson, 1982)	Symptoms linked to classification (eg. rhizopathy)
4.	Empirically derived classification (Heinrich, Cohen, Naliboff, Collins, & Bonebakker, 1985)	Empirically derived seven category classification system
5.	Quebec Task Force on Spinal Disorders (Spitzer, 1987)	See Table 2.6 (p. 45)
6.	Complaint (duration and location) based classification (Mooney, 1989)	Links symptom distribution and duration to form nine categories
7.	Mechanical and psychiatric syndromes (Coste, Paolaggi, & Spira, 1992)	Based on a relationship between clinical presentation and presence of psychiatric disease
8.	Quebec Task Force on Spinal Disorders – based (DeRosa & Porterfield, 1992)	Seven category classification system based on the Quebec Task Force classification approach
9.	Diagnostic classification (Rezaian, Spector, & Collins, 1993)	Classifies patients as having either constant (malignant) or intermittent (benign) pain with subcategories
10.	National Institute of Occupational Safety and Health Low Back Atlas (Moffroid, Haugh, Henry, & Short, 1994)	Uses assessment of physical impairments (e.g. muscle length/strength, range of motion) to classify

11.	Pathoanatomic-based classification (Laslett & van Wijmen, 1999)	12 category putative pathoanatomic classification system with multiple subcategories
12.	Spatial distribution of pain and straight leg raise test (BenDebba, Torgerson, & Long, 2000)	Assigned to one of four groups based on distribution of pain and results of straight leg raise test
13.	Pathoanatomic classification (Petersen et al., 2003)	13 category system using putative pathoanatomic labels based on clinical signs and symptoms
14.	Lifting pattern-based classification (Slaboda, Boston, & Rudy, 2006)	Classification of patients based on time series of lifting patterns
15.	Modified Quebec classification (O'Hearn, Lowry, Emerson-Kavchak, & Courtney, 2009)	Modification of original Quebec Task Force classification system to suit physical therapy practice
16.	Patho-mechanism-based classification (Schäfer, Hall, & Briffa, 2009)	Four category system for classifying low back-related leg pain

Table 2.11 Prognostic classification systems for chronic low back pain (adapted from Fairbank et al. (2011))

1.	Behavioural classification (Keefe, Bradley, & Crisson, 1990)	Four category classification system based on pain behaviours
2.	The Distress and Risk Assessment Method (Main, Wood, Hollis, Spanswick, & Waddell, 1992)	Four category system based on identifying risk of poor outcome
3.	Psychosocial-based classification (Klapow et al., 1993)	Three category classification system based on psychosocial factors
4.	Classification based on a Phase Model of Disability (Krause & Ragland, 1994)	Classification system consisting of eight consecutive phases of the development of permanent work disability
5.	Psychosocial Assessment Model (Strong, Ashton, & Stewart, 1994)	Three category classification system integrating differing aspects of LBP, including pain, attitudes and coping strategies
6.	Multidimensional pain inventory-based classification (Bergstrom, Jensen, Bodin, Linton, & Nygren, 2001)	Classed as adaptive copers, dysfunctional, or interpersonally distressed, based on the Multidimensional Pain Inventory
7.	Based on Dallas Pain Questionnaire (Ozguler et al., 2002)	Four category classification system based on responses to the Dallas Pain Questionnaire

Table 2.12 Treatment-based classification systems for chronic low back pain (adapted from Fairbank et al. (2011))

1.	Mechanical Diagnosis and Therapy (McKenzie, 1981)	Clinical assessment used to categorise patients into different treatment groups of differing mechanical loading strategies
2.	Classified by pain category (Sikorski, 1985)	Eight category system linked to location and duration of pain, and exacerbating postures and movements.
3.	Movement system impairment classification (van Dillen et al., 1998)	Five category system based on symptomatic response to lumbar movement testing
4.	Movement and control impairment (P. O'Sullivan, 2005)	Five category system based on patterns of movement and hypothesised pain mechanisms
5.	Canadian Back Institute classification system (H. Hall, McIntosh, & Boyle, 2009)	Classification based on location of dominant pain and pain behaviours.

In the limited studies in which patients with non-specific LBP have been randomised to receive either 'matched' or 'non-matched' interventions according to a clinical sub-classification system, treatment effects have generally been identified to be larger in the matched treatment cohorts and cost-savings have also been demonstrated. This is illustrated in the studies summarised below.

Long, Donelson, and Fung (2004) investigated the effect of prescribing specific exercises according to the principles of the Mechanical Diagnosis and Therapy classification approach (McKenzie, 1981) (see Table 2.12). Two-hundred and thirty patients with LBP who had previously been identified

to have a ‘directional preference’ (a favourable symptomatic response to a directionally-specific mechanical loading strategy) were randomised to receive exercises matched to their directional preference, or exercises in the opposite direction, or multi-directional exercises. Those receiving exercises matched to their identified directional preference demonstrated greater improvements in their back pain, leg pain, disability, medication use and depressive symptoms compared to the other treatment groups at the two week follow-up. No further follow-up data was collected however, and therefore it is not known if the between-group differences in improvements lasted beyond this time-frame. It is also not known if the superiority of the experimental intervention would remain if the comparison intervention was more reflective of routine clinical practice and consequently comprised of exercises individually prescribed for each patient based on their identified needs and preferences. Nevertheless, the results of this study highlight the potential for matched exercises to at least accelerate the symptomatic recovery from LBP in those patients with an identified directional preference. This sub-classification system also appears to be favoured by many physiotherapists given its reported wide use in clinical practice (Hamm et al., 2003; Poitras, Blais, Swaine, & Rossignol, 2005, 2007; Spoto & Collins, 2008).

Another treatment-based sub-classification system, incorporating different modes of interventions, has been demonstrated to lead to superior clinical improvements when applied in patients presenting to physiotherapy. The approach, termed ‘Treatment-Based Classification’, was originally described

by Delitto, Erhard, and Bowling (1995) and assigns patients with LBP into treatment groups based on their presenting signs and symptoms. This classification approach has been demonstrated to have acceptable overall inter-rater reliability (Fritz, Brennan, Clifford, Hunter, & Thackeray, 2006; Henry, Fritz, Trombley, & Bunn, 2012), however classifications may be unclear for up to one third of patients due to the system classifications not being entirely mutually exclusive or comprehensively exhaustive (Stanton et al., 2011; Stanton, Hancock, Apeldoorn, Wand, & Fritz, 2013).

In the first RCT to evaluate the effectiveness of this sub-classification system, Fritz, Delitto, and Erhard (2003) randomised 78 patients with acute LBP to receive treatment according to the classification system, or to guideline recommended care including advice to remain active, aerobic exercise and muscle reconditioning exercise. The results of this study demonstrated superior clinical improvements in disability (11 point between-group difference on 100 point scale) and return to work outcomes at four weeks in the group receiving ‘matched’ treatment. However, no statistically significant between-group differences persisted at the 12 month follow-up. The median costs of care were 23% lower in the ‘matched’ intervention group. It is possible that the findings of this study simply reflect treatment effects related to the difference in interventions received in the 2 groups, as opposed to any effect produced by ‘matching’ patients to specific interventions. This limitation was identified and addressed in a subsequent RCT (Brennan et al., 2006).

Brennan et al. (2006) conducted a follow-up RCT to that of Fritz et al. (2003) in which 123 patients with acute LBP were assessed at baseline concerning their 'status' on the Treatment-Based Classification system and subsequently randomised to receive interventions within that approach (manipulation, stabilisation exercise, specific exercise). The main analysis sought to explore differences in outcomes between those randomised to receive either 'matched' or 'unmatched' interventions according to their baseline status. Those receiving 'matched' treatments experienced larger improvements in disability at four weeks (6.6 points on 100 point scale, 95%CI 0.7 – 12.5) and at one year (8.3 points, 95%CI 2.5 – 14.1) compared to those randomly assigned to 'unmatched' interventions. Additionally, no significant main effect across time was identified for either 'treatment group' or 'classification group'. This indicates that differences were not simply due to the treatment a patient received or to the subgroup to which a patient was assigned, but instead it was dependent upon a patient receiving their appropriate 'matched' intervention.

A consideration in interpreting the results of this study is that in clinical practice, patients are not generally randomly allocated to receive a given intervention, but instead clinicians use complex decision-making processes to inform their treatment decisions (I. Edwards, Jones, Carr, Braunack-Mayer, & Jensen, 2004; Higgs, Jones, Loftus, & Christensen, 2008; M. A. Jones & Rivett, 2004). The results of the Brennan et al. (2006) RCT therefore need to be considered within the context of the comparative group being randomly allocated to one of three interventions, and not being allocated

using clinician judgement. The clinical value of the Treatment-Based Classification system, and indeed all sub-classification approaches, is best demonstrated by evidence that its application results in greater beneficial clinical consequences than would otherwise be achieved using traditional decision-making processes. Such a study design is a strength of the RCTs (Foster et al., 2014; J. C. Hill, Whitehurst, et al., 2011; Whitehurst, Bryan, Lewis, Hill, & Hay, 2012) that have sought to evaluate the benefit of the STarT Back tool.

The STarT Back tool is a brief questionnaire and algorithm that informs treatment decision-making based on discriminating between groups of patients with LBP with differing degrees of risk of developing a poor outcome (J. C. Hill et al., 2008). In contrast to most other sub-classification approaches which have been derived based almost entirely on expert opinion and biologic plausibility, the development of the STarT Back tool was informed by the empirical identification of modifiable baseline prognostic predictors of a poor disability outcome (J. C. Hill et al., 2008). The tool has been recommended to be used to inform decisions regarding which patients presenting in primary care should be referred to consult with a physiotherapist and whether they require cognitive-behavioural interventions (Sowden et al., 2011).

When applied in primary care, stratifying patient care using the STarT Back tool has been demonstrated to result in slightly larger improvements in disability at four, six and 12 months, small improvements in the overall cost of

care, and less time off work compared to contemporary best practice care (Foster et al., 2014; J. C. Hill, Whitehurst, et al., 2011; Whitehurst et al., 2012). Other studies have however raised questions as to whether the application of the tool is likely to generalise to other settings, given the tool's mixed predictive performance in cohorts of patients presenting to physiotherapy and chiropractic clinics (Beneciuk et al., 2013; Beneciuk, Fritz, & George, 2014; Field & Newell, 2012; Fritz, Beneciuk, & George, 2011). Whether the implementation of the STarT Back tool results in improved clinical outcomes in these settings remains a testable hypothesis, and importantly, is not dependent of the tool's prognostic predictive accuracy.

A relatively more recent approach to the sub-classification of LBP relates to the development of clinical prediction rules (CPRs). A CPR is a clinical tool designed to be used with in individual patient, and is comprised of variables from the history, physical examination, and other investigations that have been statistically identified to predict a given diagnosis or outcome (McGinn et al., 2008). CPRs may be viewed as one of several overlapping methods that have arisen to help sub-classify patients with LBP to optimise clinical outcomes and facilitate targeted treatment (Foster et al., 2013). CPRs designed to assist in the assessment and management of LBP are the focus of this program of research, and are discussed in detail in the following chapter.

2.6 Summary

LBP has likely always afflicted mankind. It is exceptionally common and the leading cause of global health burden. The economic footprint of LBP is also significant, particularly with consideration of the indirect costs related to work absenteeism, reduced productivity, early retirement and the reduced ability to perform usual non-paid activities. Currently, it is not possible to identify a specific cause of LBP in the large majority of patients presenting for care. Clinical practice guidelines reflect this in advocating the use of a diagnostic triage whereby patients are classified as either having (1) confirmed or suspected serious or specific pathology; (2) radicular syndrome; or (3) non-specific LBP. Patients with non-specific LBP are not however considered to a homogeneous clinical population, but are instead comprised of subgroups of patients with differing prognoses and likelihoods of treatment response to given interventions. Numerous approaches have been developed and advocated to assist in the sub-classification of patients with non-specific LBP. Most are predominantly based on expert opinion and biologic plausibility, and there is generally little concordance amongst them. In the limited circumstances in which sub-classification approaches have been investigated for their ability to positively impact clinical practice, the emergent evidence has generally been favourable. Most recent approaches to the sub-classification of LBP include the development of CPRs, which will be discussed in detail in the following chapter.

CHAPTER 3

LITERATURE REVIEW: CLINICAL PREDICTION RULES

3.1 Introduction

A clinical prediction rule (CPR) may be used to sub-classify patients presenting with LBP, and has been defined as “a clinical tool that quantifies the individual contributions that various components of the history, physical examination and basic laboratory results make towards the diagnosis, prognosis, or likely response to treatment in an individual patient” (McGinn et al., 2008, p. 493). This chapter will provide a brief overview of the history of the use of statistics in healthcare decision-making and the concept and importance of considering knowledge within a probabilistic framework. The concept of CPRs will be detailed including the processes and methodological considerations pertinent to their development. The challenges and complexities relating to CPRs as an innovation within contemporary clinical practice will also be discussed. Finally, clinician behaviours and priorities in relation to CPRs will be examined.

3.2 An overview of statistics in healthcare decision making

Decision-making in ancient medicine was predominantly based upon philosophies and practices that integrated mysticism with observations of

human phenomena (Koenig, 2000). Supernatural explanations were central to early primitive medicine, with treatments often intended to exorcise demonic forces (Rodin, 1962). More rational medical belief systems emerged later in Ancient Egyptian (Breasted, 1930) and Greek cultures with notable contributions from Hippocrates and Galen and their bodily humors theory of disease causation (Coxe, 1846). Galen's texts subsequently served as the unchallengeable authority on medicine in Western Europe for the next one and a half millennia until the scientific revolution of the Renaissance (Fullerton & Silverman, 2009). During this period, new scientific discoveries and insights in medicine lead to the collapse of the established orthodox views of the time and gave way to the explicit use of the Scientific Method to inform healthcare decision-making (Brock, 1916).

In a landmark publication, Ledley and Lusted (1959) first highlighted the opportunities for considering healthcare decision-making within a quantified probabilistic framework (O'Connor & Sox, 1991). It was postulated that medical diagnoses may be mathematically calculated by deriving the conditional probability of each differential diagnosis to determine which is the most likely (Ledley & Lusted, 1959). The axiom of this reasoning approach is founded upon the mathematical concept of knowledge existing as a probability continuum reflecting one’s degree of belief (Laplace, 1902). A patient’s diagnosis, prognosis or likely treatment response is therefore considered with regard to its quantified degree of certainty given the available information (Lusted, 1975). Within this framework, new knowledge quantifiably updates existing beliefs proportional to the strength of each piece

of information in a mathematical relationship first described by Thomas Bayes in the 18th century (Bayes, 1763). This statistical approach to reasoning, referred to as 'Bayesian', is applied prolifically in modern society across numerous fields that involve complex decision-making, including finance, insurance, physics, astronomy, meteorology, computer science, education, law, security, and notably, healthcare (McGrayne, 2011).

In recent decades, the traditional sufficiency of subjective judgment, biologic plausibility, expert opinion, practice customs and unmethodical experience to inform decision-making in healthcare has become increasingly challenged (Eddy, 2011). Originally termed 'scientific medicine' and later revised to 'evidence-based medicine' due to unfavourable reaction from some in the medical community (Guyatt, 2008), healthcare decision-making is increasingly founded upon an enlightened awareness of the nature, quality and hierarchy of evidence (Guyatt et al., 1992). While the value of statistical literacy for effective healthcare practice has been long-recognized (Anon, 1937; Ledley & Lusted, 1959), the evidence-based movement has conscientiously proliferated the awareness and employment of statistics, such as disease and outcome prevalence (Laupacis et al., 1994; Richardson et al., 1999), diagnostic test accuracy (Jaeschke, Guyatt, Sackett, & Evidence-Based Medicine Working Group, 1994) and the quantification of treatment effect (Guyatt, Sackett, Cook, & Evidence Based Medicine Working Group, 1994), among many others.

In contemporary healthcare practice, the application of statistics facilitates the transformation of data into evidence-based diagnostic, prognostic and treatment decisions (Horvitz, 2010). Such an approach to decision-making has been labelled 'probabilistic reasoning' (Doust, 2009; Richardson & Wilson, 2008), and is a system two analytic strategy within the dominant theory of human decision-making known as the Dual Process Theory (Croskerry, 2009; Eva, 2005).

3.3 The clinical-statistical controversy

The conscientious application of statistics and probabilities to inform decisions in healthcare has been associated with controversy since its inception in the early 20th century (Grove & Lloyd, 2006). The so called 'clinical-statistical controversy' (Dana & Thomas, 2006) concerns the relative merits, accuracy and risks associated with clinician judgement and statistical prediction. The debate gained notoriety following the publication of Paul Meehl's self-proclaimed 'disturbing little book' in 1954, which after examining the limited available evidence at that time within clinical psychology, reached the conclusion that all else being equal, statistical predictions were generally more accurate than unassisted clinician judgement in predicting human behaviour (P. E Meehl, 1954; P. E. Meehl, 1986). Of the 20 studies considered in Meehl's review, only one study found clinician judgement to be more accurate compared to statistical prediction.

Meehl's original conclusion has been recurrently supported by more recent evidence. Several reviews in the fields of psychology and medicine have been conducted evaluating the relative accuracy of judgements concerning human behaviour made by either statistical prediction or clinician judgement using the same information (Ægisdóttir et al., 2006; Bishop & Trout, 2002; Dawes, Faust, & Meehl, 1989; Grove & Meehl, 1996; Grove, Zald, Lebow, Snitz, & Nelson, 2000). All of these reviews have identified that the body of evidence demonstrates that statistical predictions are generally more accurate than unassisted clinician judgements in these tasks. The consistency of this finding has led some to call it the 'Golden Rule of Predictive Modelling' – "when based on the same evidence, the predictions of SPRs (statistical prediction rules) are at least as reliable, and are typically more reliable, than the predictions of human experts" (Bishop & Trout, 2002, p. S198).

The meta-analysis of Ægisdóttir et al. (2006) examined the comparative accuracy of predictions made within the field of counselling psychology by mental health clinicians and those made using statistical procedures. Sixty-seven studies were included in this review, of which 49 evaluated a statistical procedure that had undergone validation. The meta-analysis of studies in which the statistical procedures in question had been cross-validated identified an overall small effect (Cohen's d = 0.14, 95%CI 0.12 – 0.17) favouring statistical prediction. It was further identified that the degree of predictive advantage of statistical methods over clinician judgement was significantly influenced by numerous factors including: the type of prediction

task; clinician familiarity with the setting (contrary to the hypothesis, there was more advantage for statistical prediction if the clinician was in their familiar setting); and the type of statistical formula (greatest advantage for statistical prediction with linear formulas). In contrast, providing the clinician with information concerning the base rate of the dependent outcome, and making the statistical formula available to the clinician did not influence the degree of observed advantage of statistical prediction over clinician judgement. The stratified analysis did however demonstrate that statistical prediction outperformed 'expert' clinician judgement to a lesser degree than their more novice counterparts. Indeed, when considering just the seven included studies that compared expert clinician judgement to statistical prediction, the confidence interval of the effect size (d = 0.05, 95%CI -0.03 to 0.14) suggested no difference in their respective predictive accuracy.

The superior predictive accuracy of statistical methods over unassisted human judgement does not appear to be unique to the clinical psychology context. A meta-analysis conducted by Grove et al. (2000) examined 136 studies which evaluated the relative predictive performance of human judges and 'mechanical procedures' (mathematical formulas or actuarial tables) in predicting a range of human behaviours, psychological or medical diagnoses, or prognoses. The types of predicted outcomes in the included studies were remarkably diverse and included: the prediction of magazine advertising sales, the diagnosis of appendicitis, success in military training, length of hospitalisation, probation success, suicide attempt, diagnosis of myocardial infarction, academic performance and surgical outcomes, to name just a few.

In 63 (46%) studies, the findings identified greater accuracy using mechanical procedures. In a further 65 studies (48%), there was negligible difference in the predictive performance between each method. In just eight studies (6%), the findings found greater accuracy in the prediction made by human experts compared to that made by mechanical prediction. An analysis of the eight studies favouring clinician judgement in this review failed to identify any commonalities or characteristics that may account for their conflicting observations (Grove & Meehl, 1996).

A criticism of many studies that have sought to compare the predictive accuracy of clinician judgement and statistical prediction concerns the creation of an artificial environment in which the clinician is only privy to the same quantified information that is used in the statistical prediction (Holt, 1970). Consequently, additional and qualitative information that may typically be applied in the clinical context to form judgements are not used in these studies due to their absence in the statistical prediction model. This hypothesis was investigated within the meta-analysis of Ægisdóttir et al. (2006) and, as expected, the amount of information available to the clinician was identified to significantly influence the relative predictive accuracy of each method. However, the direction of the effect was the reverse to what had been hypothesised. In the 15 studies in which the mental health clinician(s) had more information than was used in the statistical prediction model, statistical predictions were found to have larger relative predictive superiority over clinician judgement (d = 0.13, 95%CI 0.09 – 0.16) compared to those studies in which the information available to clinicians and the

statistical procedure was the same (d = 0.06, 95%CI 0.01 – 0.11). Thus, accessing more information to that which is considered within the statistical prediction tool may not necessarily improve clinician predictive performance. This phenomenon has been referred to as the 'illusion of knowledge', and is a cognitive bias in human judgement whereby supplementary information increases the confidence of a prediction, but has a deleterious effect on its accuracy (C. C. Hall, Ariss, & Todorov, 2007).

Cognitive biases in human judgement have been postulated as a possible explanation as to why statistical methods have been demonstrated to have greater predictive accuracy compared to unassisted human judgement across a variety of different tasks (Grove et al., 2000). There are many different types of cognitive biases that have been postulated to contribute to diagnostic error in medical practice (Croskerry, 2003) and these are summarised in Table 3.1.

Table 3.1 Types of cognitive biases in human judgement as they relate to the process of diagnosis (adapted from Croskerry (2003))

Aggregate bias	The belief that aggregate data does not apply to individual patients, or the false belief that one's patients are atypical
Anchoring	Locking onto salient features early in the diagnostic process and failing to adjust the initial impression in light of additional information
Ascertainment bias	When thinking is shaped by prior expectation
Availability	Erroneously believing something is more likely based on the readiness to which it comes to mind
Base-rate neglect	Failure to include the true base-rate prevalence of a disease or outcome into decision-making
Commission bias	Tendency to invoke action due to the erroneous belief that only action may prevent harm
Confirmation bias	Tendency to seek confirmatory evidence and ignore negating evidence
Diagnosis momentum	Whereby an initial diagnostic hypothesis becomes stuck to a patient through intermediaries
Feedback sanction	Error in judgement is not identified due to lack of timely feedback processes
Framing effect	Judgements are influenced by the context in which they are framed
Fundamental attribution error	Tendency to blame patients for their illnesses and not examine the factors that may have been responsible
Gambler's fallacy	False belief that an otherwise independent event is somehow influenced by previous observations
Gender bias	False belief that gender is a determining factor in a given diagnosis contrary to the evidence
Hindsight bias	Knowledge of outcome influences perception of past events and interferes with learning
Multiple alternatives bias	Too many options on a differential diagnosis may create excessive uncertainty
Omission bias	The tendency to not act based on the principle of non-maleficence
Order effects	Tendency to predominately remember the beginning and the end of a series of information
Outcome bias	Tendency to falsely opt for diagnoses that have better outcomes, based on hope.
Overconfidence bias	Believing one knows more than what is actually known, and placing unfounded faith in opinion
Playing the odds	Tendency to opt for a benign diagnosis as it is more common than a more serious presentation

Posterior probability error	Opposite of Gambler's fallacy, and relates to the false belief that a sequence of otherwise independent events will continue based on prior observations
Premature closure	Accepting a diagnosis before it has been verified
Psych-out error	Tendency for patients with psychiatric disorders to be more vulnerable to misdiagnosis
Representativeness restraint	Seeking only prototypical signs and symptoms of disease, and therefore more likely to miss atypical variants of the presentation
Search satisfying	Failure to address secondary problems once the primary problem has been identified
Sutton's slip	Failure to consider alternatives other than the most obvious
Sunk costs	Unwillingness to depart from a previously favoured diagnosis due to the investment placed in it
Triage cueing	Biases incurred as a result of triaging processes that affect subsequent management
Unpacking principle	Failure to gain all pertinent information to inform decision-making
Vertical line failure	Failure to think laterally and consider other alternatives
Visceral bias	Errors resulting from the influence of affect, both positive and negative
Yin-Yang out	Withdrawing further diagnostic effort from the belief that nothing further can be done

Human cognitive biases have been identified as a significant contributor to medical diagnostic error. Graber, Franklin, and Gordon (2005) critically evaluated 100 cases of delayed, missed or incorrect diagnoses within the discipline of Internal Medicine across five tertiary level medical centres. The causes of each error were classified by a three-person team as being due to non-fault, system-related and/or cognitive errors. In almost three-quarters of the cases evaluated, cognitive errors were identified to be a contributing factor. Cognitive errors were the *sole* attributable cause of diagnostic error in 28% of cases. Notably, most of the cognitive errors were not related to

knowledge, but rather to the processing and synthesis of the available information.

Such errors in clinical problem solving are thought, at least in part, to be a consequence of limitations in the human cognitive capacity (Elstein & Schwarz, 2002). Simon (1990) described this as the principle of 'bounded rationality' – decision-making is limited by human behaviour being only partly rational, thereby causing limitations in information processing and complex problem solving, thus requiring the use of suboptimal approximation methods and heuristics. The need for fast and efficient decision-making 'short-cuts' and cognitive biases are believed, at least in part, to have arisen adaptively through our evolutionary history as a result of their intrinsic advantages for survival (Johnson, Blumstein, Fowler, & Haselton, 2013). Such adaptive cognitive processes may however be suboptimal in many modern decision-making contexts and their identification is frequently cited as central to reducing errors in medical practice (Croskerry, 2009; Ely, Graber, & Croskerry, 2011; Graber et al., 2005; Graber, Gordon, & Franklin, 2002; E. P. Hicks & Kluemper, 2011).

The 'clinical-statistical controversy' may however represent a false dichotomy, in a manner similar to that of evidence-based medicine and clinical expertise (Parker, 2005). The use of statistical procedures to inform decisions is reliant upon a skilled individual's ability to judge the appropriateness of its application, awareness of its limitations and assumptions, and the accurate interpretation of its results (Dawes et al.,

1989; Swets, Dawes, & Monahan, 2000a, 2000b). P. E Meehl (1954) first highlighted the critical role of the skilled individual in the application of statistical prediction models in what is known as the ‘broken leg countervailing’ – a prediction model that may normally perform well under usual circumstances (e.g. a model that predicts someone's attendance at the movies given the day of the week) will require human adjustment in the light of additional information not accounted for in the model that will influence the predicted outcome (e.g. in the rare case that someone has broken their leg they are much less likely to attend the movies) (Grove & Meehl, 1996).

The ‘human adjustment’ of statistical prediction models has been used very successfully for some time in the field of meteorology. Most modern weather forecasts are quite accurate (Met Office, 2014) and principally informed by statistical prediction models in a process known as Numerical Weather Prediction (Zhang & Pu, 2010). The advent of Numerical Weather Prediction is credited as one of the greatest scientific achievements of the 20th century and has been responsible for large societal and economic benefits (McCaslin, Nakazawa, Swinbank, & Toth, 2010; Thorpe & Petersen, 2006). Importantly however, weather forecasts are not the crude output from meteorological statistical prediction models. Instead, predictions are updated and adjusted by local meteorologists based on their knowledge of the risks, biases and performance of each model, and their knowledge of pertinent information not adequately considered within the statistical models (Novak et al., 2011; Swets et al., 2000b). The 'human' addition to the weather forecast has been demonstrated to produce more accurate forecasts than unadjusted

numerical weather predictions in many circumstances (Carter & Polger, 1986; P. J. McCarthy, Ball, & Purcell, 2007; Novak et al., 2011; Roebber & Bosart, 1996).

Rather than being a slave to a mathematical formula, it is suggested that clinicians using statistical prediction models integrate the objective data produced from such tools with all other existing information to facilitate their decision-making (Swets et al., 2000a). That is, statistical predictions do not *form* a clinical decision, but instead, *inform* a clinical decision. Several different types of statistical prediction tools have been developed ranging from simple actuarial tables to more computationally complex approaches, such as artificial neural networks (Baxt, 1995; P. E Meehl, 1954). Irrespective of the type, all statistical prediction tools use statistical analysis of prior cases with known outcomes to identify the quantified relationship between predictor variables and a particular diagnosis or outcome, such that they may be used to make future predictions (Swets et al., 2000b). The remainder of this chapter will focus specifically on a type of statistical prediction tool most commonly referred to as a 'clinical prediction rule' (CPR).

3.4 Clinical prediction rules

A clinical prediction rule (CPR) has been defined as a "a clinical tool that quantifies the individual contributions that various components of the history, physical examination and basic laboratory results make towards the diagnosis, prognosis, or likely response to treatment in an individual patient"

(McGinn et al., 2008, p. 493). Common synonyms include ‘clinical prediction guides’ (McGinn et al., 2008; US National Library of Medicine, 2009), ‘clinical prediction tools’ (Randolph et al., 1998), ‘clinical decision rules’ (Osmond et al., 2010), ‘clinical decision guides’ (Schneider et al., 2014) and ‘clinical decision tools (Thiruganasambandamoorthy et al., 2014). Differences in the nomenclature used to describe CPRs may reflect clinician preferences. The results of a mail survey of 1,769 Emergency Physicians across five countries identified that the word ‘rule’ was the preferred term for less than 10% of clinicians (Graham et al., 2001). The term ‘guideline’ was preferred by the majority in the US (74%), UK (81%) and Canada (72%), whilst the term ‘criteria’ was the most popular in Spain (63%). Such findings may reflect a preference for less authoritarian words to the term ‘rule’, and may also provide insight regarding clinician attitudes toward the perceived function of such tools in the clinical environment.

Currently, there is no Medical Subject Heading (MeSH) specific to CPRs, and their identification in the medical literature is consequently complex (Geersing et al., 2012; Holland, Wilczynski, & Haynes, 2005; Ingui & Rogers, 2001; Keogh et al., 2011; Wong, Wilczynski, Haynes, Ramkissoonsingh, & Hedges Team, 2003). A search for common synonyms of CPRs (listed above) in the titles or abstracts of studies in Medline, Embase, Amed and PsychInfo from their inception to December 2014, identifies 1,920 unique records. When graphed across time, it is apparent that the CPR literature is growing exponentially (see Figure 3.1). To facilitate clinician awareness and access to CPRs, an international web-based registry of CPRs relevant to

primary care has been developed in collaboration with the Cochrane Primary Health Care Field (Keogh et al., 2011; The Cochrane Collaboration, 2012). The registry includes 434 CPRs that most commonly relate to the assessment and management of cardiovascular, respiratory and musculoskeletal disease (Keogh et al., 2014).

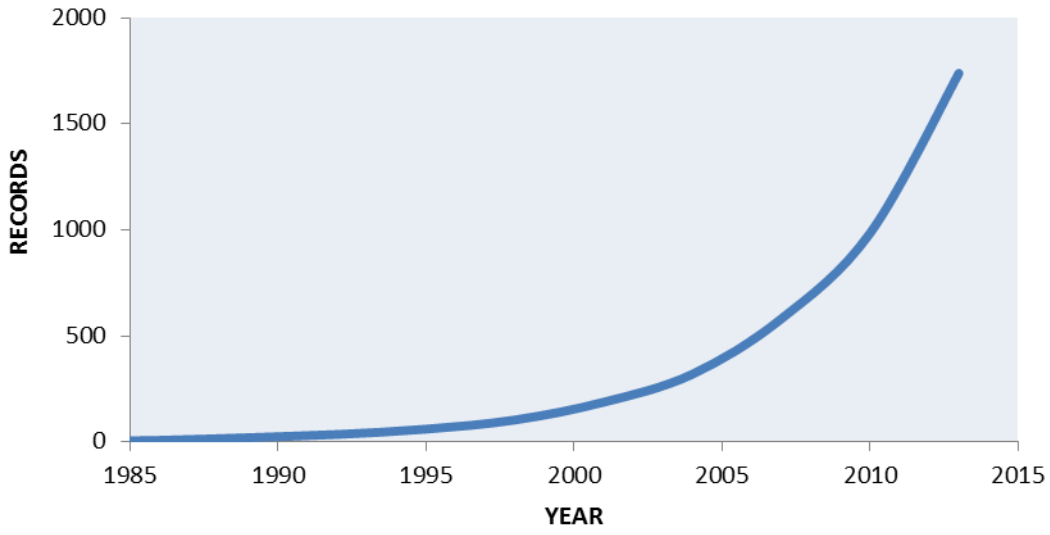


Figure 3.1 **Growth of the clinical prediction rule medical literature. Unique records containing synonyms of 'clinical prediction rule' in the title or abstract across Medline, Embase, PsychInfo, and Amed**

CPRs may be conceptualised as a method of incorporating research evidence into clinical decision-making (Beattie & Nelson, 2006). They are clinical tools comprised of the most *parsimonious* set of variables that have been empirically identified to predict a meaningful diagnosis or outcome (Childs & Cleland, 2006). Variables are commonly components of the history,

physical examination and/or other tests or investigations that may be reliably collected within a standard clinical encounter (Laupacis, Sekar, & Stiell, 1997). Some forms of CPRs enable the calculation of the probability of a given outcome or diagnosis, whilst others function to directly inform a specific course of action (Reilly & Evans, 2006). It is generally considered that CPRs may be of greatest utility when developed to assist in complex clinical decisions (McGinn et al., 2000). Accordingly, CPRs that function to assist in the sub-classification of patients with LBP into more homogenous subgroups are considered to have significant potential to benefit clinical practice (Fritz, 2009).

Three major types of CPRs have been identified in the medical literature: diagnostic, prognostic and prescriptive (C. Cook, 2008). Diagnostic CPRs function to inform clinical decisions regarding an individual patient's diagnosis or present classification/status. An example of a diagnostic CPR is the Ottawa Knee Rule (Stiell, Greenberg, et al., 1995). This five item tool is designed to help inform decisions regarding which patients presenting to an Emergency Department following an acute knee injury require an x-ray. A patient's status on this CPR is determined by considering the presence or absence of five clinical variables (Table 3.2). In the absence of all five clinical variables, the likelihood of a knee fracture is remote (Bachmann, Haberzeth, Steurer, & ter Riet, 2004) and consequently an x-ray of the knee is unlikely to yield valuable clinical information.

Table 3.2 Example of a diagnostic clinical prediction rule: The Ottawa Knee Rule (Stiell, Greenberg, et al., 1995)

1. Age ≥ 55 years
2. Tenderness at head of fibula
3. Isolated tenderness of patella
4. Inability to flex knee to 90°
5. Inability to bear weight (twice onto each limb regardless of limping), both immediately and in the Emergency Department

Prognostic CPRs differ to their diagnostic counterparts with respect to their dependence upon the dimension of time. Prognostic CPRs function to inform clinical judgements regarding future outcomes or events, such as an individual's pain severity or likelihood of returning to work in six months' time. An example of a prognostic CPR is the $ABCD^2$ score, which is a five item tool that functions to identify the risk of a patient sustaining a stroke within the days following a transient ischaemic attack (Johnston et al., 2007). A patient's risk classification on the $ABCD^2$ is determined by calculating a patient's score using knowledge of their age, blood pressure, presence or absence of motor deficit or speech impairment, duration of symptoms, and the presence or absence of diabetes. The risk of stroke within seven days of a transient ischaemic attack has been demonstrated to reliably correlate with a patient's risk classification on this CPR (Galvin, Geraghty, Motterlini, Dimitrov, & Fahey, 2011), and clinical practice guidelines have used $ABCD^2$ scores to inform patient-specific management recommendations (National Institute for Health and Care Excellence, 2008; National Stoke Foundation, 2010; Stroke Foundation of New Zealand, 2008).

Table 3.3 Example of a prognostic clinical prediction rule: The ABCD^2 score (Johnston et al., 2007)

1.	**A**ge	≥ 60 years	1 point
2.	**B**lood pressure	systolic > 140 mmHg or diastolic > 90mmHg	1 point
3.	**C**linical features	unilateral weakness	2 points
		speech impairment without unilateral weakness	1 point
4.	**D**uration of symptoms	10 – 59 minutes	1 point
		≥ 60 minutes	2 points
5.	**D**iabetes	present	1 point

Prescriptive CPRs are the third major type of these tools and function to sub-classify patient populations by matching patients to treatments based on their predicted responsiveness to that treatment, independent of a diagnostic classification (Foster et al., 2013). As such, prescriptive CPRs inform clinical decisions regarding treatment selection (C. Cook, 2008), and can be conceptualised as a special form of prognostic CPR that specifically relate to treatment effects. The treatment effect is the difference in outcome that is achieved by one intervention in comparison to that achieved by an alternative or control intervention (Kamper et al., 2010). Prescriptive CPRs are thus comprised of treatment effect modifiers (also known as effect moderators) – these are the baseline variables that differentiate patient subgroups which experience differing magnitudes of treatment effect (Kraemer, Frank, & Kupfer, 2006). Such variables are subsequently distinct from prognostic variables, which predict outcomes independent of treatment (J. C. Hill & Fritz, 2011).

A patient's status on a treatment effect modifier predicts the relative benefit they will likely achieve from one intervention compared to another. Figure 3.2 illustrates this relationship. Hancock, Herbert, and Maher (2009) provide the following helpful example – a patient's type of stoke (ischaemic vs haemorrhagic) is a treatment effect modifier for anticoagulant therapy: that is, if the patient has an ischaemic stroke they are likely to experience benefit, however if the stroke is haemorrhagic the therapy may actually worsen their outcome as compared to the benefit or otherwise achieved from an alternative intervention. Treatment effect modifiers are identified in randomised clinical trials by exploring interaction effects between candidate baseline variables and treatment groups (Hancock, Herbert, et al., 2009; Sun, Briel, Walter, & Guyatt, 2010). The sample sizes required for such trials are however very large. To adequately power a study to detect an interaction effect, the sample size needs to be approximately four times that required to detect an overall treatment effect of the same magnitude (Brookes et al., 2004).

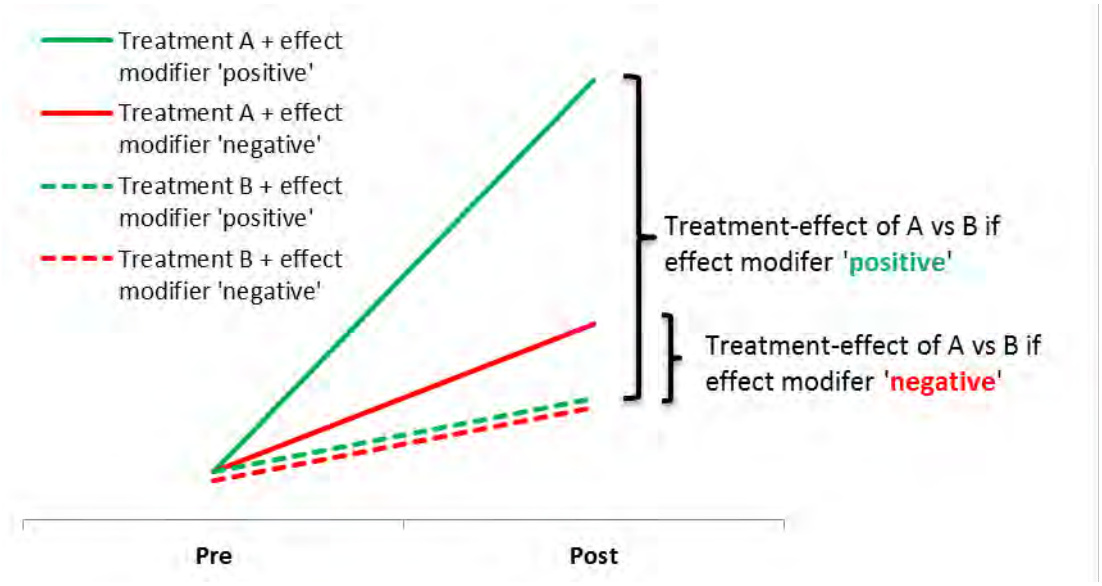


Figure 3.2 Illustration of a treatment effect that is modified by a patient's status on a baseline variable

3.5 Development of clinical prediction rules

The development of a CPR occurs across three main stages: derivation, validation and impact analysis (Figure 3.3) (Childs & Cleland, 2006; McGinn et al., 2000; McGinn et al., 2008). Each stage functions to develop and investigate a specific aspect of a CPR and has crucial implications upon its ability to be applied in clinical practice, as will be discussed in section 3.7. The following subsections describe the processes involved in each of the main stages of a CPR's development.

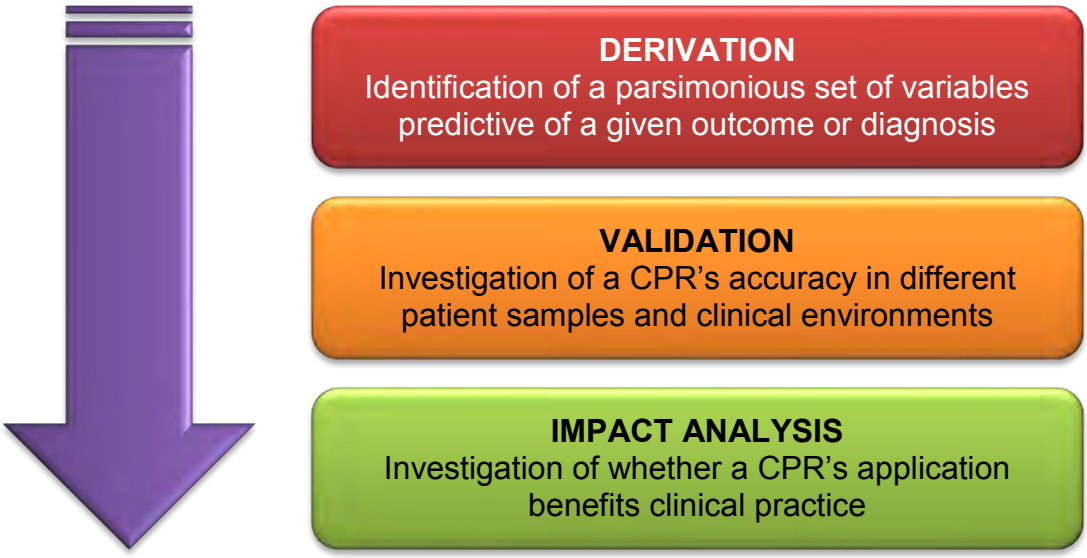


Figure 3.3 Stages in the development of a clinical prediction rule (adapted from Childs and Cleland (2006))

3.5.1 Derivation

The first step in the development of a CPR is derivation. This process commences with the identification of a meaningful problem for which the development of a CPR may be perceived as clinically useful. Considerations that help inform the need for a CPR include the complexity of clinical decision-making, the accuracy of unassisted clinician judgement, clinician attitudes, variations in practices and the hypothesised potential for a tool to beneficially impact practice by improving patient outcomes or improving resource efficiencies (Fritz, 2009; Stiell & Wells, 1999).

The study design required to derive a CPR is dependent upon the type of CPR under development. Diagnostic CPRs are derived in cross-sectional studies; prognostic CPRs are derived in longitudinal cohort studies; and prescriptive CPRs require randomised controlled trials (Hancock, Herbert, et al., 2009; J. C. Hill & Fritz, 2011). In all instances, a meaningful, valid and clearly defined dependent outcome that is able to be reliably measured requires selection (Stiell & Wells, 1999). A small number of candidate predictor variables also need to be selected *a priori* and considered within the context of their hypothesised predictive performance, validity and reliability, and their practicality and availability within the clinical environment (C. Cook et al., 2010; Lubetzky-Vilnai, Ciol, & McCoy, 2014; Seel, Steyerberg, Malec, Sherer, & Macciocchi, 2012). Clinical judgement, literature reviews, focus groups, and questionnaires have been used to select candidate predictor variables in some CPR derivation studies (Dionne et al., 2005; Hewitt, Hush, Martin, Herbert, & Latimer, 2007; Heymans et al., 2009; Heymans et al., 2007).

The patient population sampled in CPR derivation studies needs to represent the spectrum of patients to which the tool is likely to be applied (Stiell & Wells, 1999). Generally, large sample sizes are required to satisfy the assumptions of the statistical techniques that are used and also to generate greater precision of the findings (Childs & Cleland, 2006). Larger sample sizes are particularly required when investigating an outcome with a very low prevalence (e.g. cancer in patients with LBP), testing large numbers of

candidate predictors, and when investigating treatment effect modifiers (Babyak, 2004; Brookes et al., 2004).

Once data collection is complete, statistical analysis is used to identify the candidate variables that have a significant predictive relationship with the dependent outcome. There are several different techniques that have been used to derive CPRs in the medical literature. Table 3.4, adapted from Grobman and Stamilio (2006) and Adams and Leveson (2012), provides an overview of these techniques and their relative advantages and disadvantages.

Table 3.4 Techniques used to develop clinical prediction rules (adapted from Grobman and Stamilio (2006) and Adams and Leveson (2012))

TECHNIQUE	ADVANTAGES	DISADVANTAGES
Univariate analysis	Simple to develop. Easy to use.	Predictors may not be independent. Weightings are arbitrary. Less accurate.
Multivariable analysis	Improved accuracy.	Slightly more complicated to develop.
Nomograms	Improved accuracy. Easy to use.	More complicated to develop.
Classification and regression trees (recursive partitioning)	Easy to use. Enables development of rules that are optimised for sensitivity or specificity.	Can often be less accurate than other techniques. Does not work well for continuous variables. Prone to overfitting.
Artificial neural network	Improved accuracy over time with new data. Identifies complex non-linear relationships and interactions.	More complicated to develop. Prone to overfitting. Hard to apply in most clinical settings.

Univariate analysis, whereby the relationships between each predictor variable and the dependent outcome are examined separately, is the simplest technique but has several limitations. Most notably, it does not account for the relationship amongst candidate predictor variables. Multivariable analysis overcomes this limitation by examining the independent relationship of each predictor variable with the target outcome, and it also enables the assignment of variable weightings based on the interpretation of the regression coefficients (Laupacis et al., 1997). Various forms of multivariable analysis have been commonly used to derive CPRs (Bouwmeester et al., 2012) and in some cases automated methods of variable selection (e.g. forward stepwise, backward deletion, best subset) are applied. However, given the increased chance of identifying spurious associations using automated procedures, these approaches may not be well suited for CPR development and may best be reserved for exploratory analysis (Babyak, 2004; Katz, 2003). Multivariable models are generally well suited to construct nomograms, which are graphical calculating tools that facilitate the application of otherwise complicated mathematical equations (Grobman & Stamilio, 2006).

Classification and regression trees (which are a type of recursive partitioning) are another approach used to derive CPRs. This analysis uses non-parametric statistical procedures to identify mutually exclusive and exhaustive subgroups based on the variables that predict the dependent outcome (Lemon, Roy, Clark, Friedmann, & Rakowski, 2003). Recursive

partitioning accounts for interactions between predictor variables (E. F. Cook & Goldman, 1984; Dionne et al., 1997) and is subsequently better suited for deriving CPRs from datasets with interacting variables than logistic regression (Katz, 2006). This approach is also considered to be well suited in instances where a CPR requires optimisation of either the sensitivity or specificity (Stiell & Wells, 1999).

Artificial neural networks require advanced computational resources and are another approach used to develop CPRs. Artificial neural networks are inherently statistically more flexible than regression approaches and, all else being equal, provide models that better fit the study data (Kattan, 2002). However, as a consequence they are also more vulnerable to overfitting, thus potentially reducing the likelihood that these approaches will perform well outside of the derivation study data (Tu, 1996).

To illustrate the development of a CPR, the Ottawa Knee Rule (Table 3.2, p.97) will be used as an example (Stiell, Greenberg, et al., 1995). A need for a tool to help decide which patients require an x-ray was based on the finding that whilst almost three-quarters of patients presenting with acute knee injury to an emergency department were referred for radiology, only 5% were identified to have a fracture (Stiell, Wells, McDowell, et al., 1995). This contributes to increased costs of care, increased waiting times and unnecessary radiation exposure. It was also identified that experienced clinicians believed that the probability of a fracture was less than 10% in the majority of patients sent for radiology (Stiell, Wells, McDowell, et al., 1995).

Consequently, a prospective study was conducted involving 1,047 adult patients with acute knee injuries presenting to one of two university hospital emergency departments in Ottawa, Canada. The dependent outcome was any fracture of the knee seen on plain x-ray, and was determined blinded to knowledge of the candidate predictor variables. For ethical reasons, patients thought not to require a knee x-ray were not sent for radiology but were followed-up via a telephone questionnaire with the aim of detecting any missed fractures. Twenty-three candidate predictor variables were selected based on clinician judgement, literature review and pilot study data. Explicit definitions of each variable were provided to clinicians in a handout. Following data collection, recursive partitioning was used to derive the CPR. The tool was developed to optimise sensitivity, given that a missed fracture would be of greater consequence than an unnecessary x-ray. Many different models were identified to fit the data and the research team decided to select the model that gave the greatest specificity and used the fewest number of variables, whilst maintaining 100% sensitivity (Table 3.2, p.97). The accuracy of the Ottawa Knee Rule in the derivation study was: sensitivity = 100% (95%CI 95% - 100%), and specificity = 54% (95%CI 51% - 57%).

3.5.2 Validation

A CPR models the study dataset from which it was derived (Beattie & Nelson, 2006). Consequently, it may not always perform well when applied outside of this original context (Justice, Covinsky, & Berlin, 1999). Validation is the second stage of a CPR's development and functions to examine the

internal validity and generalizability of the derived tool in new patient populations and clinical environments (McGinn et al., 2008). Validation of a CPR is therefore not something achievable within a single study, but rather an attribute that arises across multiple investigations (Hancock, Herbert, et al., 2009).

Methodological issues within a derivation study that challenge the internal validity of a CPR will have consequences upon the tool's ability to perform well in other studies (C. Cook, 2008). However, there are at least three reasons why even a robustly derived CPR may not necessarily perform well outside of the original study (McGinn et al., 2000). These are:

1. Chance associations. It is possible that some statistically significant relationships identified in the derivation study are purely due to chance. Consequently, it is unlikely that such associations will hold true in new datasets, thus reducing the predictive performance of a CPR.
2. Differences related to the patient population or clinical environment. It is possible that some of the predictive relationships identified in the derivation study are unique to the patient sample or clinician group under investigation. As such, derivation study findings may not generalise to other patient and clinician populations.
3. Differences related to the implementation of a CPR. Inconsistencies may arise with regard to the operational definitions of predictor and dependent variables, as well as the accurate application and

interpretation of the rule. These will influence a CPR's predictive performance.

Statistical validation (e.g. split samples, bootstrapping) will only account for the first of these threats (McGinn et al., 2000). As such, prospective studies involving different patients, clinicians and clinical settings are required to validate a CPR. 'Narrow validation' refers to the process by which a CPR is tested for its ability to replicate its predictive performance in similar patients and settings to the original derivation study (Kamper et al., 2010; Keogh et al., 2014; McGinn et al., 2000). The findings of such studies give insight into the variability of the predictive accuracy of a CPR in a specific patient population (Kent, Keating, & Leboeuf-Yde, 2010). 'Broad validation', by contrast, examines the generalizability of a CPR to different settings and patient populations unlike those in used in the derivation study (Kamper et al., 2010; Keogh et al., 2014; McGinn et al., 2000).

Toll, Janssen, Vergouwe, and Moons (2008) further delineate between the temporal, geographic and domain validation of a CPR. Temporal validation refers to the replication of a CPR's performance over time, with little change to the patient population sampled or other elements of the clinical setting. Geographic validation refers to the investigation of a CPR's performance in similar patient populations, but in different clinical environments. Finally, domain validation, which is considered to provide the strongest evidence of generalizability, refers to the assessment of a CPR's performance in different

clinical environments and in different patient populations that differ non-randomly to that of the derivation sample.

Several studies have contributed to the validation of the Ottawa Knee Rule (Bachmann et al., 2004). Ketelslegers et al. (2002) investigated the performance of this tool when applied by clinicians with differing levels of training in an emergency teaching centre in Brussels, Belgium. Medical students and surgical residents were trained in the accurate implementation of the CPR by the research team. The 261 patients recruited in this study were assessed with regard to their status on the Ottawa Knee Rule. Blinded outcome assessment for the presence of a fracture was determined by x-ray (84%) or by telephone or face to face follow-up. The results of this study demonstrated that the Ottawa Knee Rule had a sensitivity of 100% (95%CI 99% - 100%) and a specificity of 32% (95%CI 26% - 38%). No difference in the predictive accuracy of the CPR was identified between medical students and surgical residents, thus providing evidence of generalizability of the tool to different clinician populations of varying experience. The finding of 100% sensitivity of the tool is also consistent with that of the derivation study, and provides further evidence of the predictive performance of the CPR in identifying patients presenting with acute knee injury who are unlikely to benefit from radiological assessment.

3.5.3 Impact analysis

The final stage of a CPR's development is called 'impact analysis' and is the investigation of whether a tool's application in clinical practice results in

meaningful beneficial consequences, such as improved outcomes or resource efficiencies (Childs & Cleland, 2006). This step is important as even a well-validated CPR may not necessarily outperform unassisted clinician judgement. Further, if a CPR is difficult to use or if there are other factors that impede its implementation, it may not necessarily be successfully adopted in clinical practice (McGinn et al., 2000).

The best study design to conduct an impact analysis is a randomised controlled trial, whereby the outcomes produced from the use of a CPR are able to be rigorously evaluated (Toll et al., 2008). Randomisation may be at the level of the patient, clinician or the facility, with the latter helping to minimise potential contamination (Wallace et al., 2011). Before and after designs are often a more feasible approach to assessing the impact of the use of a CPR, however the evidence from such designs is weaker than that produced from a randomised control trial due to the greater potential for bias (Childs & Cleland, 2006; Reilly & Evans, 2006).

In addition to exploring the effectiveness of a CPR on patient outcomes and resource consumption, it may also be useful to investigate changes in clinician practice behaviours, clinicians' acceptance of the tool and patient satisfaction (Beattie & Nelson, 2006; Childs & Cleland, 2006; McGinn et al., 2000; Stiell & Wells, 1999). Clinician acceptance of a CPR may be assessed using the 12-item Ottawa Acceptability of Decision Rules Instrument (Brehaut et al., 2010). Qualitative assessment of the perspectives of study participants may also be advantageous to gain greater understanding regarding the

modifiable aspects of a CPR's implementation that may facilitate its successful clinical application (Wallace et al., 2011).

Continuing the Ottawa Knee Rule example, Stiell et al. (1997) used a before and after non-randomised controlled trial to evaluate the impact of the clinical application of this CPR. Two control and two intervention hospitals were used in this two year study, with the intervention hospitals applying the Ottawa Knee Rule in the last year of the study period. Following the implementation of the CPR in the intervention hospitals, there was a 20.5% absolute reduction in the use of knee x-rays (77.6% to 57.1%). Over the same time period, the use of knee x-rays in the control hospitals decreased by just 1% (76.9% to 75.9%). Those patients not receiving knee radiography spent an average of 33 minutes less time in the emergency department and their overall costs of care were US$103 less. During the period of use of the Ottawa Knee Rule in the intervention hospitals, clinicians overruled the CPR in 6.9% of cases. The main reasons for this related to patient preferences (either wanting or not wanting an x-ray) and clinician judgement. Almost all patients (95.7%) who did not receive a knee x-ray during the period of Ottawa Knee Rule application reported to be satisfied with their episode of care. The sensitivity of the CPR in this study was 100% (95%CI 94% - 100%) and the specificity was 48% (95%CI 45% - 51%).

3.6 Methodological considerations

The development of a CPR, irrespective of its type, requires consideration of a number of methodological standards specific to its stage of development. Such standards are an extension to the various methodological requisites that are specific to the underlying study design. A 23-item quality checklist has been developed to help guide the derivation of prescriptive CPRs (C. Cook et al., 2010), however no universally accepted validated tool exists to help inform the development of all other forms of CPRs at their respective stages of development (Fritz, 2009). Nevertheless, many publications within the medical literature provide commentary regarding the appropriate methodological considerations relevant to the development of CPRs. Table 3.5, Table 3.6, and Table 3.7 provide an overview of the relevant methodological considerations highlighted within five well-cited publications on the derivation, validation and impact assessment of CPRs respectively (Beattie & Nelson, 2006; Childs & Cleland, 2006; Laupacis et al., 1997; McGinn et al., 2000; Stiell & Wells, 1999).

Table 3.5 Methodological considerations common to the derivation of all forms of clinical prediction rules

	Laupacis et al. (1997)	Stiell and Wells (1999)	McGinn et al. (2000)	Beattie and Nelson (2006)	Childs and Cleland (2006)
Prospective design	■	■			
Outcomes defined	■	■	■	■	■
Outcome clinical important	■	■		■	■
Blinded outcome assessment	■	■	■	■	■
All important predictors included	■		■	■	■
Predictive variables clearly defined	■	■	■		
Blinded predictor assessment	■	■	■	■	■
Assessment of the reliability of the predictive variables	■	■			
Important patient characteristics described	■	■		■	
Inclusion criteria explicitly stated		■			
Representative sample		■		■	
Complete follow-up				■	
Study site described	■	■			
Justification for the number of study subjects		■	■	■	■
At least 10 outcome events per independent variable in the rule		■	■	■	■
Important predictors present in a significant proportion of the study population			■		
Mathematical techniques described	■	■	■		
Multivariate analysis	■	■	■	■	■
Results of the rule described	■	■	■	■	■
Clinically sensible/reasonable	■	■	■	■	■
Easy to use	■	■			
Probability of diagnosis or outcome described	■	■		■	■
Course of action described	■	■			
Estimation of potential impact of use		■			

Table 3.6 Methodological considerations common to the validation of all forms of clinical prediction rules

	Laupacis et al. (1997)	Stiell and Wells (1999)	McGinn et al. (2000)	Beattie and Nelson (2006)	Childs and Cleland (2006)
Prospective validation in new patient population	■	■	■	■	■
Different clinical setting to derivation study		■	■	■	■
Different clinicians to derivation study		■	■	■	■
Representative sample		■	■		
The rule is applied accurately		■	■		■
Complete follow-up		■	■		
Blinded outcome assessment			■		
Blinded predictor assessment			■		
Accuracy of the rule in the validation study sample described		■	■		
Justification of the validation study sample size		■			
Assessment of the inter-observer reliability of the rule	■	■			
Assessment of clinicians' perceived ease of use of the rule		■			
Rule is refined when indicated		■			
Estimation of potential impact of use		■			

Table 3.7 Methodological considerations common to the impact assessment of all forms of clinical prediction rules

	Laupacis et al. (1997)	Stiell and Wells (1999)	McGinn et al. (2000)	Beattie and Nelson (2006)	Childs and Cleland (2006)
Effects of clinical use prospectively measured	■	■	■	■	■
Assessment of changes to clinician behaviour/practice		■	■	■	■
Assessment of rules' ability to improve outcomes			■	■	■
Effect upon efficiency assessed		■	■	■	■
Accuracy of the rule is described		■	■		
Clinician acceptance is assessed		■	■		
Patient satisfaction is assessed		■			■

3.7 Readiness for application in clinical practice

The stage of a CPR's development has direct implications upon its readiness to be applied in clinical practice. This is due to the structured process of a CPR's development enabling progressively greater confidence in the tool's accuracy and generalizability (Childs & Cleland, 2006). McGinn et al. (2000) proposed a four level hierarchical framework to determine the degree to which a CPR may be used to confidently inform clinical decisions based upon its stage of development:

1. CPRs that have been derived, but have not yet undergone validation, are not considered within this framework to be ready to be applied in clinical practice. As discussed in section 3.5.2 (p. 106), there are many reasons why even a rigorously derived CPR may not perform well

outside of the original study data. McGinn et al. (2000) suggest that clinicians may wish to consider which variables were and were not identified to have a significant predictive relationship to the target outcome or diagnosis within a derivation study to cautiously inform their clinical practice. Clinicians need to be wary, however, that such relationships may simply reflect chance associations or may be specific to the unique characteristics of the derivation study's patient sample, clinicians or setting.

2. CPRs that have undergone 'narrow validation' (examination of the tool's performance in a population and setting very similar to that in which the CPR was derived) may be cautiously applied with some confidence in their predictive accuracy in the limited instances where a clinician's caseload closely approximates that of the validation and derivation studies. This stage of a CPR's development does not however provide evidence that it may accurately perform outside of this limited context. Additionally, the application of a CPR at this stage of development, even within this limited context, may not necessarily result in improved patient outcomes or other improvements in clinical care.
3. CPRs that have undergone 'broad validation' (examination of the tool's performance in heterogeneous patient populations and different settings to that used in the derivation study) may be applied with some confidence in their predictive accuracy across various clinical settings. This stage of CPR development does not however provide evidence that the use of a CPR will have beneficial clinical consequences.

4. CPRs that have undergone impact analysis may be applied with confidence that their application in clinical practice is likely to result in improved patient outcomes and/or resource efficiencies, while maintaining quality of care and patient satisfaction.

3.8 Barriers to the adoption of clinical prediction rules

Even a well-developed CPR that has been demonstrated to positively influence patient outcomes may not necessarily be adopted in clinical practice. A good example of this concerns the implementation of the Ottawa Knee Rule.

As discussed in section 3.5, the Ottawa Knee Rule (Stiell, Greenberg, et al., 1995) is a validated CPR that has been demonstrated to safely and effectively reduce the need for unnecessary knee radiology in patients presenting to hospital emergency departments with acute knee injury (Bachmann et al., 2004; Ketelslegers et al., 2002; Stiell et al., 1997). This evidence has also informed clinical practice guidelines concerning the appropriate use of knee radiology (American College of Radiology, 2011; Bussieres, Taylor, & Peterson, 2007). However, studies have demonstrated that the implementation of the Ottawa Knee Rule is suboptimal (Beutel et al., 2012; Graham et al., 2001; M. J. O'Sullivan & O'Sullivan, 2006).

Most recently, Beutel et al. (2012) retrospectively reviewed the medical records of 260 patients presenting with acute knee injury to one of three US emergency departments in 2009. By comparing the recorded use of knee radiology to a patient's status on the Ottawa Knee Rule identified by examining baseline clinical findings, the author's concluded that the compliance with the tool was 63%. It is not known, however, if the Ottawa Knee Rule was actually explicitly used to inform decisions regarding knee radiology, thus suggesting that the reported compliance rate may represent the best case scenario. Beutel et al. (2012) also surveyed 47 US emergency department physicians regarding their self-reported use of the Ottawa Knee Rule. Less than a quarter of study participants reported that they used the tool most of the time with another third indicating that they never used it at all. Encouragingly, these usage rates are higher than that reported in a similar clinician population more than a decade earlier. In a survey study involving US emergency physicians conducted between 1997 and 1998, just 9% of US respondents indicated that they used the Ottawa Knee Rule most of the time (Graham et al., 2001). This low usage rate may at least in part be explained by the rule only being published two years earlier, and self-reported awareness of the CPR being just 53% of the respondents.

In contrast to the body of evidence regarding the use of the Ottawa Knee Rule, relatively little is known about the use of CPRs developed for LBP. The best available evidence concerning the use of LBP CPRs is a survey of 535 US physical therapists who use thrust manipulation to treat patients with LBP (Learman, Showalter, & Cook, 2012). In this study, 40% of participants

reported using a five item CPR (Flynn et al., 2002) to inform their clinical decision-making. Male physical therapists, participants with specialist board certification and those attending a greater number of manual therapy courses were identified to be more likely to indicate that they used the CPR. Another much smaller study of US physical therapists who completed a survey before participating in a post-graduate continuing education program, found that just 7% of this cohort were familiar with the same five item CPR for lumbopelvic manipulation (Willett, Johnson, & Jones, 2011). In this study, only 21% of participants had previously used spinal manipulation to treat LBP, which could possibly explain to the relatively low familiarity with the CPR.

The research conducted on the implementation of orthopaedic CPRs in emergency medicine has provided valuable insight regarding the range of barriers to their adoption (Beutel et al., 2012; Brehaut et al., 2006; Brehaut et al., 2005; Eagles, Stiell, Clement, Brehaut, Taljaard, et al., 2008; Graham et al., 2001; Graham et al., 1998; M. J. O'Sullivan & O'Sullivan, 2006; Stiell et al., 2006). Similar to those identified barriers to the use of practice guidelines, outcome measures and evidence-based practice (Abrams et al., 2006; Côté et al., 2009; Jette et al., 2003), identified barriers to the use of CPRs have been demonstrated to be multi-factorial and incorporative of interacting individual and system-level factors.

In the study of Beutel et al. (2012), emergency physicians were asked to self-report barriers to the use of the Ottawa Knee Rule. The most commonly identified barriers predominantly related to meeting patient expectations and

preferences, as well as meeting system requirements for orthopaedic referrals. Other reported barriers indicate physician disagreement with the CPR, including a lack of confidence in some of its variables, concerns regarding its effectiveness, and a perception that the rule is an oversimplification of clinical decision-making. Negative attitudes toward the use of statistical tools such as CPRs to inform healthcare decisions are frequently identified in the medical literature, seemingly irrespective of the specific tool in question or the healthcare discipline investigated (Grove & Meehl, 1996; Liao & Mark, 2003; Plüddemann et al., 2014; Stiell et al., 2006). A deep understanding of the range of barriers, including clinician attitudes, is plausibly important to optimising the implementation of CPRs.

The methods used in many studies investigating barriers to the adoption of CPRs have predominantly involved survey questions with close-ended response formats (Beutel et al., 2012; Brehaut et al., 2006; Eagles, Stiell, Clement, Brehaut, Taljaard, et al., 2008; Graham et al., 2001; Graham et al., 1998; Runyon, Richman, & Kline, 2007; Stiell et al., 2006). It is plausible that qualitative research methods may provide additional insights, including a more in-depth understanding of the range and nature of potential barriers to the adoption of CPRs (Petty, Thomson, & Stew, 2012a). Such methods have been employed in previous research aiming to understand the barriers to the adoption of research evidence within physiotherapy clinical practice (Barnard & Wiles, 2001; Dannapfel, Peolsson, & Nilsen, 2013; Hannes, Staes, Goedhuys, & Aertgeerts, 2009; Harting, Rutten, Rutten, & Kremers, 2009).

A framework of knowledge, attitudes and practices/behaviours (Cabana et al., 1999; Legare, Ratte, Gravel, & Graham, 2008) has been proposed as an appropriate model to consider the potential barriers to the implementation of CPRs (Abboud & Cabana, 2001). This framework has been applied to help identify barriers to the use of other clinical innovations including clinical practice guidelines (Larson, 2004; Pogorzelska & Larson, 2008; Schouten et al., 2007) and clinical protocols (Barlow et al., 2008; Dennison, Mendez-Tellez, Wang, Pronovost, & Needham, 2007; Rubinson, Wu, Haponik, & Diette, 2005). Table 5.1 (p. 179) provides a synthesis of the potential barriers concerning the use of musculoskeletal CPRs from the existing body of literature (Beutel et al., 2012; Brehaut et al., 2006; Brehaut et al., 2005; Eagles, Stiell, Clement, Brehaut, Taljaard, et al., 2008; Graham et al., 2001; Graham et al., 1998; Stiell et al., 2006) using a framework of knowledge, attitudes and practices/behaviours (Cabana et al., 1999; Legare et al., 2008). It is plausible that similar barriers may be identified regarding the adoption of CPRs for LBP in physiotherapy practice, and this is the subject of the research study detailed in Chapter 5 of this thesis. Identifying and targeting such barriers may enable the development of CPRs and accompanying implementation strategies with the greatest potential to positively influence clinical practice.

3.9 Clinicians' priorities and the need for a CPR

Given the time and resources required to develop a CPR, it is important that from the outset of their development that they aim to address the perceived

needs of their intended target ‘consumers’: clinicians (Graham et al., 2006; Plüddemann et al., 2014). Soliciting input from clinicians throughout the development of a CPR has been identified as an important strategic approach to overcoming some of the potential barriers to their eventual clinical implementation (Reilly & Evans, 2006).

There are several methods that have been employed in the medical literature to investigate the need for a CPR. In the case of the Ottawa Knee Rule, the research team established a need for the CPR by investigating the magnitude of the clinical problem (the high frequency of unnecessary knee radiology in patients with acute knee injury), measuring clinicians’ predictions of the outcome (predicted probability of a fracture), and measuring clinicians’ degree of comfort with adopting an alternate practice behaviour (not ordering an x-ray) (Stiell, Wells, McDowell, et al., 1995). In this study, it was identified that although 74% of patients were referred for knee radiology, just 5% were found to have a fracture. In three-quarters of instances, clinicians predicted the probability of a fracture to be ≤ 10%, and would have felt comfortable in not ordering an x-ray in just over half (56%). These findings was subsequently used to justify the derivation of the Ottawa Knee Rule to help inform decisions about which patients may safely avoid the need for knee radiology (Stiell, Greenberg, et al., 1995).

Other studies have used a more direct approach. Perry et al. (2009) surveyed 1,149 emergency physicians across four countries to investigate the need for a CPR that functions to identify which patients with acute

headache and without neurological deficit presenting to an emergency department require further investigations to exclude subarachnoid haemorrhage. In this study, participants were explicitly asked whether they would consider using a highly sensitive and well-validated CPR for this purpose, if such a tool were to be developed. Participants were further asked how sensitive a CPR would need to be in order for them to consider using it for this purpose. The findings of this study indicated that the vast majority (96%) of respondents would consider using a CPR designed to identify patients with headache requiring further investigations. Participants in this study also indicated that such a tool would require near perfect sensitivity to be considered useful (median response = 99%, IQR 98% - 99%). In addition to justifying the development of this CPR, knowledge regarding the required sensitivity of such a tool has practical implications regarding the determination of the required sample size of the derivation study (Buderer, 1996; S. Jones, Carley, & Harrison, 2003).

Another approach used to determine the need for CPRs has been to more broadly investigate clinician priorities for such tools within their practice. In a separate survey of emergency physicians accompanying that administered by Perry et al. (2009) outlined above, Eagles, Stiell, Clement, Brehaut, Kelly, et al. (2008) asked participants to rank the perceived usefulness of 26 hypothetical CPRs. A ‘top 10’ list was subsequently generated by the research team with relatively little variance in the priorities of emergency physicians identified when stratified by the four countries represented in the survey. This was particularly evident when considering the broader functional

themes of the hypothesised CPRs (e.g. paediatric patients requiring further investigations, admissions for potentially serious medical conditions, etc.). Although an open-ended question regarding the need for other CPRs was provided to participants in this survey, only 14% of respondents provided a response. It is plausible that research methods not incorporating the close-ended ranking of a list of hypothesised CPRs may elicit further or differing clinician priorities for such tools (Schuman & Presser, 1979; Schuman & Scott, 1987).

Despite the growing number of CPRs for LBP that have been developed to date, physiotherapists' priorities for such tools have not been established. It is not known if the range of CPRs for LBP that have been derived are likely to be considered useful by practising clinicians, or if there are other clinical problems for which CPRs are likely to be considered helpful that have not yet been identified. Further, it is not known if the modifiable properties of a CPR (e.g. the number and type of variables, scoring mechanism, clinical interface etc.) are likely to be influential to the tool's implementation. The research study detailed in Chapter 6 of this thesis aims to help address these knowledge gaps by investigating the types and characteristics of LBP CPRs that are considered important by practising physiotherapists working within an Australian musculoskeletal context.

3.10 Summary

CPRs represent one branch of an evolving approach to clinical practice that conscientiously incorporates quantified research evidence into clinical decision-making. LBP has been explicitly identified as an ideal target for such tools given the complexity of clinical judgements resulting from the inherent heterogeneity of the condition and also its numerous assessment and management alternatives. It is widely hypothesised that the development of CPRs for LBP has the potential to lead to substantial patient and system-level gains. Prior to the program of research detailed in this thesis, relatively little was known about the range and nature of CPRs that have been already developed for LBP, and whether such tools were sufficiently developed to enable their confident application in clinical practice. This knowledge gap is addressed in the studies detailed in Chapters 4, 8 and 9.

Well-developed CPRs for LBP that have been demonstrated to positively influence clinical outcomes may not necessarily be adopted in clinical practice for a variety of complex and interacting reasons. Recognising these barriers and incorporating such knowledge into the development of a CPR, and the design of associated implementation strategies, are widely considered to be critical to optimising the appropriate use of a CPR in clinical practice. Potential barriers and facilitators to the successful implementation of CPRs for LBP have not been previously investigated, and are explored in the study detailed in Chapter 5 within an Australian physiotherapy context.

Finally, ensuring that CPRs match the perceived needs of clinicians is considered to be essential for the development of tools with the highest potential to meaningfully impact clinical practice. Physiotherapists' priorities for the development of CPRs for LBP have not been previously investigated. Consequently, this is the subject of the research study detailed in Chapter 6.

CHAPTER 4

CLINICAL PREDICTION RULES IN THE PHYSIOTHERAPY MANAGEMENT OF LOW BACK PAIN: A SYSTEMATIC REVIEW

This chapter has been published in a peer-reviewed scientific journal (Appendix 4):

Haskins, R., Rivett, D. A., & Osmotherly, P. G. (2012). Clinical prediction rules in the physiotherapy management of low back pain: a systematic review. *Manual Therapy*, *17*(1), 9-21.

The work presented in this manuscript was completed in collaboration with the co-authors (Appendix 1).

Overview

This is the first of five studies conducted in this program of research. At the time when this study was designed, only one review had been previously published on the topic of CPRs relevant to musculoskeletal physiotherapy practice, although this was limited to CPRs for physical therapy interventions in the derivation phase of development (Beneciuk et al., 2009). Consequently, the range of CPRs that had been developed to assist in the physiotherapy management of LBP was not known, and their readiness for

clinical application at that time had also not been evaluated. This study was therefore designed and implemented to help address this knowledge gap and to form the foundation for the subsequent studies in this research program.

4.1 Abstract

Objective:

To identify, appraise and determine the clinical readiness of diagnostic, prescriptive and prognostic Clinical Prediction Rules (CPRs) in the physiotherapy management of Low Back Pain (LBP).

Data Sources:

MEDLINE, EMBASE, CINAHL, AMED and the Cochrane Database of Systematic Reviews were searched from 1990 to January 2010 using sensitive search strategies for identifying CPR and LBP studies. Citation tracking and hand-searching of relevant journals were used as supplemental strategies.

Study Selection:

Two independent reviewers used a two-phase selection procedure to identify studies that explicitly aimed to develop one or more CPRs involving the physiotherapy management of LBP. Diagnostic, prescriptive and prognostic studies investigating CPRs at any stage of their development, derivation, validation, or impact-analysis, were considered for inclusion using a priori criteria. 7453 unique records were screened with 23 studies composing the final included sample.

Data Extraction:

Two reviewers independently extracted relevant data into evidence tables using a standardised instrument.

Data Synthesis:

Identified studies were qualitatively synthesized. No attempt was made to statistically pool the results of individual studies. The 23 scientifically admissible studies described the development of 25 unique CPRs, including 15 diagnostic, 7 prescriptive and 3 prognostic rules. The majority (65%) of studies described the initial derivation of one or more CPRs. No studies investigating the impact phase of rule development were identified.

Conclusions:

The current body of evidence does not enable confident direct clinical application of any of the identified CPRs. Further validation studies utilizing appropriate research designs and rigorous methodology are required to determine the performance and generalizability of the derived CPRs to other patient populations, clinicians and clinical settings.

4.2 Introduction

A Clinical Prediction Rule (CPR) is “a clinical tool that quantifies the individual contributions that various components of the history, physical examination and basic laboratory results make towards the diagnosis, prognosis, or likely response to treatment in an individual patient” (McGinn et al., 2008). These tools aim to facilitate clinical decision-making in the assessment and treatment of individual patients (Beattie & Nelson, 2006) and are thought to

be of greatest potential when they are developed and utilised for clinical conditions that involve complex clinical decision making.

Low Back Pain (LBP) is a common and costly complaint (Andersson, 1998; Riihimaki, 1996; Walker, 1999) that has been specifically identified as an ideal target for CPRs due to its heterogeneous population and numerous treatment alternatives (Fritz, 2009). Clinical trials (Brennan et al., 2006; Fritz et al., 2003; Long et al., 2004) have highlighted the benefits of LBP classification systems that aim to 'match' interventions according to the particular sub-group of patients. Concordantly, there has been a surge in the number of publications that discuss the development and application of CPRs that are relevant to the assessment and treatment of LBP (Beneciuk et al., 2009; May & Rosedale, 2009; Stanton, Hancock, Maher, & Koes, 2010). However, before a CPR can be confidently incorporated into clinical practice, it must undergo a process of development that investigates the rule's performance, generalizability, and influence upon clinical outcomes and/or resource consumption.

Numerous publications have discussed the common methodological standards that should apply to the development of CPRs (Beattie & Nelson, 2006; Childs & Cleland, 2006; C. Cook, 2008; Laupacis et al., 1997; McGinn et al., 2000; McGinn et al., 2008; Randolph et al., 1998; Stiell & Wells, 1999; Wasson, Sox, Neff, & Goldman, 1985), although the specific criteria often differ between studies. It is, however, commonly accepted that a hierarchical process of rule development is utilised (McGinn et al., 2000), initially

commencing with derivation of the rule, and then progressing to a process of validation and then subsequent investigation of its clinical impact.

CPRs that have been derived, but not yet validated are not considered ready for clinical use (McGinn et al., 2000; McGinn et al., 2008; Reilly & Evans, 2006). Even rigorously derived rules may reflect chance associations between variables and the target condition or outcome, or they may be unique to the studied population or other characteristics of that clinical setting (McGinn et al., 2008). This is reflected in the finding that most CPRs perform less accurately in subsequent studies involving different patients (Toll et al., 2008). Despite these limitations, it has been suggested that derived CPRs may inform clinical practice by providing clinicians with an understanding of some of the most important predictors of a given target condition or outcome (McGinn et al., 2008).

The process of validation investigates a rule's performance and generalizability to other patient populations, clinicians and clinical settings. Importantly, the validation of a CPR cannot be accomplished by a single study, but requires a process involving a series of studies that test the internal and external validity of the rule across a broad range of clinical environments (Hancock, Herbert, et al., 2009). Narrow validation of a CPR involves investigating the performance of the rule in a similar patient population and similar clinical setting to the derivation study. A CPR that has been demonstrated to perform well in such a setting is considered to be

ready for cautious clinical application to patients that are representative of the studied population (McGinn et al., 2000; McGinn et al., 2008).

Confidence in the rule's accuracy improves as it is progressively investigated in various other settings comprising different clinicians and patients with differing prevalence of disease or injury and with differing responsiveness to treatment. CPRs that demonstrate consistent and strong performance in this process of broad validation are considered ready to be applied in clinical practice with confidence in their accuracy (McGinn et al., 2000).

It is not appropriate, however, to assume that the clinical application of a rigorously-validated rule will result in improved clinical care. Impact-analysis is the process of CPR development that involves testing a rule's ability to positively influence clinical outcomes and/or resource consumption, and change clinicians' behaviour (McGinn et al., 2008). Ideally, this involves a direct comparison to usual clinical care or judgement (Toll et al., 2008). Rules that are demonstrated to be highly accurate and perform well across multiple clinical environments may actually be no more accurate, or even worse, than unassisted clinician judgement. Rigorously-validated CPRs that have been demonstrated to produce beneficial clinical consequences via impact-analysis can be confidently incorporated into clinical practice (McGinn et al., 2000; McGinn et al., 2008; Reilly & Evans, 2006).

Before clinicians can consider incorporating the growing number of CPRs into their practice, a determination of their readiness for clinical application is

required. Previous systematic reviews of CPRs relevant to physiotherapy (Beneciuk et al., 2009; May & Rosedale, 2009; Stanton et al., 2010) have focused upon the identification of prescriptive rules that facilitate treatment decision-making by identifying variables that moderate the magnitude of the treatment-effect. These reviews have specifically excluded studies concerning diagnosis and prognosis, thereby preventing a complete assessment of the available CPRs a physiotherapist may consider in their clinical management of LBP. Further, the quality appraisal systems used in these reviews have not been reflective of the consensus of the common methodological standards for CPR development.

As no universally-accepted standardised tool currently exists for the methodological appraisal of studies of CPRs (Fritz, 2009), previous systematic reviews have used a variety of means to evaluate the quality of included studies. Some reviews have utilised standardised tools that were developed to appraise prognostic (Beneciuk et al., 2009) and diagnostic studies (Bachmann et al., 2004; Hess et al., 2008). Criticism in this approach has focused upon recognising that methodological standards for the development of CPRs differ to that of other types of studies (Stanton, Maher, & Hancock, 2009). Other reviews (Dahri & Loewen, 2007; May & Rosedale, 2009; Stanton et al., 2010; Wisnivesky et al., 2005) have developed checklists based upon previously proposed methodological standards. A potential problem with this approach is that the proposed methodological standards differ between texts, leading to the possible inclusion of extraneous criteria or the possible exclusion of important criteria dependent

upon the text(s) selected. For example, although Stiell and Wells (1999) highlight the importance of a representative sample in the derivation phase of a rule's development, this criterion is omitted from other well-cited texts (Laupacis et al., 1997; McGinn et al., 2000).

The aim of the present review was to identify, appraise and determine the clinical readiness of CPRs in the physiotherapy management of LBP.

4.3 Methods

4.3.1 Data sources and searches

A systematic literature search of MEDLINE, EMBASE, CINAHL, AMED and the Cochrane Database of Systematic Reviews from 1990 to January 2010 limited to articles available in English was conducted. A sensitive search strategy for CPRs (Ingui & Rogers, 2001) that has been used in previous systematic reviews (Beneciuk et al., 2009; Dahri & Loewen, 2007; May & Rosedale, 2009) was employed in combination with the search strategy recommended by the Cochrane Back Group (2009) for identifying articles relevant to LBP (Table 4.1). Citation tracking and hand-searching of relevant journals were used as supplemental search strategies.

Table 4.1 Database search strategies

	MEDLINE, EMBASE and the Cochrane Database of Systematic Reviews
1	Validat$.mp. or Predict$.ti. or Rule$.mp.
2	(Predict$ and (Outcome$ or risk$ or model$)).mp.
3	((History or Variable$ or Criteria or Scor$ or Characteristic$ or Finding$ or Factor$) and (Predict$ or Model$ or Decision$ or Identif$ or Prognos$)).mp.
4	Decision$.mp. and ((Model$ or Clinical$).mp. or Logistic Models/)
5	(Prognostic and (History or Variable$ or Criteria or Scor$ or Characteristic$ or Finding$ or Factor$ or Model$)).mp.
6	4 or 1 or 3 or 2 or 5
7	dorsalgia.ti,ab.
8	exp Back Pain/
9	backache.ti,ab.
10	exp Low Back Pain/
11	(lumbar adj pain).ti,ab.
12	coccyx.ti,ab.
13	coccydynia.ti,ab.
14	sciatica.ti,ab.
15	sciatica/
16	spondylosis.ti,ab.
17	lumbago.ti,ab.
18	11 or 7 or 9 or 17 or 12 or 15 or 14 or 8 or 16 or 10 or 13
19	6 and 18
20	("1990" or "1991" or "1992" or "1993" or "1994" or "1995" or "1996" or "1997" or "1998" or "1999" or "2000" or "2001" or "2002" or "2003" or "2004" or "2005" or "2006" or "2007" or "2008" or "2009").yr.
21	19 and 20
	CINAHL, AMED
S22	S13 and S21
S21	S12 and S20
S20	S15 or S16 or S17 or S18 or S19
S19	prognostic and (history or variable* or criteria or scor* or characteristic* or finding* or factor* or model*)
S18	decision* and (model* or clinical* or mh Logistic Models)
S17	(history or variable* or criteria or scor* or characteristic* or finding* or factor*) and (predict* or model* or decision* or identif* or prognos*)
S16	predict* and (outcome* or risk* or model*)
S15	Validat* or ti Predict* or Rule*
S14	S12 and S13
S13	yr 1990 or yr 1991 or yr 1992 or yr 1993 or yr 1994 or yr 1995 or yr 1996 or yr 1997 or yr 1998 or yr 1999 or yr 2000 or yr 2001 or yr 2002 or yr 2003 or yr 2004 or yr 2005 or yr 2006 or yr 2007 or yr 2008 or yr 2009
S12	S1 or S2 or S3 or S4 or S5 or S6 or S7 or S8 or S9 or S10 or S11
S11	ti lumbago or ab lumbago
S10	ti spondylosis or ab spondylosis

S9	mh sciatica
S8	ti sciatica or ab sciatica
S7	ti coccydynia or ab coccydynia
S6	ti coccyx or ab coccyx
S5	ti (lumbar n0 pain) or ab (lumbar n0 pain)
S4	mh Low Back Pain+
S3	ti backache or ab backache
S2	mh Back Pain+
S1	ti dorsalgia or ab dorsalgia

4.3.2 Study selection

For a study describing the development of a CPR to be included in the review it had to meet the following criteria:

Studies needed to explicitly aim to develop one or more CPRs involving the physiotherapy management of LBP. The operational definition of a CPR for this study was that defined by McGinn et al. (2008). Although it has been suggested that there should be a minimum of three variables in a CPR (Laupacis et al., 1997; Stiell & Wells, 1999), previous systematic reviews (Tamariz et al., 2004; Wisnivesky et al., 2005) have included studies with two or more predictor variables. To ensure all relevant studies were identified, this review used the more liberal definition of a CPR as that containing two or more predictor variables.

Substantial practice variation between low back pain treatment providers (Kent & Keating, 2005; Werner & Indahl, 2005) including marked differences in the methods chosen to assess this condition (Kent, Keating, & Taylor, 2009) makes it arguably inappropriate to assume that the selection and

assessment of potential predictor variables will generalise across disciplines. Thus, it was determined a priori that for a study to be included, the assessment of potential predictor variables was required to be performed by a physiotherapist to ensure their direct relevance to the primary research aim.

Consistent with the definition of a CPR employed in this review (McGinn et al., 2008), predictor variables were required to be independently meaningful. Diagnostic, prescriptive and prognostic studies investigating CPRs at any stage of their development (McGinn et al., 2000), derivation, validation, or impact-analysis, were included.

No restriction was placed upon the type of potential predictor variables (eg. history items, imaging modalities, physical examination items, psychological variables etc) under investigation in the studies considered for inclusion. Further, no restriction was placed upon the clinical setting or the type of patients with LBP under investigation in studies considered for eligibility in this review.

Identified studies were downloaded into an electronic reference management system (EndNote, version X2.0.1, Thomson Reuters, California, USA) and duplicates were removed.

Two reviewers performed the first-stage screening of titles and abstracts based upon the stated eligibility criteria. Any study denoted eligible by either

reviewer was progressed to the second-stage of eligibility screening. Additionally, studies identified by citation tracking and hand-searching of relevant journals were progressed to the second-stage. The full-text of included studies was obtained and examined by two reviewers. During this second-stage of screening, concordance between reviewers determined inclusion, with disagreements resolved by consensus, or if needed by a third reviewer.

4.3.3 Data extraction and quality assessment

A standardised instrument was used for data extraction. Information collected from each study included the country of origin, the number of rules developed, study design, stated objective, and details of the patient population. The reviewers also investigated whether included studies specifically used the term “clinical prediction rule”. The hierarchy of evidence for CPRs (McGinn et al., 2000) was initially employed to determine which stage of CPR development an article was describing. Studies were subsequently defined as derivation, validation or impact-analysis.

Consistent with the aim of the present review, the quality of the included studies were evaluated against the well-cited methodological standards that are employed by researchers in the development of all forms of CPRs. These criteria reflect the necessary methodological requirements to develop any form of a CPR and should be considered as an extension to the various methodological requisites that are specific to the underlying study design. In the absence of an appropriate standardised tool and to avoid the limitations

of unsystematically selecting criteria from previous reports, we initially identified the key texts describing the methodological standards common to the development of all forms of CPRs, including those used in previous systematic reviews. From these texts, five (Beattie & Nelson, 2006; Childs & Cleland, 2006; Laupacis et al., 1997; McGinn et al., 2000; Stiell & Wells, 1999) were selected based upon their inclusion in previous reviews, their number of citations in MEDLINE and EMBASE and their relevance to the research aim. Criteria that were represented in two or more of the five selected texts were included in the methodological appraisal of the included studies. This review employed definitions of the accepted CPR quality criteria that have been previously published (Beattie & Nelson, 2006; Childs & Cleland, 2006; Laupacis et al., 1997; McGinn et al., 2000; Stiell & Wells, 1999). A checklist was subsequently developed for each of the three phases of rule development. The research designs of the included studies were anticipated to be extensively heterogeneous ranging from randomised controlled and observational intervention studies, to cross-sectional diagnostic investigations and longitudinal prognostic studies. Consequently, no attempt was made to appraise and contrast the included studies against the methodological standards that are specific to their unique underlying research design.

Two reviewers independently appraised the methodological quality of the included studies. Each criterion was evaluated independently with concordance between examiners determining the appropriate outcome. Disagreement was resolved by consensus and if needed, by a third reviewer.

For a criterion to be marked as being met, studies must have entirely fulfilled the requirements of that criterion with no occasions of disparity. For example, in studies that aimed to develop two or more CPRs, all rules within the study must have achieved the requirements of that criterion for it to be considered met. Criteria marked as 'unclear' or 'not met' were consolidated to enable the dichotomisation of each criterion as 'met' or 'not met'.

The research design of studies investigating predictors of responsiveness to intervention were specifically evaluated for their ability to identify treatment-effect modifiers. These variables, also known as 'moderators', are the baseline characteristics that identifies subgroups of patients with differing treatment effect-sizes for a given intervention (Kraemer et al., 2006; Kraemer & Gibbons, 2009; Kraemer, Wilson, Fairburn, & Agras, 2002; MacKinnon & Luecken, 2008; Turner, Holtzman, & Mancl, 2007) . Recent commentary in the rehabilitation literature (Hancock, Herbert, et al., 2009) has highlighted the inadequacy of single-arm research designs in identifying the variables that influence a patient's responsiveness to an intervention. Controlled trials are required in all stages of prescriptive CPR development to discriminate between the non-specific prognostic factors associated with clinical outcome, and the specific treatment-effect modifying variables that help further guide clinical decision making. The distinction between single-arm prescriptive CPR studies and prognostic CPR studies was determined by the stated clinical aim of the CPR in each study.

4.3.4 Data synthesis and analysis

Due to the anticipated heterogeneity of the included studies, no attempt was made to statistically pool the results of individual studies.

Between-rater agreement was evaluated for each stage of the screening process and for the methodological appraisal of the included studies. The absolute and chance-corrected degrees of agreement (κ) with 95% confidence intervals were calculated for both stages of the screening procedure. Between group comparisons were analysed following exploratory data analysis and relevant parametric or non-parametric tests were applied. All statistical analyses were conducted using Stata 11.0 (StataCorp, Texas, USA).

4.4 Results

4.4.1 Study selection

The database search strategy yielded 10,202 studies. Another twelve studies were identified via hand-searching of relevant journals and citation-tracking of included studies. Following the removal of duplicate records, 7,453 records were screened via title and abstract with 381 records progressing to the second stage of screening. The full-text copies of these studies were located and reviewed with 23 studies composing the final included sample. The reasons for exclusion are highlighted in Figure 4.1 (below).

The absolute agreement between raters for the first and second-round screening procedures was 96.6% and 94% respectively. The chance-corrected degree of agreement was observed to be “moderate” (Sackett, Hayes, Guyatt, & Tugwell, 1991) for both procedures with κ = 0.49 (95%CI 0.43 – 0.55) for the screening by titles and abstracts, and κ = 0.53 (95%CI 0.35 – 0.72) for the screening by full-text. All but one episode of disagreement between raters was resolved by consensus, with the remaining study ruled to be included by the third reviewer.

4.4.2 Characteristics of included studies

The majority of included studies (n=15) originated from the USA. Three studies were conducted in Australia and two in The Netherlands. The remaining three studies were conducted in Singapore, Spain and the United Kingdom. Although the search strategy enabled the inclusion of studies from 1990, the earliest year of publication of the included sample was 2002. The majority of included studies developed just one CPR, although some studies investigated up to five rules in one publication.

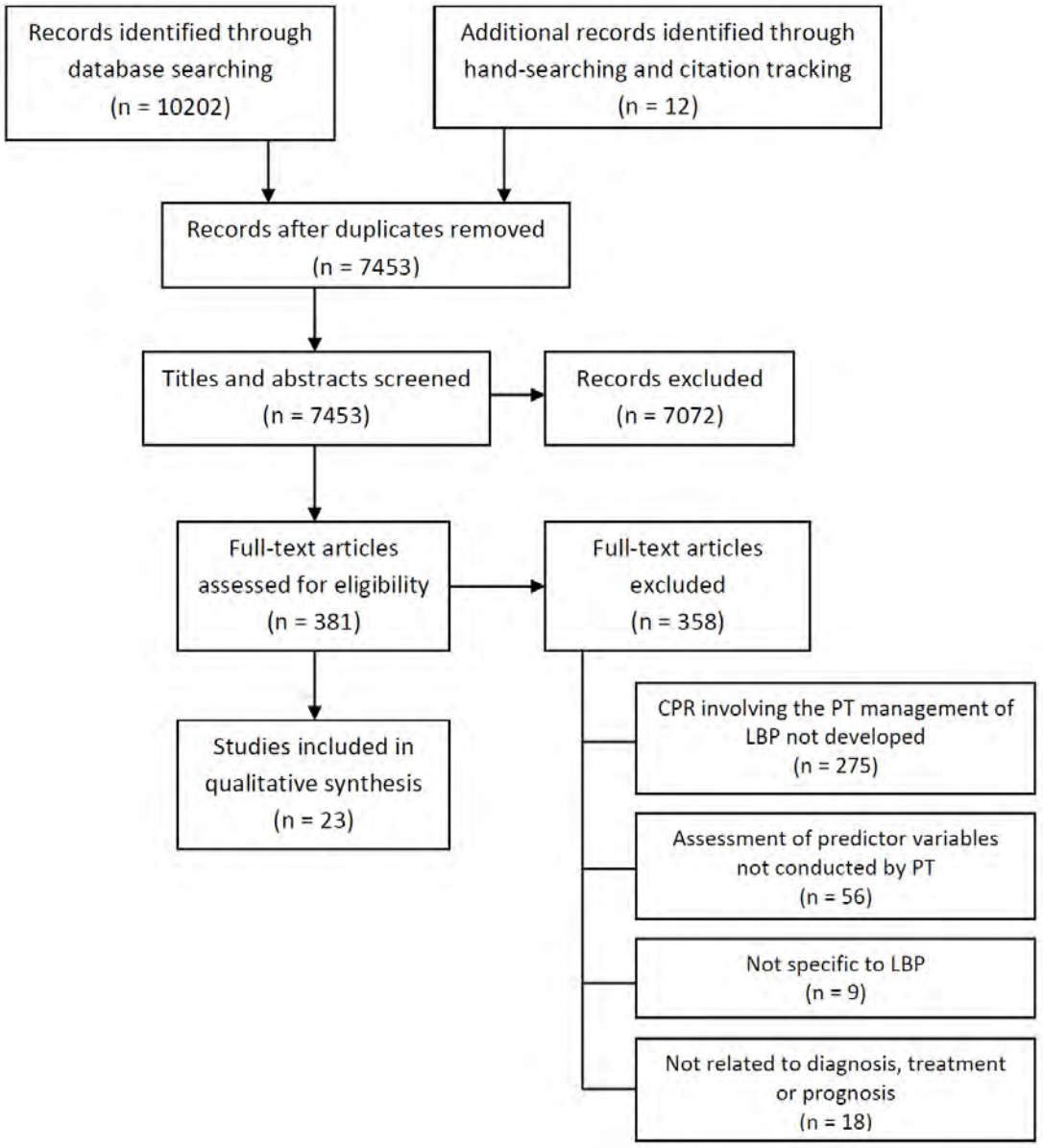


Figure 4.1 Flow chart of search strategy and study selection

Fifteen derivation and eight validation studies compose the included sample. No studies investigating the impact phase of rule development were identified. Fourteen studies describe CPRs used to influence treatment decision-making. Ten (43%) of the included studies relate to the prediction of clinical outcome with the use of spinal manipulation. Seven studies concern

diagnosis and only two prognostic studies were included. Across the 23 included publications, 25 unique CPRs are described including 15 diagnostic, 7 prescriptive and 3 prognostic rules. Table 4.2 (p. 145), Table 4.3 (p. 150) and Table 4.4 (p. 157) detail the identified diagnostic, prescriptive and prognostic CPRs respectively, and the relevant studies that have contributed to their development.

Table 4.2 Diagnostic clinical prediction rules included in qualitative synthesis

CPR	Variables	Publication	Stage of rule development	Sample	Results / outcome
Radiographic instability	Lumbar Flexion > 53°, lack of hypomobility with intervertebral motion testing (2 variables)	Fritz, Piva, and Childs (2005)	Derivation	n=49, LBP +/- leg pain, referred for imaging on suspicion of instability, mean 39.2 years old, 57% female, median 78 days of symptoms, 57% prevalence of target condition.	If 2 variables positive, +LR[1] = 12.8 (95%CI 0.79-211.6). If 1 variable positive, +LR = 4.3 (95%CI 1.8-10.6).
Diskogenic pain CPR1	CP[2], PPE[3], VABLE[4], Ext Loss[5] (4 variables)	Laslett, Aprill, et al. (2006)	Derivation	n=216, LBP+/- leg pain, referred to specialist diagnostic centre, mean 44.2 years old, 43% female, mean 158 weeks of symptoms, 35% prevalence of target condition. Only 107 patients received reference standard.	If 1 or more variables positive, then +LR = 1.9 (95%CI 1.1-3.2) and -LR[6] = 0.37 (95%CI 0.21-0.65). If 2 variables positive, then +LR = 6.7 (95%CI 0.95-50) and -LR = 0.73(0.61-0.97)

[1] +LR = positive likelihood ratio
[2] CP = centralization phenomenon
[3] PPE = persistent low back pain between episodes of acute low back pain
[4] VABLE = subjective report of 'vulnerability ' when in the semi-stooped position or when performing twisting actions
[5] Ext Loss = visual estimation of moderate or major loss of lumbar extension range of movement
[6] -LR – negative likelihood ratio

CPR	Variables	Publication	Stage of rule development	Sample	Results / outcome
Diskogenic pain CPR2	No CP, PPE, VABLE, Ext Loss (4 variables)	Laslett, Aprill, et al. (2006)	Derivation	n=216, LBP+/- leg pain, referred to specialist diagnostic centre, mean 44.2 years old, 43% female, mean 158 weeks of symptoms, 35% prevalence of target condition. Only 107 patients received reference standard.	If 2 variables positive, then sensitivity = 37% (95%CI 24-50) and specificity = 100% (95%CI 82-100). LR's not calculated due to 100% specificity.
Diskogenic pain CPR3	PPE, VABLE, Ext Loss (3 variables)	Laslett, Aprill, et al. (2006)	Derivation	n=216, LBP+/- leg pain, referred to specialist diagnostic centre, mean 44.2 years old, 43% female, mean 158 weeks of symptoms, 35% prevalence of target condition. Only 107 patients received reference standard.	If 2 variables positive, then +LR = 6.5 (95%CI 0.9-46.3) and -LR = 0.77 (95%CI 0.66-0.9).
SIJ mediated pain CPR1	Distraction, Compression, Thigh thrust, Gaenslen's (right), Gaenslen's (left), Sacral Thrust (6 variables)	Laslett, Aprill, McDonald, and Young (2005)	Derivation	n=48, buttock pain +/- LBP +/- leg pain, referred to specialist diagnostic centre with suspicion of SIJ pain, mean 42.1 years old, 67% female, mean 32 months of symptoms, 33% prevalence of target condition.	If 3 or more variables positive, then +LR = 4.29 (95%CI 2.34-8.58) and -LR = 0.8 (95%CI 0.14-0.37)
SIJ mediated pain CPR2	Distraction, Thigh Thrust, Compression, Sacral Thrust (4 variables)	Laslett, Aprill, et al. (2005)	Derivation	n=48, buttock pain +/- LBP +/- leg pain, referred to specialist diagnostic centre with suspicion of SIJ pain, mean 42.1 years old, 67% female, mean 32 months of symptoms, 33% prevalence of target condition.	If 2 positives, then +LR = 4 (95%CI 2.13-8.08) and -LR = 0.16 (95%CI 0.04-0.47)

CPR	Variables	Publication	Stage of rule development	Sample	Results / outcome
SIJ mediated pain CPR3	Distraction, Thigh Thrust, Gaenslen's test, Compression, Sacral Thrust (5 variables)	Laslett et al. (2003)	Derivation	n=43 (subset of patients from Laslett et al 2005 using different reference standard), buttock pain +/- LBP +/- leg pain, referred to specialist diagnostic centre with suspicion of SIJ pain, insufficient data to report precise demographic details, 26% prevalence of target condition.	If 3 or more positives, then +LR = 4.16 (95%CI 2.16-8.39) and -LR = 0.12 (95%CI 0.02-0.49).
SIJ mediated pain CPR4	No CP/peripheralisation, Distraction, Thigh Thrust, Gaenslen's test, Compression, Sacral Thrust (6 variables)	Laslett et al. (2003)	Derivation	n=34 (subset of patients from Laslett et al 2005 using different reference standard), buttock pain +/- LBP +/- leg pain, referred to specialist diagnostic centre with suspicion of SIJ pain, insufficient data to report precise demographic details, 32% prevalence of target condition.	If no CP/periphalisation and if 3 or more positives of remaining variables, then +LR = 6.97 (95%CI 2.7-20.27) and -LR = 0.11 (95%CI 0.02-0.44)
SIJ mediated pain CPR5	Distraction, Compression, Thigh Thrust, Patrick sign, Gaenslen's test (5 variables)	van der Wurff, Buijs, and Groen (2006)	Derivation	n=60, buttock pain +/- leg pain, referred for invasive procedures, mean 51 years old, 78% female, mean 98 months of symptoms, 45% prevalence of target condition.	If 3 or more positives, then +LR = 4.02 (95%CI 2.04-7.89) and -LR 0.19 (95%CI 0.07-0.47)

CPR	Variables	Publication	Stage of rule development	Sample	Results / outcome
Z-jt mediated pain CPR1	Age ≥ 50, symptoms best walking, symptoms best sitting, onset pain is paraspinal, MSPQ[7] > 13, ext/rot test[8], no CP (7 variables)	Laslett, McDonald, et al. (2006)	Derivation	n=120, LBP +/- leg pain, referred to specialist diagnostic centre with suspicion of z-jt pain, mean 43 years old, 46% female, mean 158 weeks of symptoms, 11% prevalence of target condition.	If 4 or more positives, then +LR = 7.6 (95%CI 4.5-13.7) and -LR = 0.0 (95%CI 0.0-0.35)
Z-jt mediated pain CPR2	Age ≥ 50, symptoms best walking, symptoms best sitting, onset pain is paraspinal, MSPQ > 13, ext/rot test (6 variables)	Laslett, McDonald, et al. (2006)	Derivation	n=120, LBP +/- leg pain, referred to specialist diagnostic centre with suspicion of z-jt pain, mean 43 years old, 46% female, mean 158 weeks of symptoms, 11% prevalence of target condition.	If 2 or more positives, then +LR = 1.6 (95%CI 1.5-1.8) and -LR = 0.0 (95%CI 0.0-0.69).
Z-jt mediated pain CPR3	Age ≥ 50, symptoms best walking, symptoms best sitting, onset pain is paraspinal, MSPQ > 13 (5 variables)	Laslett, McDonald, et al. (2006)	Derivation	n=120, LBP +/- leg pain, referred to specialist diagnostic centre with suspicion of z-jt pain, mean 43 years old, 46% female, mean 158 weeks of symptoms, 11% prevalence of target condition.	If 1 or more positives, then +LR = 1.4 (95%CI 1.3-1.5) and -LR = 0.0 (95%CI 0.0-0.95).

[7] MSPQ = Modified Somatic Perception Questionnaire
[8] Ext/Rot test = Extension/Rotation test

CPR	Variables	Publication	Stage of rule development	Sample	Results / outcome
Z-jt mediated pain CPR4	Age≥50, symptoms best walking, symptoms best sitting, onset pain is paraspinal, ext/rot test (5 variables)	Laslett, McDonald, et al. (2006)	Derivation	n=120, LBP +/- leg pain, referred to specialist diagnostic centre with suspicion of z-jt pain, mean 43 years old, 46% female, mean 158 weeks of symptoms, 11% prevalence of target condition.	If 2 or more positives, then +LR = 2.0 (95%CI 1.8-2.5) and -LR = 0.0 (95%CI 0.0-0.49).
Z-jt mediated pain CPR5	Age≥50, symptoms best walking, symptoms best sitting, onset pain is paraspinal, ext/rot test (5 variables)	Laslett, McDonald, et al. (2006)	Derivation	n=120, LBP +/- leg pain, referred to specialist diagnostic centre with suspicion of z-jt pain, mean 43 years old, 46% female, mean 158 weeks of symptoms, 11% prevalence of target condition.	If 3 or more positives, then +LR = 9.7 (95%CI 5.0-18.8) and -LR = 0.17 (95%CI 0.05-0.6).
Vertebral fracture	Female sex, age > 70, significant trauma, prolonged use of corticosteroids (4 variables)[9]	Henschke et al. (2009)	Derivation	n=1172, acute LBP +/- leg pain patients presenting to a primary care provider, mean 44 years old, 47% female, 59% had duration of less than one week, 0.7% prevalence of target condition.	If 2 or more positives, then +LR = 15.5 (95%CI 7.2-24.6). If 3 or more positives, then +LR = 218.3(95%CI 45.6-953.8).

[9] Predictor variables not exclusively assessed by physiotherapists. Physiotherapists = 72.6%, general practitioners = 22.8%, chiropractors = 4.6%.

Table 4.3 Prescriptive clinical prediction rules included in qualitative synthesis

CPR	Variables	Publication	Stage of rule development	Sample	Results / outcome	Methodological notes
Spinal manipulation	duration of symptoms < 16 days, FABQ-W[10] < 19, at least 1 hip with > 35° IR ROM[11], hypomobility with lumbar spring testing, no symptoms distal to knee (5 variables)	Flynn et al. (2002)	Derivation	n=71, LBP+/- leg pain, baseline ODQ[12] score ≥ 30%, referred to physiotherapy, mean 37.6 years old, 41% female, mean 42 days of symptoms, 45% prevalence of target outcome.	If 4 or more positives, then +LR[13] = 24.38 (95%CI 4.63-139.41)	Single-arm design. Therefore unable to identify treatment-effect modifiers.
		Childs, Fritz, Piva, and Erhard (2003)	Validation	n=2 (case reports), 54 and 26 year old males, LBP and buttock pain respectively. One patient met 5 CPR criteria, the other patient met just 1 (or 2) criteria.	Only the patient with all 5 criteria positive experienced dramatic improvement in pain and disability following manipulation.	Research design prevents identification of treatment-effect modifiers.

[10] FABQ-W = Fear Avoidance Beliefs Questionnaire Work Subscale
[11] IR ROM = internal rotation range of movement
[12] ODQ = Oswestry Disability Questionnaire
[13] +LR = positive likelihood ratio

CPR	Variables	Publication	Stage of rule development	Sample	Results / outcome	Methodological notes
		Childs et al. (2004)	Validation	n=131 (RCT), LBP +/- leg pain, baseline ODQ score ≥ 30%, referred to physiotherapy, mean 33.9 years old, 42% female, median 27 days of symptoms, 29% prevalence of target outcome at 1/52 and 50% at 4/52.	Significant 3 way-interaction between CPR status (≥4/5= positive), Rx-group and time for pain and disability. For dichotomized outcome (success/failure) the interaction between CPR status and Rx-group strongly predicted success. For patients receiving manipulation, CPR positive status had +LR =13.2 (95%CI 3.4-52.1). For patients CPR positive the NNT with manipulation = 1.3 (95%CI 1.1-1.9)	RCT. Therefore treatment-effect modifiers able to be identified.
		Childs, Flynn, and Fritz (2006)	Validation	n=131 (RCT), LBP +/- leg pain, baseline ODQ score ≥ 30%, referred to physiotherapy, mean 33.9 years old, 42% female, median 27 days of symptoms.	Aimed to investigate if CPR status is predictive of a worsening in disability. No patient that was CPR positive and received manipulation worsened, preventing appropriate statistical analysis.	Secondary analysis of 2004 RCT. Therefore treatment-effect modifiers able to be identified.

CPR	Variables	Publication	Stage of rule development	Sample	Results / outcome	Methodological notes
		Cleland, Fritz, Whitman, Childs, and Palmer (2006)	Validation	n=12 (case series), LBP, ODQ score ≥ 30%, referred to physiotherapy, all CPR positive (≥4/5= positive), mean 39 years old, 42% female, median 19 days of symptoms.	Aimed to investigate generalizability of CPR status to another high-velocity thrust manipulation procedure. 11 out of 12 patients (92%) achieved the target outcome of 'success' at 1/52 following intervention.	All patients CPR positive, therefore unable to determine rule performance. Research design prevents identification of treatment-effect modifiers.
		Hancock, Maher, Latimer, Herbert, and McAuley (2008)	Validation	n=239 (RCT), LBP < 6/52 duration, presenting to general practitioner, mean 40.7 years old, 44% female, mean 9 days of symptoms.	Non-significant 3-way interaction between Rx-group, CPR status (≥4/5= positive) and time for pain (p=0.805) and disability (p=0.6). Patients that were CPR positive had better pain and disability outcomes independent of treatment group.	RCT. Therefore treatment-effect modifiers able to be identified. Spinal manipulative technique differed to derivation study. Only 5% of sample received high-velocity thrust manipulation.

CPR	Variables	Publication	Stage of rule development	Sample	Results / outcome	Methodological notes
		Cleland et al. (2009)	Validation	n=112 (RCT), LBP +/- leg pain, attending an outpatient physiotherapy clinic, modified ODQ baseline score >25%, all CPR positive (≥4/5 = positive), mean 40.3 years old, 52% female, median 45 days of symptoms.	Aimed to investigate the generalizability of CPR to another high-velocity thrust manipulation procedure and a non-thrust manipulative technique. No difference between the 2 high-velocity thrust procedures in pain and disability at any time point. Outcomes poorer in the non-thrust group.	All patients CPR positive, therefore unable to determine rule performance. RCT. Therefore treatment-effect modifiers able to be identified.
Spinal manipulation - pragmatic rule	duration of symptoms < 16 days, no symptoms distal to knee (2 variables)	Fritz, Childs, and Flynn (2005)	Derivation	n=141 (data from 2 previous studies), LBP+/- leg pain, baseline ODQ score ≥ 30%, referred to physiotherapy, mean 35.5 years old, 49% female, median 22 days of symptoms, 45% prevalence of target outcome.	If both criteria positive, then +LR = 7.2 (95%CI 3.2-16.1).	Single-arm design. Therefore unable to identify treatment-effect modifiers.

CPR	Variables	Publication	Stage of rule development	Sample	Results / outcome	Methodological notes
		Fritz, Brennan, and Leaman (2006)	Validation	n=215 (retrospective review of clinical database), occupational LBP, receiving Rx in outpatient physiotherapy clinic, all CPR positive (2/2 = positive), mean 35.9 years old, 32% female, mean 5.3 days of symptoms.	66.5% received manipulation (49.8% thrust and 16.7% non-thrust). Patients receiving manipulation experienced greater reductions in pain and disability with treatment, compared to those not receiving manipulation.	Research design prevents the identification of treatment-effect modifiers. All patients CPR positive, therefore unable to determine rule performance.
		Hallegraeff, de Greef, Winters, and Lucas (2009)	Validation	n=64 (RCT), acute LBP, all CPR positive (2/2 = positive), mean 39 years old, 45% female, 31% had symptoms less than 1/52.	Significant interaction for disability at 2.5 weeks between CPR status (including the additional criterion of age > 35 years) and Rx-group. No significant interactions for pain or lumbar spinal mobility.	All patients CPR positive (by derivation study criteria), therefore unable to determine rule performance. RCT. Therefore treatment-effect modifiers able to be identified. Analysis performed with the additional CPR criterion of age > 35 years.

CPR	Variables	Publication	Stage of rule development	Sample	Results / outcome	Methodological notes
Lumbar traction	FABQ-W < 21, no neurological deficit, age > 30, non-manual work job status (4 variables)	Cai, Pua, and Lim (2009)	Derivation	n=129, diagnosis related to the lumbosacral spine +/- leg pain, referred from orthopaedics to physiotherapy, mean 30.9 years old, 16% female, mean 40 weeks of symptoms, 19% prevalence of target outcome.	If 3 or more positives, then +LR = 3.04 (95%CI 2.04-4.53). If all 4 positive, then +LR = 9.36 (95%CI 3.13-28.0).	Single-arm design. Therefore unable to identify treatment-effect modifiers.
Stabilisation exercise - success	Age < 40 years, average SLR[14] > 91°, aberrant movement present, positive prone instability test (4 variables)	G. E. Hicks, Fritz, Delitto, and McGill (2005)	Derivation	n=54, LBP +/- leg pain, referred to outpatient physiotherapy clinics, mean 42.4 years old, 57% female, mean 41 days of symptoms, 33% prevalence of target outcome (success).	If 3 or more positives, then +LR = 4.0 (95%CI 1.6-10.0). If 2 or more positives, then +LR = 1.9 (95%CI 1.2-2.9).	Single-arm design. Therefore unable to identify treatment-effect modifiers.
Stabilisation exercise - failure	Prone instability test, aberrant movement, hypermobility, FABQ physical activity subscale > 8 (4 variables)	G. E. Hicks et al. (2005)	Derivation	n=54, LBP +/- leg pain, referred to outpatient physiotherapy clinics, mean 42.4 years old, 57% female, mean 41 days of symptoms, 72% prevalence of target outcome (not failure).	In the absence of 2 or more positives (ie. 1 or 0 positives), then -LR[15] = 0.18 (95%CI 0.08-0.38).	Single-arm design. Therefore unable to identify treatment-effect modifiers.

[14] SLR = straight leg raise
[15] -LR = negative likelihood ratio

CPR	Variables	Publication	Stage of rule development	Sample	Results / outcome	Methodological notes
McKenzie approach (MDT[16])	< 12/52 duration, centralization or abolition of symptoms with MDT loading strategies (2 variables)	May, Gardiner, Young, and Klaber-Moffett (2008)	Derivation	n=102 (secondary analysis of single-arm of RCT), back and neck pain patients referred by GP's to Physiotherapy, study sample demographics not provided.	For those patients with back pain, the presence of both predictor variables gave a probability of success ('liberal' definition provided in study) of 68.9%. The absence of both variables gave a probability of success of 10%.	Single-arm design. Therefore unable to identify treatment-effect modifiers.
Specific exercise program for Ankylosing Spondylitis	SF-36 Physical Role > 37, SF-36 Bodily Pain > 27, BASDAI[17] > 31 (3 variables)	Alonso-Blanco, Fernandez-de-las-Penas, and Cleland (2009)	Derivation	n=35, patients with AS referred to physiotherapy clinic, mean 45.7 years old, 20% female, mean 9.7 years of symptoms, 46% prevalence of target outcome.	If 2 or more positives, then +LR = 11.2 (95%CI 1.7-76.0). If 3 or more positives, then +LR = 2.6 (95%CI 1.6-4.0).	Single-arm design. Therefore unable to identify treatment effect modifiers.

[16] MDT = mechanical diagnosis and therapy
[17] BASDAI = Bath Ankylosing Spondylitis Disease Activity Index

Table 4.4 Prognostic clinical prediction rules included in qualitative synthesis

CPR	Variables	Publication	Stage of rule development	Sample	Results / outcome	Notes
6 month pain outcome for acute/subacute LBP	Baseline pain intensity (0-10 NRS[18]), CP[19] (present = 1, absent = 0) (2 variables)	George, Bialosky, and Donald (2005)	Derivation	n=28 (secondary analysis of subgroup in earlier clinical trial), LBP < 60 days duration, aged 18-55 years, demographic details of this subgroup not reported.	6 month pain intensity (0-10 NRS) = 0.97 + 0.27(Pain 0-10 NRS) - 1.6 (CP).	Analysis limited to only those patients that were classified for 'specific exercise'.
6 month disability outcome for acute/subacute LBP	Baseline disability (ODQ[20]), FABQ-W[21], CP (present = 1, absent = 0) (3 variables)	George et al. (2005)	Derivation	n=28 (secondary analysis of subgroup in earlier clinical trial), LBP < 60 days duration, aged 18-55 years, demographic details of this subgroup not reported.	6 month disability (ODQ) = 4.4 + 0.24 (ODQ) + 0.34(FABQ-W) - 10 (CP).	Analysis limited to only those patients that were classified for 'specific exercise'.

[18] NRS = numerical rating scale
[19] CP = centralisation phenomenon
[20] ODQ = Oswesty Disability Questionnaire
[21] FABQ-W = Fear Avoidance Beliefs Questionnaire Work Subscale

CPR	Variables	Publication	Stage of rule development	Sample	Results / outcome	Notes
Time to recovery from acute LBP	baseline pain ≤ 7/10, duration of current episode ≤ 5 days, and ≤ 1 previous episodes (3 variables)	Hancock, Maher, Latimer, Herbert, and McAuley (2009)	Derivation	n=239 (RCT), LBP +/- leg pain < 6/52, presenting to GPs, mean age 40.7 years, 44% female, mean 9 days of symptoms.	If 3 variables positive, then median days to recovery (from baseline assessment) = 6 (95%CI 4-8). If no variables are positive, then median days to recovery = 22 (95%CI 11-33).	All arms of study included in analysis.

4.4.3 Qualitative appraisal of included studies

Quality scoring for the derivation and validation studies is provided in Table 4.5 (p. 160) and Table 4.6 (p. 162) respectively. “Substantial” (Sackett et al., 1991) between-rater agreement was observed for the quality scoring with an absolute degree of agreement of 88.7% ($\kappa = 0.74$, 95%CI 0.66-0.81). Three episodes of disagreement required resolution by a third reviewer, with the remaining disagreements being resolved by consensus.

Five of the 14 publications (36%) concerning prescriptive CPRs used a randomised controlled-study design that would permit the identification of treatment-effect modifiers.

Although all included studies satisfied the operational definition of a CPR, not all articles specifically used the term. Of the 23 included studies, only 15 (65%) explicitly used the term “clinical prediction rule” when describing the clinical tool being developed. It was more common for prescriptive studies to use the term “clinical prediction rule”, compared to diagnostic and prognostic studies ($p<0.001$).

Table 4.5 Derivation study quality appraisal

	Diagnostic								Prescriptive					Prognostic	
	Fritz, Piva, et al. (2005)	Henschke et al. (2009)	Laslett, Aprill, et al. (2006)	Laslett, Aprill, et al. (2005)	Laslett, McDonald, et al. (2006)	Laslett et al. (2003)	van der Wurff et al. (2006)	Alonso-Blanco et al. (2009)	Cai et al. (2009)	Flynn et al. (2002)	Fritz, Childs, et al. (2005)	G. E. Hicks et al. (2005)	May et al. (2008)	George et al. (2005)	Hancock, Maher, et al. (2009)
Prospective design	✓	✓	✓	✓	✓[22]	✓	✓	✓	✓	✓	✓[22]	✓	✓[22]	✓[22]	✓
Outcomes defined	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Outcome clinical important	✓	✓	✓	No[23]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Blinded outcome assessment	✓	No	No	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	No
All important predictors included	No	✓	No	No	No	No	No	No	No	✓	No	No	No	✓	✓
Predictive variables clearly defined	✓	No	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Blinded predictor assessment	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Assessment of the reliability of the predictive variables	✓	No	No	✓	No	✓	✓	✓	✓	✓	✓	✓	✓	✓	No
Important patient characteristics described	✓	✓	✓	✓	✓	No	✓	✓	✓	✓	✓	✓	No	✓	✓
Representative sample	No	✓	No	No	No	No	No	✓	✓	✓	✓	✓	✓	No	✓
Study site described	No	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Justification for the number of study subjects	No	✓	No	No	No	No	No	No	No	No	No	No	No	No	No
≥ 10 outcome events per independent variable in the rule	✓	No	✓	No[24]	No	No[24]	No[24]	No	No	No	No[24]	No[24]	No	No[25]	✓

[22] Secondary analysis of prospectively derived data.
[23] Single diagnostic injection not consistent with current SIJ diagnostic criterion standard (Szadek, van der Wurff, van Tulder, Zuurmond, & Perez, 2009)
[24] Multivariable regression not performed for all prediction rules.
[25] More than one rule presented. Not all rules satisfy criterion.

	Diagnostic								Prescriptive					Prognostic	
	Fritz, Piva, et al. (2005)	Henschke et al. (2009)	Laslett, Aprill, et al. (2006)	Laslett, Aprill, et al. (2005)	Laslett, McDonald, et al. (2006)	Laslett et al. (2003)	van der Wurff et al. (2006)	Alonso-Blanco et al. (2009)	Cai et al. (2009)	Flynn et al. (2002)	Fritz, Childs, et al. (2005)	G. E. Hicks et al. (2005)	May et al. (2008)	George et al. (2005)	Hancock, Maher, et al. (2009)
Mathematical techniques described	✓	No	✓	✓	No	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Multivariable analysis	✓	✓	✓	No	✓	No	No	✓	✓	✓	No	No[24]	✓	✓	✓
Results of the rule described	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Clinically sensible/reasonable	No	✓	No	✓	✓	✓	✓	No	No	✓	✓	✓	No	✓	✓
Easy to use	✓	✓	No[25]	✓	No[25]	No[25]	✓	✓	✓	✓	✓	✓	✓	✓	✓
Probability of diagnosis or outcome described	✓	✓	No	No	✓	No	✓	✓	✓	✓	✓	✓	✓	No	✓
Course of action described	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
Specifically uses the term “Clinical Prediction Rule”	No	No	No	No	✓	No	No	✓	✓	✓	✓	✓	✓	No	✓

Table 4.6 Validation study quality appraisal

	Prescriptive							
	Childs et al. (2003)	Childs et al. (2004)	Childs et al. (2006)	Cleland et al. (2006)	Cleland et al. (2009)	Fritz, Brennan, and Leaman (2006)	Hancock, Maher, Latimer, et al. (2008)	Hallegraeff et al. (2009)
Prospective validation in new patient population	No[26]	✓	✓	✓	✓	No	✓	✓
Different clinical setting to derivation study	No[26]	No	No	✓	✓	✓	✓	✓
Different clinicians to derivation study	No[26]	No[26]	No[26]	✓	✓	✓	✓	✓
Representative sample	No	✓	✓	No	✓	No	✓	No
The rule is applied accurately	No	✓	✓	✓	✓	✓	✓	No
Complete follow-up	✓	✓	✓	✓	✓	✓	✓	✓
Accuracy of the rule in the validation study sample described	No	✓	No[27]	No	No	No	✓	No
Assessment of the interobserver reliability of the rule	No	No	No	No	No	No	No[26]	No
Specifically uses the term “Clinical Prediction Rule”	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes

[26] Unclear. Insufficient information.
[27] Absence of target outcome in subgroup preventing appropriate statistical analysis

4.5 Discussion

There has been a rapid growth in the number of studies reporting upon the development of CPRs in the physiotherapy literature. This trend mirrors that seen in Medicine, particularly in the fields of Emergency and Intensive Care and may be reflective of a progressive move towards models of clinical decision-making that are increasingly data-driven and firmly founded upon the process of scientific enquiry. The quest to identify meaningful sub-groups of patients will have important implications for clinical practice, particularly for presentations, such as LBP, which are confounded by their degree of heterogeneity and numerous treatment alternatives.

To our knowledge, the present review is the first to systematically locate, appraise and determine the clinical readiness of diagnostic, prescriptive and prognostic CPRs involving the physiotherapy management of LBP in all phases of their development. Twenty-five unique CPRs were identified encompassing a diverse range of factors. While the growth in this research is arguably important for LBP treatment providers, this observed large variation in CPR themes may reflect the current lack of understanding of clinicians' priorities for CPRs. Investigation of the areas of perceived clinical need for CPRs would facilitate the development of rules with the greatest potential to positively influence clinical practice (Eagles, Stiell, Clement, Brehaut, Kelly, et al., 2008).

Previous systematic reviews of CPRs in the physical rehabilitation literature (Beneciuk et al., 2009; May & Rosedale, 2009; Stanton et al., 2010) have included four studies involving the physiotherapy management of LBP which were excluded in the present review. Two studies (Fritz et al., 2007; Fritz, Whitman, Flynn, Wainner, & Childs, 2004) included in earlier reviews have investigated the characteristics that are associated with treatment outcomes. However, as both studies did not develop a clinical tool that may be applied to an individual patient they did not meet the present review's eligibility criteria. One excluded study (Brennan et al., 2006) was determined to have investigated a classification system while the other excluded study (Teyhen, Flynn, Childs, & Abraham, 2007) was limited to describing the arthrokinematic characteristics of a subgroup that were positive on a previously derived CPR.

4.5.1 Summary of evidence

Based upon the findings of the present review, the available evidence does not support the direct clinical application of any of the identified CPRs for LBP at this time. Of the 25 unique CPRs identified, only two have progressed to the process of validation and no rule has been investigated for its ability to positively influence clinical outcomes and/or resource consumption.

The 5-item spinal manipulation CPR derived by Flynn et al. (2002) in a single-arm study design is one of the CPRs that has been further investigated in a series of validation studies. Recent commentary in the literature (Allison, 2009; C. Cook et al., 2010; Hancock, Herbert, et al., 2009)

and in two Physical Therapy podcasts (Fritz, Hancock, Herbert, & Riddle, 2009a, 2009b) have discussed the limitations of single-arm study designs in the development of prescriptive CPRs. The lack of a control group enables the identification of non-specific prognostic variables but is unable to investigate the moderators of treatment-effect. Controlled-study designs utilizing tests of interactions are required to identify on whom and under what circumstances treatments produce different outcomes (Hancock, Herbert, et al., 2009; Kraemer et al., 2002). Accordingly, it has been suggested that the subsequent study undertaken by Childs et al. (2004) is most appropriately considered a derivation study and not a validation study. This is because it was the first controlled-study that enabled the investigation of the CPR as a treatment response modifier, in contrast to a non-specific prognostic factor (Hancock, Herbert, et al., 2009).

Of the remaining validation studies that have aimed to develop the 5-item spinal manipulation CPR in new cohorts of patient populations, only two (Cleland et al., 2009; Hancock, Maher, Latimer, et al., 2008) have used a controlled-study design. Cleland et al. (2009) aimed to examine the generalizability of the CPR to different thrust and non-thrust manipulative techniques. The generalizability of a CPR to other procedures is most appropriately determined by controlled-study designs that investigate if a patient's status on the rule significantly moderates the effect-size of an intervention (Assmann, Pocock, Enos, & Kasten, 2000; Kraemer et al., 2006; Kraemer et al., 2002; MacKinnon & Luecken, 2008; Turner et al., 2007). However, as the patient population in this study were all positive on the

spinal manipulation CPR, the performance of the rule in identifying those with a difference in treatment responsiveness remained untested. Finally, in the well-designed validation study by Hancock, Maher, Latimer, et al. (2008), the spinal manipulation CPR was found to perform no better than chance in identifying patients likely to respond to this intervention. Positive status on the rule, however, was found to be a non-specific prognostic factor. One of the many possible explanations for the observed findings noted by these researchers (Hancock, Maher, & Herbert, 2008; Hancock, Maher, Latimer, et al., 2008) and others (Hebert & Perle, 2008) is the difference in treatment provided in this study compared to the original derivation studies (Childs et al., 2004; Flynn et al., 2002), with high-velocity thrust manipulative techniques only being used on a very small proportion of the patients in this study.

The 2-item pragmatic spinal manipulation CPR derived by (Fritz, Childs, et al., 2005) was based upon the collated results of two previous studies (Childs et al., 2004; Flynn et al., 2002) used to develop the 5-item rule. This abbreviated form of the spinal manipulation CPR was found to strongly identify those patients with a good outcome following treatment. However, as no control group was included in the derivation, the variables may represent prognostic factors that may have no specific relationship with the intervention provided. Two subsequent studies (Fritz, Brennan, & Leaman, 2006; Hallegraeff et al., 2009) attempting to validate this rule restricted their patient populations to only those that were positive on the pragmatic spinal manipulation CPR. As previously noted, without the inclusion of patients that

are also negative on the rule, a prescriptive CPR's performance is unable to be rigorously investigated. Consequently, the body of evidence does not yet enable confidence in the direct clinical application of either the 5-item or 2-item spinal manipulation CPRs in identifying subgroups of patients with differences in responsiveness to this intervention.

The 23 rules that have been derived, but not yet proceeded to validation may inform clinical practice by providing clinicians with an understanding of some of the most important predictors of a given target condition or outcome (McGinn et al., 2008). However, even in this limited application clinicians must exercise due caution as predictor variables may simply reflect chance associations or unique characteristics of the studied population or setting. Further, prescriptive predictor variables identified through single-arm study designs may not identify the relevant features that modify the effect of a given intervention, but instead reflect non-specific prognostic factors (Hancock, Herbert, et al., 2009).

It has been argued that the biologic plausibility of predictor variables be carefully considered throughout the derivation of a CPR to minimise the likelihood of including factors that reflect chance associations with the target outcome (Childs & Cleland, 2006; Fritz et al., 2009b; Raney et al., 2009). However, the primary function of a CPR is to accurately predict a target outcome and not to identify the determinants of that outcome. The composite of factors that together accurately predict a given outcome are of most value, regardless of whether this relationship is confounded by other variables

(Katz, 2006). To illustrate this point, consider that although carrying a cigarette lighter will not cause lung cancer, it may accurately predict a greater likelihood of developing the disease (Katz, 2006). Excluding predictive variables that are not believed at the time to be causally related to the target outcome may result in the development of CPRs with inferior predictive accuracy. Consequently, the process of rigorous validation of derived CPRs is the most suitable method to identify and exclude those variables that previously reflected chance associations with the target outcome (McGinn et al., 2008).

4.5.2 Methodological quality

Substantial variation was observed in the methodological quality of the fifteen included derivation studies. In addition to the previously mentioned research-design limitations of many prescriptive CPR studies, other common methodological shortcomings included the omission of important predictor variables, not providing a justification for the sample size and not including an appropriate number of outcome events per independent predictor when performing multivariable regression analysis.

Including the most probable predictor variables in the investigation aims to ensure that important relevant factors are not omitted (Laupacis et al., 1997). However, this needs to be balanced with restricting the analysis to a pre-determined small number of variables, ideally for only one outcome, to reduce the likelihood of eliciting findings that are due to chance and random error (Assmann et al., 2000). Researchers should consider examining the

results of secondary-analyses of randomised controlled trials and the findings of single-arm treatment studies to help guide the selection of variables (Fritz et al., 2009a).

Only one of the included derivation studies explicitly justified the size of the studied population. Larger sample sizes enable more precise estimates of a rule's predictive power, which in turn enhances confidence in its clinical application (Childs & Cleland, 2006; McGinn et al., 2008). A further consideration is that the investigation of treatment-effect modifiers in prescriptive CPRs requires much larger sample sizes in comparison to identifying main effects between treatment groups. Simulation studies have demonstrated that a study with an 80% power of detecting a given overall effect would require four times the number of subjects to maintain this power in detecting an interaction effect of the same magnitude (Brookes et al., 2004).

Researchers developing CPRs need to carefully consider the prevalence of the target outcome or condition when determining the sample size to ensure that there is a sufficient number of outcome events to satisfy the assumptions implicit to the statistical analysis. Seventy percent of the included derivation studies that used multivariable regression analysis did not have an adequate number of outcome events per independent variable in the model. Guidelines for the development of multivariable logistic regression and Cox proportional hazard models advocate a minimum of ten outcome events per independent variable to reduce the likelihood of identifying erroneous associations and to

improve the precision of the findings (Concato, Feinstein, & Holford, 1993). For multiple linear regression, it is recommended that there should be at least ten patients for every variable selected (Lewis, 2007).

Similar to the variance observed in the derivation studies, the methodological quality for the eight included validation studies varied substantially. No validation study included in this review investigated the inter-observer reliability of the CPR. Guidelines on the validation of CPRs have recommended that researchers examine the inter-observer reliability of the rule, at least within a subset of the study population, to ensure consistency in the interpretation of a patient's status on the rule (Laupacis et al., 1997; Stiell & Wells, 1999).

4.5.3 Study limitations

The search strategy employed in this review has been demonstrated to have high sensitivity for the detection of CPR studies (Ingui & Rogers, 2001) and has been used in other systematic reviews (Beneciuk et al., 2009; Dahri & Loewen, 2007; May & Rosedale, 2009). However, due to inconsistent nomenclature used to describe these clinical tools, it is plausible that not all potentially eligible studies were identified.

The primary aim of this review was the identification and appraisal of CPRs in the physiotherapy management of LBP. Due to substantial between-discipline practice differences in the assessment of LBP (Kent et al., 2009), it was determined a priori that for a study to be included, the assessment of

potential predictor variables was required to be performed by a physiotherapist. This eligibility criterion resulted in the exclusion of studies that had developed CPRs using other LBP treatment providers for the assessment of predictor variables. While outside the scope of the present review, the value and validity of such CPRs for physiotherapy practice arguably merits investigation.

The sensitive operational definition of a CPR used in this review enabled the inclusion of studies that may not have explicitly used the term "clinical prediction rule". Consequently, the methodological standards that would be considered by researchers explicitly aiming to develop a CPR may not have been considered in the design of these other studies. As the quality appraisal tool used in this review reflects these well-cited standards for CPR development, it is perhaps not surprising that a large variation of quality was observed between those studies that did and did not explicitly use the term “clinical prediction rule”.

The methodological appraisal tool used in this review was developed via a systematic process that aimed to minimise bias in the selection of appropriate quality criteria. While we believe this approach represents an improvement upon that used in previous systematic reviews of CPRs, our checklist has not been formally validated, and consequently the results need to interpreted with caution. The degree of between-rater agreement was high for the majority of the quality criteria, however, it is clear that some variables particularly those relating to the appraisal of validation studies would benefit

from measures to further improve rater concordance. An important consideration is that the quality criteria used in this review reflects the well-cited methodological standards that are common to diagnostic, prescriptive and prognostic forms of CPRs. Although this approach appropriately reflects the primary aim of this review and enables a qualitative comparison of the included studies, it is acknowledged that the omission of appraisal criteria that are specific to the development of each particular form of CPR may represent a potential limitation of the present study. Recently, a quality checklist for prescriptive derivation-based CPRs (the QUADCPR) has been developed using Delphi methods (C. Cook et al., 2010). While this checklist will require further investigation of its reliability and validity, and is not advocated for the retrospective appraisal of CPR studies, it constitutes an important contribution in providing clear methodological guidelines for developing future studies aiming to derive prescriptive rules.

4.6 Conclusions

This review is the first to systematically locate, appraise and determine the clinical readiness of diagnostic, prescriptive and prognostic CPRs involving the physiotherapy management of LBP in all phases of their development. Twenty-five unique rules were identified across fifteen derivation and eight validation studies. No impact studies were located. The current body of evidence does not enable confident direct clinical application of any of the identified CPRs. Further validation studies utilizing appropriate research designs and rigorous methodology are required to determine the

performance and generalizability of the derived CPRs to other patient populations, clinicians and clinical settings.

CHAPTER 5

PHYSIOTHERAPISTS' KNOWLEDGE, ATTITUDES AND PRACTICES REGARDING CLINICAL PREDICTION RULES FOR LOW BACK PAIN

This chapter has been published in a peer-reviewed scientific journal (Appendix 4):

Haskins, R., Osmotherly, P. G., Southgate, E., & Rivett, D. A. (2014). Physiotherapists' knowledge, attitudes and practices regarding clinical prediction rules for low back pain. *Manual Therapy*, *19*(2), 142-151.

The work presented in this manuscript was completed in collaboration with the co-authors (Appendix 1). The ethical approval for the study reported in this chapter appears in Appendix 2.

Overview

This is the second of five studies in this program of research. The findings of study one (Chapter 4, p.127) highlighted that a growing number of CPRs relevant to the physiotherapy management of LBP had commenced development. Very little was known, however, about whether these emergent tools were being applied in clinical practice by physiotherapists, and how they were conceptualised and integrated within the broader clinical reasoning

framework. Further, physiotherapists' acceptance of CPRs in the management of LBP, and the range of barriers and facilitators to their implementation were not yet understood. The study reported in this chapter was conducted to gain a greater understanding regarding the range of factors that may influence the implementation of LBP CPRs within physiotherapy clinical practice.

5.1 Abstract

Clinical Prediction Rules (CPRs) have been developed to assist in the physiotherapy management of Low Back Pain (LBP) although little is known about the factors that may influence their implementation in clinical practice. This study used qualitative research methodology to explore the knowledge, attitudes and practices/behaviours of physiotherapists in relation to these tools. Four semi-structured focus groups involving 26 musculoskeletal physiotherapists were conducted across three Australian geographic regions. A fictitious LBP case scenario was developed and used to facilitate group discussion. Participant knowledge of CPRs was found to be mixed, with some clinicians never having previously encountered the term or concept. LBP CPRs were often conceptualised as a formalisation of pattern recognition. Attitudes towards CPRs expressed by study participants were wide-ranging with several facilitating and inhibiting views identified. It was felt that more experienced clinicians had limited need of such tools. Only a small number of participants expressed that they had ever used LBP CPRs in clinical practice. To optimise the successful adoption of a LBP CPR,

researchers should consider avoiding the use of the term ‘rule’ and ensure that the tool and its interface are uncomplicated and easy to use. Understanding potential barriers, the needs of clinicians and the context in which CPRs will be implemented will help facilitate the development of tools with the highest potential to positively influence physiotherapy practice.

5.2 Introduction

The identification of meaningful sub-groups of patients with low back pain is a priority area for LBP research and is believed to have the potential to lead to substantial improvements in patient care (Borkan & Cherkin, 1996; Costa et al., 2013; Foster et al., 2009; Henschke, Maher, Refshauge, et al., 2007). Although the idea of sub-grouping patients with LBP is not new (C. McCarthy et al., 2004; Riddle, 1998), more recently greater emphasis has been placed upon the use of statistical procedures to identify the factors that delineate patients with LBP with differing prognoses and degrees of responsiveness to certain interventions. One such sub-grouping mechanism is the clinical prediction rule (CPR).

A CPR is a clinical tool that is used to inform decision-making by quantifying the probability of a given outcome, diagnosis or treatment response using a parsimonious set of factors from the history, physical examination and other investigations (McGinn et al., 2008). In recent years a growing number of CPRs relevant to physiotherapy have been derived for LBP presentations for a wide variety of diagnostic, prognostic and prescriptive functions (Beneciuk

et al., 2009; Haskins et al., 2012; May & Rosedale, 2009; Stanton et al., 2010). At this time however, it is not clear if these tools are consistent with the perceived needs of physiotherapists or will be accepted by them.

Limited evidence suggests that LBP CPRs may be accepted and used by some US physical therapists. A recent US study found that 40% of surveyed physical therapists who routinely employ lumbar thrust manipulation report using a CPR (Learman et al., 2012). Outside of a US context, however, there is no discernible research data on physiotherapists' awareness or use of LBP CPRs. Awareness of Emergency Medicine CPRs has been demonstrated to vary internationally and to be highest in the countries in which the tools have been developed (Eagles, Stiell, Clement, Brehaut, Taljaard, et al., 2008; Graham et al., 2001). As most LBP CPRs relevant to physiotherapy practice have been developed in the US (Haskins et al., 2012), it is likely that awareness and use of these tools in other countries may be much lower.

In addition to limited awareness, previous research has highlighted that once CPRs have been validated and demonstrated to positively impact clinical practice, there are a number of individual and system level barriers that may impede their successful adoption (Beutel et al., 2012; Brehaut et al., 2006; Brehaut et al., 2005; Eagles, Stiell, Clement, Brehaut, Taljaard, et al., 2008; Graham et al., 2001; Graham et al., 1998; Stiell et al., 2006). Table 5.1 (p. 179) provides an overview of the literature-informed potential barriers to the adoption of LBP CPRs in physiotherapy practice based on the current body of evidence using a framework of knowledge, attitudes and

practices/behaviours (Cabana et al., 1999; Legare et al., 2008). This framework has been used in previous research to help identify the barriers to the adoption of other clinical innovations, such as clinical practice guidelines (Larson, 2004; Pogorzelska & Larson, 2008; Schouten et al., 2007) and clinical protocols (Barlow et al., 2008; Dennison et al., 2007; Rubinson et al., 2005), and has been recommended as an appropriate framework to investigate the barriers to the use of CPRs (Abboud & Cabana, 2001). Recognition of the facilitators and barriers to the use of LBP CPRs will enable the development of tailored strategies that may assist the adoption of these tools into practice (Bero et al., 1998; Cabana, Rushton, & Rush, 2002; Grol & Wensing, 2004; Mehta, 2004; National Institute of Clinical Studies, 2006).

Table 5.1 Literature-informed potential barriers to the adoption of LBP CPRs in physiotherapy practice

Theme	Subtheme	Potential barrier	Description
Knowledge	Awareness	Lack of awareness	Unaware of the existence of LBP CPRs
	Familiarity	Lack of familiarity	Insufficient knowledge of the content of LBP CPRs to enable their application
	Forgetting	Forgetting	Inadvertently omitting to implement LBP CPRs
Attitudes	Agreement in general	Too 'cookbook'	Perception that LBP CPRs oversimplify the complexities of the clinical encounter
		Dislike of the term 'rule'	Aversion to using LBP CPRs due to the term 'rule' implying an authoritative influence on decision-making
		Challenge to autonomy	Perception that LBP CPRs are a threat to professional autonomy
		Biased synthesis	Perception that the development of the tool was biased
		Not practical	Perception that LBP CPRs are unclear or impractical to follow
		Unspecified overall lack of agreement with using the tool	Lack of agreement with LBP CPRs in general
	Expectancy	No perceived benefit to patient outcomes	Perception that using LBP CPRs will not lead to improved patient outcomes
		No perceived benefit to health care processes	Perception that using LBP CPRs will not lead to improved health care processes
	Self-efficacy	Lack of self-efficacy	Belief that one cannot use LBP CPRs
	Motivation	Lack of motivation / Inertia of current practice	Lack of motivation to use LBP CPRs or to change one's habits

Theme	Subtheme	Potential barrier	Description
Practices / Behaviours	Patient factors	Lack of consistency with patient preferences	Perceived inability to reconcile patient preferences with the use of LBP CPRs
	Factors associated with LBP CPRs as an innovation	Lack of triability	Perception that LBP CPRs cannot be tried or experimented with
		Lack of compatibility	Perception that LBP CPRs are not consistent with one's own approach
		High complexity	Perception that LBP CPRs are difficult to understand and use
		Lack of observability	Lack of the visibility of the results of using LBP CPRs
		Not communicable	Perception that it is not possible to communicate with colleagues about LBP CPRs to reach a mutual understanding
		Increased uncertainty	Perception that the use of LBP CPRs will increase uncertainty
		Not modifiable	Lack of flexibility to modify or adapt LBP CPRs
	Environmental factors	Lack of time	Insufficient time to use LBP CPRs
		Lack of resources	Insufficient resources to use LBP CPRs
		Organisational constraints	Insufficient support from the organisation to use LBP CPRs
		Lack of reimbursement	Insufficient reimbursement for using LBP CPRs
		Increased medicolegal liability	Perceived increased risk of legal actions arising from using LBP CPRs

Although considerable work has been invested in the development of LBP CPRs for physiotherapy practice, very little is known about how they will be integrated within the complex thinking and decision-making processes of clinical reasoning (I. Edwards et al., 2004). Limited evidence suggests that clinicians using LBP CPRs may not necessarily use them in isolation but rather consider them within the context of all other available information to inform their decision-making (Learman et al., 2012). Understanding the ways in which physiotherapists apply LBP CPRs in the clinical setting will also be informative to designing strategies to optimise their use.

What physiotherapists know about LBP CPRs, as well as their attitudes and practices in relation to these tools remains largely unknown but will underpin their successful adoption into clinical practice (National Institute of Clinical Studies, 2006). Qualitative research methodology seeks to construct meaning and knowledge through the understanding of human experience (Petty et al., 2012a) and provides an appropriate avenue to gain deep understanding and greater insight into the factors that influence LBP CPR implementation in physiotherapy. The generation of such knowledge is anticipated to inform strategies that may optimise the development of LBP CPRs with the greatest potential to positively impact physiotherapy practice.

5.3 Methods

5.3.1 Design

Qualitative Descriptive design is intended to provide a clear description of a specific phenomenon or experience from the perspective of research participants (Magilvy & Thomas, 2009). It is an approach that seeks to identify and explore rich straight description on particular topics using language reflective of that used by participants and with minimal interpretative meaning inferred by the researcher (Neergaard, Olesen, Andersen, & Sondergaard, 2009; Sandelowski, 2000, 2010). Qualitative Descriptive design was deemed an appropriate approach to gain firsthand insight into the knowledge, attitudes and practices/behaviours of physiotherapists in relation to LBP CPRs. The investigation of these domains is a well-recognised approach used to examine the barriers to the adoption of evidence in practice (Lang, Wyer, & Haynes, 2007).

5.3.2 Participants

Purposive sampling (Greenwood & Parsons, 2000) is a sampling technique that involves the selective recruitment of participants who may provide the best insight into the research questions. This sampling technique was used in this study to recruit physiotherapists of varying degrees of experience who manage patients with low back pain, in both private and public sectors, across metropolitan and regional areas of New South Wales, Australia. It was

considered by the research team that this sample is likely to be a target consumer group of LBP CPRs.

Potential participants were identified and recruited using phonebook listings and an online professional search tool. Additionally, an advertisement for the study was included within a professional email bulletin sent to all members of the Australian Physiotherapy Association.

5.3.3 Data collection

Four focus groups each lasting between 1.5 and 2 hours were conducted. Each group consisted of between 5 to 11 participants and was moderated by a member of the research team. A focus group schedule (Krueger & Casey, 2009) of activities and questions was developed and informed by research exploring the knowledge, attitudes and practices/behaviours of emergency physicians in relation to CPRs (Brehaut et al., 2006; Brehaut et al., 2005; Eagles, Stiell, Clement, Brehaut, Taljaard, et al., 2008; Graham et al., 2001; Graham et al., 1998; Stiell et al., 2006). A questioning route was developed centred upon addressing the following key research questions:

1. What is the knowledge of musculoskeletal physiotherapists in regard to CPRs?
2. What are the attitudes of musculoskeletal physiotherapists toward CPRs for LBP?
3. What are the self-reported practices/behaviours of musculoskeletal physiotherapists in relation to LBP CPRs?

A semi-structured format for each focus group was used. Table 5.2 (p. 185) details the foci, activities, prompts and approximate time spent on each section of the focus group.

A fictitious LBP case scenario (Figure 5.1) was developed based upon a previously published case study (Glynn & Weisbach, 2010) and adapted to include the predictor variables of 25 CPRs for LBP identified in a recent systematic review (Haskins et al., 2012). The credibility and consistency of the case scenario was checked by four specialist musculoskeletal physiotherapists and Fellows of the Australian College of Physiotherapists (Australian Physiotherapy Association, 2013) before being used in the focus groups.

After conducting the first two focus groups, the research team decided to provide participants in the remaining focus groups with a one-page summary (Figure 5.2, p. 191) detailing a common definition and example of each type of CPR. This was instigated in response to the research team's recognition that participant knowledge about LBP CPRs was diverse and it was believed that discussion concerning attitudes and practices/behaviours may be facilitated by providing a brief standardised summary to all participants. Qualitative research is often characterised by the simultaneous collection and analysis of data, thereby enabling researchers to adjust their avenue of investigation to build greater knowledge where opportunities are identified (Krueger & Casey, 2009; Sandelowski, 2000).

Table 5.2 Schedule of activities for focus groups

Section	Foci	Activities	Prompts	Approximate time
Introduction	Participant backgrounds, clinical experience, work setting, experience managing patients with LBP.	Participants introduce themselves to the group.	Please tell the group a little about yourself? Where do you work? How long have you been working as a physiotherapist? What proportion of your caseload are LBP patients?	10 minutes
Case study	Knowledge, attitudes and practices regarding CPRs for LBP. Clinical decision making in the assessment and management of LBP.	Participants are asked to read a fictitious LBP case study and discuss their perspectives on the assessment and management of that patient.	What are your thoughts regarding this patient's diagnosis/prognosis/management? What information is important in helping you make these decisions? Is anyone aware of any CPRs that could be used? Would anyone consider using a CPR? Which one(s) and why?	30 minutes
CPRs for LBP	Knowledge, attitudes and practices regarding CPRs for LBP.	Group discussion on CPRs for LBP. Participants in focus groups 3 and 4 received one-page summary about CPRs following discussion about knowledge of the topic (Figure 5.2).	What do you understand about the term 'CPR'? Which CPRs have you heard of? Can you describe any CPRs? What are your experiences with using LBP CPRs? How would you incorporate LBP CPRs into your clinical reasoning? How do you feel about CPRs for LBP? What are the barriers to using LBP CPRs?	40 minutes
Priorities for LBP CPR development	Participant priorities for the development of LBP CPRs.	Group discussion on participants' priorities for LBP CPR development.	In the management of LBP, are there any areas of your practice that may benefit from a CPR? Which types of LBP CPRs would be most useful? What characteristics do they need to have to be useful? What advice would you give to researchers who are developing LBP CPRs?	30 minutes

CASE SCENARIO

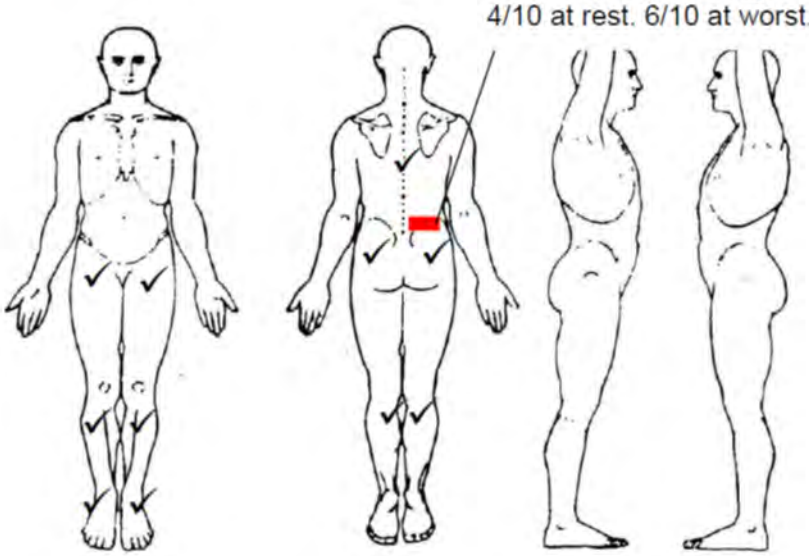
Summary:

32 year old male computer programmer complaining of a 10 day history of right sided low back pain.

SUBJECTIVE EXAMINATION

Body chart

Pain A
Constant/varying deep ache
4/10 at rest. 6/10 at worst.



Nil p+n/numb

Behaviour

Aggravating Factors
- Forward bending
- Sitting for greater than 15 minutes

Easing Factors
- Rest
- Best when lying on back with knees bent (crook-lying)

24-Hour Pattern
- Nil diurnal variation. Activity dependent.

Irritability
- Bending forward immediately causes 6/10 pain which eases to 4/10 within a few seconds after returning to a neutral standing posture

History

Current History

- Pain first commenced 10 days ago
- Pain spontaneously occurred a few hours after gardening (involved heavy lifting and bending)
- Nil change in presentation since first onset

Past History

- 3 previous episodes of mild low back pain over the past 5 years
- Each previous episode has resolved spontaneously after 2-3 days without treatment. Nil persisting symptoms between episodes.
- Otherwise, nil significant

Social History

- Lives with wife. No kids.
- Nil regular sporting activities.
- Has not taken any time off work due to low back pain.

Treatment History

- Nil to date

Special Questions

General Health

- Nil significant. Nil surgery.

Medications

- Panadol. Up to 8 tablets daily since onset, resulting in "mild relief".
- Nil steroids.

Investigations

- Nil to date.

Weight

- Stable.

Spinal Cord Screening

- Nil paraethesia/anaesthesia
- Nil gait abnormalities

Cauda Equina Screening

- Nil change to bladder or bowel function
- Nil paraesthesia/anaesthesia in saddle region

Self Report Outcome Measures

Outcome Measure	Score	Notes
Modified Oswestry Disability Index	34%	Higher score = greater disability
Numeric Pain Rating Scale: Current	4 / 10	Higher score = greater current pain
Numeric Pain Rating Scale: Worst in past 24/24	6 / 10	Higher score = greater pain in past 24/24
Numeric Pain Rating Scale: Best in past 24/24	4 / 10	Higher score = greater pain in past 24/24
Fear Avoidance Belief Questionnaire: Work Subscale	4 / 42	Higher score = greater fear and avoidance of work-related activities
Fear Avoidance Belief Questionnaire: Physical Activity Subscale	7 / 24	Higher score = greater fear and avoidance of physical activities
Modified Somatic Perception Questionnaire	5 / 39	Higher score = heightened somatic awareness

PHYSICAL EXAMINATION

Observation

Standing: Slight forward head posture with protracted scapulae. Decreased kyphosis at mid-thoracic spine. Increased lumbar lordosis. Nil lateral shift. Nil other significant findings.

Sitting: Slight forward head posture. Nil other significant findings.

Physiological Movements

Lumbar Spine

Flexion: Mild restriction of movement (43°). Pain 6/10. No aberrant movement pattern. Patient reports to feel 'vulnerable' at the beginning of the movement.

Extension: Moderate restriction of movement (12°). Nil change in pain.

(R) Lateral Flexion: Mild restriction of movement (25°). Pain 5/10.

(L) Lateral Flexion: Moderate restriction of movement (18°). Pain 6/10.

(R) Side Glide: Mild restriction of movement. Pain 5/10.

(L) Side Glide: Moderate restriction of movement. Pain 6/10.

Rotation: L = R. Nil restriction of movement. Nil change in pain.

Extension/Rotation: L = R. Nil change in pain.

Hip

Internal Rotation: L = R. 40°. Nil change in pain.

Repeated Movements

Flexion in Standing x 10: Nil effect on symptoms or mechanical presentation

Extension in Standing x 10: Nil effect on symptoms or mechanical presentation

Flexion in Lying x 10: Nil effect on symptoms or mechanical presentation

Extension in Lying x 10: Nil effect on symptoms or mechanical presentation

Neurological Testing

Motor: Nil abnormalities detected

Sensation: Nil abnormalities detected

Reflexes (KJ & AJ): Nil abnormalities detected

Neurodynamic Testing

Straight Leg Raise: L = R. 70° of hip flexion. Nil change in symptoms.

Slump Test: L = R. Nil restriction of movement. Nil change in symptoms.

Sacroiliac Joint Pain Provocation Tests

Distraction: nil change in pain.

Thigh Thrust: L = R. nil change in pain.

Gaenslen's Test: L = R. nil change in pain.

Compression: nil change in pain.

Sacral Thrust: nil change in pain.

Patrick Sign: L = R. nil change in pain.

Muscular/Motor Assessment

Transversus Abdominis – Prone test with pressure biofeedback unit

- Pressure reduced by 3mmHg (70mmHg – 67mmHg) x 10sec with normal respiratory pattern.

- Nil spinal movement during test and nil palpable abdo bulging.
- Consistent test response across 10 repeated tests.

Segmental Lumbar Multifidus – Prone palpation test

- Nil palpable deficits

Special Tests

Prone Instability Test: negative for pain reduction.

Palpation

Passive Accessory Intervertebral Movement Testing

- Generally, hypomobile throughout thoracic and lumbar spine
- Familiar pain reproduced with central and right unilateral PA testing over L_4 and L_5 to a similar degree
- Nil evidence of hypermobility at any spinal segment

Figure 5.1 Case scenario

CLINICAL PREDICTION RULES FOR LOW BACK PAIN

DIAGNOSTIC DECISION MAKING ***Diagnostic CPRs***	
Possible functions: 1. To help clinicians quantify the likelihood of a particular diagnosis given the presence or absence of certain signs and symptoms. 2. To help identify which patients do not require further testing for a particular diagnosis given the presence or absence of certain signs and symptoms.	Example: Diagnosis of SIJ mediated pain[1] Positive on CPR if 3 or more of the following tests reproduce the patient's complaint: 1. Distraction 2. Thigh thrust 3. Gaenslen's test 4. Compression 5. Sacral thrust. +LR = 4.16 (95%CI 2.16-8.39) -LR = 0.12 (95%CI 0.02-0.49)
TREATMENT DECISION MAKING ***Prescriptive CPRs***	
Possible function: 1. To help clinicians identify which patients have a higher likelihood of success for a given intervention in comparison to an alternate intervention	Example: Success with lumbopelvic manipulation compared with aerobic and strengthening exercises[2] Positive on CPR if 4 or more of the following tests reproduce the patient's complaint: 1. Duration of current episode < 16 days 2. No symptoms below the knee 3. Fear-avoidance belief questionnaire work subscale < 19 4. At least 1 hypomobile lumbar segment 5. At least 1 hip with more than 35° of internal rotation ROM. For those positive on CPR, NNT = 1.3 (95%CI 1.1-1.9)
PROGNOSTICATION ***Prognostic CPRs***	
Possible functions: 1. To help clinicians quantify the likely clinical outcome for an individual given the presence or absence of certain signs and symptoms 2. To help clinicians identify which patients may not require intervention	Example: Time to recovery from acute low back pain[3] Positive on CPR if all 3 of the following are present: 1. Baseline pain ≤ 7/10 2. Duration of current episode ≤ 5 days 3. ≤ 1 previous episodes. For those positive on CPR, median days to recovery = 6 days (95%CI 4-8)

CPR = clinical prediction rule, CI = confidence interval, NNT = number needed to treat

[1] Laslett M, Young SB, Aprill CN, McDonald B. Diagnosing painful sacroiliac joints: a validity study of a McKenzie evaluation and sacroiliac provocation tests. Australian Journal of Physiotherapy 2003;49(2):89-97

[2] Childs JD, Fritz JM, Flynn TW, Irrgang JJ, Johnson KK, Majkowski GR, et al. A clinical prediction rule to identify patients with low back pain most likely to benefit from spinal manipulation: a validation study. Annals of Internal Medicine 2004; 141(12):920-8.

[3] Hancock MJ, Maher CG, Latimer J, Herbert RD, McAuley JH. Can rate of recovery be predicted in patients with acute low back pain? Development of a clinical prediction rule. European Journal of Pain 2009;13(1):51-5.

Figure 5.2 Examples of clinical prediction rules for low back pain

5.3.4 Data analysis

The audio file from each digitally recorded focus group was transcribed and analysed using Thematic Networks (Attride-Stirling, 2001). Focus group transcriptions were uploaded to NVivo (version 9, QSR International Pty Ltd, Victoria, Australia) and pseudonyms were substituted for participant names and places. Two processes were used to code the data. Transcripts were read several times by the first author and then segments of text were coded based upon the identification of recurrent themes (Morse, Barrett, Mayan, Olson, & Spiers, 2008). All recurrent themes identified by the first author were coded during this process independent of personal beliefs concerning their relationship with the study's research questions. This was done to enhance the trustworthiness of the data analysis process by ensuring that the full research team were involved in the selection of themes that best related to the study's research questions. Following this, the first author coded data according to a list of potential barriers to the adoption of CPRs identified in the literature and presented in Table 5.1 (p. 179).

Themes identified from both rounds of coding were examined by the research team and those that best related to the study's research questions were included. Themes that were considered to be overlapping were combined and re-coded to produce smaller set of mutually exclusive themes. Clusters of themes with commonality were arranged into organising themes (Attride-Stirling, 2001). Organising themes were grouped together based on the research question they addressed.

Member checking was used to enhance the trustworthiness of the study's findings (Krefting, 1991; Petty, Thomson, & Stew, 2012b). This process involved e-mailing a one-page collated summary of the research team's interpretation of key themes from the focus groups to participants and inviting their feedback on any and all aspects of the summary (Mays & Pope, 2000). The feedback provided by participants was considered by all members of the research team in regards to whether it confirmed or challenged the research team's interpretation of the findings. The trustworthiness of the study's findings was also improved with the use of peer debriefing, whereby the selection and organisation of included themes, and the consideration of participant feedback, were discussed and agreed upon by all members of the research team (Creswell & Miller, 2000; Petty et al., 2012b).

5.4 Findings

Four focus groups involving a total of 26 participants were conducted. Participant characteristics are summarised in Table 5.3 below. The participant sample were predominantly male (77%), worked in a private setting (81%) and had an average of 15.5 years (SD 11) of clinical experience. Three new graduate physiotherapists participated in the study. Nine participants were previously known to the first author through various professional networks. The use of pseudonyms and peer debriefing throughout data analysis minimised any risk of bias that could result from existing researcher-participant relationships.

Table 5.3 Participant characteristics

Participant pseudonym	Gender	Clinical experience (years)	Place of entry-level qualification	Current work setting	Focus group
David	Male	3	ACT, Australia	Public hospital	1
Colleen	Female	31	NSW, Australia	Private practice	1
Jason	Male	24	NSW, Australia	Private practice	1
Mark	Male	14	NSW, Australia	Private practice	1
Kevin	Male	6	NSW, Australia	Public hospital	1
Courtney	Female	34	NSW, Australia	Private practice	2
Donald	Male	29	NSW, Australia	Private practice	2
Logan	Male	19	Germany	Private practice	2
Corey	Male	23	NSW, Australia	Private practice	2
Lachlan	Male	27	NSW, Australia	Private practice	2
Bill	Male	23	NSW, Australia	Private practice	3
Rupert	Male	8	NSW, Australia	Public hospital	3
Henry	Male	17	NSW, Australia	Public hospital	3
Fred	Male	7	NSW, Australia	Private practice	3
Cameron	Male	18	United Kingdom	Public hospital	3
Christian	Male	3	NSW, Australia	Private practice	4
Neil	Male	3	NSW, Australia	Private practice	4
Clayton	Male	1	NSW, Australia	Private practice	4
Kathleen	Female	1	NSW, Australia	Private practice	4
Tyrone	Male	6	NSW, Australia	Private practice	4
Erik	Male	28	NSW, Australia	Private practice	4
Melanie	Female	1	NSW, Australia	Private practice	4
Hugh	Male	25	Victoria, Australia	Private practice	4
Harold	Male	21	NSW, Australia	Private practice	4

Participant pseudonym	Gender	Clinical experience (years)	Place of entry-level qualification	Current work setting	Focus group
Charlene	Female	missing	NSW, Australia	Private practice	4
Yvette	Female	missing	United Kingdom	Private practice	4

The first round of coding led to the identification of 62 recurrent themes. Sixteen (62%) of the literature-informed themes (Table 5.1) were identified within the transcribed text. Integrating the two coding processes led to the development of 27 non-overlapping themes relevant to the study's research questions which were arranged into 7 organising themes (Table 5.4 below). Participant feedback on the summarized themes was primarily confirmatory and did not lead to substantial modifications.

Table 5.4 Summary of themes

Themes	Organising Themes	Research Questions
Awareness of CPRs is varied	Awareness and familiarity	Knowledge about CPRs
Familiarity with CPRs is varied		
Conceptualisation of CPRs is varied	Conceptualisation	
CPRs are the formalisation of existing reasoning processes		
CPRs are evidence-based practice	Facilitative attitudes	Attitudes toward CPRs
CPRs enable greater confidence in making predictions		
CPRs may help inform decision-making		
CPRs may help novice clinicians		
CPRs may positively challenge traditional reasoning strategies		
Numeric data may be helpful		
CPRs are complicated	Inhibitive attitudes	
CPRs are or could become fads		
CPRs could cause intellectual laziness		
CPRs have limited generalisability		
CPRs may challenge clinicians' autonomy		
CPRs may not work because treatment techniques are too varied		
CPRs oversimplify the complexities of a clinical presentation		
Dislike of the word 'rule'		
Existing CPRs are not yet ready to be applied		
LBP is too complicated for CPRs		
No personal need for a CPR		
Some CPRs are used without knowledge that they are CPRs	Current practices / behaviours	Practices / behaviours and implementation issues
Use of CPRs is varied		
CPRs may function as second opinions or as a safety net	CPRs within the clinical reasoning process	
CPRs should not be used in isolation		
CPRs should only be applied to patients for which they have been developed		
Third party payers may use CPRs	Third party payer issues	

5.4.1 Knowledge

Two organising themes related to the research question regarding physiotherapists' knowledge of CPRs were identified.

Awareness of and familiarity with LBP CPRs

Participants reported mixed awareness of CPRs with some participants (n=5) not having previously encountered the term or concept. A CPR developed to identify patients with LBP who are more likely to respond favourably to spinal manipulation (Flynn et al., 2002) was the most commonly recognised CPR, although many of the criteria that constituted the tool were not commonly identifiable by participants.

Fred: *I have a vague recollection (of CPRs) from uni but I must admit I'm pretty ignorant of them.*

Kevin: *I know there's lots of studies on clinical prediction rules with low back pain and there's even some meta-analyses of those studies.*

Conceptualisation of LBP CPRs

Parallels were identified between CPRs and patient management paradigms such as Mechanical Diagnosis and Therapy (McKenzie & May, 2003), with the sub-classification of patients into smaller and more homogenous groups. Most believed that CPRs were simply the formalisation of clinical reasoning

strategies like pattern recognition that physiotherapists commonly use. For example:

> *Jason:* *It's just really formalising and detailing something that we do all the time…. It's what your experience is developing and that's developed for you to be able to look at something and say, "Look, I think this is what this is and I know if I go down this path with it I'm going to be likely to get a good outcome"…It's just an informal thing, they're part of our art.*

5.4.2 Attitudes

A wide range of attitudes toward LBP CPRs were expressed. The identified themes were clustered into two organising themes based on their facilitative or inhibitive influence on the implementation of CPRs in clinical practice.

Facilitative attitudes towards LBP CPRs

CPRs were viewed positively by some participants as consistent with evidence-based practice. The conscientious use of numerical data and probabilities to inform decision-making was welcomed by some participants, although it was acknowledged that CPRs could be applied clinically without reference to numbers. Statistically derived data were considered by some to be valuable in challenging existing models and assumptions. It was

considered that CPRs may be helpful in informing clinical decision-making for LBP, in particular by enabling clinicians to have greater confidence in their predictions.

Christian: *I do think that it (using CPRs) is evidence based practice…One little thing that we pick up from a clinical prediction rule might inform something that we might change in our practice.*

Jason: *I think that it's (developing and using CPRs is) a big step in the right direction for us as clinicians…We'd probably have more confidence in being able to say to people "Look, if we do this (treatment) for people with your sorts of signs and symptoms, we get a good outcome..".*

The value of CPRs for LBP was considered to be clinician-dependent, with most participants expressing a view that novice clinicians with limited experience may benefit the most.

Tyrone: *It (CPRs) certainly would have helped as a new grad six years ago… If I had those (CPRs), it would have made it a lot easier…*

Kathleen: *I think for a new grad these clinical prediction rules are excellent…*

Inhibitive attitudes towards LBP CPRs

Some participants viewed CPRs as overly complicated and seldom generalisable to the patients that they treat.

Corey: *…my scant reading of them (CPRs) is that they're too complicated and not trustworthy enough..*

Cameron: *You know I've seen a few of them come off the market now, the manip(ulation), the stabilisation, even the directional preference one which I find a more common utility, but even that, I'm still finding unfortunately my patients don't fit any of these.*

The term 'rule' was viewed negatively by a number of participants and there was a perception by some that CPRs oversimplified the complexities of a clinical presentation and the clinical reasoning process. Variability in patients, as well as the way in which treatments such as manipulation are applied, was considered by some to adversely affect the utility of CPRs for LBP presentations. Clinical experience was believed by some to obviate the need for LBP CPRs.

Rupert: *You've spent twenty-three years developing your own algorithm and you go, that would be a thousand times more complicated than anything that gets put into this thing and it's not about this tool trying to replace that but*

then you also think oh but hang on, I can sort of do this automatically almost, why do I change to that.

Some physiotherapists believed that CPRs for LBP were not sufficiently developed at this time to enable confident application in the clinical setting and that their premature adoption could have negative implications for the profession.

Cameron: *… I'm just a bit worried about some of this (CPR research) being the next big thing and being part of the vernacular in every day clinical practice before it's been tested in multiple populations.*

David: *My thoughts are that they're not absolute at this stage… I like the overall idea but I think there' a long way to go (before CPRs can be used).*

There was a perception that CPRs could become the next professional 'fad' and be viewed by some clinicians as a sort of magical panacea, despite the current lack of evidence of a positive impact.

Bill: *It reminds me of a patient with arthritis searching for a cure and if someone is proposing that this particular thing is fantastic they'll jump on the bandwagon and do it because really there's not really any great answer for arthritis, arthritis pain, there's not really any great answer*

for low back pain and how physios are treating low back pain. We want it, we're looking for it, it's not there.

Rupert: *I don't want to feel like I've just drunk the Kool Aid… It's almost too easy.*

Many clinicians expressed a concern regarding the potential of CPRs to cause 'intellectual laziness', as well as negatively impacting upon the autonomy of the clinician.

Fred: *You could start to get intellectually sloppy, you know, 'clinical reasoning sloppiness'.*

Cameron: *What about a society that we're working in in 30 years' time where these prediction models say you must go this direction and you can't have that treatment and you have to have this treatment… "I'm sorry doctor or physio you don't have that latitude and that freedom" …*

5.4.3 Practices / behaviours and implementation issues

Three organising themes related to the research question concerning the practices/behaviours of physiotherapists in relation to LBP CPRs.

Current practices / behaviours

A small minority (n=3) of participants expressed that they would have used a CPR within their assessment and management of the fictitious LBP case scenario. Of those previously familiar with CPRs for LBP, only a small number (n=7) of participants acknowledged that they had ever used them to inform their decision-making in clinical practice. A greater number expressed that they had used CPRs for non-LBP presentations, with the most commonly cited rules being the Ottawa ankle (Stiell et al., 1992) and knee (Stiell, Greenberg, et al., 1995) rules, and Wells et al's CPR for deep vein thrombosis (Wells et al., 1995). A CPR for the diagnosis of sacroiliac joint mediated pain (Laslett et al., 2003) was reported to be used by several participants both for the fictitious case scenario and within routine clinical practice, although it was not always recognised by participants as a type of CPR.

Melanie: *Well, at uni they don't even teach you - for example that SIJ (sacroiliac joint), that that's a clinical prediction rule. That's just how you assess an SIJ. That's not a clinical prediction rule.*

CPRs within the clinical reasoning process

Clinicians believed that when CPRs are used in clinical practice they should not be used in isolation, but rather used within the suite of clinical reasoning

processes physiotherapists typically employ including the consideration of patient expectations and preferences.

Kathleen: *I think we've still got to use your own clinical judgement and your own intuition as well. I think these (CPRs) need to be used to complement all of that. I don't think we can just rely solely on clinical prediction rules.*

Some participants considered that CPRs best serve clinical practice as second opinions or as 'safety nets', and are able to be overruled by the clinician.

Colleen: *They're (CPRs) confirming things that you might be a little bit unsure about.*

Henry: *The rule might say one thing but we've already decided that we may not necessarily do that.*

Many stressed the importance of restricting the use of CPRs to the patient populations for which they were intended.

Donald: *…if those rules are built up around a certain type of patient and someone expects us to apply them to every sort of patient, you're asking for chaos, it's not going to work.*

Third party payer issues

A recurring theme identified across the focus groups was that CPRs may be used by third party payers, such as insurance companies and government funded health services. Some perceived this as beneficial and thought this may help minimize over-servicing and the use of ineffective treatment modalities. The majority of participants however, considered the use of CPRs by third party payers as predominantly negative and believed that this would restrict clinician autonomy and preclude the incorporation of patient preferences into decision-making.

Hugh: *I just don't want to get painted into a corner where if I don't treat according to these clinical prediction rules WorkCover (government insurance) might say…"Well, that's not evidence based. That's not gold standard treatment. Why are we paying you to treat this person when it's not following your clinical prediction rule?"*

Kevin: *WorkCover (government insurance) would use a rule to cut people off from funding… They'd use it in the worst possible way.*

5.5 Discussion

The findings of this study have highlighted that a range of factors related to clinician knowledge, attitudes and practices / behaviours may influence the adoption of LBP CPRs into physiotherapy clinical practice. Many of these

factors share similarities with the identified barriers to the adoption of other innovations in physiotherapy, including the use of clinical practice guidelines (Côté et al., 2009), outcome measures (Abrams et al., 2006) and the application of evidence-based practice (Jette et al., 2003).

Knowledge of CPRs, in terms of awareness and familiarity, was quite mixed among the participants in this study, with some not having previously encountered this term or concept. This might suggest that as an innovation, CPRs have not as yet permeated into the mainstream conversation of practising clinicians at least within parts of NSW, Australia. Previous research suggests that awareness of CPRs may be highest in the countries in which the tools have been developed (in this case predominantly the US) (Eagles, Stiell, Clement, Brehaut, Taljaard, et al., 2008; Graham et al., 2001) and subsequently the knowledge about LBP CPRs of participants in this study may plausibly contrast to that of clinicians in those regions. Addressing knowledge gaps about LBP CPRs may be an important first step in any strategy designed to enhance the adoption of these tools.

CPRs were seen by many as the formalisation of a traditional reasoning process used by experienced clinicians and this view may have influenced some of the attitudes expressed by participants in this study. Attitudes were notably diverse and encompassed positions that may be seen as both facilitating and inhibitive to the implementation of LBP CPRs. A key belief that emerged across the focus groups was that the benefit of using these tools was experience-dependent. That is, novice physiotherapists may benefit from

the use of LBP CPRs, however more experienced clinicians had limited need of such tools. This belief may have substantial implications for the adoption of LBP CPRs in physiotherapy practice and warrants timely investigation.

For a LBP CPR to be successfully incorporated into physiotherapy practice, the findings of this study suggest there are modifiable characteristics that may enhance its acceptability. While participants in this study expressed they would not blindly adhere to a CPR independent of and naïve to other clinical information and decision-making processes, the word 'rule' had negative connotations and was considered by some to be an implementation barrier. Avoiding the term 'rule' may therefore be a simple but important strategy in improving the use of CPRs. Similar to this study's findings, previous research in the field of Emergency Medicine has found that less than 10% of physicians prefer the term 'rule' when describing these tools (Graham et al., 2001). Less authoritarian terms like 'tool' or 'guideline' may be more palatable and perhaps more consistent with the intended function of CPRs – that is, to help inform decision-making, not dictate decision-making (Swets et al., 2000a). The findings of this study also highlight the importance of ensuring that CPRs appear uncomplicated and easy to use. This may include making sure that all aspects of the tool are clear and unambiguous (Brehaut et al., 2010), using graphical aids where appropriate (Björk, Ekelund, & Ohlsson, 2012) and soliciting input from practicing clinicians throughout their development (Reilly & Evans, 2006).

Participants in this study infrequently incorporated CPRs into their assessment or management of LBP. This stands in contrast to a recent US study that found that 40% of surveyed physical therapists who routinely employ thrust manipulation report using a CPR (Learman et al., 2012). Learman et al. (2012) further identified that clinicians who reported to use a CPR were no more likely to perform manipulation in the presence of contraindications than those who do not use the tool. That is, clinicians did not blindly 'obey' a CPR but rather considered it within the context of all other presenting information. Physiotherapists in the present study reported a similar attitude toward the use of CPRs and believed that the optimal use of these tools was nested within the suite of clinical reasoning strategies clinicians typically employ. Further, physiotherapists felt strongly that third-party payers, naïve to all of the available information, should be prevented from using CPRs to direct the clinician to provide particular forms of therapy.

A limitation of the current study is that the findings represent the thoughts and opinions of study participants and may not be generalisable to other populations. Readers should carefully consider the methods and analytic strategies used in this research when considering the degree to which the findings may be transferable to their own setting (Krueger & Casey, 2009).

5.6 Conclusions

This is the first study outside of a US context to explore the knowledge, attitudes and practices of physiotherapists in regards to LBP CPRs. Most of

the participants in this study reported to be aware of LBP CPRs however very few reported to have ever used them to inform their decision-making. Barriers to the use of LBP CPRs identified in this study included a negative connotation associated with the term ‘rule’, a perception that CPRs are overly-complex and infrequently applicable, clinical experience obviating the need for such tools, and the potential threat to clinical autonomy and for misuse by third-party payers. Study participants felt that LBP CPRs were best used within the suite of clinical reasoning processes physiotherapists typically employ and considered as second opinions or safety nets that were be able to be overruled by the clinician. Consideration of these views may inform strategies that will optimise the development of LBP CPRs with the highest potential to positively influence physiotherapy practice and implementation strategies that will optimise their adoption into clinical practice.

CHAPTER 6

AUSTRALIAN PHYSIOTHERAPISTS' PRIORITIES FOR THE DEVELOPMENT OF CLINICAL PREDICTION RULES FOR LOW BACK PAIN: A QUALITATIVE STUDY

This chapter has been published in a peer-reviewed scientific journal (Appendix 4):

Haskins, R., Osmotherly, P. G., Southgate, E., & Rivett, D. A. (2014). Australian physiotherapists' priorities for the development of clinical prediction rules for low back pain: a qualitative study. *Physiotherapy*, *101*(1), 44-49.

The work presented in this manuscript was completed in collaboration with the co-authors (Appendix 1). The ethical approval for the study reported in this chapter appears in Appendix 2.

Overview

The focus of the study presented in this chapter is upon physiotherapists' priorities regarding the development of CPRs for LBP. This is the third of the five studies that comprise this thesis and was conducted concurrently with

study two (Chapter 5, p.174). As discussed in section 3.9 (p. 121), it has been proposed that CPRs should function to address the perceived needs of their intended target users from the outset of their development. All else being equal, it is hypothesised that CPRs will be more likely to be considered useful and subsequently more likely to be implemented if they match the identified needs of clinicians.

However, the findings of Study 1 (Chapter 4) highlighted that the CPRs developed for LBP within the physiotherapy profession to date have wide-ranging and divergent clinical functions. It is possible that this may reflect a current lack of understanding regarding the specific needs and preferences of clinicians for such tools. This study aims to help address this gap and extends upon the previous work detailed in this thesis by examining a critical aspect of CPR development that may influence their capacity to improve the physiotherapy management of LBP.

6.1 Abstract

Objective:

To identify the types of clinical prediction rules (CPRs) for low back pain (LBP) that Australian physiotherapists wish to see developed and the characteristics of LBP CPRs that physiotherapists believe are important.

Design:

Qualitative study using semi-structured focus groups.

Setting:

Metropolitan and regional areas of New South Wales, Australia.

Participants:

Twenty-six physiotherapists who manage patients with LBP (77% male, 81% private practice).

Results:

Participants welcomed the development of prognostic forms of LBP CPRs. Tools that assist in identifying serious spinal pathology, likely responders to interventions, patients who are likely to experience an adverse outcome, and patients not requiring physiotherapy management were also considered useful. Participants thought that LBP CPRs should be uncomplicated, easy to remember, easy to apply, accurate and precise, and well-supported by research evidence. They should not contain an excessive number of variables, use complicated statistics, or contain variables that have no clear logical relationship to the dependent outcome. It was considered by participants that LBP CPRs need to be compatible with traditional clinical reasoning and decision-making processes, and sufficiently inclusive of a broad range of management approaches and common clinical assessment techniques.

Conclusion:

There were several identified areas of perceived need for LBP CPR development and a range of characteristics such tools need to encompass to be considered clinically meaningful and useful by physiotherapists in this study. Targeting and incorporating the needs and preferences of

physiotherapists is likely to result in the development of tools for LBP with the greatest potential to positively impact clinical practice.

6.2 Introduction

Clinical Prediction Rules (CPRs) are an aid to support clinical decision-making (Reilly & Evans, 2006). They are generally a simple predictive tool designed to be used with individual patients (Beattie & Nelson, 2006). Unlike other forms of decision aids, CPRs most commonly provide a clinician with the quantified probability of a patient having a certain diagnosis or achieving a particular prognostic outcome (McGinn et al., 2008). CPRs come in many different formats and have been developed for a wide range of clinical problems. In some instances, CPRs provide an approach to stratified patient care, enabling treatments to be targeted to particular patient subgroups (Foster et al., 2013). Over the past decade a growing number of CPRs have been developed within physiotherapy, with many relating to the management of Low Back Pain (LBP) (Beneciuk et al., 2009; Haskins et al., 2012; May & Rosedale, 2009; Stanton et al., 2010). To date, such tools are remarkably diverse with little consistency in the type of clinical problems they aim to address. While the growth in the development of LBP CPRs is arguably important for the physiotherapy profession, the wide-ranging diversity in these tools may reflect a current lack of awareness about what clinicians actually want or need.

There has been substantial dialogue in the recent physiotherapy literature regarding the appropriate methodology required to derive, validate and assess the impact of CPRs (Hancock, Herbert, et al., 2009; Haskins, Osmotherly, Tuyl, et al., 2014; Kamper et al., 2010; Kent, Hancock, Petersen, & Mjøsund, 2010; Kent, Keating, et al., 2010; Nee & Coppieters, 2011). In contrast, there is a lack of literature about the types of problems for which CPRs should be developed or the characteristics and features they need to encompass to be considered useful by physiotherapists.

Given the substantial resources and time required to develop these tools, there is a need right from the preliminary stages of their development to ensure that CPRs will be accepted by clinicians and viewed as useful in addressing an important clinical problem (Eagles, Stiell, Clement, Brehaut, Kelly, et al., 2008). Investigating and explicitly addressing clinician needs in the preliminary development of a CPR may be an important step in supporting the effective translation of CPR research evidence into clinical practice.

The aim of this study was to explore and describe the types and characteristics of LBP CPRs that are considered important by practicing physiotherapists working in the musculoskeletal field.

6.3 Method

6.3.1 Design

A qualitative descriptive method was employed to gain insight into physiotherapists' priorities for CPR development in relation to LBP. This method is intended to provide a clear description of a specific phenomenon or experience from the perspectives of research participants (Magilvy & Thomas, 2009; Neergaard et al., 2009; Sandelowski, 2010). It concentrates on thematic analysis which seeks to identify common threads across participant perspectives in qualitative data (Vaismoradi, Turunen, & Bondas, 2013). Focus groups are commonly used in the early stages of product development to gain insight into the target consumers' thoughts and feelings about that product (Krueger & Casey, 2009). This approach is arguably well suited for exploring the needs and preferences of practising physiotherapists who are target clinical consumers of LBP CPRs. Four semi-structured focus groups, each lasting 1.5-2 hours and consisting of between 5-11 participants, were conducted across three geographic regions of New South Wales, Australia incorporating both metropolitan and regional areas. The first focus group was moderated by the third author (female, PhD, senior lecturer, experienced moderator) and the following three groups were moderated by the first author (male, B.Phty(Hons), physiotherapist, student of qualitative research methods). Groups were conducted outside of business hours to facilitate recruitment and held on a locally-based university campus or in a

private function centre. Ethical approval for the study was granted by the University of Newcastle's Human Research Ethics Committee.

Previous research in the field of Emergency Medicine (Brehaut et al., 2010; Eagles, Stiell, Clement, Brehaut, Kelly, et al., 2008) informed the development of a focus group schedule of questions. Participants were asked about areas of their practice with patients with LBP they thought may benefit from a CPR and the characteristics such tools require to be useful and meaningful. Clinicians were asked to share their beliefs on how LBP CPRs are most appropriately incorporated within physiotherapy practice and about any advice they would give to researchers who were considering developing LBP CPRs. Each focus group was recorded using a digital voice recorder. The audio file from each group was transcribed and used for data analysis.

6.3.2 Participants

Participants were recruited according to a purposive sampling framework (Greenwood & Parsons, 2000) that would reflect the likely clinical consumers of LBP CPRs. Participants were selected according to the following characteristics: registered practising physiotherapist; working in public or private practice; having a caseload inclusive of patients with LBP; and proficiency in English. The study design deliberately included clinicians with a range of clinical experience from recent graduates to those with several decades of practice. Public listings were used to identify and recruit potentially eligible participants, in addition to an advertisement within an

electronic bulletin e-mailed to all members of the Australian Physiotherapy Association.

6.3.3 Data analysis

Focus group transcriptions were uploaded to NVivo (version 9, QSR International Pty Ltd, Victoria, Australia) and pseudonyms were substituted for participant names and places. Transcripts were read several times by the first author and then segments of text were inductively coded. Inductive coding is data-driven and is different from deductive coding in that there is minimal attempt to interpret the data through pre-existing categories derived from the literature (Fade & Swift, 2011). Clusters of basic themes with commonality were arranged into organising themes (Figure 6.1 and Figure 6.2) (Attride-Stirling, 2001). Ongoing analysis with inductive coding and the development of these thematic levels (Morse et al., 2008) occurred over the course of the focus groups. Thematic saturation (Fade & Swift, 2011) occurred at the fourth focus group. Organising themes were then related to the study's research questions to develop a smaller number of organising themes for each of the research questions (Attride-Stirling, 2001). Ongoing analysis informed a decision to include a one page explanation (Figure 5.2, p. 191) describing CPRs for focus groups 3 and 4 to help direct participants to the topic of the focus group, that is CPRs.

Physiotherapists' priorities for LBP CPR development	
Organising Themes	**Basic Themes**
Diagnosis	Physiotherapists want LBP CPRs that enable the early and accurate identification of serious spinal pathology
	There is limited desire for the development of CPRs that facilitate the sub-classification of non-specific LBP by pathoanatomic diagnosis
Intervention	Physiotherapists want LBP CPRs that predict non-success, worsening or no need for intervention
	Physiotherapists want LBP CPRs that accurately identify likely responders to intervention
Prognosis	Physiotherapists have strong desire for LBP CPRs that accurately predict a patient's probable prognosis

Figure 6.1 Summary of the types of LBP CPRs physiotherapists wish to see developed

To enhance credibility and confirmability, peer debriefing with other members of the research team was used throughout all stages of the data analysis (Creswell & Miller, 2000; Petty et al., 2012b). Three members of the research team (RH, PO, DR) have previous research experience and a declared interest in the development of LBP CPRs. Member checking was employed by sending a summary of the research team's interpretation of the key study themes to participants and inviting and integrating their feedback into the study's findings (Mays & Pope, 2000). Dependability was enhanced by confirmation of the accuracy of audio transcriptions (Slade, Molloy, & Keating, 2009) and cross-checking with field notes.

Physiotherapists' requirements of LBP CPRs	
Organising Themes	**Basic Themes**
Application considerations	LBP CPRs need to be simple, practical and easily applied
	LBP CPRs should be developed and applied to clearly defined presentations
	LBP CPRs need to be compatible with traditional clinical reasoning and decision-making strategies
Credibility and meaningfulness	LBP CPRs need to make sense and should contain predictor variables that have a clear logical relationship with the dependent outcome
	LBP CPRs need to be meaningful and clinically relevant
Performance expectations	Physiotherapists require confidence that use of a LBP CPR will lead to improved patient outcomes
	LBP CPRs need to be accurate to be useful

Figure 6.2 Summary of physiotherapists' desired characteristics of LBP CPRs

6.4 Results

6.4.1 Participants

Twenty-six physiotherapists participated in this study including three new-graduates (Table 6.1).The majority of participants were male (77%) and were working in a private clinical setting (81%). The average amount of clinical experience was 16 years (SD 11). Nine participants were previously known to the first author through various professional networks. No participants were previously known to the third author who moderated the first focus group. The use of pseudonyms and peer debriefing throughout data analysis

minimised any risk of bias that could result from existing researcher-participant relationships.

Table 6.1 Participant characteristics

Participant pseudonym	Current work setting	Sex	Clinical experience (years)
Lauren	Private practice	Female	1
Sophie	Private practice	Female	1
William	Private practice	Male	1
Adam	Private practice	Male	3
Ethan	Private practice	Male	3
Phil	Public hospital	Male	3
Clive	Public hospital	Male	6
Mick	Private practice	Male	6
Brett	Private practice	Male	7
Jon	Public hospital	Male	8
Graham	Private practice	Male	14
Sam	Public hospital	Male	17
Tim	Public hospital	Male	18
Francis	Private practice	Male	19
Jerry	Private practice	Male	21
Craig	Private practice	Male	23
Jeff	Private practice	Male	23
Dave	Private practice	Male	24
Adrian	Private practice	Male	25
George	Private practice	Male	27
Terry	Private practice	Male	28
Brian	Private practice	Male	29
Elaine	Private practice	Female	31
Emma	Private practice	Female	34
Julie	Private practice	Female	Not provided
Mary	Private practice	Female	Not provided

6.4.2 Types of LBP CPRs that physiotherapists wish to see developed

Three organising themes were derived for the research question concerning the types of LBP CPRs physiotherapists wish to see developed (Figure 6.1 above).

Diagnosis

Participants (n=9) expressed a desire for CPRs to be developed that would enable them to accurately diagnose serious spinal pathology, as the following quote illustrates.

> *Clive:* *It would be really nice if we could have a clinical prediction rule that would rule out the really heavily nasty stuff that can exist in spines… A little case story that I use to highlight is that I know a girl recently who's had a severe spinal cord injury with a history of having years and years of chiropractic treatment…There's nothing wrong that the chiropractor did in terms of his treatment …Unfortunately it was a tumour and it wasn't a musculoskeletal thing. But they can present as musculoskeletal things and to have a clinical prediction rule that was really good at being able to tell the difference would be nice.*

There was limited identified need for the development of CPRs that facilitate the sub-classification of non-specific LBP by pathoanatomic diagnosis. A few

participants (n=5) welcomed such tools, with similarities drawn with the classification of peripheral joint presentations. Other participants (n=13), however, believed that they would have little impact on clinical practice as pathoanatomic diagnoses are not commonly influential in physiotherapy management decisions, as the following quote illustrates.

Jeff: *I definitely wouldn't say that I use it (CPRs) in terms of identifying pathology, which is the medical model, which I don't think applies very well to physiotherapy. I think we're better to categorise people in groups of mechanical presentations.*

Intervention

Participants (n=10) wanted CPRs that would predict which patients with LBP would worsen from a particular intervention, or who would not benefit from physiotherapy treatment.

Emma: *It's also (helpful for CPRs to inform) when should we not treat this patient, and send them to someone, say they're a surgical candidate or whatever. That's how I would use a clinical prediction rule.*

Brian: *Rules that tell us when to lay off them (not treat patients with LBP) are good.*

LBP CPRs that identify patients who are more likely to achieve a successful outcome from a particular intervention were considered useful by the majority of participants (n=16) and were thought to be helpful in informing treatment decision-making.

Adrian: *It would be great to have an idea that if you do this treatment on this type of thing that you're going to get a great response. That would be fantastic rather than farting around (aimlessly continuing) with the stuff that you did for 10 years.*

Jerry: *I think if we had a prediction rule for treatment that said that this one generally gets better with this treatment…then that definitely would prejudice my treatment (decision making) for sure.*

Prognosis

Participants (n=20) commonly expressed a view that CPRs which can accurately predict a patient's probable prognosis would be very valuable. Several physiotherapists (n=8) expressed that determining a patient with LBP's likely prognosis was particularly challenging. Therefore, participants wanted CPRs that would predict time to recovery from a presentation, time to return to work, time to return to normal physical activity, likelihood of persisting symptoms, likelihood of requiring surgery, and the likelihood of experiencing a recurrence.

Brian: *Rules that talk about when they (patients with LBP) should return to work or not return to work - that would probably be a big (priority).*

Jeff: *That's the horrible question that is asked to us. I think it's the hardest question of all, "how long?"*

Tim: *I think that's where these prognostic models can actually help us.*

6.4.3 Physiotherapists' desired characteristics of LBP CPRs

Three organising themes were derived for the research question concerning physiotherapists' required characteristics of LBP CPRs to maximize their clinical utility (Figure 6.2 above).

Application considerations

Participants felt that LBP CPRs must be simple and easy to apply for the benefit of both clinicians and patients (n=13). Many considered it was important that tools focused on being practical and not being overly complicated by too many variables or complex statistics. The interface of the prediction tool needed to be compatible with the clinical environment and preferably the CPR could be able to be easily memorised. Some (n=5) also expressed a view that predictor variables should be able to be obtained in a timely fashion using existing routine clinical measures and not require the need for sophisticated equipment.

Adam: *I think keeping them simple so practitioners can use them but also patients, in the sense that we're involving them in decision making.*

Ethan: *I'd have to be able to use it straight up without looking at the computer and without researching it.*

Physiotherapists also expressed a view that LBP CPRs would be more clinically useful if they were developed for very clear, well-defined patient presentations.

George: *I would say pick a clearly defined presentation, a typical well-defined presentation and work on that particular presentation, a predictor rule for that.*

Francis: *You have to really have a very specific subgroup and then you can make perhaps a rule that applies for that subgroup only, and that it makes sense.*

It was commonly reported by participants (n=19) that LBP CPRs need to be compatible with their clinical reasoning and decision-making processes. CPRs were unlikely to be considered useful if they were viewed to be based on discordant management paradigms or insufficiently encompassing of a preferred management philosophy. Participants felt it was important that the developers of CPRs investigate the types of assessment techniques clinicians typically use and believe to be important in decision-making.

Terry: *It depends how broad based your clinical predication rules are. Whose models are they based on or which treatment or assessment philosophy are they based on?*

Emma: *We follow clear clinical reasoning paths and…it has to fit into your clinical reasoning.*

Credibility and meaningfulness

Participants expressed a view that the selection of predictor variables in LBP CPRs should be based on sound clinical reasons and not solely on statistical procedures. A logical relationship (e.g., biological plausibility) between predictor variables and the dependent outcome was considered by some (n=6) to be very important, and if lacking, a potential threat to a rule's acceptance and implementation.

George: *To me it doesn't fit a particular model (hip rotation variable in spinal manipulation CPR), it just seems to be a statistical aberration that's popped up and they think oh, that's interesting. It would be like saying well if you've got red hair you're going to respond better to a manipulation than someone with brown hair. So it's got to match in with a clinical reasoning model I think.*

It was commonly (n=17) believed that LBP CPRs need to address meaningful and important problems to be considered clinically useful. Examples of such problems identified by participants have been summarised in Figure 6.1.

Tim: *From a face validity point of view, unless us, as clinicians see them as adding value to our day in day out practice we're unlikely to adopt them at this stage.*

Performance expectations

Participants commonly (n=16) reported a view that evidence of a positive benefit on patient outcomes was required before a LBP CPR could be confidently used by physiotherapists in clinical practice.

Phil: *The idea (of LBP CPRs) I think sounds good if I can guarantee that I'm always getting the best likelihood of doing the best thing for this patient.*

Tim: *…I'm still waiting for some real juicy stuff that's going to be quick and easy to use but have a profound effect on our practice as well, consistently outperform me...*

The accuracy and precision of predictions made by LBP CPRs was also considered by participants to be very important. Many participants (n=14) felt

that LBP CPRs needed to be highly accurate to be considered useful for clinical practice.

Ethan: *I'd have to be able to be very, very confident, so really high (likelihood) ratios, to then apply it.*

Tim: *You could have all the gut experience and thirty years' experience in the world but if this prediction model works…then you'd use it no matter who you think you were…because it can consistently outperform you…*

6.5 Discussion

The successful translation of new knowledge into practice requires consideration and incorporation of the needs of clinicians (Graham et al., 2006). This study is the first to explore the priorities of physiotherapists regarding the development of LBP CPRs. It has highlighted several areas of perceived need and a range of characteristics required of such tools to be useful to practicing physiotherapists.

One of the key findings from this study was that participants very commonly (n=20) believed that prognostic CPRs would be helpful to their clinical practice. However, of the three major types of LBP CPRs (diagnostic, prescriptive and prognostic) that have been derived in physiotherapy to date, prognostic tools number the fewest (Haskins et al., 2012). In contrast, relatively few participants (n=5) expressed a need for tools that facilitate the

pathoanatomic sub-classification of non-specific LBP, but these have been derived in the greatest number (Haskins et al., 2012). Consistent with the identified needs of physiotherapists in this study, timely validation and impact assessment of existing derived prognostic LBP CPRs (George et al., 2005; Hancock, Maher, et al., 2009) may be warranted. Additionally, physiotherapists in this study would welcome CPRs for prognostic outcomes, such as the likely time to return to work and the probability of experiencing a recurrence.

The results of this study support the ongoing development of diagnostic LBP CPRs that facilitate the early and accurate identification of serious spinal pathology, such as vertebral fracture (Henschke et al., 2009). In addition, the ongoing development of prescriptive LBP CPRs that function to identify those patients with a relatively higher likelihood of success from a given intervention was supported by participants. Notably, physiotherapists in this study (n=10) also expressed a need for tools that can identify patients who are likely to worsen from a given intervention, and also those who may not require physiotherapy management. To date, relatively little emphasis in the physiotherapy CPR literature has been placed upon the identification of these latter groups.

Participants in this study identified several modifiable properties of LBP CPRs that may enhance their clinical utility and meaningfulness. Tools that are uncomplicated, easy to remember, easy to apply and well-supported by research evidence were considered most useful. In contrast, negative

attributes reported by participants in this study related to having a large number of variables, use of complicated statistics, or the inclusion of variables that have no clear logical relationship to the dependent outcome. Study participants expressed a view that LBP CPRs need to be compatible with traditional clinical reasoning and decision-making processes and sufficiently inclusive of a broad range of management approaches and common clinical assessment techniques. The accuracy and precision of LBP CPRs was also identified to be a consideration for some physiotherapists in this study (n=14).

Some limitations of this study should be acknowledged. The findings represent the thoughts and opinions of study participants who practice physiotherapy in metropolitan and regional areas in New South Wales, Australia, and may not necessarily generalise to other clinician populations (Krueger & Casey, 2009). It was not possible to determine the degree to which the study sample is reflective of the wider potential participant population. The sample predominantly included males (77%) and those working within a private setting (81%), and it is not known whether such characteristics are related to the views expressed in this study. Particular care would therefore seem warranted in generalising the study's findings to non-comparable clinician populations. The sampling strategy aimed to include the intended clinical consumers of LBP CPRs and as such, no restrictions were placed on the baseline knowledge of study participants in regards to the study topic. It is plausible that participants who had greater prior awareness and familiarity with LBP CPRs may have held different views

on the subject compared to those less knowledgeable. However, the primary aim of this research was to undertake an exploration of the range of views held by practicing physiotherapists and it was therefore determined a priori to not screen for inclusion based on subject knowledge. Finally, the question schedule was not formally evaluated prior to its implementation in the focus groups.

It is anticipated that the findings of this study will help researchers to develop LPB CPRs that have the greatest potential to positively influence physiotherapy clinical practice. Given the large time and resource commitment required to develop CPRs, targeting and incorporating the identified needs and preferences of the intended clinical users from the preliminary stages of development merits consideration, and may plausibly enhance the translation of research findings into clinical practice.

CHAPTER 7

UNCERTAINTY IN CLINICAL PREDICTION RULES: THE VALUE OF CREDIBLE INTERVALS

This chapter has been published in a peer-reviewed scientific journal (Appendix 4):

Haskins, R., Osmotherly, P. G., Tuyl, F., & Rivett, D. A. (2014). Uncertainty in Clinical Prediction Rules: The Value of Credible Intervals. *Journal of Orthopaedic & Sports Physical Therapy*, *44*(2), 85-91.

The work presented in this manuscript was completed in collaboration with the co-authors (Appendix 1).

Overview

The precision and accuracy of a CPR was identified in Study 3 (Chapter 6, p. 210) to have important implications regarding its perceived usefulness and likelihood to be implemented in a clinical context. However, the systematic review of CPRs relevant to the physiotherapy management of LBP (Chapter 4, p. 127) highlighted that the precision of posterior probability estimates were seldom reported. Further, while formal and approximation methods had been established to inform the calculation of uncertainty intervals for posterior probabilities, these methods had received little recognition within

the physiotherapy literature. Consequently, the Clinical Commentary detailed in this chapter sought to highlight the importance of, and provide a technical guide to, the calculation of uncertainty intervals for posterior probabilities. The primary calculation method detailed in this chapter was also applied within the subsequent studies of this thesis detailed in Chapter 8 (p. 256) and Chapter 9 (p. 304).

7.1 Synopsis

Decision making in physical therapy is increasingly informed by evidence in the form of probabilities. Prior beliefs concerning diagnoses, prognoses and treatment effects are quantitatively revised by the integration of new information derived from the history, physical examination and other investigations in a well-recognised application of Bayes' Theorem. Clinical prediction rule development studies commonly employ such methodology to produce quantified estimates of the likelihood of patients having certain diagnoses or achieving given outcomes. To date, the physical therapy literature has been limited to the discussion and calculation of the point-estimate of such probabilities. The degree of precision associated with the construction of posterior probabilities, which requires consideration of both uncertainty associated with pre-test probability and uncertainty associated with test accuracy, remains largely unrecognised and unreported. This paper provides an introduction to the calculation of the uncertainty interval, known as a credible interval, around posterior probability estimates. The method for calculating the credible interval is detailed and illustrated with example data

from 2 clinical prediction rule development studies. Two relatively quick and simple methods for approximating the credible interval are also outlined. It is anticipated that knowledge of the credible interval will have practical implications for the incorporation of probabilistic evidence in clinical practice. Consistent with reporting standards for interventional and diagnostic studies, it is equally appropriate that studies reporting posterior probabilities calculate and report the level of precision associated with these point-estimates.

7.2 Introduction

Decision making in healthcare is increasingly informed by the incorporation of knowledge considered within a probabilistic framework (Ledley & Lusted, 1959; O'Connor & Sox, 1991). New information derived from the history, physical examination and other investigations is used to revise prior beliefs about the likelihood of a given diagnosis or outcome by a magnitude proportional to the relative strength of that information (Davidson, 2002; Fritz & Wainner, 2001). The application of such methods, particularly within a diagnostic context, has been termed ‘probabilistic reasoning’ (Doust, 2009; Richardson & Wilson, 2008). Within a probabilistic framework, perfect predictions are not anticipated and error is knowingly accepted (Einhorn, 1986). The goal of probabilistic reasoning is therefore not to predict outcomes with certainty, but rather to generate predictions that are more often 'less wrong' than those generated by other methods.

Prior beliefs (prevalence of a diagnosis or outcome), also known as 'pre-test probability', and new information (outcome of test with known sensitivity and specificity) may be mathematically integrated to produce the quantified probability of a given diagnosis or outcome (known as a 'posterior probability' or 'post-test probability') using a well-known application of Bayes' Theorem (Equation 7.1)(Bayes, 1763).

Equation 7.1 Calculation of the posterior probability using Bayes' Theorem

$$P(\text{diagnosis or outcome}) = \frac{\text{Prevalence}*\text{Sensitivity}}{\text{Prevalence}*\text{Sensitivity} + (1-\text{Prevalence})*(1-\text{Specificity})}$$

Clinical prediction rules (CPR) are a common application of probabilistic reasoning in healthcare and function to produce estimates of the likelihood of a target diagnosis, prognosis or treatment outcome which in turn inform clinical decision making (Haskins et al., 2012; McGinn et al., 2008). For example, G. E. Hicks et al. (2005) derived a CPR that functions to identify patients experiencing low back pain with a higher likelihood of achieving success from an 8 week stabilization exercise program. They found that when 3 or more factors were present at baseline including a positive prone instability test, aberrant movements, average straight leg raise greater than 91°, and age greater than 40 years, the probability of achieving a successful treatment outcome increased from 33% to 67%. A second example from the physical therapy literature is a CPR designed to identify patients with low

back pain more likely to benefit from lumbopelvic thrust manipulation. Flynn et al. (2002) identified that the probability of success from this intervention increased from 45% to 95% when 4 or more factors were present including duration of symptoms less than 16 days, at least 1 hip with more than 35° of internal rotation, lumbar hypomobility, no symptoms below the knee, and a score of less than 19 on the work subscale of the Fear Avoidance Beliefs Questionnaire (Waddell, Newton, Henderson, Somerville, & Main, 1993).

7.3 Uncertainty in clinical prediction: credible intervals

To date, the physical therapy literature has been limited to the discussion and calculation of single values ('point estimates') of posterior probabilities. When uncertainty is considered, it is frequently limited to test accuracy (sensitivity, specificity, likelihood ratio), with rare consideration of the precision of the pre-test probability. The degree of precision associated with the construction of the posterior probability that is frequently reported in many CPR development studies, remains largely unrecognised and unreported. This stands in contrast to the near-uniform reporting of confidence intervals around estimates of treatment effect and diagnostic test accuracy seen in the modern scientific literature (Bossuyt, Reitsma, Bruns, Gatsonis, Glasziou, Irwig, Lijmer, et al., 2003; Schulz, Altman, Moher, & Consort Group, 2010). While it is accepted that confidence intervals reported in interventional studies have important implications for clinical practice (Stratford, 2010), no such considerations have been given to the need for uncertainty intervals for

posterior probability estimates published in the physical therapy CPR literature.

The precision of the posterior probability estimate requires consideration of the uncertainty associated with prevalence, as well as the uncertainty associated with test accuracy (Baron, 1994; Bianchi, Alexander, & Cash, 2009). The resultant uncertainty interval within a Bayesian framework is known as a ‘credible interval’ (CrI) and gives the range in which the ‘true’ likelihood of a given diagnosis or outcome lies for a specified probability level (Mossman & Berger, 2001). Conceptually, it is the probability of a probability. For instance, the range of values expressed in a 95% CrI has a 95% chance of containing the true probability of a given outcome based on the information available.

Figure 7.1 (below) illustrates the relationship between pre-test probability and post-test probability for 6 likelihood ratio values. The slope of each curve, and therefore the influence of a 1% change in pre-test probability (x-axis) on post-test probability (y-axis) is defined by the formula given in Equation 7.2 (p. 239).

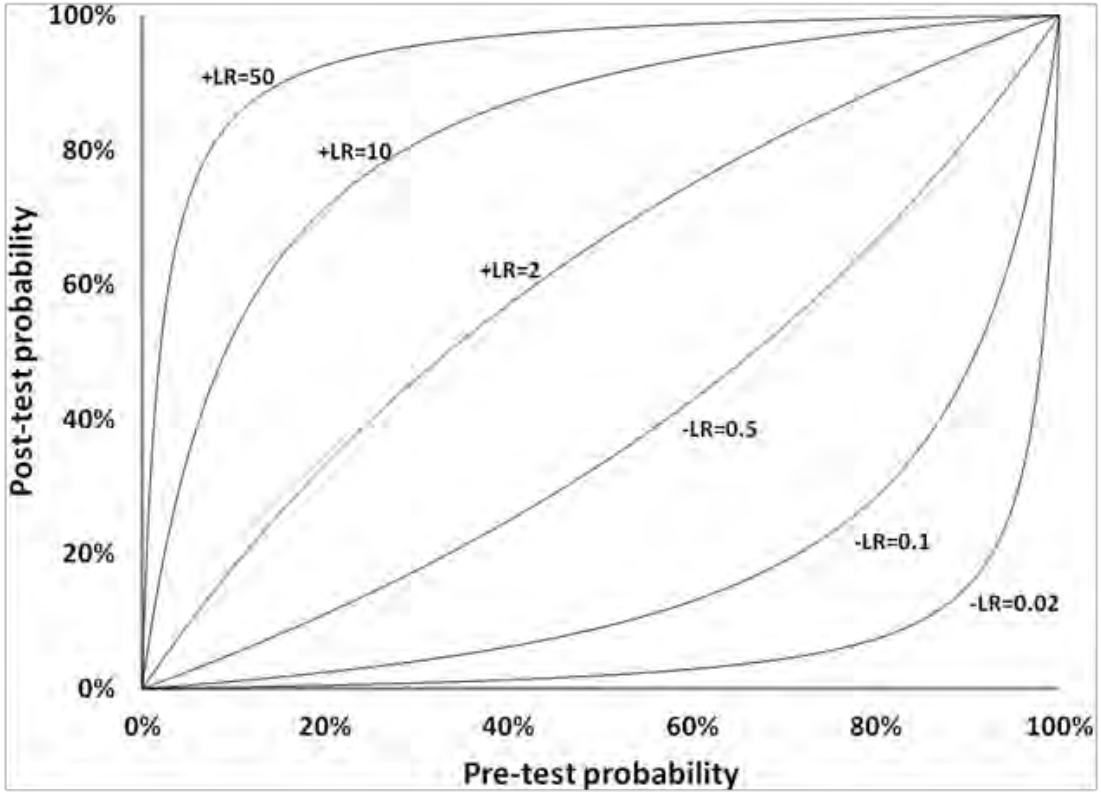


Figure 7.1 **The relationship between pre-test probability and post-test probability for specified likelihood ratios. The slope of each curve illustrates the degree to which variation in the pre-test probability influences the post-test probability.**

Equation 7.2 Calculating the slope of the curve which illustrates the relationship between pre-test probability, likelihood ratio, and post-test probability

$$\text{Slope} = \frac{\text{Likelihood ratio}}{(\text{Pre-test probability}*(\text{Likelihood ratio} - 1) + 1)^2}$$

Where the slope of this illustrated relationship is steep, small variations in the pre-test probability have a large influence on the post-test probability. Conversely, where the slope is relatively flat, consideration of uncertainty of the pre-test probability will only have a relatively small impact on the posterior probability uncertainty interval (Bianchi et al., 2009). The degree to which uncertainty in the likelihood ratio impacts the posterior probability is dependent upon the magnitude of the pre-test probability. Therefore, incorporation of uncertainty of both the pre-test probability and the likelihood ratio is required to calculate the degree of uncertainty of the posterior probability.

7.4 Calculating the credible interval

An appropriate method to calculate the CrI for a posterior probability estimate of a binary outcome is the Objective Bayesian Method using Monte Carlo simulation. This method takes into consideration uncertainty related to both pre-test probability and test accuracy. Mossman and Berger (2001) identified this approach as having superior performance properties compared to other

methods of uncertainty interval calculation. Detailed below is a step-by-step guide as to how clinicians and researchers can perform this calculation using Microsoft Excel (2010, Microsoft Corporation, WA, USA) and the statistical freeware R (http://www.r-project.org/). Each calculation step is illustrated using 2 separate examples from the physical therapy CPR literature. Of note, we have substituted the use of Jeffreys' prior used in Mossman and Berger's original method with the uniform Bayes-Laplace prior due to its additional ability to perform well in instances where there are no false positives (Tuyl, Gerlach, & Mengersen, 2008). There are essentially 6 steps;

7.4.1 Step 1

Identify the numerator (x) and denominator (n) for prevalence (P_0), sensitivity (P_1) and 1-specificity (P_2). In many circumstances, it may be appropriate to source prevalence and test accuracy data from separate studies. In the case of a single study, these data may often be derived from a standard 2 x 2 contingency table (Table 7.1 below). If a contingency table has not been provided in the published paper, it can sometimes be calculated and constructed by the reader as in the 2 examples below.

Table 7.1 Two-by-two table to obtain prevalence, sensitivity and 1-specificity

	Outcome present	Outcome not present
Test positive	A	B
Test negative	C	D

Prevalence = (A+C) / (A+B+C+D)

Sensitivity = A / (A+C)

1-Specificity = B / (B+D)

Example 1:

Based on the published results of G. E. Hicks et al. (2005) for determining the likelihood of success for a stabilization exercise program for low back pain for when 3 or more variables were present at baseline, a 2 x 2 contingency table may be derived (Table 7.2 below). The following calculations may be derived from this table:

$P_0 = x_0 / n_0 = 18 / 54$ (33%)

$P1 = x_1 / n_1 = 10 / 18$ (56%)

$P2 = x_2 / n_2 = 5 / 36$ (14%)

Example 2:

Using the results of Flynn et al. (2002) for identifying the likelihood of treatment success from lumbopelvic manipulation for low back pain for when

4 or more variables are present, a 2 x 2 contingency table may be derived (Table 7.3 below) that enables the following calculations of prevalence (P_0), sensitivity (P_1) and 1-specificity (P_2):

$P_0 = x_0 / n_0 = 32 / 71$ (45%)

$P_1 = x_1 / n_1 = 20 / 32$ (63%)

$P_2 = x_2 / n_2 = 1 / 39$ (3%)

Table 7.2 Two-by-two contingency table derived from the data by G. E. Hicks et al. (2005)

	Treatment successful (≥50% improvement[28])	**Treatment not successful** (<50% improvement)
CPR positive (≥3 variables present[29])	10	5
CPR negative (<3 variables present)	8	31

7.4.2 Step 2

The beta distribution is a type of continuous probability distribution that is employed in Bayesian analysis. The shape parameters 'a' and 'b' that will be used to construct beta distributions (β(a, b)) for prevalence, sensitivity and 1-specificity are calculated using Equation 7.3 below.

[28] Percentage change in baseline and 8 week Oswestry Low Back Pain Disability Questionnaire score

[29] Age < 40 years, average straight leg raise > 91°, aberrant movement present, positive prone instability test

Equation 7.3 Defining the shape parameters of the beta distributions for prevalence, sensitivity and 1-specificity

$$\beta(a, b) = \beta(x_i + 1, n_i - x_i + 1)$$

Table 7.3 Two-by-two contingency table derived from the data by Flynn et al. (2002)

	Treatment successful (>50% improvement[30])	**Treatment not successful** (≤50% improvement)
CPR positive (≥4 variables present[31])	20	1
CPR negative (<4 variables present)	12	38

Example 1:

Continuing our example for stabilization exercises for low back pain:

Prevalence: $\beta(a, b) = \beta(18+1, 54-18+1) = \beta(19, 37)$

Sensitivity: $\beta(a, b) = \beta(10+1, 18-10+1) = \beta(11, 9)$

1-Specificity: $\beta(a, b) = \beta(5+1, 36-5+1) = \beta(6, 32)$

[30] Percentage change in Oswestry Low Back Pain Disability Questionnaire score over 3 sessions

[31] Symptom duration < 16 days, at least 1 hip with >35° internal rotation, hypomobility with lumbar spring testing, no symptoms distal to the knee, Fear-Avoidance Beliefs Questionnaire work subscale score < 19

Example 2:

Using the data for lumbopelvic thrust manipulation for low back pain:

Prevalence: $\beta(a, b) = \beta(32+1, 71-32+1) = \beta(33, 40)$

Sensitivity: $\beta(a, b) = \beta(20+1, 32-20+1) = \beta(21, 13)$

1-Specificity: $\beta(a, b) = \beta(1+1, 39-1+1) = \beta(2, 39)$

7.4.3 Step 3

Use a random number generator to select a large number of values (N) (Mossman and Berger (2001) suggest 10,000) from each beta distribution. This can be performed using the statistical freeware R with the following command to facilitate simple importing into Microsoft Excel (replace 'a' and 'b' with the relevant shape parameters calculated in Step 2):

```
write.csv(rbeta(10000,a,b))
```

7.4.4 Step 4

Combine each series of randomly drawn values for prevalence (P_0), sensitivity (P_1) and 1-specificity (P_2) using the formula provided in Equation 7.4 below.

Equation 7.4 Combining randomly drawn values for prevalence (P_0), sensitivity (P_1), and 1-specificity (P_2)

$$\frac{P_0 * P_1}{P_0 * P_1 + (1-P_0) * P_2}$$

Note that using the point-estimates of P_0, P_1 and P_2 in the above will provide the point estimate of the posterior probability.

Example 1:

For the stabilization exercise example:

$$\frac{0.33*0.56}{0.33*0.56 + (1-0.33)*0.14} = 67\%$$

Example 2:

For the lumbopelvic thrust manipulation example:

$$\frac{0.45*0.63}{0.45*0.63 + (1-0.45)*0.03} = 95\%$$

7.4.5 Step 5

Next, sort the resultant array of values in ascending order.

7.4.6 Step 6

To determine the lower and upper boundaries of a 2-tailed 95% CrI, find the values corresponding to N*0.025 and N*0.975 respectively in the sorted array.

Example 1:

For the stabilization exercise example, the 95% CrI = 41% - 85%.

Example 2:

For the lumbopelvic thrust manipulation example, the 95% CrI = 77% - 99%.

Thus, for the stabilization exercise example data on the likelihood of success of this intervention for low back pain the corresponding 2-tailed 95% CrI is 41% to 85%. This means that, given the data, we can be 95% certain that the true probability of success from this exercise program for a patient presenting with 3 or more of the identified baseline predictors is between 41% and 85%. For the lumbopelvic thrust manipulation CPR example, the 95% CrI corresponding to positive status on this rule is 77% to 99%.

For those more experienced in using R, steps 3 - 6 of the method outlined above can be performed entirely within R using the syntax commands listed below. Computationally, this method is very quick and permits the use of an extremely large value for N (eg. 1 million) to produce even more precise

estimates of the posterior uncertainty interval as the calculation remains in vector form and does not require exporting to a spreadsheet.

```
N=1000000
P0=rbeta(N,x0+1,n0-x0+1)
P1=rbeta(N,x1+1,n1-x1+1)
P2=rbeta(N,x2+1,n2-x2+1)
Post=P0*P1/(P0*P1+(1-P0)*P2)
quantile(Post,c(0.025,0.975))
```

Figure 7.2 below illustrates the appropriate syntax commands and output for this method in R using the data from the CPR examples provided above.

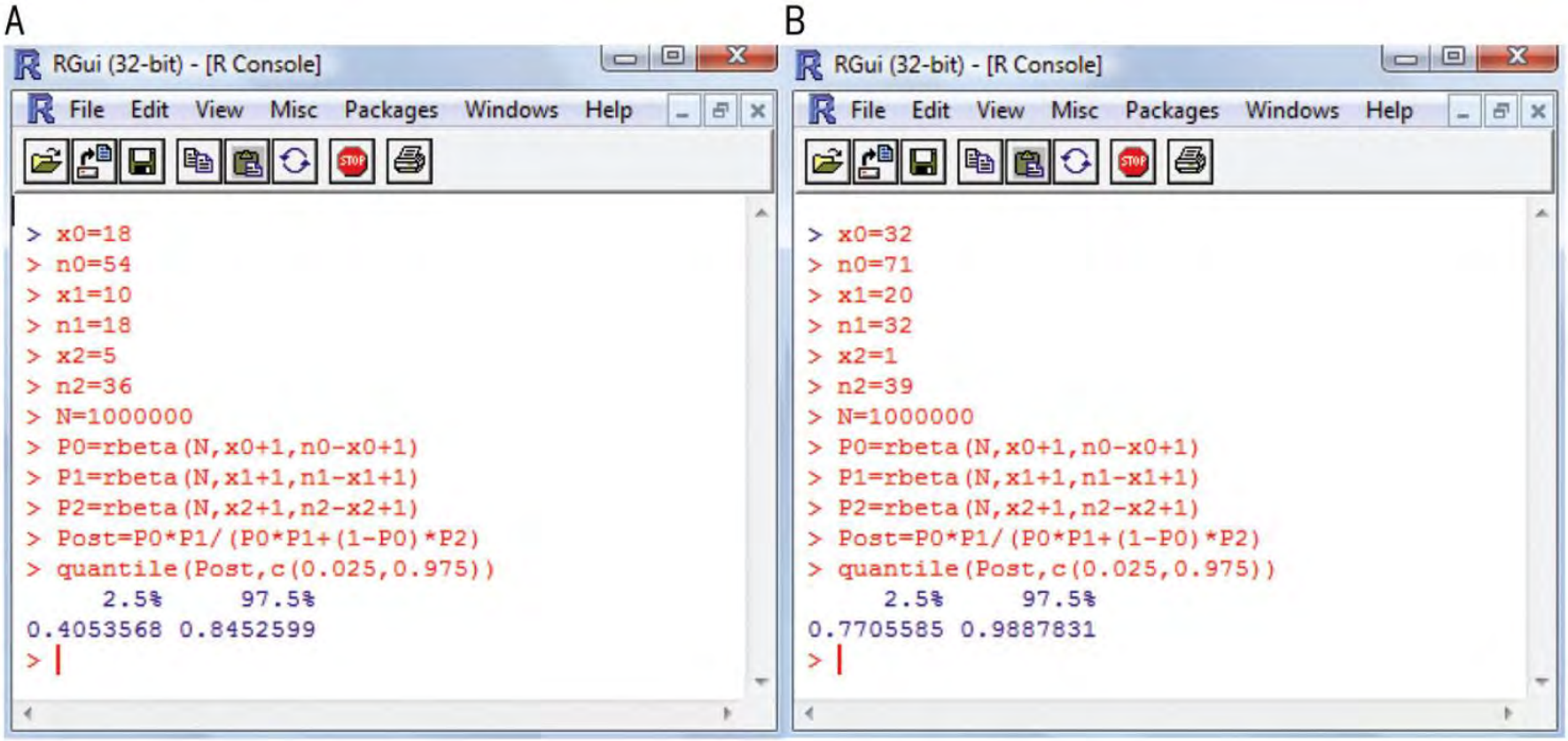


Figure 7.2 **R syntax commands and output for the calculation of the posterior probability 95% credible interval using the objective Bayesian method with Monte Carlo simulation. (A) Stabilization-exercise CPR example data. (B) Lumbopelvic thrust manipulation CPR example data.**

7.5 Quick and simple approximations

In situations where it may be appropriate to gather prevalence, sensitivity and specificity data from within a single study (this excludes case-control designs), 2 methods may provide quick and relatively simple approximations of the uncertainty interval of the posterior probability that may be sufficiently accurate for clinical purposes in the majority of instances.

Tuyl (2009) proposed that adequate intervals may be constructed by using the beta(True Positive + 1, False Positive + 1) distribution. To calculate the 95% interval using this method, the following formulae may be used in Microsoft Excel:

lower boundary: =beta.inv(0.025,True Positive + 1, False Positive + 1)

upper boundary: =beta.inv(0.975,True Positive + 1, False Positive +1)

A useful adjunct to this method is that the mode of this distribution will correspond to the point-estimate of the posterior probability. The intervals generated by this method for both the stabilization exercise example data and the lumbopelvic thrust manipulation example data are almost identical (< 1% different) to those generated by the Objective Bayesian Method using Monte Carlo simulation.

In the instance of zero false positives, a one-sided interval appears preferable such that the point-estimate (100%) will be included within the uncertainty interval. To construct a one-tailed 95% CrI in Microsoft Excel in the case of zero false positives, the following formulae may be applied:

lower boundary: =beta.inv(0.05,True Positive + 1,1)

upper boundary: = 1

An alternative method is to calculate the binomial proportion confidence interval of the positive predictive value (True Positives / (True Positives + False Positives)). The following command in the statistical freeware R will produce the relevant 95% interval using the Clopper-Pearson method (Clopper & Pearson, 1934):

binom.test(True Positives, True Positives + False Positives)

This method produces uncertainty intervals that tend to be more conservative than the aforementioned methods (Newcombe, 1998) but may nevertheless be adequate for clinical decision-making purposes. For the stabilization exercise example data, this method provides an interval of 38% - 88%, and for the thrust manipulation example data an interval of 76% - 100%.

7.6 A brief discussion on selected alternative methods

It may seem intuitive to many to calculate the boundaries of the uncertainty interval of the posterior probability by combining the lower boundary of the pre-test probability confidence interval with the lower boundary of the likelihood ratio confidence interval, and repeating this procedure for the upper boundaries. This method, however, produces uncertainty intervals that are generally much more conservative than the aforementioned procedures. For the stabilization exercise example data, this method produces an interval of 30% - 90%, and for the lumbopelvic thrust manipulation example data, an interval of 63% - 100%. Given the computational steps involved in completing this method, it is also unlikely to be considered simpler than the approximation methods previously outlined.

As already discussed, in some instances uncertainty in the pre-test probability will only have a very small influence on the uncertainty of the post-test probability. In such cases, simply combining the point-estimate of the pre-test probability with the upper and lower boundaries of the likelihood ratio confidence interval will produce an uncertainty interval similar to that produced by the methods already outlined. For the stabilization exercise example data, this method produces an interval of 45% - 83%, and for the lumbopelvic thrust manipulation example data, an interval of 74% - 99%. This method, however, may only give a suitable approximation when small variations in the pre-test probability do not greatly influence the post-test probability across any portion of the pre-test probability uncertainty interval.

The influence of the pre-test probability on the post-probability at a given likelihood ratio value may be calculated at its upper and lower uncertainty intervals using Equation 7.2 presented above (p. 239). Alternatively, it can be approximated by considering the slope of the corresponding parts of the likelihood ratio curves presented in Figure 7.1 (p. 238). Given the calculations required to check the appropriateness of this approximation method it will be quicker and simpler in most cases to use the methods previously recommended.

7.7 Future directions

The additional knowledge of the CrI surrounding a posterior probability point estimate has important implications concerning the appropriate clinical application of probabilistic evidence into practice. Rather than sole reliance on the point estimate, consideration of its precision as highlighted by the corresponding CrI, provides substantially greater depth of information which will assist decision-making. Analogous to the use of confidence intervals reported for treatment effect-sizes (Stratford, 2010), the CrI of a posterior probability may be used to inform clinical decision-making by considering its position relative to a 'threshold level of certainty' (Fritz & Wainner, 2001) required by a clinician to make a decision such as commencing a particular treatment program or confirming or negating a diagnostic hypothesis. Such thresholds are dependent on the relative risks and benefits associated with that decision. In the context of treatment decision-making, a clinician may have a relatively low threshold of certainty required for a treatment that is low

cost and low risk, and may have a higher threshold of certainty required for a treatment that is associated with a higher risk of an adverse outcome and is more expensive and time consuming.

In the stabilization exercise CPR example provided, it is plausible that the lower boundary of the uncertainty interval (41%) may be below a 'treatment threshold' (Fritz & Wainner, 2001) many clinicians and patients would consider sufficient for an 8 week treatment program that requires 16 supervised sessions and daily home exercises. Within a probabilistic framework, further confirmatory evidence from the history and physical examination would be required for clinicians and patients to have confidence in the benefit that may be achieved from this intervention. To their credit, G. E. Hicks et al. (2005) explicitly state that the results of their study about the likelihood of success from a program of stabilization exercises represent the preliminary step in the development of a CPR and have not recommended that the derived tool be implemented in clinical practice. We have used the data from their derivation study to illustrate how clinical decisions informed by posterior probabilities may be influenced by the additional knowledge of the CrI.

In our second CPR example concerning the probability of treatment success from lumbopelvic thrust manipulation, the published results of Flynn et al. (2002) enable the calculation of a CrI of 77%-99%. In contrast to the stabilization exercise program example, it is plausible that many clinicians and patients may perceive that the lower boundary of this uncertainty interval

(77%) is above a suitable ‘treatment threshold’ (Fritz & Wainner, 2001) for this intervention given its relatively low cost, short timeframe and low risk of serious adverse events (Shekelle, Adams, Chassin, Hurwitz, & Brook, 1992).

The primary method of calculating the CrI outlined in this paper (Objective Bayesian Method using Monte Carlo simulation) enables the incorporation of data from more than 1 study. Given that several independent studies are required in the development of a CPR (McGinn et al., 2008), meta-analysis of such data may be helpful in the construction of more precise estimates of the posterior probability interval. For example, a recent systematic review (Billington, Fahey, & Galvin, 2012) of a CPR designed to help identify patients at risk of falling, used a meta-analysis of the included studies to calculate more precise estimates of the tool’s sensitivity and specificity. These data were then used to help calculate more precise estimates of the post-test probability of a fall in patients who were either positive or negative on that CPR. To the best of our knowledge, published guidelines specific to the meta-analysis of CPR studies do not yet exist, however, many considerations concerning the appropriateness of pooling CPR accuracy data may be plausibly generalized from guidelines on the meta-analysis of diagnostic tests. Such considerations include the methodological rigor of included studies as well as between-study variations in study design, study participants, CPR application and the assessment of the reference standard (Bossuyt & Leeflang, 2008; Irwig et al., 1994; C. M. Jones & Athanasiou, 2009; Macaskill, Gatsonis, Deeks, Harbord, & Takwoingi, 2010; Reitsma et al., 2009).

7.8 Conclusion

This commentary focuses on the precision of posterior probability estimates in clinical prediction rule development studies. This is a necessary, but not sufficient, consideration in the application of such evidence into clinical practice. Other factors beyond the scope of this commentary including study design, methodological quality, validation and impact analysis require equally due consideration and have been well described in the existing physical therapy literature (Beattie & Nelson, 2006; Childs & Cleland, 2006; Hancock, Herbert, et al., 2009; Kamper et al., 2010; Kent, Hancock, et al., 2010; Kent, Keating, et al., 2010; Nee & Coppieters, 2011). Consistent with reporting standards for interventional (Schulz et al., 2010) and diagnostic (Bossuyt, Reitsma, Bruns, Gatsonis, Glasziou, Irwig, Lijmer, et al., 2003) studies, we believe that it is equally appropriate that studies reporting posterior probabilities, as commonly practiced in CPR development studies, calculate and report the level of precision associated with these point-estimates.

CHAPTER 8

DIAGNOSTIC CLINICAL PREDICTION RULES FOR SPECIFIC SUBTYPES OF LOW BACK PAIN: A SYSTEMATIC REVIEW

This chapter has been published in a peer-reviewed scientific journal (Appendix 4):

Haskins, R., Osmotherly, P. G., & Rivett, D. A. (2015). Diagnostic clinical prediction rules for specific subtypes of low back pain. *Journal of Orthopaedic & Sports Physical Therapy, 45*(2), 61-76.

The work presented in this manuscript was completed in collaboration with the co-authors (Appendix 1).

Overview

This is the fourth of five studies in this research program. The findings of studies 2 and 3 (Chapter 5, p. 174; and Chapter 6, p. 210) identified that although Australian physiotherapists may value some diagnostic LBP CPRs, there is mixed awareness and limited familiarity with tools currently under development. Such knowledge gaps will impede the translation of this evidence into practice. Identifying CPRs under development, however, is quite challenging given the range of nomenclature used to describe such

tools, and the current absence of a specific medical subject heading (see section 3.4 p. 93). Further, determining the degree to which CPRs for LBP have been sufficiently developed to enable their application in practice requires an up to date consideration of the full range of studies that have contributed to their development.

At the time when the present study was initiated, the only study that had sought to synthesise the evidence concerning diagnostic forms of CPRs for LBP was the previous systematic review presented in Chapter 4 (p. 127) of this thesis. However that study was focused upon CPRs developed within the physiotherapy profession, and therefore does not encapsulate the full range of CPRs that would be relevant to a physiotherapist in their management of a patient with LBP. A substantially greater volume of LBP CPR development studies had been also published since the search undertaken for the first systematic review (up to January 2010), further justifying an update of the earlier review.

There were several additional reasons that were identified for performing an update and expansion of the earlier review. Advancements in the methods by which to optimally retrieve relevant CPR studies from the medical literature had been made, thus providing an avenue to conduct a more sensitive search. Further, more recent literature regarding the range of methodological considerations pertinent to the development of CPRs provided an opportunity to conduct a more comprehensive quality appraisal of identified studies. The findings of Study 3 (Chapter 6, p. 210) highlighted that the predictive

precision of a CPR may have important implications regarding its perceived usefulness. Consequently, an opportunity was identified to apply the methods outlined in the Clinical Commentary detailed in Chapter 7 (p. 232) to calculate and report uncertainty intervals for posterior probabilities in an updated systematic review.

The study presented in this chapter aims to provide a comprehensive synthesis of the evidence base for diagnostic forms of CPRs for LBP. The earlier studies in this program of research have informed the specific elements of the review, such that it may serve as an informative resource for both clinicians managing patients with LBP and researchers seeking to develop CPRs for LBP that may meaningfully benefit clinical practice.

8.1 Abstract

Study Design:

Systematic review.

Objectives:

To identify diagnostic clinical prediction rules (CPRs) for low back pain (LBP) and to assess their readiness for clinical application.

Background:

Significant research has been invested into the development of CPRs that may assist in the meaningful subgrouping of patients with LBP. To date, very little is known about diagnostic forms of CPRs for LBP, which relate to the

present status or classification of an individual, and whether they have been developed sufficiently to enable their application in clinical practice.

Methods:

A sensitive electronic search strategy using 7 databases was combined with hand-searching and citation tracking to identify eligible studies. Two independent reviewers identified relevant studies for inclusion using a 2-staged selection process. The quality appraisal of included studies was conducted by 2 independent raters using QUADAS-2 and checklists comprised of accepted methodological standards for the development of CPRs.

Results:

Of 10,014 studies screened for eligibility, the search identified that 13 diagnostic CPRs for LBP have been derived. Amongst those, 1 tool for identifying lumbar spinal stenosis and 2 tools for identifying inflammatory back pain have undergone validation. No impact analysis studies were identified.

Conclusion:

Most diagnostic CPRs for LBP are in their initial development phase and cannot be recommended for use in clinical practice at this time. Validation and impact analysis of the diagnostic CPRs identified in this review is warranted, particularly for those tools which meet an identified unmet need of clinicians who manage patients with LBP.

8.2 Introduction

Patients with low back pain (LBP) are generally considered to consist of smaller subgroups that differ meaningfully with regard to their symptomology, prognosis, and response to various treatments (Kent & Keating, 2004). It is hypothesised that such heterogeneity within the LBP patient population may contribute to the relatively modest effect sizes generally observed in most high quality clinical trials (Balague et al., 2012; Foster et al., 2011). The identification of LBP subgroups has been a research priority for several years (Borkan & Cherkin, 1996; Costa et al., 2013; Foster et al., 2009; Henschke, Maher, Refshauge, et al., 2007) and several classification approaches have been proposed (Karayannis et al., 2012; Kent & Keating, 2005; C. McCarthy et al., 2004; Riddle, 1998). Traditionally, such classification approaches have been predominantly based upon expert opinion and biologic plausibility with notably little concordance among them. More recently, there has been greater focus upon empirically derived subgrouping methods including the development of clinical prediction rules (CPRs) (Beattie & Nelson, 2006; Foster et al., 2013).

A CPR is a clinical tool designed to assist decision making for individual patients by combining elements from the history, physical examination, and other investigations to make predictions regarding a patient's diagnosis, prognosis, or likely response to a particular treatment (Beattie & Nelson, 2006; Childs & Cleland, 2006; McGinn et al., 2008). A CPR is initially derived using multivariable statistical procedures to identify which aspects of a

patient's presentation are independently related to a certain diagnosis or outcome. The tool then undergoes a process of validation whereby it is applied in new groups of patients in different settings to evaluate its ability to accurately predict that same diagnosis/outcome. Validated CPRs subsequently undergo impact analysis by which they are tested to determine whether their clinical application leads to improved patient outcomes or efficiencies in resource consumption (McGinn et al., 2000; McGinn et al., 2008; Toll et al., 2008).

The stage of CPR development has important implications regarding the CPR's appropriateness to be applied in clinical practice. A CPR that has not undergone validation is not recommended for use in practice as it may reflect chance statistical associations or be specific to the patient sample or setting from which it was derived (McGinn et al., 2000). It is generally accepted that a CPR that has been prospectively validated in new patient cohorts across broad clinical settings may be applied in practice in similar patient populations with confidence in its known predictive accuracy (McGinn et al., 2000; McGinn et al., 2008). However, an important consideration is that though its predictive accuracy (the amount of agreement between the results from an index test and those from a reference standard)(Bossuyt, Reitsma, Bruns, Gatsonis, Glasziou, Irwig, Moher, et al., 2003) may be known, it should not be assumed that a validated CPR's accuracy will be superior to unassisted clinician judgment, or that application of the rule will result in beneficial clinical consequences. Impact analysis is required before a CPR can be applied in clinical practice with confidence that its implementation will

likely result in improved patient care (Beattie & Nelson, 2006; McGinn et al., 2000; McGinn et al., 2008; Toll et al., 2008).

In the management of patients with LBP, the benefits of the clinical application of a well-developed diagnostic CPR may include: reducing the need for unnecessary tests; identifying patients who are likely to benefit from referral to other services and/or further investigations; improving the efficiency of the clinical assessment; reducing the costs of care; enabling more timely initiation of treatment; and informing treatment decision making. Such benefits have been demonstrated or hypothesised for diagnostic CPRs developed for other musculoskeletal presentations (Auleley et al., 1997; Kocher, Mandiga, Zurakowski, Barnewolt, & Kasser, 2004; Stiell et al., 1994; Stiell, Wells, Laupacis, et al., 1995; Stiell et al., 1997; Sutlive et al., 2008; Waldrop, 2006).

Previous systematic reviews of LBP CPRs have predominantly focused on prognostic tools designed to predict outcomes such as improvements in clinical status following treatment (Beneciuk et al., 2009; Lubetzky-Vilnai et al., 2014; May & Rosedale, 2009; Patel, Friede, Froud, Evans, & Underwood, 2013; Stanton et al., 2010; van Oort et al., 2012). To the best of our knowledge, only 1 previous review (Haskins et al., 2012) has included diagnostic forms of LBP CPRs, however that study was limited to tools developed within physical therapy practice. Consequently, little is known about the current state of research on diagnostic LBP CPRs and their readiness to be applied in clinical practice. Therefore the aim of the present

review was to identify diagnostic forms of LBP CPRs and to appraise their readiness for application in clinical practice.

8.3 Methods

8.3.1 Literature search

A systematic search of the literature was conducted to identify derivation, validation, and impact analysis studies investigating diagnostic forms of CPRs relevant to the non-surgical management of adults with LBP. The database search strategy (Table 8.1) incorporated search strings identified to have high sensitivity for prediction model studies in combination with disease-specific filters for back related disorders (Bombardier et al., 2014; Geersing et al., 2012; Holland et al., 2005; Ingui & Rogers, 2001). Components of this search strategy have been used in previous systematic reviews for prognostic CPRs (Beneciuk et al., 2009; Haskins et al., 2012; May & Rosedale, 2009; van Oort et al., 2012). Medline, Embase, the Cochrane Central Register of Controlled Trials, PyschINFO, CINAHL, AMED, and the Index to Chiropractic Literature were searched from their inception to July 2013. Identified records were downloaded into reference management software (EndNote X6.0.1, Thomson Reuters, New York, NY) and duplicates were removed. Hand searching and citation tracking were used as supplementary search strategies.

Table 8.1 Search strategy

Medline via OVID (1946 - July 2013)	
1	dorsalgia.ti,ab OR exp Back Pain/ OR backache.ti,ab OR exp Low Back Pain/ OR (lumbar adj pain).ti,ab OR coccyx.ti,ab OR coccydynia.ti,ab OR sciatica.ti,ab OR sciatic neuropathy/ OR spondylosis.ti,ab OR lumbago.ti,ab OR back disorder$.ti,ab
2	(Validat$ OR Predict$.ti. OR Rule$) OR (Predict$ AND (Outcome$ OR Risk$ OR Model$)) OR ((History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$) AND (Predict$ OR Model$ OR Decision$ OR Identif$ OR Prognos$)) OR (Decision$ AND (Model$ OR Clinical$ OR Logistic Models/)) OR (Prognostic AND (History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$ OR Model$))
3	Stratification OR ROC Curve/ OR Discrimination OR Discriminate OR c-statistic OR c statistic OR "Area under the curve" OR AUC OR Calibration OR Indices OR Algorithm OR Multivariable
4	1 AND (2 OR 3)
5	limit 4 to english
6	limit 5 to humans
Embase via OVID (1947 - July 2013)	
1	dorsalgia.mp. OR back pain.mp. OR exp LOW BACK PAIN/ OR exp BACKACHE/ OR (lumbar adj pain).mp. OR coccyx.mp. OR coccydynia.mp. OR sciatica.mp. OR exp ISCHIALGIA/ OR spondylosis.mp. OR lumbago.mp. OR back disorder$.ti,ab.
2	predict:.tw. OR exp methodology OR validat:.tw.
3	1 AND 2
4	limit 3 to english
5	limit 4 to humans
6	limit 5 to exclude medline journals
Cochrane Central Register of Controlled Trials via OVID (1898 - July 2013)	
1	exp Back Pain/ OR back ache OR exp Low Back Pain/ OR (lumbar adj pain) OR coccyx OR coccydynia OR sciatica OR spondylosis OR exp Spine/ OR exp Spinal Diseases/ OR lumbago OR discitis OR (disc adj degeneration) OR (disc adj prolapse) OR (disc adj herniation) OR spinal fusion OR spinal neoplasms OR (facet adj joints) OR exp Intervertebral Disk/ or postlaminectomy OR arachnoiditis OR (failed adj back) OR exp Cauda Equina/ OR (lumbar adj vertebra$) OR (spinal adj stenosis) OR (slipped adj (disc$ or disk$)) OR (degenerat$ adj (disc$ or disk$)) OR (stenosis adj (spine or root or spinal)) OR (displace$ adj (disc$ or disk$)) OR (prolap$ adj (disc$ or disk$)) OR exp Sciatic Neuropathy/ OR sciatic$ OR back disorder$ OR (back adj pain)
2	(Validat$ OR Predict$.ti. OR Rule$) OR (Predict$ AND (Outcome$ OR Risk$ OR Model$)) OR ((History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$) AND (Predict$ OR Model$ OR Decision$ OR Identif$ OR Prognos$)) OR (Decision$ AND (Model$ OR Clinical$ OR Logistic Models/)) OR (Prognostic AND (History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$ OR Model$))

3	Stratification OR ROC Curve/ OR Discrimination OR Discriminate OR c-statistic OR c statistic OR "Area under the curve" OR AUC OR Calibration OR Indices OR Algorithm OR Multivariable
4	1 AND (2 OR 3)
5	limit 4 to medline records
6	limit 4 to embase records
7	4 NOT (5 OR 6)
PsychINFO via OVID (1806 - July 2013)	
1	back pain/ OR lumbar spinal cord/ OR (low adj back adj pain).mp OR (back adj pain).mp OR spinal column/ OR (lumbar adj2 vertebra$).mp OR coccyx.mp OR sciatica.mp OR lumbago.mp OR dorsalgia.mp OR back disorder$.mp OR ((disc or disk) adj degenerat$).mp OR ((disc or disk) adj herniat$).mp OR ((disc or disk) adj prolapse$).mp OR (failed adj back).mp
2	(Validat$ OR Predict$.ti. OR Rule$) OR (Predict$ AND (Outcome$ OR Risk$ OR Model$)) OR ((History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$) AND (Predict$ OR Model$ OR Decision$ OR Identif$ OR Prognos$)) OR (Decision$ AND (Model$ OR Clinical$ OR Logistic Models/)) OR (Prognostic AND (History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$ OR Model$))
3	Stratification OR ROC Curve/ OR Discrimination OR Discriminate OR c-statistic OR c statistic OR "Area under the curve" OR AUC OR Calibration OR Indices OR Algorithm OR Multivariable
4	1 AND (2 OR 3)
5	limit 4 to english
6	limit 5 to human
CINAHL via EBSCO (1937 - July 2013)	
1	"dorsalgia" OR (MH "Back Pain+") OR (MH "Low Back Pain") OR "backache" OR (lumbar W1 pain) OR (lumbar N5 pain) OR (MH "Coccyx") OR (MH "Sciatica") OR "sciatica" OR "coccyx" OR "coccydynia" OR "back disorder*" OR (MH "Lumbar Vertebrae") OR (lumbar N2 vertebra) OR (MH "Thoracic Vertebrae") OR (MH "Spondylolisthesis") OR (MH "Spondylolysis") OR "lumbago"
2	(validat* OR ti predict* OR rule*) OR (predict* AND (outcome* OR risk* OR model*)) OR ((history OR variable* OR criteria OR scor* OR characteristic* OR finding* OR factor*) AND (predict* OR model* OR decision* OR identif* OR prognos*)) OR (decision* AND (model* OR clinical* OR MH "logistic regression+")) OR (prognostic AND (history OR variable* OR criteria OR scor* OR characteristic* OR finding* OR factor* OR model*))
3	stratification OR mh "ROC Curve" OR discrimination OR discriminate OR c-statistic OR c statistic OR "Area under the curve" OR AUC OR calibration OR indices OR algorithm OR multivariable
4	S1 AND (S2 OR S3)
5	applied limit to English
6	applied limit to humans
7	applied limit to exclude Medline records
AMED via OVID (1985 - July 2013)	

1	dorsalgia.ti,ab OR exp Back Pain/ OR backache.ti,ab OR exp Low Back Pain/ OR (lumbar adj pain).ti,ab OR coccyx.ti,ab OR coccydynia.ti,ab OR sciatica.ti,ab OR sciatic neuropathy/ OR spondylosis.ti,ab OR lumbago.ti,ab OR back disorder$.ti,ab
2	(Validat$ OR Predict$.ti. OR Rule$) OR (Predict$ AND (Outcome$ OR Risk$ OR Model$)) OR ((History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$) AND (Predict$ OR Model$ OR Decision$ OR Identif$ OR Prognos$)) OR (Decision$ AND (Model$ OR Clinical$ OR Logistic Models/)) OR (Prognostic AND (History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$ OR Model$))
3	Stratification OR ROC Curve/ OR Discrimination OR Discriminate OR c-statistic OR c statistic OR "Area under the curve" OR AUC OR Calibration OR Indices OR Algorithm OR Multivariable
4	1 AND (2 OR 3)
5	limit 4 to english
Index of Chiropractic Literature (1981 - July 2013)	
1	Subject:"Back" OR Subject:"Back Injuries" OR Subject:"Back Pain" OR Subject:"Low Back Pain" OR Subject:"Lumbar" OR Subject:"Lumbosacral Region" OR Subject:"Sciatica" OR All Fields:sciatica OR Subject:"Coccyx" OR Subject:"Sacroiliac Joint" OR Subject:"Sacrum"
2	(Validat* OR Predict* OR Rule*) OR (Predict*AND (Outcome* OR Risk* OR Model*)) OR ((History OR Variable* OR Criteria OR Scor* OR Characteristic* OR Finding* OR Factor*) AND (Predict* OR Model* OR Decision* OR Identif* OR Prognos*)) OR (Decision* AND (Model* OR Clinical* OR "Logistic Model*")) OR (Prognostic AND (History OR Variable* OR Criteria OR Scor* OR Characteristic* OR Finding* OR Factor* OR Model*))
3	Stratification OR "ROC Curve" OR Discrimination OR Discriminate OR c-statistic OR "c statistic" OR "Area under the curve" OR AUC OR Calibration OR Indices OR Algorithm OR Multivariable
4	S2 OR S3
5	S1 AND S4

8.3.2 Study selection

A CPR was operationally defined as “a clinical tool that quantifies the individual contributions that various components of the history, physical examination, and basic laboratory results make towards the diagnosis, prognosis, or likely response to treatment in an individual patient” (McGinn et al., 2008). The eligibility criteria used in this review are summarised in Table 8.2. Diagnostic CPRs were operationally defined as relating to the present

status or classification of an individual, which included, but was not limited to, pathoanatomic diagnoses. Studies developing CPRs that function to inform predictions related to future outcomes or treatment effects were excluded from this review. Study eligibility was not restricted by year of publication, stage of CPR development, types of diagnostic predictor variables under consideration (eg. physical tests, history items, laboratory or imaging tests etc.), or the professional discipline(s) involved in the tool's development.

Table 8.2 Study eligibility criteria

Inclusion criteria	
1	Reports on the derivation, validation, and/or impact analysis of 1 or more diagnostic clinical prediction rules related to the non-surgical management of adults with low back pain
2	The clinical prediction rule under development contains 2 or more predictor variables
3	The clinical prediction rule under development was initiated by a formal derivation process in which a larger pool of candidate predictor variables was refined to a smaller set of variables based on their identified independent predictive value using formal multivariable statistical procedures
4	A tool is clearly presented in sufficient detail that may be applied by a clinician to inform a diagnosis for an individual patient
5	Published in English
Exclusion criteria	
1	Limited to the investigation of modifiable and/or determinant predictor variables
2	Clinical prediction rule not capable of directly contributing to patient care
3	Conference proceedings/abstracts, dissertations, commentaries, reviews, editorials, letters, study protocols, n=1 designs (case reports), books, book chapters, book reviews, practice guidelines

The titles and abstracts of identified records were initially screened for eligibility by 2 independent reviewers, with potentially eligible studies

identified by either reviewer progressing to the next stage of study selection (P. Edwards et al., 2002; Higgins & Green, 2011). A screening of the full text of potentially eligible studies was subsequently conducted by both reviewers, and studies were determined to be eligible for inclusion by concordance of the 2 reviewers, with a third independent reviewer providing the final judgement in instances where disagreement could not be resolved by consensus.

8.3.3 Data extraction

Data extraction was conducted using an electronic spreadsheet. The intended function of the CPR, stage of development, patient population, reference standard, outcome prevalence, predictor variables, statistical analysis, tool format, accuracy, reporting of uncertainty intervals, and posterior probabilities were recorded for each included study. The CPR development framework proposed by McGinn et al. (2000) was used to categorise each study as derivation, validation or impact analysis.

8.3.4 Statistical analysis

Interrater agreement for quality appraisal and each stage of study selection was calculated and reported as the absolute and chance-corrected (κ) degree of agreement. 2x2 contingency tables were extracted, calculated, or approximated when reported study findings permitted. Sensitivity, specificity, likelihood ratios, and posterior probabilities and their associated uncertainty intervals were subsequently calculated from these data and reported for each

study in instances where these were not directly conveyed in the original publication. The objective Bayesian method using Monte Carlo simulation was used to derive uncertainty intervals for posterior probabilities (Haskins, Osmotherly, Tuyl, et al., 2014; Mossman & Berger, 2001).

8.3.5 Quality appraisal

The Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool (Whiting et al., 2011) was applied by 2 independent reviewers with consensus determining criteria status. A third reviewer provided the final decision in instances where consensus was not reached.

Included studies were further appraised in regards to the methodological standards of CPR development specific to each phase of development. In the absence of a validated tool for this purpose, an updated version of a quality appraisal tool used in a previous systematic review on CPRs was applied (Table 8.3 and Table 8.4) (Haskins et al., 2012). The items selected for inclusion in this tool reflect well-recognised methodological standards for the development of CPRs (Beattie & Nelson, 2006; Bouwmeester et al., 2012; Childs & Cleland, 2006; C. Cook et al., 2010; Haskins, Osmotherly, Tuyl, et al., 2014; Laupacis et al., 1997; Lubetzky-Vilnai et al., 2014; McGinn et al., 2000; Seel et al., 2012; Stiell & Wells, 1999; van Oort et al., 2012). Two independent reviewers applied this tool, determining quality criterion status by concordance and resolving disagreements by consensus, or if needed a third reviewer.

Table 8.3 Derivation study quality appraisal criteria

Domain	Criteria	Definition
Design	1. Prospective design	The study is conducted forwards in time.
	2. Study site described	The nature of the study site is described in sufficient detail to enable comparison to other settings.
Participants	3. Justification for the number of study subjects is reported	Justification is provided for the number of subjects enrolled into the study.
	4. Representative sample	The reported method of patient selection is free of bias so that study subjects encompass a wide clinical and demographic spectrum and are representative of all patients seen at the site with the designated condition.
	5. Important patient characteristics described	The study subjects are well described in terms of inclusion criteria, method of selection, and clinical and demographic characteristics.
Predictor variables	6. Selection of candidate predictor variables justified	The selection of candidate predictor variables is justified with appropriate reasoning and may include previous literature, psychometric properties, clinical reasoning, and/or expert opinion.
	7. Blinded predictor assessment	The assessment of the predictor variables is determined without knowledge of the reference standard. If the study was prospective and the predictor variables were clearly collected prior to determination of the reference standard, then assessment can be considered to be blind. If the study was retrospective and the authors did not mention blinding, it can be assumed that it was not blinded.
	8. Predictor variables have demonstrated reliability	Predictor tests are reported to be reliable (kappa >=0.60 or ICC >=0.70) either through previous report or through report within the findings of the study.
Reference standard	9. Reference standard has demonstrated reliability and validity	The reference standard is reported to have demonstrated reliability and validity. Literature is cited to support the reference standard and psychometric characteristics of the reference standard are reported.
	10. Blinded assessment of the reference standard	Interpretation of the reference standard is reported to be determined without knowledge of the status of the predictor variables. If a study does not comment upon whether the reference standard was categorised without knowledge of the predictor variables, it can be assumed that it was not blinded.

Domain	Criteria	Definition
Analysis	11. Mathematical techniques described	The mathematical techniques employed are adequately described.
	12. Reporting and handling of missing data described	Missing data (eg. values per participant, missing values per predictor, lost to follow-up) and how it was handled (eg. omitted, imputation) are reported.
	13. At least 10 outcome events per independent variable in the final multivariable model	There are at least 10 outcome events per independent variable in the final multivariable model. The number of outcome events is defined in proportional hazards analysis by the count of 'failure' events. In logistic regression the number of outcome events is the smaller number of binary outcomes of the dependent variable. For linear regression models there should be at least 10 patients per variable in the final model.
	14. At least 10 outcome events per candidate predictor variable	As per item 13, except the number of candidate predictor variables replaces the number of independent variables in the final model.
	15. Collinearity of predictor variables tested	Collinearity of predictor variables were examined such as testing pairwise correlations or the variance inflation factor.
	16. Continuous predictor variables are kept continuous in the multivariable analysis	Continuous predictor variables were kept as continuous variables in the multivariable analysis.
CPR performance	17. Uncertainty in the accuracy of the CPR is described	Uncertainty intervals are reported for accuracy statistics of the CPR.
	18. Uncertainty in the posttest probability is described	Uncertainty intervals are reported for posttest probabilities.
	19. CPR performance is non-paradoxical	The performance of the CPR behaves logically, such that the probability of a given diagnosis does not decrease at any point with increasing positive status on that tool.

Table 8.4 Validation study quality appraisal criteria

Domain	Criteria	Definition
Design	1. Prospective validation in new patient population	The study is conducted forwards in time in a different population to the derivation study.
	2. Different clinical setting to derivation study	The CPR is tested in a clinical setting that is different to the derivation study.
Participants	3. Representative sample	The method of patient selection is free of bias so that study subjects encompass a wide clinical and demographic spectrum and are representative of all patients seen at the site with the designated condition.
CPR application	4. The rule is applied accurately	The rule is applied exactly as described in the derivation study.
	5. Assessment of the reliability of the rule	The reliability of the interpretation of a rule is explicitly measured using at least a representative subset of the study sample.
Follow-up	6. Complete follow-up	There is complete follow-up. All patients are subjected to the reference standard to determine their true outcome compared to that predicted by the rule.
	7. Reporting and handling of missing data described	Missing data (eg. values per participant, missing values per predictor, lost to follow-up) and how it was handled (eg. omitted, imputation) is reported.
CPR performance	8. Accuracy of the rule described	The accuracy/performance of the rule is described.
	9. Uncertainty in the accuracy of the CPR is described	Uncertainty intervals are reported for accuracy statistics of the CPR.
	10. Uncertainty in the posttest probability is described	Uncertainty intervals are reported for posttest probabilities.

8.3.6 Qualitative synthesis

Study findings and the risk of methodological bias of each CPR development study were used to qualitatively synthesise the evidence identified in this review. Recommendations concerning the readiness of CPRs for clinical

application reflect accepted standards within a well-recognised hierarchical approach to CPR development (Beattie & Nelson, 2006; Childs & Cleland, 2006; McGinn et al., 2000; McGinn et al., 2008; Reilly & Evans, 2006). The Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines were used to prepare the manuscript (Liberati et al., 2009). The review did not qualify for protocol registration as a consequence of the research question not being specific to a particular intervention, exposure, or outcome measure.

8.4 Results

8.4.1 Search results

Figure 8.1 details the results of the study selection process. There were 10,014 unique records initially screened for eligibility, 151 of which progressed to the second round. The full-text of these 151 records was reviewed and 15 publications (Apeldoorn et al., 2012; A. Braun, Saracbasi, Grifka, Schnitker, & Braun, 2011; Chan, Inrig, Molloy, Stone, & Derzko-Dzulynsky, 2012; C. Cook et al., 2011; Fritz, Piva, et al., 2005; Gregg, Dean, & Schneiders, 2009; Kato et al., 2009; Konno et al., 2007; Revel et al., 1998; Roman et al., 2010; Roux et al., 2007; Rudwaleit, Metter, Listing, Sieper, & Braun, 2006; Scholz et al., 2009; Sieper, van der Heijde, et al., 2009; Sugioka, Hayashino, Konno, Kikuchi, & Fukuhara, 2008) were subsequently included in the review. The most common reason for exclusion was not satisfying the review's operational definition of a CPR.

Agreement between raters was ‘moderate’ (Landis & Koch, 1977) for both stages of study selection: κ = 0.56 (95%CI 0.48, 0.65) for the screening by titles and abstracts, and κ = 0.56 (95%CI 0.34, 0.78) for the screening by full-text. The absolute degree of interrater agreement was 99% and 92% for the first and second stages respectively. Nine of the 12 episodes of disagreement in the second stage of screening were resolved by consensus, with the remaining 3 studies later included by the third reviewer.

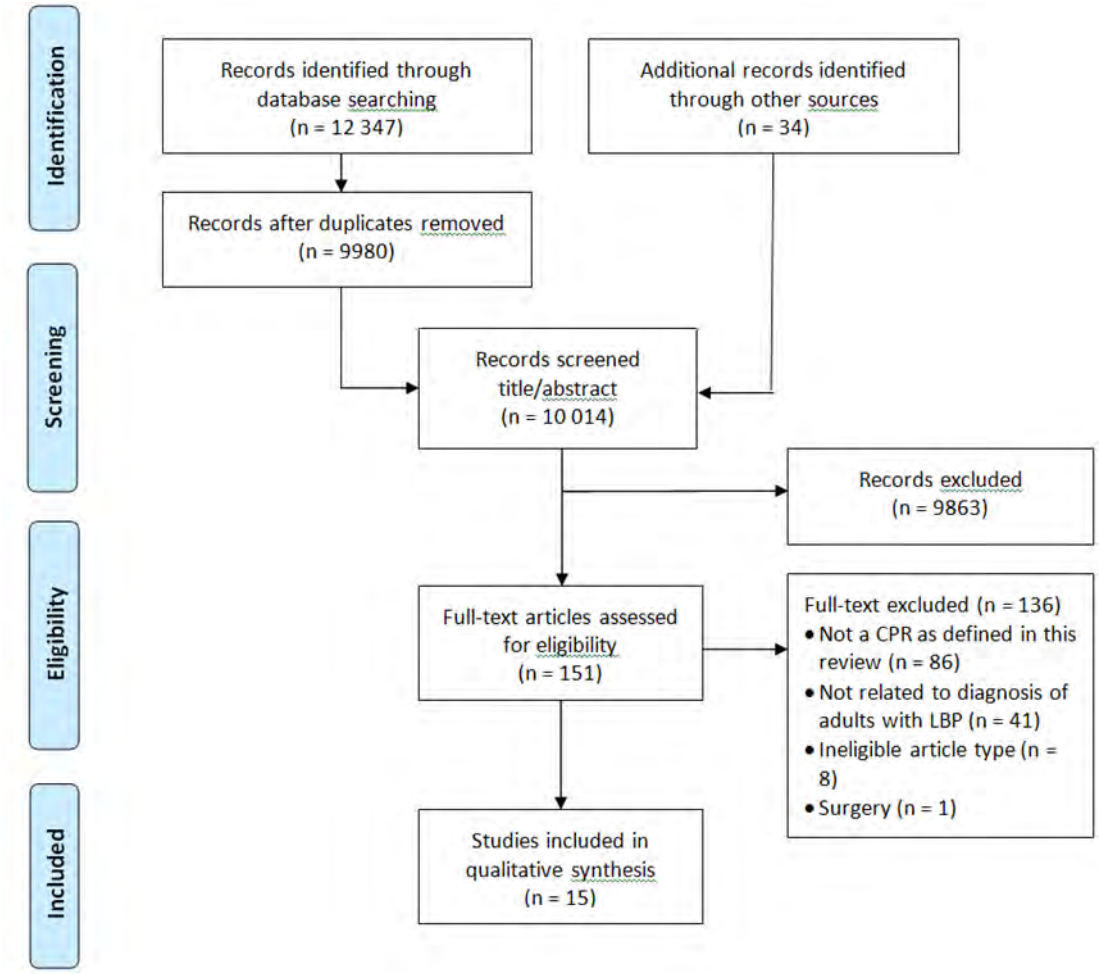


Figure 8.1 **Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) study selection flowchart for systematic review of diagnostic clinical prediction rules for low back pain**

8.4.2 Characteristics of included studies

The 15 publications included in this review reported on 18 studies concerning the development of 13 diagnostic LBP CPRs. Five studies describe the validation of 3 CPRs. No impact analysis studies were identified.

Table 8.5 summarizes the identified diagnostic LBP CPRs and the studies contributing to their development. The clinical presentations that the CPRs aim to assist identifying are spinal stenosis (n=3), vertebral fracture (n=2), inflammatory back pain (n=2), spondyloarthritis (n=1), zygapophyseal joint mediated pain (n=1), radicular LBP (n=1), radiographic instability (n=1), spondylolysis (n=1), and psychological disturbance (n=1). Eight CPRs use a predictor count format, whereby a clinician would sum the number of dichotomized predictors that were present to determine a patient's status on the rule. The remaining 5 CPRs use a score chart format, whereby a clinician would first assign a score for a patient's status on each predictor variable, and then use the sum score to calculate the patient's status on the CPR.

Table 8.5 Characteristics of the diagnostic low back pain clinical prediction rules included in this review

Diagnosis supported	CPR	Function	Predictors	Format	Study	Patients	Reference standard	Accuracy	Post-test probability
Spinal stenosis	1.	Identification of patients with probable lumbar spinal stenosis among those presenting with leg pain or numbness	1. age; 2. absence of diabetes; 3. intermittent claudication; 4. exacerbation of symptoms when standing up; 5. symptom improvement when bending forward; 6. symptoms induced by having patient bend forward; 7. symptoms induced by having patient bend backward; 8. good peripheral artery circulation; 9. abnormal Achilles tendon reflex; 10. positive SLR test	Score chart	Derivation Konno et al. (2007)	n=468, patients with pain or numbness in the lower extremities presenting to university hospitals and medical centres, mean age 65 years (SD 14), 54% female, 50% < 6 months duration of symptoms, 47% prevalence of dependent outcome	Diagnosis of lumbar spinal stenosis by an orthopaedic surgeon and verified by study coordinator (dichotomous)	For scores ≥7 points, sens = 0.93; spec = 0.72; +LR = 3.3; -LR = 0.1.	For scores -2 to 5 = 6%; scores 6 to 8 = 39%; scores 9-11 = 72%; scores 12-16 = 99%. Post-test probability for scores ≥7 points not reported but calculated from study data to be 75%. 95% CrI for scores ≥7 points calculated from study data to be 69.4%, 79.7%.
					Validation Kato et al. (2009)	n=118, presenting to hospital orthopaedic outpatient clinics due to lower extremity symptoms, mean age 68 years (range 12-96), 47% female, 49% prevalence of dependent outcome	Diagnosis of lumbar spinal stenosis by panel consensus of 4 spine surgeons using information from radiology, clinical examination, and CPR (dichotomous)	For scores ≥7 points, sens = 0.95; spec = 0.40. LRs not reported, but calculated to be +LR = 1.6 (95%CI 1.3, 2.0); -LR = 0.13 (0.04, 0.41).	Post-test probability for scores ≥7 points not reported but calculated from study data to be 60%. 95% CrI for scores ≥7 points calculated from study data to be 50.1%, 69.9%.

Diagnosis supported	CPR	Function	Predictors	Format	Study	Patients	Reference standard	Accuracy	Post-test probability
	2.	Identification of patients with probable lumbar spinal stenosis from self-reported patient information	1. older age; 2. duration of symptoms > 6 months; 3. improvement of symptoms when bending forward; 4. no improvement of symptoms when bending backward; 5. occurrence of symptoms when standing up; 6. symptoms occurring when walking are improved by resting; 7. urinary incontinence	Score chart	Derivation Sugioka et al. (2008)	n=468, patients with pain or numbness in the lower extremities presenting to primary care clinics, mean age 65 years (SD 14), 46% female, duration of symptoms not reported for overall sample, randomly allocated to training set (n=374) and validation set (n=94), 47% prevalence of dependent outcome in training set, 50% prevalence of dependent outcome in validation set	Diagnosis of lumbar spinal stenosis by an orthopaedic surgeon and verified by study coordinator (dichotomous)	For scores ≥5 points, sens = 0.81; spec = 0.58 in training set, and sens = 0.75; spec = 0.51 in validation set. LRs not reported, but calculated to be +LR = 1.9 (95% CI 1.6, 2.3); -LR = 0.33 (0.24, 0.46) in training set; and +LR = 1.5 (1.1, 2.1); -LR = 0.50 (0.29, 0.88) in validation set.	For scores ≥5 points = 63% in training set, and 60% in validation set. 95% CrI for scores ≥5 points calculated from study data to be 56.4%, 69.0% in training set and 47.4%, 72.0% in validation set.

Diagnosis supported	CPR	Function	Predictors	Format	Study	Patients	Reference standard	Accuracy	Post-test probability
	3.	To indicate the likelihood of the presence of lumbar spinal stenosis	1. bilateral symptoms; 2. leg pain more than back pain; 3. pain during walking/standing; 4. pain relief upon sitting; 5. age > 48 years.	Count of predictors	Derivation C. Cook et al. (2011)	n=1448, patients with a primary low back pain complaint presenting to tertiary care institution, mean age 55 years (calculated), 59% female, duration of symptoms not reported, 40% prevalence of dependent outcome	Diagnosis of lumbar spinal stenosis by an orthopaedic surgeon (dichotomous)	For 4 predictors present sens = 0.06 (0.05, 0.07); spec = 0.98 (0.98, 0.99); +LR = 4.6 (95%CI 2.4, 8.9); -LR = 0.95 (0.94, 0.97).	Post-test probability for 4 predictors present = 76%. 95% CrI for 4 predictors present approximated from study data to be 61.8%, 85.4%.
Vertebral fracture	4.	Identification of the need for radiography in postmenopausal women with osteoporosis presenting with back pain	1. age; 2. intensity of back pain; 3. height loss (reported height at age 25 - present height); 4. history of low-trauma non-vertebral fracture; 5. sudden occurrence of pain; 6. thoracic localisation of pain	Score chart	Derivation Roux et al. (2007)	n=410, patients with osteoporosis aged 65-85 presenting to a rheumatologist with back pain, mean age 74 years (SD 6), 100% female, mean 61 months (SD 97) of back pain symptoms, 52% prevalence of dependent outcome	Radiographic evidence of 1 or more vertebral fractures (dichotomous)		Scores ≥7, probability of fracture ≥43%. Scores ≤2, probability of fracture <20%

Diagnosis supported	CPR	Function	Predictors	Format	Study	Patients	Reference standard	Accuracy	Post-test probability
	5.	Identification of patients with probable osteoporotic vertebral compression fracture	1. age > 52 years; 2. no presence of leg pain; 3. BMI ≤ 22; 4. does not exercise regularly; 5. female gender	Count of predictors	Derivation Roman et al. (2010)	n=1448, patients with a primary low back pain complaint presenting to tertiary care institution, mean age 55 years (calculated), 59% female, duration of symptoms not reported, 3% prevalence of dependent outcome	Diagnosis of lumbar compression fracture or wedge deformity by standard radiographs or CT scan (dichotomous)	If 2 of 5 present, sens = 0.95 (95%CI 0.83, 0.99); spec = 0.34 (0.33, 0.34); -LR = 0.16 (0.04. 0.51). If 4 of 5 present, sens = 0.37 (0.24, 0.51); spec = 0.96 (0.95, 0.97); +LR = 9.6 (3.7, 14.9).	Post-test probability for 2 of 5 negative not reported but approximated from study data to be 0.4% and 95% CrI for 2 predictors not positive approximated to be 0.1%, 1.5%. Post-test probability for 4 of 5 present = 20%. 95%CrI for 4 predictors positive approximated to be 12.0%, 30.4%.
Inflammatory back pain	6.	Identification of patients with probable inflammatory back pain	1. morning stiffness > 30 minutes; 2. improvement in back pain with exercise but not with rest; 3. awakening because of back pain during the second half of the night only; 4. alternating buttock pain	Count of predictors	Derivation Rudwaleit et al. (2006)	n=213, non-consecutive convenience sample of patients with chronic back pain already diagnosed with either AS (n=101, mean age 36 years (SD 8), 36% female, mean 13 years (SD 9) of symptoms) or mechanical low back pain (n=112, mean age 39 years (SD 8), 41%	Diagnosis of AS by a rheumatologist or other specialist prior to the study, using the modified New York criteria (dichotomous)	For 2 or more predictors present, sens = 0.70 (95%CI 0.61, 0.78); spec = 0.81 (0.73, 0.87); +LR = 3.7 (2.5, 5.6); -LR = 0.4 (0.3, 0.5).	Case-control design, thereby prohibiting prevalence and post-test probability estimates.

Diagnosis supported	CPR	Function	Predictors	Format	Study	Patients	Reference standard	Accuracy	Post-test probability
						female, mean 12 years (SD 10) of symptoms)			
					Validation Sieper, van der Heijde, et al. (2009)	n=20, patients selected by rheumatologists with suspected axial SpA, mean age 41 years (SD 11), 60% female, mean 7 years (SD 6) duration of symptoms	Diagnosis of inflammatory back pain by rheumatologist with expertise in AS/SpA (dichotomous)	For 2 or more predictors present, sens = 0.84; spec = 0.63. LRs not reported, but calculated to be +LR = 2.2 (95%CI 1.5, 3.3); -LR = 0.26 (0.14, 0.48).	Analysis conducted using 109 judgements from 13 assessors across 20 patients. Prevalence and post-test probability estimates unable to be calculated.
					Validation Sieper, van der Heijde, et al. (2009)	n=648, patients with chronic back pain of unknown origin that began before 45 years of age presenting to a rheumatologist, mean age 34 years, 55% female, mean duration of symptoms 7.3 years, approximated prevalence of dependent outcome is 66%	Diagnosis of inflammatory back pain by rheumatologist (dichotomous)	For 2 or more predictors present, sens = 0.70; spec = 0.81. LRs not reported, but approximated to be +LR = 3.8 (95%CI 2.8, 5.0); -LR = 0.37 (0.31, 0.43).	Post-test probability for 2 or more predictors present not reported but approximated from study data to be 88%. 95% CrI approximated from study data to be 84.1%, 91.0%.

Diagnosis supported	CPR	Function	Predictors	Format	Study	Patients	Reference standard	Accuracy	Post-test probability
					Validation Chan et al. (2012)	n=25, patients with anterior uveitis, demographics not reported for this subgroup of the larger study (n=141), prevalence of dependent outcome calculated to be 52%	Diagnosis of inflammatory back pain by a rheumatologist (dichotomous)	For 2 or more predictors present, sens = 0.92; spec = 0.67. LRs not reported, but calculated to be +LR = 2.8 (95%CI 1.2, 6.3); -LR = 0.12 (0.02, 0.79).	Post-test probability for 2 or more predictors present not reported but calculated from study data to be 75%. 95% CrI calculated from study data to be 49.7%, 89.9%.
	7.	Identification of patients with probable inflammatory back pain	1. age at onset < 40 years; 2. insidious onset; 3. improvement with exercise; 4. no improvement with rest; 5. pain at night with improvement upon getting up	Count of predictors	Derivation Sieper, van der Heijde, et al. (2009)	n=20, patients selected by rheumatologists with suspected axial SpA, mean age 41 years (SD 11), 60% female, mean 7 years (SD 6) duration of symptoms	Diagnosis of inflammatory back pain by rheumatologist with expertise in AS/SpA (dichotomous)	For 4 or more predictors present, sens = 0.77; spec = 0.92. LRs not reported, but calculated to be +LR = 9.2 (95%CI 3.6, 23.9); -LR = 0.25 (0.16, 0.40).	Analysis conducted using 109 judgements from 13 assessors across 20 patients. Prevalence and post-test probability estimates unable to be calculated.

Diagnosis supported	CPR	Function	Predictors	Format	Study	Patients	Reference standard	Accuracy	Post-test probability
					Validation Sieper, van der Heijde, et al. (2009)	n=648, patients with chronic back pain of unknown origin that began before 45 years of age presenting to a rheumatologist, mean age 34 years, 55% female, mean duration of symptoms 7.3 years, approximated prevalence of dependent outcome is 66%	Diagnosis of inflammatory back pain by rheumatologist (dichotomous)	For 4 or more predictors present, sens = 0.80; spec = 0.72. LRs not reported, but approximated to be +LR = 2.9 (95%CI 2.3, 3.6); -LR = 0.28 (0.23, 0.35).	Post-test probability for 4 or more predictors present not reported but approximated from study data to be 85%. 95% CrI approximated from study data to be 81.0%, 88.0%.
Spondyloarthritis	8.	Identification of patients with axial SpA in primary care	1. age at onset ≤ 35 years; 2. alternating buttock pain; 3. improvement by NSAIDs within 48 hours; 4. waking up in the second half of the night; 5. improvement with movement and not by rest	Count of predictors	Derivation A. Braun et al. (2011)	n=322, patients with low back pain > 2 months presenting to orthopaedic surgeons, most with ≥1 inflammatory back pain symptom, mean age 36 years (SD 8), 51% female, mean 44 months (SD 38) duration of symptoms, 35% prevalence of dependent outcome	Diagnosis of axial SpA as determined by a rheumatologist (dichotomous)	For 4 or more predictors present, sens = 0.48; spec = 0.86; +LR = 3.4; -LR = 0.6.	Post-test probability for 4 or more predictors present not reported but calculated from study data to be 65%. 95% CrI calculated from study data to be 54.2%, 74.4%.

Diagnosis supported	CPR	Function	Predictors	Format	Study	Patients	Reference standard	Accuracy	Post-test probability
Zygapophyseal joint mediated pain	9.	Identification of patients with probable lumbar zygapophyseal joint mediated pain	1. pain well relieved by recumbency; 2. presence of 5 of more of the following 7 variables; a. age>65 years; b. pain not exacerbated by coughing; c. not worsened by hyperextension; d. not worsened by forward flexion; e. not worsened when rising from flexion; f. not worsened by extension-rotation; g. well-relieved by recumbency	Count of predictors	Derivation Revel et al. (1998)	n=42 (lidocaine group), patients with chronic low back pain non-responsive to conservative management referred for zygapophyseal injections, full-study demographics (n=80): mean age 58 years (range 34-87), 68% female, mean 79 weeks duration of symptoms, 31% prevalence of dependent outcome	Reduction in low back pain measured on 100mm VAS by more than 75%, 30 minutes following intra-articular injection of lidocaine into lower lumbar zygapophyseal joints (dichotomous)	For both predictors present, sens = 0.92; spec = 0.80. LRs not reported, but approximated to be +LR = 4.5 (95%CI, 2.2, 9.3); -LR = 0.10 (0.02, 0.64).	Post-test probability for both predictors present not reported but approximated from study data to be 67%. 95% CrI approximated from study data to be 42.2%, 83.3%.

Diagnosis supported	CPR	Function	Predictors	Format	Study	Patients	Reference standard	Accuracy	Post-test probability
Radicular low back pain	10.	Differentiation of radicular and axial low back pain	1. radicular pain in the SLR test; 2. abnormal response to cold temperature; 3. abnormal response to pinprick; 4. abnormal response to blunt pressure; 5. decreased response to vibration; 6. dysaethesia; 7. temporal summation; 8. burning or cold quality of pain; 9. abnormal response to brush movement; 10; ongoing pain; 11. skin changes	Score chart	Derivation Scholz et al. (2009)	n=137 (described as validation set in study) patients with chronic low back pain referred to neurosurgical outpatient triage clinic, demographics for radicular low back pain group = median age 45 (range 20-82), 55% female, median 1 year (range 0.3-34) of symptoms; for axial low back pain group = median age 55 years (range 24-78), 56% female, median 5 years (range 0.3-46) of symptoms, 55% prevalence of dependent outcome	Radicular low back pain as determined by team consisting of rheumatologist, neurosurgeon and physiotherapist (dichotomous)	Score ≥4 points, sens = 0.92 (95%CI 0.83, 0.97); spec = 0.97 (0.89, 1.0). LRs not reported, but calculated to be +LR = 28.5 (7.3, 111.7); -LR = 0.08 (0.04, 0.18).	For scores ≥4 points = 97% (90%, 100%). 95% CrI for scores ≥4 points calculated from study data to be 90.3%, 99.1%.

Diagnosis supported	CPR	Function	Predictors	Format	Study	Patients	Reference standard	Accuracy	Post-test probability
Radiographic instability	11.	Identification of patients with probable radiographic instability	1. lack of hypomobility during intervertebral testing; 2. lumbar flexion ROM > 53°	Count of predictors	Derivation Fritz, Piva, et al. (2005)	n=49, patients with low back pain referred for imaging on suspicion of instability, mean age 39 years (SD 11), 57% female, median 78 days (variability not reported) of symptoms, 57% prevalence of dependent outcome	2 segments with either rotational or translational instability or 1 segment with both translational and rotational instability on radiographic assessment (dichotomous)	For both predictors present, sens = 0.29 (95%CI 0.13, 0.46); spec = 0.98 (0.91, 1.0); +LR = 12.8 (0.79, 211.6); -LR = 0.72 (0.55, 0.94).	For both predictors present = 93%. 95% CrI for both predictors present calculated from study data to be 66.6%, 99.7%.
Spondylolysis	12.	Identification of patients with probable active spondylolysis	1. male gender; 2. age ≤ 20 years	Count of predictors	Derivation Gregg et al. (2009)	n=82, patients with low back pain with suspected spondylolysis referred for SPECT bone scan, 66% under 20 years of age, 48% female, 52% symptom duration < 3 months, 32% prevalence of dependent outcome	Diagnosis of active spondylolysis by SPECT bone scan (dichotomous)	For both predictors present, sens = 0.62; spec = 0.84; +LR = 3.66; -LR = 0.27.	Post-test probability for both predictors present not reported but calculated from study data to be 64%. 95% CrI calculated from study data to be 43.8%, 79.5%.

Diagnosis supported	CPR	Function	Predictors	Format	Study	Patients	Reference standard	Accuracy	Post-test probability
Psychological disturbance	13.	Identification of patients with chronic low back pain who might benefit from additional psychological assessment	1. number of Waddell signs (0-8); 2. pain drawing score (0-100); 3. no directional preference; 4. daily use of pain medication for chronic low back pain	Score chart	Derivation Apeldoorn et al. (2012)	n=194, patients with low back pain > 3 months attending outpatient rehabilitation, mean age 44 years (SD 11), 63% female, median 84 months duration of symptoms (IQR 24-180), 53% prevalence of dependent outcome	Relevant psychological disturbance as determined by a psychologist (dichotomous)		Score ranges matched to post-test probability ranges; <21 = <20%; 21-29 = 20-30%; 30-35 = 30-40%; 36-41 = 40-50%; 42-48 = 50-60%; 49-54 = 60-70%; 55-62 = 70-80%; 63-74 = 80-90%; >74 = 90-100%.

8.4.3 Qualitative appraisal of included studies

Table 8.6 summarizes the quality appraisal of included studies using QUADAS-2. The most frequent potential risk of bias concerned the reference standard. In particular, evidence supporting the reliability and validity of the reference standard was often lacking and its measurement was non-blinded in several studies. All but 4 studies (Chan et al., 2012; C. Cook et al., 2011; Gregg et al., 2009; Roman et al., 2010) used a prospective research design. The overall interrater agreement for quality appraisal using the QUADAS-2 was 'moderate' (Landis & Koch, 1977) (κ = 0.60; 95%CI 0.45, 0.75; absolute agreement, 83%).

The methodological appraisal of derivation studies is detailed in Table 8.7. Within these studies, a lack of assessment of predictor collinearity, lack of reporting of uncertainty intervals for posterior probabilities, insufficient outcome events per number of candidate predictor variables, dichotomization of continuous predictor variables, and insufficient evidence concerning the reliability of predictor variables were common sources of potential bias. Interrater concordance for quality appraisal using the tool developed for CPR derivation studies was 'substantial' (Landis & Koch, 1977) (κ = 0.65; 95%CI 0.56, 0.73; absolute agreement, 79%).

Table 8.6 Risk of potential bias of included studies as appraised using QUADAS-2

Study	CPR reference number	Stage of development	Risk of bias				Applicability concerns		
			Patient selection	Index tests	Reference standard	Flow and timing	Patient selection	Index tests	Reference standard
Apeldoorn et al. (2012)	13	Derivation	Low risk	Low risk	High risk	High risk	Low risk	Low risk	Low risk
A. Braun et al. (2011)	8	Derivation	Low risk	Low risk	High risk	High risk	Low risk	Low risk	Low risk
Chan et al. (2012)	6	Validation	High risk	Low risk	High risk	Low risk	Low risk	Low risk	Low risk
C. Cook et al. (2011)	3	Derivation	Low risk	High risk	High risk	Low risk	Low risk	Low risk	Low risk
Fritz, Piva, et al. (2005)	11	Derivation	Unclear	High risk	Low risk	Low risk	Low risk	Low risk	Low risk
Gregg et al. (2009)	12	Derivation	High risk	Low risk	Unclear	Unclear	Low risk	Low risk	Low risk
Kato et al. (2009)	1	Validation	Unclear	Low risk	High risk	Low risk	Low risk	Low risk	Low risk
Konno et al. (2007)	1	Derivation	Low risk	Low risk	High risk	Low risk	Low risk	Low risk	Low risk
Revel et al. (1998)	9	Derivation	Low risk	Low risk	Unclear	Unclear	Low risk	Low risk	Low risk
Roman et al. (2010)	5	Derivation	Low risk	High risk	Unclear	Unclear	Low risk	Low risk	Low risk
Roux et al. (2007)	4	Derivation	Unclear	Unclear	Low risk	Low risk	Low risk	Low risk	Low risk

Study	CPR reference number	Stage of development	Risk of bias				Applicability concerns		
			Patient selection	Index tests	Reference standard	Flow and timing	Patient selection	Index tests	Reference standard
Rudwaleit et al. (2006)	6	Derivation	High risk	High risk	Low risk	Low risk	Low risk	Low risk	Low risk
Scholz et al. (2009)	10	Derivation	Unclear	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
Sieper, van der Heijde, et al. (2009)	7	Derivation	High risk	Low risk	High risk	High risk	Low risk	Low risk	Low risk
Sieper, van der Heijde, et al. (2009)	6	Validation	High risk	Low risk	High risk	High risk	Low risk	Low risk	Low risk
Sieper, van der Heijde, et al. (2009)	6	Validation	Low risk	Low risk	High risk	Low risk	Low risk	Low risk	Low risk
Sieper, van der Heijde, et al. (2009)	7	Validation	Low risk	Low risk	High risk	Low risk	Low risk	Low risk	Low risk
Sugioka et al. (2008)	2	Derivation	Low risk	Low risk	High risk	Low risk	Low risk	Low risk	Low risk

Table 8.7 Methodological appraisal of included derivation studies

Study	CPR reference number	Prospective design	Study site described	Justification for number of participants	Representative sample	Important patient characteristics described	Candidate predictor variables justified	Blinded predictor assessment	Predictor variables have demonstrated reliability	Reference standard valid and reliable	Blinded assessment of reference standard	Mathematical techniques described	Reporting and handling of missing data	10 outcome events per variable in final model	10 outcome events per candidate variable	Collinearity of predictor variables assessed	Predictor variables kept continuous	Uncertainty in CPR accuracy described	Uncertainty in posttest probability described	Non-paradoxical performance
Apeldoorn et al. (2012)	13	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Partly	Yes	No	No	Yes	N/A	No	Yes
A. Braun et al. (2011)	8	Yes	Partly	No	Yes	Yes	No	Yes	No	No	No	Yes	Partly	Yes	Yes	No	N/A	No	N/A	Yes
C. Cook et al. (2011)	3	No	Yes	Partly	Yes	Yes	Yes	No	No	Partly	No	Yes	No	Yes	Yes	No	No	Yes	No	Yes
Fritz, Piva, et al. (2005)	11	Yes	No	No	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Partly	Yes	No	Yes	No	Yes	No	Yes
Gregg et al. (2009)	12	No	Yes	No	No	Yes	Yes	No	Partly	Yes	No	Yes	No	Yes	No	No	N/A	No	N/A	Yes
Konno et al. (2007)	1	Yes	Partly	No	Yes	Yes	No	Yes	No	No	No	Yes	Partly	Yes	No	No	N/A	No	No	Yes
Revel et al. (1998)	9	Yes	No	No	Partly	Yes	Yes	Yes	No	No	No	Yes	No	No	No	No	N/A	No	N/A	N/A
Roman et al. (2010)	5	No	Yes	Partly	Yes	Yes	Yes	No	No	Partly	No	Yes	No	No	No	No	No	Yes	No	No
Roux et al. (2007)	4	Yes	Partly	No	Partly	Yes	Partly	Yes	No	Partly	No	Yes	Partly	Yes	No	No	No	N/A	No	Yes
Rudwaleit et al. (2006)	6	Yes	Yes	Partly	No	Yes	Yes	No	No	Partly	Yes	Yes	No	Yes	No	No	N/A	Yes	No	Yes
Scholz et al. (2009)	10	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No	No	Yes	Yes	No	No	No	N/A	Yes	Yes	Yes
Sieper, van der Heijde, et al. (2009)	7	Yes	Yes	No	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	No	No	No	N/A	No	N/A	Yes
Sugioka et al. (2008)	2	Yes	Partly	Yes	Yes	Yes	No	Yes	No	No	No	Yes	Partly	Yes	No	No	No	No	No	Yes

Table 8.8 Methodological appraisal of included validation studies

Study	CPR reference number	Prospective validation in new sample	Different clinical setting	Representative sample	The rule is applied accurately	Reliability of the rule is assessed	Complete follow-up	Reporting and handling of missing data	Accuracy of the rule described	Accuracy uncertainty described	Posttest probability uncertainty described
Chan et al. (2012)	6	No	Yes	No	Yes	No	Yes	Yes	Yes	No	N/A
Kato et al. (2009)	1	Yes	Yes	Partly	Yes	No	Yes	Yes	Yes	No	N/A
Sieper, van der Heijde, et al. (2009)	6	Yes	Yes	No	Yes	No	Partly	Partly	Yes	No	N/A
Sieper, van der Heijde, et al. (2009)	6	Yes	Yes	Yes	Yes	No	No	Partly	Yes	No	N/A
Sieper, van der Heijde, et al. (2009)	7	Yes	Yes	Yes	Yes	No	No	Partly	Yes	No	N/A

Table 8.8 provides a summary of the methodological appraisal of included validation studies. A lack of assessment concerning the reliability of CPR interpretation, lack of reporting of uncertainty intervals for accuracy data, incomplete follow-up, and non-representative samples were the most commonly identified potential sources of bias. There was 'moderate' (Landis & Koch, 1977) overall agreement between raters for the methodological appraisal of validation studies (κ = 0.54; 95%CI 0.33, 0.75; absolute agreement, 73%).

8.5 Discussion

The primary finding of this review is that the majority of diagnostic LBP CPRs have not yet been developed beyond the initial derivation phase and therefore cannot be recommended for use in clinical practice at this time. Clinicians may however wish to consider which variables were and were not identified to be independently predictive of the target diagnosis within each CPR derivation study to cautiously inform their clinical decision making (McGinn et al., 2008). Clinicians should be aware however that such findings may simply reflect chance associations or may be unique to the patient sample or setting in which the study was conducted.

Three diagnostic LBP CPRs (Figure 8.2) were identified to have undergone validation in this review: a tool to assist in the identification of patients with spinal stenosis (Kato et al., 2009; Konno et al., 2007) and 2 tools designed to assist in the identification of patients presenting with inflammatory back pain

(Chan et al., 2012; Rudwaleit et al., 2006; Sieper, van der Heijde, et al., 2009).

8.5.1 Lumbar spinal stenosis

Three CPRs were identified to have been derived to assist in the identification of lumbar spinal stenosis, 1 of which underwent validation. Older age, pain with standing/walking, and relief with sitting/bending were identified as common independent predictors across all 3 CPRs. The validated tool is a 10-item CPR derived from a sample of 468 patients presenting to 1 of 72 clinics and hospitals with primary symptoms of pain or numbness in the legs (Konno et al., 2007). This CPR was subsequently validated in a prospective study involving 118 patients with lower extremity symptoms who presented to 1 of 10 hospital outpatient orthopaedic clinics (Kato et al., 2009). Based on a score of ≥7 points on the CPR, the +LR in the derivation study was reported to be 3.3 and calculated to be 1.6 (95%CI 1.3, 2.0) in the validation study. This implies that the tool may only have limited power in increasing the diagnostic probability of lumbar spinal stenosis. Using the validation study data, positive status on the tool shifts the 49% pre-test probability of lumbar spinal stenosis to a post-test probability of 60% (95% credible interval (CrI): 50%, 70%).

Japanese Society for Spine Surgery and Related Research clinical prediction rule to assist in the identification of lumbar spinal stenosis (Konno et al 2007, Kato et al 2009)		
Variable	**Status**	**Score**
Age (years)	< 60	0
	60 – 70	1
	> 70	2
Co-morbidity of diabetes	Present	0
	Absent	1
Intermittent claudication (from history)	Absent	0
	Present	3
Exacerbation of symptom(s) when standing up (from history)	Absent	0
	Present	2
Symptom(s) improvement when bending forward (from history)	Absent	0
	Present	3
Symptom(s) induced by having a patient bend forward (from physical examination)	Present	-1
	Absent	0
Symptom(s) induced by having a patient bend backward (from physical examination)	Absent	0
	Present	1
Peripheral artery circulation (from physical examination)	Dorsalis pedis artery not easily palpable, or ankle brachial index < 0.9	0
	Dorsalis pedis artery easily palpable, or ankle brachial index ≥ 0.9	3
Achilles tendon reflex (from physical examination)	Normal	0
	Absent or low	1
Straight leg raise test (from physical examination)	Positive	-2
	Negative	0
Sum score of ≥ 7 commonly used to denote positive status on this tool		

'Berlin criteria' to assist in the identification of inflammatory back pain (Rudwaleit et al 2006, Sieper et al 2009, Chan et al 2012)
Morning stiffness > 30 minutes duration
Improvement in back pain with exercise but not with rest
Awakening because of back pain during the second half of the night only
Alternating buttock pain
Presence of 2 or more variables commonly used to denote positive status on this tool

'IBP according to experts' clinical prediction rule to assist in the identification of inflammatory back pain (Sieper et al 2009)
Age at onset < 40 years
Insidious onset
Improvement with exercise
No improvement with rest
Pain at night with improvement upon getting up
Presence of 4 or more variables commonly used to denote positive status on this tool

Figure 8.2 Summary and scoring of 3 diagnostic clinical prediction rules for specific subtypes of low back pain that have undergone validation

In contrast, the magnitude of the -LR reported in the derivation study (0.1) and calculated for the validation study (0.13, 95%CI 0.04, 0.41) suggests that it may have a relatively larger influence in downwardly revising the probability of the diagnosis given a patient's negative status on the rule. Using the validation study data, the post-test probability of a patient who is negative on the rule is 11% (95%CrI: 4%, 28%). Given this uncertainty interval however, it is probable that in many clinical contexts additional negating evidence may be needed.

This review failed to identify any impact analysis studies. It is conceivable, however, that an improved ability to identify patients with lumbar spinal stenosis may have important treatment implications. For example, surgical intervention has been demonstrated to provide superior clinical improvements compared to conservative management strategies in selected patients identified to have lumbar spinal stenosis (Kovacs et al., 2011; Weinstein et al., 2010). An important consideration however in the interpretation of any single or multi-item diagnostic test for lumbar spinal stenosis is that there is no universally accepted 'gold standard'. It is a clinical diagnosis and often based on expert opinion that incorporates knowledge of clinical variables, imaging, and/or outcomes from surgery (de Graaf et al., 2006; de Schepper et al., 2013; Kreiner et al., 2013). Inconsistencies in the interpretation of the presence or absence of lumbar spinal stenosis was identified in all studies included in this review which used more than 1 rater to determine the reference standard (Kato et al., 2009; Konno et al., 2007;

Sugioka et al., 2008). Such variability in the interpretation of the reference standard raises concerns regarding the meaningfulness of any test or tool designed to predict that reference standard. It is well recognized that there is a strong need for consensus on the criteria that may serve as a reference standard to classify spinal stenosis (de Schepper et al., 2013).

8.5.2 Inflammatory back pain

Inflammatory back pain is associated with a group of related disorders collectively referred to as the spondyloarthritides and include ankylosing spondylitis, psoriatic spondyloarthritis, reactive spondyloarthritis, spondyloarthritis associated with inflammatory bowel disease, and undifferentiated spondyloarthritis (Baraliakos & Braun, 2011; J. Braun & Inman, 2010; Harper & Reveille, 2009; Healy & Helliwell, 2005; Keith, 2012; Rudwaleit, van der Heijde, Khan, Braun, & Sieper, 2004). The early identification of these disorders is integral to their successful management (J. Braun & Sieper, 2007; Ozgocmen & Khan, 2012; Rudwaleit, Listing, Brandt, Braun, & Sieper, 2004), however radiological manifestations often take several years to develop (Rudwaleit, van der Heijde, et al., 2004). As such, several screening tools to identify patients with inflammatory back pain have been developed (Burgos-Vargas & Braun, 2012; Sieper, Rudwaleit, et al., 2009; Weisman, 2012), 2 of which satisfied the operational definition of a CPR used in this review. Both of these tools were identified to have undergone validation. Importantly, such tools function to help identify patients with probable inflammatory back pain who are likely to benefit from further investigation and referral (Golder & Schachna, 2013; Kain et al., 2008;

Sieper, 2012). They are not intended to diagnose spondyloarthritis and current classification criteria do not require an individual to have inflammatory back pain to be diagnosed as having axial spondyloarthritis (Rudwaleit et al., 2009). While inflammatory back pain is commonly considered to be a characteristic feature of spondyloarthritis, it is only modestly sensitive and specific to that presentation (J. Braun & Inman, 2010).

The 'Berlin criteria' comprise a 4-item CPR that was derived using a case-control design involving a convenience sample of patients previously diagnosed as having either ankylosing spondylitis or mechanical LBP (Rudwaleit et al., 2006). The CPR was subsequently validated in a convenience sample of 20 patients with suspected axial spondyloarthritis and in a multinational study involving 648 consecutive patients with chronic back pain presenting to a rheumatologist (Rudwaleit et al., 2009; Sieper, van der Heijde, et al., 2009). A further validation study included in this review examined the performance of the 'Berlin criteria' in 25 patients with anterior uveitis presenting to an ophthalmologist (Chan et al., 2012). The point-estimate of the +LR across the derivation and 3 validation studies for the cut-off point of 2 or more predictors being present ranges from 2.2 to 3.8. This indicates that positive status on the rule may have a small but sometimes important influence on the likelihood of the presence of inflammatory back pain (Jaeschke et al., 1994). Using data from a large multinational validation study of consecutive patients presenting to rheumatologists (Rudwaleit et al., 2009; Sieper, van der Heijde, et al., 2009), positive status on the CPR would shift the pre-test probability of inflammatory back pain from approximately

66% to 88% (95%CrI: 84%, 91%), thus highlighting its potential value in this setting.

'IBP according to experts' is a 5-item CPR that shares similar predictor variables to the 'Berlin criteria'. It was derived using judgements from 13 rheumatologists with expertise in spondyloarthritis in a convenience sample of 20 patients suspected of having the disorder (Sieper, van der Heijde, et al., 2009). The reference standard was the presence of inflammatory back pain as determined by the expert rheumatologists and was established with knowledge of the candidate predictor variables. The derived tool was subsequently validated concurrently with the 'Berlin criteria' in a large multinational study involving consecutive patients with chronic back pain presenting to rheumatologists (Rudwaleit et al., 2009; Sieper, van der Heijde, et al., 2009). Using a cut-off point of 4 or more predictors being present, the 'IBP according to experts' tool was identified to be slightly more sensitive (80% versus 70%) and slightly less specific (72% versus 81%) than the 'Berlin criteria' in identifying patients with inflammatory back pain.

No studies were identified in this review which examined the performance of either the 'Berlin' or 'IBP according to experts' criteria in a primary care setting. The prevalence of inflammatory back pain in patients with back pain presenting to primary care is estimated to be considerably lower (<15%) than in those presenting to a rheumatologist (Hamilton, Macgregor, Warmington, Pinch, & Gaffney, 2014; Underwood & Dawes, 1995). Given that test accuracy differs in populations of varying disease prevalence, it is probable

that the performance of both CPRs may differ non-predictably in a primary care setting to that observed in a rheumatology setting (Feinstein, 2002; Leeflang, Bossuyt, & Irwig, 2009). Notwithstanding this limitation, screening tools and referral strategies involving the assessment of inflammatory back pain have been demonstrated to be helpful in primary care in aiding the earlier diagnosis of spondyloarthritis (Brandt et al., 2007; Hermann, Giessauf, Schaffler, Ofner, & Graninger, 2009; Poddubnyy et al., 2011; Sieper et al., 2013). The evidence considered within this review however does not permit recommendations concerning whether the 'Berlin' or 'IBP according to experts' criteria are the optimal tools for such purposes, or whether their application results in improved clinical outcomes or resource efficiencies. Nevertheless, the 'IBP according to experts' criteria have been endorsed and adopted by organizations in at least 2 countries to enhance public and health professional awareness of spondyloarthritis (Arthritis & Osteoporosis New South Wales, 2014; Arthritis Care, 2011; National Ankylosing Spondylitis Society, 2014).

8.5.3 Methodological considerations

The findings of this review have highlighted several opportunities to improve the methodological quality of future diagnostic CPR development studies. For a CPR to accurately identify patient subgroups outside of the study from which it was derived the reference standard must be measurably valid and reliable. While LBP sub-presentations often lack a definitive 'gold standard', the use of reproducibly identifiable reference standards that represent the broad consensus of a presentation's classification are arguably integral to the

development of a CPR. For the same reason, the selection of candidate predictor variables should ideally be confined to those that can be reliably assessed.

The study sample should be sufficiently large to ensure at least 10 outcome events per candidate predictor variable (Bouwmeester et al., 2012; van Oort et al., 2012). Univariate screening of candidate predictor variables to select items for inclusion in a multivariable model is common but may not reduce the risk of overfitting (Babyak, 2004). All else being equal, larger sample sizes will also result in narrower accuracy and posterior probability uncertainty intervals which need to be consistently reported in CPR development studies (Haskins, Osmotherly, Tuyl, et al., 2014). Collinearity of predictor variables in regression models should be assessed as non-trivial correlations may give invalid results (Lubetzky-Vilnai et al., 2014). Finally, continuous predictor variables (eg, range of motion, age) should be kept as such until at least after the multivariable analysis, as their transformation into categorical variables results in poorer performing models and influences which variables are identified as significantly related to the reference standard (Lubetzky-Vilnai et al., 2014; Schellingerhout, Heymans, de Vet, Koes, & Verhagen, 2009).

8.5.4 Limitations

Study eligibility is notably sensitive to how a CPR is operationally defined. The definition used in this review was designed to reflect the most common use of the term ‘clinical prediction rule’ in the literature and incorporates

minimum accepted methodological standards. The present review excluded studies investigating the diagnostic accuracy of multi-test regimes whereby the individual tests were not selected from a larger pool of candidate predictor variables using multivariable analysis. This lead to the exclusion of 6 studies (Henschke et al., 2009; Laslett, Aprill, et al., 2006; Laslett, Aprill, et al., 2005; Laslett, McDonald, et al., 2006; Laslett et al., 2003; van der Wurff et al., 2006) included in an earlier review of diagnostic CPRs for back pain (Haskins et al., 2012).

There is no validated standardised tool to appraise the methodological quality of diagnostic forms of CPRs for each stage of their development. As such, quality assessment items were selected in this review based upon their incorporation of well cited methodological standards, methodological considerations identified in the recent CPR literature, and their use in previous systematic reviews. The derivation and validation study quality appraisal tools used in this review have not been validated. It would be inappropriate to calculate a 'sum score' for each study or to otherwise quantitatively synthesize the findings of the quality appraisal analysis summarized in this review.

8.6 Conclusions

There has been significant research investment into the identification of meaningful subgroups of patients presenting with LBP. Thirteen diagnostic CPRs have been derived for this purpose to date, although just 3 of these

tools have undergone validation. No impact analysis studies have been conducted, and therefore no evidence-informed statements can be provided at this time regarding the ability of diagnostic CPRs to beneficially impact clinical practice in the management of patients presenting with LBP.

Validation of the derived CPRs identified in this review is indicated, particularly for those tools which meet an identified unmet need of clinicians who manage patients with LBP. Further validation and impact analysis of a 10-item CPR designed to aid the identification of patients presenting with lumbar spinal stenosis is also warranted. Research is also needed to validate the ‘Berlin criteria’ and ‘IBP according to experts’ CPRs in primary care settings and to investigate whether their clinical application results in improved patient outcomes or beneficial improvements to resource consumption.

CHAPTER 9

VALIDATION AND IMPACT ANALYSIS OF PROGNOSTIC CLINICAL PREDICTION RULES FOR LOW BACK PAIN IS NEEDED: A SYSTEMATIC REVIEW

This chapter has been published in a peer-reviewed scientific journal (Appendix 4):

Haskins, R., Osmotherly, P. G., & Rivett, D. A. (2015). Validation and impact analysis of prognostic clinical prediction rules for low back pain is needed: a systematic review. *Journal of Clinical Epidemiology, 68*(7), 821-832.

The work presented in this manuscript was completed in collaboration with the co-authors (Appendix 1).

Overview

This is the final study comprising this program of research and was conducted concurrently with Study 4 (Chapter 8, p. 256). One of the key findings from Study 3 (Chapter 6, p. 210) was that physiotherapists within that study indicated a preference for the development of prognostic forms of CPRs for LBP for a range of varying functions. The results of Study 2

(Chapter 5, p. 174) however also indicated that physiotherapists' awareness and familiarity with such tools may be quite limited. A resource for clinicians and researchers that seeks to identify the range of prognostic forms of CPRs for LBP and appraise their current readiness for clinical application would consequently be of significant value.

Prior to the present study, attempts to synthesise the available literature base in this area had been limited to CPRs designed for specific interventions, a particular health profession, or to a particular stage of CPR development. Given the narrower scope of these reviews, it is highly probable that many relevant prognostic CPRs for LBP had not yet been identified or assessed for their clinical appropriateness to be implemented in practice. This presents challenges both for clinicians considering applying such tools in practice, and for researchers aiming to derive or progress their development. Also, given the near exponential growth of research in this area in recent years (Figure 3.1, p. 95), such reviews require frequent updating to remain relevant.

The present study is the most comprehensive review of prognostic forms of CPRs relevant to the management LBP ever conducted. It is anticipated that the study's findings will provide a platform for future research in this area, in addition to informing current clinical practices. As discussed in the overview of Study 4 (p. 256), several additional opportunities were identified in performing an update and expansion of the systematic review presented in Chapter 4 (p. 127). Briefly, these included: greater sensitivity of the electronic search strategy; more comprehensive quality appraisal of included studies;

and the calculation and reporting of uncertainty intervals for posterior probabilities (Clinical Commentary, Chapter 7).

9.1 Abstract

Objective:

To identify prognostic forms of clinical prediction rules (CPRs) related to the non-surgical management of adults with low back pain (LBP) and to evaluate their current stage of development.

Study Design and Setting:

Systematic review using a sensitive search strategy across 7 databases with hand-searching and citation tracking.

Results:

10,005 records were screened for eligibility with 35 studies included in the review. The included studies report on the development of 30 prognostic LBP CPRs. The majority of the identified CPRs are in their initial phase of development. Three CPRs were found to have undergone validation – the Cassandra rule for predicting long-term significant functional limitations and the 5-item and 2-item Flynn manipulation CPRs for predicting a favourable functional prognosis in patients being treated with lumbopelvic manipulation. No studies were identified that investigated whether the implementation of a CPR resulted in beneficial patient outcomes or improved resource efficiencies.

Conclusion:

The majority of the identified prognostic CPRs for LBP are in the initial phase of development and are consequently not recommended for direct application in clinical practice at this time. The body of evidence provides emergent confidence in the limited predictive performance of the Cassandra rule and the 5-item Flynn manipulation CPR in comparable clinical settings and patient populations.

9.2 Introduction

The stratification of patients into meaningful subgroups is a priority area of low back pain (LBP) research (Costa et al., 2013). Identifying patients with LBP with differing prognoses and targeting interventions based on the relative likelihood of treatment benefit provides individual and population level benefits, including improved patient outcomes and efficiencies in resource consumption (Brennan et al., 2006; Foster et al., 2014; J. C. Hill, Whitehurst, et al., 2011; Long et al., 2004; Whitehurst et al., 2012). Clinical prediction rules (CPRs) are one of several overlapping methods proposed to facilitate such stratification (Foster et al., 2013).

CPRs are simple statistical prediction tools designed to be used with individual patients that comprise a small number of clinical variables that have been identified to be independently predictive of a given diagnosis, outcome or treatment effect (Randolph et al., 1998). Prognostic forms of CPRs consist of non-specific prognostic variables that inform predictions

concerning future outcomes such as pain, disability and return to work. Such tools are therefore well-suited for screening and prioritising patients for interventions, and informing advice provided to patients and other parties regarding anticipated prognoses (Foster et al., 2013; J. C. Hill & Fritz, 2011; Moons, Royston, Vergouwe, Grobbee, & Altman, 2009). Prescriptive CPRs are a special type of prognostic CPR that inform predictions regarding the relative treatment-effect a patient may experience from an intervention. The variables that comprise a prescriptive CPR are treatment-effect modifiers, which are the baseline variables that differentiate patient subgroups who experience differing magnitudes of treatment-effect (Hancock, Herbert, et al., 2009; Kraemer et al., 2006; Stanton et al., 2010). Thus, prescriptive CPRs function to inform clinical decisions regarding treatment selection (C. Cook, 2008; J. C. Hill & Fritz, 2011).

The development of a CPR broadly occurs across three main phases, whereby the tool is initially derived, then prospectively validated in new patient cohorts, and finally evaluated for its ability to positively impact clinical practice (Childs & Cleland, 2006). The validation of a CPR is important as predictor variables may simply reflect chance statistical associations or the CPR may be specific to the study sample or setting in which it was derived (McGinn et al., 2008). CPRs that have been demonstrated to perform consistently across different patient groups and across broad clinical settings may be applied in practice with confidence in their accuracy (McGinn et al., 2000). Impact analysis is an important final step in the development of a CPR as it evaluates whether the implementation of a validated CPR is likely to

have meaningful beneficial consequences (McGinn et al., 2000; Toll et al., 2008). Such benefits may include more accurate selection and prioritisation of patients requiring intervention, improved patient outcomes and reduced costs of care (Foster et al., 2013; J. C. Hill & Fritz, 2011; Reilly & Evans, 2006).

The limited data concerning the use of CPRs for LBP in clinical practice suggests that many clinicians have an awareness of such tools and consider their application in their clinical decision-making (Haskins, Osmotherly, Southgate, et al., 2014; Learman et al., 2012; Sparks, McGehee, Buettner, & Scott, 2010; Willett et al., 2011). Consequently, the identification of the range of existing prognostic CPRs for LBP, and an appraisal of their appropriateness to be applied in clinical practice at this time, is potentially of significant clinical benefit. Previous systematic reviews of CPRs relevant to the non-surgical management of LBP have limited their scope to tools designed for specific interventions (Beneciuk et al., 2009; Lubetzky-Vilnai et al., 2014; May & Rosedale, 2009; Patel et al., 2013; Stanton et al., 2010), a particular health profession (Beneciuk et al., 2009; Haskins et al., 2012; Lubetzky-Vilnai et al., 2014; Patel et al., 2013; van Oort et al., 2012), or to a particular stage of CPR development (Beneciuk et al., 2009; Lubetzky-Vilnai et al., 2014; Patel et al., 2013). It is probable that many prognostic CPRs related to the non-surgical management of LBP have not yet been identified in systematic reviews to date.

Therefore, the aim of this systematic review was to identify prognostic forms of CPRs related to the non-surgical management of adults with LBP and to evaluate their current stage of development. It is anticipated that the evidence identified in this review will be informative to clinicians managing patients with LBP and to researchers involved in the development of LBP CPRs.

9.3 Methods

This systematic review sought to include studies reporting on the derivation, validation or impact analysis of one or more prognostic or prescriptive CPRs related to the non-surgical management of adults with LBP. A CPR was operationally defined as “a clinical tool that quantifies the individual contributions that various components of the history, physical examination and basic laboratory results make towards the diagnosis, prognosis, or likely response to treatment in an individual patient” (McGinn et al., 2008). Eligibility criteria were developed by the research team to address the review’s research question and are summarised in Table 9.1. No restrictions were placed on the year of study publication, stage of CPR development, types of predictor variables under consideration (eg. physical tests, history items, psychosocial factors etc), types of non-surgical management interventions, or the professional disciplines involved in the development of a CPR. CPRs were included independent of whether they were developed specifically for patients receiving a particular non-surgical intervention.

The database search strategy (Table 9.2) incorporated search strings identified to have high sensitivity for prognostic prediction model studies (Geersing et al., 2012; Holland et al., 2005; Ingui & Rogers, 2001) and disease-specific filters for back related disorders (Bombardier et al., 2014). Seven databases were searched from their inception to July 2013; Medline (1946-); Embase (1947-); Cochrane Central Register of Controlled Trials (1898-); PsychINFO (1806-); CINAHL (1937-); AMED (1985-); and Index to Chiropractic Literature (1981-). Identified records were downloaded into EndNote (Thomson Reuters) and duplicates were removed. Citation tracking and hand-searching were conducted as supplementary search strategies.

Table 9.1 Study eligibility criteria

Inclusion criteria
1. Reports on the derivation, validation and/or impact analysis of one or more prognostic or prescriptive CPRs related to the non-surgical management of adults with LBP
2. The CPR under development contains 2 or more predictor variables
3. The CPR under development was initiated by a formal derivation process in which a larger pool of candidate predictor variables was refined to a smaller set of variables based on their identified independent predictive value using formal multivariable statistical procedures
4. A tool is clearly presented in sufficient detail that may be applied by a clinician to predict a prognostic outcome or likelihood of treatment response in an individual patient
5. Published in English
Exclusion criteria
1. Limited to the investigation of modifiable and/or determinant predictor variables
2. CPR not capable of directly contributing to patient care
3. Conference proceedings/abstracts, dissertations, commentaries, reviews, editorials, letters, study protocols, n=1 designs (case reports), books, book chapters, book reviews, practice guidelines

Table 9.2 Search strategy

Medline via OVID (1946 - July 2013)	
1	dorsalgia.ti,ab OR exp Back Pain/ OR backache.ti,ab OR exp Low Back Pain/ OR (lumbar adj pain).ti,ab OR coccyx.ti,ab OR coccydynia.ti,ab OR sciatica.ti,ab OR sciatic neuropathy/ OR spondylosis.ti,ab OR lumbago.ti,ab OR back disorder$.ti,ab
2	(Validat$ OR Predict$.ti. OR Rule$) OR (Predict$ AND (Outcome$ OR Risk$ OR Model$)) OR ((History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$) AND (Predict$ OR Model$ OR Decision$ OR Identif$ OR Prognos$)) OR (Decision$ AND (Model$ OR Clinical$ OR Logistic Models/)) OR (Prognostic AND (History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$ OR Model$))
3	Stratification OR ROC Curve/ OR Discrimination OR Discriminate OR c-statistic OR c statistic OR "Area under the curve" OR AUC OR Calibration OR Indices OR Algorithm OR Multivariable
4	1 AND (2 OR 3)
5	limit 4 to english
6	limit 5 to humans
Embase via OVID (1947 - July 2013)	
1	dorsalgia.mp. OR back pain.mp. OR exp LOW BACK PAIN/ OR exp BACKACHE/ OR (lumbar adj pain).mp. OR coccyx.mp. OR coccydynia.mp. OR sciatica.mp. OR exp ISCHIALGIA/ OR spondylosis.mp. OR lumbago.mp. OR back disorder$.ti,ab.
2	predict:.tw. OR exp methodology OR validat:.tw.
3	1 AND 2
4	limit 3 to english
5	limit 4 to humans
6	limit 5 to exclude medline journals
Cochrane Central Register of Controlled Trials via OVID (1898 - July 2013)	
1	exp Back Pain/ OR back ache OR exp Low Back Pain/ OR (lumbar adj pain) OR coccyx OR coccydynia OR sciatica OR spondylosis OR exp Spine/ OR exp Spinal Diseases/ OR lumbago OR discitis OR (disc adj degeneration) OR (disc adj prolapse) OR (disc adj herniation) OR spinal fusion OR spinal neoplasms OR (facet adj joints) OR exp Intervertebral Disk/ or postlaminectomy OR arachnoiditis OR (failed adj back) OR exp Cauda Equina/ OR (lumbar adj vertebra$) OR (spinal adj stenosis) OR (slipped adj (disc$ or disk$)) OR (degenerat$ adj (disc$ or disk$)) OR (stenosis adj (spine or root or spinal)) OR (displace$ adj (disc$ or disk$)) OR (prolap$ adj (disc$ or disk$)) OR exp Sciatic Neuropathy/ OR sciatic$ OR back disorder$ OR (back adj pain)
2	(Validat$ OR Predict$.ti. OR Rule$) OR (Predict$ AND (Outcome$ OR Risk$ OR Model$)) OR ((History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$) AND (Predict$ OR Model$ OR Decision$ OR Identif$ OR Prognos$)) OR (Decision$ AND (Model$ OR Clinical$ OR Logistic Models/)) OR (Prognostic AND (History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$ OR Model$))

3	Stratification OR ROC Curve/ OR Discrimination OR Discriminate OR c-statistic OR c statistic OR "Area under the curve" OR AUC OR Calibration OR Indices OR Algorithm OR Multivariable
4	1 AND (2 OR 3)
5	limit 4 to medline records
6	limit 4 to embase records
7	4 NOT (5 OR 6)
Cochrane Central Register of Controlled Trials via OVID (1898 - July 2013)	
1	exp Back Pain/ OR back ache OR exp Low Back Pain/ OR (lumbar adj pain) OR coccyx OR coccydynia OR sciatica OR spondylosis OR exp Spine/ OR exp Spinal Diseases/ OR lumbago OR discitis OR (disc adj degeneration) OR (disc adj prolapse) OR (disc adj herniation) OR spinal fusion OR spinal neoplasms OR (facet adj joints) OR exp Intervertebral Disk/ or postlaminectomy OR arachnoiditis OR (failed adj back) OR exp Cauda Equina/ OR (lumbar adj vertebra$) OR (spinal adj stenosis) OR (slipped adj (disc$ or disk$)) OR (degenerat$ adj (disc$ or disk$)) OR (stenosis adj (spine or root or spinal)) OR (displace$ adj (disc$ or disk$)) OR (prolap$ adj (disc$ or disk$)) OR exp Sciatic Neuropathy/ OR sciatic$ OR back disorder$ OR (back adj pain)
2	(Validat$ OR Predict$.ti. OR Rule$) OR (Predict$ AND (Outcome$ OR Risk$ OR Model$)) OR ((History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$) AND (Predict$ OR Model$ OR Decision$ OR Identif$ OR Prognos$)) OR (Decision$ AND (Model$ OR Clinical$ OR Logistic Models/)) OR (Prognostic AND (History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$ OR Model$))
3	Stratification OR ROC Curve/ OR Discrimination OR Discriminate OR c-statistic OR c statistic OR "Area under the curve" OR AUC OR Calibration OR Indices OR Algorithm OR Multivariable
4	1 AND (2 OR 3)
5	limit 4 to medline records
6	limit 4 to embase records
7	4 NOT (5 OR 6)
PsychINFO via OVID (1806 - July 2013)	
1	back pain/ OR lumbar spinal cord/ OR (low adj back adj pain).mp OR (back adj pain).mp OR spinal column/ OR (lumbar adj2 vertebra$).mp OR coccyx.mp OR sciatica.mp OR lumbago.mp OR dorsalgia.mp OR back disorder$.mp OR ((disc or disk) adj degenerat$).mp OR ((disc or disk) adj herniat$).mp OR ((disc or disk) adj prolapse$).mp OR (failed adj back).mp
2	(Validat$ OR Predict$.ti. OR Rule$) OR (Predict$ AND (Outcome$ OR Risk$ OR Model$)) OR ((History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$) AND (Predict$ OR Model$ OR Decision$ OR Identif$ OR Prognos$)) OR (Decision$ AND (Model$ OR Clinical$ OR Logistic Models/)) OR (Prognostic AND (History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$ OR Model$))
3	Stratification OR ROC Curve/ OR Discrimination OR Discriminate OR c-statistic OR c statistic OR "Area under the curve" OR AUC OR Calibration OR Indices OR Algorithm OR Multivariable
4	1 AND (2 OR 3)
5	limit 4 to english
6	limit 5 to human
CINAHL via EBSCO (1937 - July 2013)	

1	"dorsalgia" OR (MH "Back Pain+") OR (MH "Low Back Pain") OR "backache" OR (lumbar W1 pain) OR (lumbar N5 pain) OR (MH "Coccyx") OR (MH "Sciatica") OR "sciatica" OR "coccyx" OR "coccydynia" OR "back disorder*" OR (MH "Lumbar Vertebrae") OR (lumbar N2 vertebra) OR (MH "Thoracic Vertebrae") OR (MH "Spondylolisthesis") OR (MH "Spondylolysis") OR "lumbago"
2	(validat* OR ti predict* OR rule*) OR (predict* AND (outcome* OR risk* OR model*)) OR ((history OR variable* OR criteria OR scor* OR characteristic* OR finding* OR factor*) AND (predict* OR model* OR decision* OR identif* OR prognos*)) OR (decision* AND (model* OR clinical* OR MH "logistic regression+")) OR (prognostic AND (history OR variable* OR criteria OR scor* OR characteristic* OR finding* OR factor* OR model*))
3	stratification OR mh "ROC Curve" OR discrimination OR discriminate OR c-statistic OR c statistic OR "Area under the curve" OR AUC OR calibration OR indices OR algorithm OR multivariable
4	S1 AND (S2 OR S3)
5	applied limit to English
6	applied limit to humans
7	applied limit to exclude Medline records
AMED via OVID (1985 - July 2013)	
1	dorsalgia.ti,ab OR exp Back Pain/ OR backache.ti,ab OR exp Low Back Pain/ OR (lumbar adj pain).ti,ab OR coccyx.ti,ab OR coccydynia.ti,ab OR sciatica.ti,ab OR sciatic neuropathy/ OR spondylosis.ti,ab OR lumbago.ti,ab OR back disorder$.ti,ab
2	(Validat$ OR Predict$.ti. OR Rule$) OR (Predict$ AND (Outcome$ OR Risk$ OR Model$)) OR ((History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$) AND (Predict$ OR Model$ OR Decision$ OR Identif$ OR Prognos$)) OR (Decision$ AND (Model$ OR Clinical$ OR Logistic Models/)) OR (Prognostic AND (History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$ OR Model$))
3	Stratification OR ROC Curve/ OR Discrimination OR Discriminate OR c-statistic OR c statistic OR "Area under the curve" OR AUC OR Calibration OR Indices OR Algorithm OR Multivariable
4	1 AND (2 OR 3)
5	limit 4 to english
Index of Chiropractic Literature (- July 2013)	
1	Subject:"Back" OR Subject:"Back Injuries" OR Subject:"Back Pain" OR Subject:"Low Back Pain" OR Subject:"Lumbar" OR Subject:"Lumbosacral Region" OR Subject:"Sciatica" OR All Fields:sciatica OR Subject:"Coccyx" OR Subject:"Sacroiliac Joint" OR Subject:"Sacrum"
2	(Validat* OR Predict* OR Rule*) OR (Predict*AND (Outcome* OR Risk* OR Model*)) OR ((History OR Variable* OR Criteria OR Scor* OR Characteristic* OR Finding* OR Factor*) AND (Predict* OR Model* OR Decision* OR Identif* OR Prognos*)) OR (Decision* AND (Model* OR Clinical* OR "Logistic Model*")) OR (Prognostic AND (History OR Variable* OR Criteria OR Scor* OR Characteristic* OR Finding* OR Factor* OR Model*))
3	Stratification OR "ROC Curve" OR Discrimination OR Discriminate OR c-statistic OR "c statistic" OR "Area under the curve" OR AUC OR Calibration OR Indices OR Algorithm OR Multivariable
4	S2 OR S3
5	S1 AND S4

Two independent reviewers selected studies for inclusion using a two-step process (P. Edwards et al., 2002; Higgins & Green, 2011). Firstly, the titles and abstracts of identified records were screened by both reviewers with studies deemed eligible by either reviewer progressing to the second stage of screening. In the second stage, the full-text of studies were screened by both reviewers with concordance determining eligibility. Episodes of disagreement were resolved by consensus and if needed, by a third independent reviewer.

A standardised tool was used for data extraction. Information regarding study design, patient population, CPR function, predictor variables, dependent outcomes, statistical analysis, tool format, tool performance, and the reporting of uncertainty intervals were recorded for each included study. The stage of tool development was defined as derivation, validation or impact analysis using a well-recognised hierarchical CPR development framework (McGinn et al., 2000). Contingency tables for dichotomized outcomes were extracted, calculated or approximated for specified cut-off points of a CPR where reported study data permitted. When not reported, the sensitivity, specificity, positive likelihood ratio (+LR) and negative likelihood ratio (-LR) with 95%CI were calculated or approximated. Uncertainty intervals for posterior probabilities were calculated where study data permitted using the objective Bayesian method using Monte Carlo simulation (Haskins, Osmotherly, Tuyl, et al., 2014).

The quality appraisal of included studies was conducted by examining the risk of bias relevant to both study design and the methodological factors

specific to CPR development. All studies were initially evaluated using standardised appraisal tools relevant to their specific research designs. Prognostic studies were appraised using the Quality In Prognosis Studies (QUIPS) tool (Hayden, Côté, & Bombardier, 2006) and randomised controlled trials were appraised using the Physiotherapy Evidence Database (PEDro) scale (Maher, Sherrington, Herbert, Moseley, & Elkins, 2003). The QUIPS tool was applied in instances where the data from an RCT had been pooled across treatment groups for the development of a CPR. There is no standardised tool to appraise the methodological quality of factors specific to the development of all forms of CPRs at each stage of their development. Consequently, an updated version of a quality appraisal tool used in a previous systematic review on this topic was applied (Haskins et al., 2012). This tool was developed to be inclusive of the commonly represented quality criteria for all forms of CPRs at each stage of their development in well-cited methodological texts (Beattie & Nelson, 2006; Childs & Cleland, 2006; Laupacis et al., 1997; McGinn et al., 2000; Stiell & Wells, 1999). These criteria were updated in this review (Table 9.3 and Table 9.4) to incorporate additional items identified in recent publications (Bouwmeester et al., 2012; C. Cook et al., 2010; Haskins, Osmotherly, Tuyl, et al., 2014; Lubetzky-Vilnai et al., 2014; Seel et al., 2012; van Oort et al., 2012) and to exclude items that formed part of the eligibility criteria of this review (e.g., use of multivariable analysis). Quality appraisal of the included studies was conducted by two independent reviewers. Concordance between reviewers determined quality criterion status, with disagreement resolved by consensus, or if needed by a third independent reviewer. Negative status for a quality criterion was

recorded in instances where a study was found not to report evidence concerning that criterion.

Table 9.3 Derivation study quality appraisal criteria

Domain	Criteria	Definition
Design	1. Prospective design	The study is conducted forwards in time.
	2. Study site described	The nature of study site is described in sufficient detail to enable comparison to other settings.
Participants	3. Justification for the number of study subjects is reported	Justification is provided for the number of subjects enrolled into the study.
	4. Representative sample	The reported method of patient selection is free of bias so that study subjects encompass a wide clinical and demographic spectrum and are representative of all patients seen at the site with the designated condition.
	5. Important patient characteristics described	The study subjects are well described in terms of inclusion criteria, method of selection and clinical and demographic characteristics.
Predictor variables	6. Selection of candidate predictor variables justified	The selection of candidate predictor variables is justified with appropriate reasoning and may include previous literature, psychometric properties, clinical reasoning, and/or expert opinion.
	7. Blinded predictor assessment	The assessment of the predictor variables is determined without knowledge of the outcome. If the study was prospective and the predictor variables were clearly collected prior to the outcome event, then assessment can be considered to be blind. If the study was retrospective and the authors did not mention blinding, it will be assumed that it was not blinded.
	8. Predictor variables have demonstrated reliability	Predictor tests are reported to be reliable (kappa >=0.60 or ICC >=0.70) either through previous report or through report within the findings of the study.
Outcomes	9. Outcome measure has demonstrated reliability and validity	The outcome measure is reported to have demonstrated reliability and validity. Literature is cited to support the outcome measure and psychometric characteristics of the outcome measure are reported.

Domain	Criteria	Definition
	10. Blinded outcome assessment	Interpretation of the outcome assessment is reported to be determined without knowledge of the status of the predictor variables. If a study does not comment upon whether the outcome was categorised without knowledge of the predictor variables, it will be assumed that it was not blinded. If the outcome is self-reported (eg. VAS), blinding is considered to be present if the participant is blinded to their status on the predictor variables.
Analysis	11. Mathematical techniques described	The mathematical techniques employed are adequately described.
	12. Reporting and handling of missing data described	Missing data (eg. values per participant, missing values per predictor, lost to follow-up) and how it was handled (eg. omitted, imputation) are reported.
	13. At least 10 outcome events per independent variable in the final multivariable model	There are at least 10 outcome events per independent variable in the final multivariable model. The number of outcome events is defined in proportional hazards analysis by the count of 'failure' events. In logistic regression the number of outcome events is the smaller number of binary outcomes of the dependent variable. For linear regression models there should be at least 10 patients per variable in the final model.
	14. At least 10 outcome events per candidate predictor variable	As per item 13, except the number of candidate predictor variables replaces the number independent variables in the final model.
	15. Collinearity of predictor variables tested	Collinearity of predictor variables were examined such as testing pairwise correlations or the variance inflation factor.
	16. Continuous predictor variables are kept continuous in the multivariable analysis	Continuous predictor variables were kept as continuous variables in the multivariable analysis.
CPR performance	17. Uncertainty in the accuracy of the CPR is described	Uncertainty intervals are reported for accuracy statistics of the CPR.
	18. Uncertainty in the posttest probability is described	Uncertainty intervals are reported for posttest probabilities.
	19. CPR performance is non-paradoxical	The performance of a CPR behaves logically, such that the probability of a given outcome does not decrease at any point with increasing positive status on that tool.

The absolute and chance-corrected degree of inter-rater agreement were calculated for each stage of study selection. A sum quality score was calculated and reported for studies appraised with the PEDro tool (de Morton, 2009). Quality appraisals using all other instruments in this review are presented descriptively. The review was not eligible for protocol registration due to the lack of a specific intervention, exposure or outcome measure under consideration. The manuscript was prepared in accordance with the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines (Liberati et al., 2009).

Table 9.4 Validation study quality appraisal criteria

Domain	Criteria	Definition
Design	1. Prospective validation in new patient population	The study is conducted forwards in time in a different population to the derivation study.
	2. Different clinical setting to derivation study	The CPR is tested in a clinical setting that is different to the derivation study.
Participants	3. Representative sample	The method of patient selection is free of bias so that study subjects encompass a wide clinical and demographic spectrum and are representative of all patients seen at the site with the designated condition.
CPR application	4. The rule is applied accurately	The rule is applied exactly as described in the derivation study.
	5. Assessment of the reliability of the rule	The reliability of the interpretation of a rule is explicitly measured using at least a representative subset of the study sample.
Follow-up	6. Complete follow-up	There is complete follow-up. All patients are subjected to the gold or criterion standard to determine their true outcome compared to that predicted by the rule.
	7. Reporting and handling of missing data described	Missing data (eg. values per participant, missing values per predictor, lost to follow-up) and how it was handled (eg. omitted, imputation) is reported.
CPR performance	8. Accuracy of the rule described	The accuracy/performance of the rule is described.
	9. Uncertainty in the accuracy of the CPR is described	Uncertainty intervals are reported for accuracy statistics of the CPR.
	10. Uncertainty in the posttest probability is described	Uncertainty intervals are reported for posttest probabilities.

9.4 Results

The database search strategy yielded 12,347 records. A further 25 records were identified via hand-searching and citation tracking. Following the removal of duplicates, 10,005 records were screened via title and abstract for eligibility with 352 records advancing to the second stage of screening. The

full-texts of these studies were reviewed and 35 were determined to be eligible (Al-Sayegh et al., 2010; Alonso-Blanco et al., 2009; Buranapanitkit, Tautakul, Lim, Geater, & Chomchan, 2003; Cai et al., 2009; Cairns, Mooney, & Crane, 1984; Childs et al., 2006; Childs et al., 2004; Cleland et al., 2009; Cleland et al., 2006; C. Cook et al., 2013; Dionne, 2005; Dionne et al., 2005; Dionne et al., 1997; Dionne et al., 2011; Flynn et al., 2002; Fritz, Brennan, & Leaman, 2006; Fritz, Childs, et al., 2005; George et al., 2005; Hallegraeff et al., 2009; Hancock, Maher, Latimer, et al., 2008; Hancock, Maher, et al., 2009; Hewitt et al., 2007; Heymans et al., 2009; Heymans et al., 2007; G. E. Hicks et al., 2005; Kovacs et al., 2012; Malmqvist et al., 2008; May et al., 2008; Roland, Morrell, & Morris, 1983; Schenk, Dionne, Simon, & Johnson, 2012; Schwind, Learman, O'Halloran, Showalter, & Cook, 2013; Stolze, Allison, & Childs, 2012; Sutlive et al., 2009; Thomas et al., 1999; Valat et al., 2000). As illustrated in Figure 9.1, the most common reason for a study's exclusion was not satisfying the study's operational definition of a CPR.

Inter-rater agreement for the first and second stages of screening was 98% and 93% respectively. The chance-corrected degree of agreement was 'moderate' (Landis & Koch, 1977) for both stages with $\kappa = 0.51$ (95%CI 0.46 – 0.57) for the screening by titles and abstracts, and $\kappa = 0.59$ (95%CI 0.45 – 0.73) for the screening by full-text. Of the 26 episodes of disagreement in the second stage of screening, all but two cases were resolved by consensus, with a third reviewer later including one of these studies.

Thirty CPRs were identified in this review and are summarized in Table 9.5 (p. 325). The majority (n=20) of the identified CPRs were derived using populations of patients receiving a specific treatment program. Functional outcomes were modelled as the dependent variable in the derivation of 15 CPRs. Work-related outcomes (n=5), pain intensity (n=4), recovery (n=3), symptom persistence (n=2), and need for surgical intervention (n=1) were used as the dependent outcomes in the derivation of the remaining CPRs. No CPR included in this review selected variables for inclusion in the tool by examining tests of interaction to identify effect modifiers. Three CPRs were identified to have undergone validation in one or more studies. No impact analysis studies were identified.

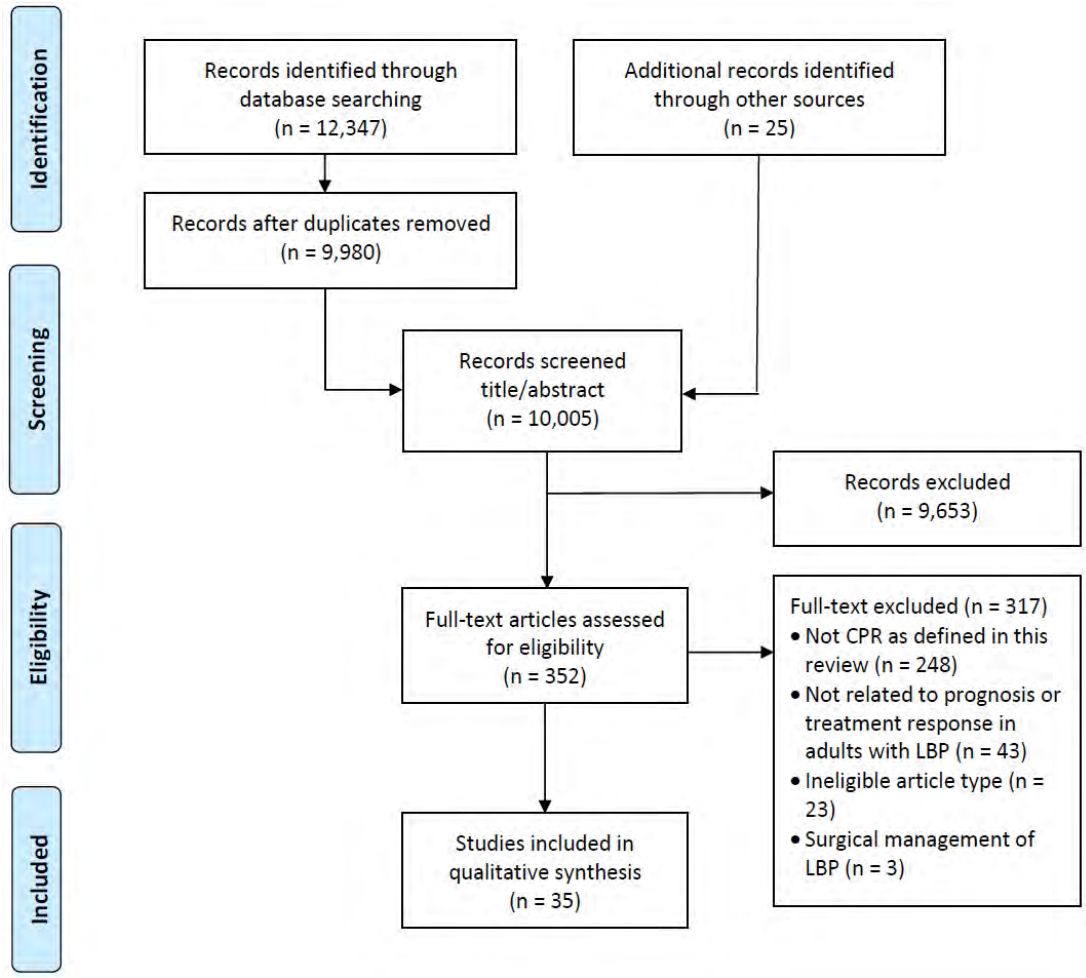


Figure 9.1 **Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) study selection flowchart for systematic review of prognostic clinical prediction rules for low back pain**

Table 9.5 Prognostic/prescriptive low back pain clinical prediction rules

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
Function	1.	Identifying patients presenting in general practice who are likely to have a low level of disability at 4 weeks	1. duration of pain <1 week; 2. SLR[32] ≥ 60°	Count of predictors	Derivation Roland et al. (1983) Single cohort	n=230 episodes of LBP[33] (from 215 patients attending GP[34] practice), 192 episodes with follow-up data at 4 weeks, mean age 41 years (dispersion not reported), 53% female, duration of symptoms not reported, 84% prevalence of dependent outcome.	Not specified	RMDQ[35] (24-items) score of < 14 points at 4 weeks (dichotomous)	Not reported, but accuracy of both variables present calculated to be, sens[36] = 0.40 (0.33-0.48); spec[37] = 0.93 (0.79-0.98); +LR[38] = 6.0 (1.6-23.3); -LR[39] = 0.64 (0.55-0.75).	2 predictors present = 97%, 1 predictor present = 81%, 0 predictors present = 61%. 95%CrI[40] calculated for both variables present to be 90.0% - 99.1%.

[32] straight leg raise
[33] low back pain
[34] general practitioner
[35] Roland Morris disability questionnaire
[36] sensitivity
[37] specificity
[38] positive likelihood ratio
[39] negative likelihood ratio
[40] 95% credible interval

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	2. Cassandra rule	Identifying patients at high risk of sustaining long-term significant functional limitations	1. Symptoms Checklist 90 Revised Depression score; 2. Symptoms Checklist 90 Revised Somatization score	Algorithm leading to 4 stratified risk groups. Dichotomized to 2 risk groups in later publications (Dionne, 2005; Dionne et al., 2011)	Derivation Dionne et al. (1997) Single cohort with split sample	Consulting primary care physician with back pain. Training set n=569, mean age 47 years (SD[41] 14), 52% female, mean 13 years (SD 13) since LBP onset, 95 lost to follow-up, 15% prevalence of dependent outcome. Validation set n=644, mean age 47 years (SD 15), 53% female, mean 12 years (SD 13) since LBP onset, 96 lost to follow-up, 16% prevalence of dependent outcome.	Not specified	Modified RMDQ (16-items) score ≥ 50% at 2 years (dichotomous)	Reported in later publications (Dionne, 2005; Dionne et al., 2011) for dichotomized risk groups (high/moderate vs low) sens = 0.86 (0.77-0.93); spec = 0.57 (0.53-0.62). LR's not reported, but approximated from study data and later publications to be +LR = 2.0 (1.7-2.3); -LR = 0.25 (0.14-0.42).	Validation set - risk of outcome for group 1 (Depression<0.444) = 5%; group 2 (Depression≥0.444 but <1.5 and Somatization < 0.333) = 4%; group 3 (Depression≥0.444 but <1.5 and Somatization ≥ 0.333) =19%; group 4 (Depression ≥1.5) = 36%. Post-test probability not reported for dichotomized risk groups, but high/moderate group risk approximated from study data and later publications to be 27%. 95%CrI for high/moderate risk group approximated to be 22.0% - 32.5%.

[41] standard deviation

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
					Validation Dionne (2005) Single cohort	n=860, adults absent from work due LBP consulting in primary care, mean age 39 years (SD 11), 42% female, 78% with recurrent or persistent back pain, 18% prevalence of dependent outcome.	Not specified	RMDQ (24-items) score of ≥ 50% at 2 years (dichotomous)	For dichotomized risk groups (high/moderate vs low) sens = 0.91; spec = 0.29. LRs not reported, but approximated to be +LR = 1.3 (1.2-1.4); -LR = 0.31 (0.19-0.52).	Risk of outcome for group 1 (Depression<0.444) = 6%; group 2 (Depression≥0.444 but <1.5 and Somatization < 0.333) = 7%; group 3 (Depression≥0.444 but <1.5 and Somatization ≥ 0.333) =14%; group 4 (Depression ≥1.5) = 27%. Post-test probability for dichotomized high/moderate group = 22%. 95%CrI for high/moderate risk group approximated to be 18.9% - 25.2%.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
					Validation Dionne et al. (2011) Single cohort	n=1262, patients presenting to an emergency department with nonspecific back pain, mean age 41 years (SD 14), 48% female, 57% with recurrent or persistent symptoms, 19% prevalence of dependent outcome.	Not specified	RMDQ (24-items) score of ≥ 50% at 2 years (dichotomous)	For dichotomized risk groups (high/moderate vs low) sens = 0.82 (0.76-0.87); spec = 0.45 (0.42-0.49); +LR = 1.50 (1.38-1.64); -LR = 0.40 (0.25-0.47). CPR[42] more sensitive (82% vs 37%) but less specific (45% vs 85%) than physician prediction.	Post-test probability for dichotomized high/moderate group = 22% (0.22-0.29). 95%CrI for high/moderate risk group approximated to be 22.1% - 28.8%. CPR modified in this study to refine it to 5-items, however, this newly derived tool does not meet the definition of a CPR in this review.

[42] clinical prediction rule

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	3.	Identifying patients receiving lumbopelvic manipulation who are likely to experience improvement	1. duration of symptoms < 16 days; 2. FABQ[43] work subscale score < 19; 3. at least 1 hip with > 35° of internal rotation ROM[44]; 4. hypomobility in the lumbar spine; 5. no symptoms distal to the knee	Count of predictors	Derivation Flynn et al. (2002) Single cohort	n=71, patients with LBP presenting to military outpatient physical therapy facilities, mean age 38 years (SD 11), 41% female, mean 42 days (SD 55) of symptoms, 45% prevalence of dependent outcome.	Up to 2 treatments of high velocity thrust lumbopelvic manipulation (up to 2 attempts on each side), 10 reps supine pelvic tilt ROM exercise, and advice to maintain usual activity level within the limits of pain.	> 50% improvement on ODQ[45] by third treatment session (up to 8 days following initial) (dichotomous)	For 4 or more predictors present, sens = 0.63 (0.45-0.77); spec = 0.97 (0.87-1.0); +LR = 24.38 (4.63-139.41).	For 4 or more predictors present = 95%. 95%CrI calculated to be 77.1% - 98.9%.

[43] fear avoidance beliefs questionnaire
[44] range of motion
[45] Oswestry disability questionnaire

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
					Validation Childs et al. (2004) RCT[46]	n=131, patients with LBP referred to physiotherapy, mean age 34 years (SD 11), 42% female, median 27 days of symptoms (IQR[47] not reported), 44% prevalence of dependent outcome in manipulation group at 1 week, and 63% at 4 weeks.	1. Manipulation group - high velocity thrust spinal manipulation and ROM exercise on first 2 sessions, advice to maintain usual activity, aerobic and strengthening exercises (5 sessions over 4 weeks); 2. Exercise group - aerobic and strengthening exercises, advice to maintain usual activity (5 sessions over 4 weeks)	1. ≥50% improvement on ODQ at 1 week (dichotomous); 2. ≥50% improvement on ODQ at 4 weeks (dichotomous); 3. ODQ score at 1 week, 4 weeks and 6 months (continuous); 4. pain (0-10 NRS[48]) at 1 week, 4 weeks and 6 months (continuous)	For 4 or more predictors present, significant 3-way interaction for rule status, treatment group, and time, for ODQ and pain. In manipulation group, having 4 or more predictors present had accuracy of +LR = 13.2 (3.4-52.1) for improvement at 1 week.	For 4 or more predictors present in manipulation group, post-test probability at 1 week = 92%. 95%CrI calculated to be 72.8% - 97.3%.

[46] randomized controlled trial
[47] interquartile range
[48] numerical rating scale

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
					Validation Childs et al. (2006) RCT secondary analysis	n=131, patients with LBP referred to physiotherapy, mean age 34 years (SD 11), 42% female, median 27 days of symptoms (IQR not reported), 99% prevalence of dependent outcome in manipulation group at 1 week, and 97% at 4 weeks.	1. Manipulation group - high velocity thrust spinal manipulation and ROM exercise on first 2 sessions, advice to maintain usual activity, aerobic and strengthening exercises (5 sessions over 4 weeks); 2. Exercise group - aerobic and strengthening exercises, advice to maintain usual activity (5 sessions over 4 weeks)	No worsening of disability defined as not having ≥6 point increase in ODQ score (dichotomous) (inverted from study)	No patients with 4 or more predictors present who received manipulation experienced worsening. Accuracy of 4 or more predictors present for prediction of outcome in manipulation group at 1 week not reported, but able to be derived from study data to be sens = 0.33 (0.23-0.45); spec = 1.0 (0.21-1.0); +LR = ∞; -LR = 0.67 (0.56 – 0.79).	For 4 or more predictors present in manipulation group, post-test probability of not being worse at 1 week = 100%. 95%CrI calculated to be 89.0% - 100%.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
					Validation Cleland et al. (2006) Case series	n=12, patients with LBP attending an outpatient physiotherapy clinic and all CPR positive (≥4 predictors present), mean age 39 years (SD 9), 42% female, median 19 days of symptoms (range 8-148), 92% prevalence of dependent outcome.	Side-lying thrust manipulation and ROM exercise x 2 sessions within 1 week	≥50% improvement on ODQ at 1 week (dichotomous)		11 of 12 (92%) participants improved at 1 week. All participants positive on CPR, therefore unable to assess rule performance.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
					Validation Hancock, Maher, Latimer, et al. (2008) RCT	n=239, patients with LBP < 6 weeks duration presenting to a GP, mean age 41 years (SD 16), 44% female, mean 9 days of symptoms (SD 9).	1. Spinal manipulative therapy (n=119), 2-3 times per week for a maximum of 12 sessions over 4 weeks, 5% received thrust manipulative techniques; 2. Placebo (n=120), detuned pulsed ultrasound, matched to active treatment group contact (both groups further randomised to receive either placebo or active diclofenac)	1. pain (11 point NRS); 2. disability (RMDQ) measured at 1,2,4 and 12 weeks (all continuous)	For 4 or more predictors present, no significant 3-way interaction between treatment group, CPR status and time for either pain (p=0.805) or disability (p=0.600). Positive rule status predicted improved pain at 2 weeks (p=0.015), and improved disability at 2 (p=0.033) and 12 weeks (0.015) independent of treatment group.	.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
					Validation Cleland et al. (2009) RCT	n=112, patients with LBP attending an outpatient physiotherapy clinic and all CPR positive (≥4 predictors present), mean age 40 years (SD 12), 52% female, median 45 days of symptoms (IQR 27-60).	1. Supine thrust manipulation group - supine high velocity thrust spinal manipulation and ROM exercise on first 2 sessions, exercise regime for next 3 sessions (5 sessions over 4 weeks); 2. Side-lying thrust manipulation group - side-lying high velocity thrust spinal manipulation and ROM exercise on first 2 sessions, exercise regime for next 3 sessions (5 sessions over 4 weeks); 3. Non-thrust manipulation group - central lower lumbar non-thrust manipulation (mobilization) and ROM exercise for first 2 sessions, exercise regime for next 3	1. ≥50% improvement on ODQ at 1 week (dichotomous); 2. ≥50% improvement on ODQ at 4 weeks (dichotomous); 3. ≥50% improvement on ODQ at 6 months (dichotomous); 4. ODQ score at 1 week, 4 weeks, and 6 months (continuous); 5. pain (0-10 NRS) at 1 week, 4 weeks and 6 months (continuous)	Significant group by time interaction for ODQ (p<0.001) and pain (p=0.001).Pair-wise comparisons indicate non-thrust group achieved inferior results to thrust manipulation groups, and no significant difference between thrust manipulation groups. Significant between group difference in proportion achieving a successful outcome at 1 week (between group difference p < 0.001); 4 weeks (between group difference p < 0.001), and 6 months (between group difference p = 0.009).	Success at 1 week, supine thrust manipulation group = 54.1%, side-lying thrust manipulation group = 52.6%, and non-thrust manipulation group = 8.1%; at 4 weeks, supine thrust manipulation group = 86.5%, side-lying thrust manipulation group = 81.6%, and non-thrust manipulation group = 18.9%; at 6 months, supine thrust manipulation group = 91.9%, side-lying thrust manipulation group = 89.5%, and non-thrust manipulation group = 67.6%. All participants positive on CPR, therefore unable to assess rule performance.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
					Validation Sutlive et al. (2009) RCT	n=60, military health care beneficiaries with LBP who are CPR positive (3 of 5 criteria present), mean age 26 years (SD 9), 48% female, 62% symptoms < 16 days.	1. lumbopelvic manipulation and pelvic tilt range of motion exercise (1 session); 2.neutral gap manipulation and pelvic tilt range of motion exercise (1 session)	1. pain (11 point NRS) at 48 hours; 2. disability (ODQ score) at 48 hours (both continuous)	No significant between group difference in the degree of improvement in pain (p=0.591) or disability (p=0.668) at 48 hours follow-up.	All participants positive on CPR, therefore unable to assess rule performance. Positive status on CPR was defined as ≥3 predictors present.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
					Validation Schenk et al. (2012) RCT	n=31 (analysis limited to n=26), patients with LBP referred to physical therapy who were positive on the CPR (analysis restricted to ≥4/5 criteria present), mean age 42 years (calculated from group demographics) (no dispersion reported) 61% female, mean 17 days (no dispersion reported) of symptoms (calculated from group demographics).	1. Mechanical Diagnosis and Therapy using directional preference established at initial session completed as home and clinic exercises; 2. High velocity thrust lumbopelvic manipulation, with 30 and 20 reps of hand-heel rock exercise at first 2 sessions respectively, and hourly home exercises in the patient's directional preference from session 3 until discharge	1. > 50% improvement on ODQ by discharge at week 4 (dichotomous); 2. ODQ score at discharge (continuous); 3. pain NRS at discharge (continuous)	No between group difference at discharge in ODQ score (p=0.31), pain (p=0.08), or the proportion improved by >50% on ODQ (p=0.16).	25% and 56% improved in manipulation group and Mechanical Diagnosis and Therapy group respectively. All participants positive on CPR, therefore unable to assess rule performance.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
					Validation C. Cook et al. (2013) Pooled results from RCT	n=149, patients with LBP attending outpatient physiotherapy, mean age 48 years (SD 15), 53% female, mean 34 weeks of symptoms (SD 99), 71 (49%) positive (≥4/5 predictors present) on CPR.	1. Thrust manipulation for first 2 sessions, then physical therapist directed care (n=76); 2. Non-thrust manipulation for first 2 sessions, then physical therapist directed care (n=73).	1. ≥50% improvement on ODQ at discharge (dichotomous); 2. ≥2.5 points of improvement on 11-point NRS at discharge (dichotomous); 3. self-reported recovery ≥75% at discharge (dichotomous); 4. total visits ≤6 sessions (dichotomous); 5. ODQ change score (continuous); 6. NRS change score (continuous); 7. total visits (continuous); 8. extent of recovery (continuous)	Positive status on CPR (≥4/5 predictors present) was an independent predictor for each of the 4 dichotomous dependent outcomes; 1. OR[49]=2.9(1.4-6.2); 2. OR=4.8(1.8-10.4); 3. OR=4.0(1.6-9.8); 4. 3.7(1.7-7.6); and for each of the 4 continuous dependent outcomes; 5. β = -4.2 (-7.7 to -0.69); 6. β = -0.98 (-1.5 to -0.47); 7. β = 0.32 (0.19-0.45); 8. β = -10.8 (-18.3 to -3.1).	Data pooled from both treatment groups, with treatment group allocation included as a covariate in each model.

[49] odds ratio

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
					Validation Schwind et al. (2013) Pooled data from RCT	n=149, patients with LPB attending outpatient physiotherapy, mean age 48 years (SD 15), 53% female, mean 34 weeks of symptoms (SD 99).	1. Thrust manipulation for first 2 sessions, then physical therapist directed care (n=76); 2. Non-thrust manipulation for first 2 sessions, then physical therapist directed care (n=73).	1. ≥ 50% improvement on ODQ by discharge; 2. ≥ 30% improvement on ODQ by discharge; 3. ≥ 17 point improvement on ODQ by discharge; 4. ≥ 10 point improvement on ODQ by discharge; 5. ≥ 5 point improvement on ODQ by discharge; 6. final ODQ score ≤ 20% (all dichotomous)	Predictors retained in each model are different depending on the cut-off point of the dependent outcome. Positive status on the CPR (≥4/5 predictors present) was retained as an independent predictor variable in 3 of the 6 multivariable predictive models. Dependent outcome 1 - positive CPR status p=0.005 (OR 2.9, 95%CI[50] 1.4-6.2); Dependent outcome 3 - positive status on CPR p=0.007 (OR 3.4, 95%CI 1.4-8.2); Dependent outcome 6 - positive status on CPR p=0.029 (OR 3.3 (95%CI 1.1-9.6).	

[50] 95% confidence interval

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	4.	Identifying patients receiving lumbopelvic manipulation who are likely to experience improvement	1. duration of symptoms < 16 days; 2. no symptoms distal to the knee	Count of predictors	Derivation Fritz, Childs, et al. (2005) Pooled data from 2 studies	n=141, patients with LBP referred to physiotherapy at predominantly military health care facilities, mean age 36 years (SD 11), 49% female, median 22 days of symptoms (range 1-2775), 45% prevalence of dependent outcome.	Up to 2 treatments of high velocity thrust lumbopelvic manipulation (up to 2 attempts on each side), 10 reps supine pelvic tilt ROM exercise, and advice to maintain usual activity level within the limits of pain.	> 50% improvement on ODQ by third treatment session (up to 8 days following initial) (dichotomous)	For both predictors present, sens = 0.56 (0.43-0.67); spec = 0.93 (0.84-0.96); +LR = 7.2 (3.2-16.1). 83.7% classification accuracy compared to 5-item CPR.	For both predictors present = 85%. 95% CrI calculated to be 71.4% - 93.0%.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
					Validation Fritz, Brennan, and Leaman (2006) Retrospective database review	n=215, patients positive on CPR (both predictors present) and received physical therapy for occupational LBP, mean age 36 years (SD 10), 32% female, mean 5.3 days of symptoms (SD 4.7).	107 (50%) received thrust manipulation in first 2 sessions, 36 (17%) received non-thrust manipulation in first 2 sessions, and 72 (33%) received no manipulation in first 2 sessions.	1. change in ODQ score; 2. change in pain score; 3. number of treatment sessions; 4. length of stay; 5.costs (all continuous)	Manipulation group (n=143) improved more than non-manipulation group (n=72) in pain (p=0.008) and disability (p=0.01) and had a shorter length of stay (p=0.02), but there was no difference in number of therapy sessions (p=0.35) or costs (p=0.94). Thrust manipulation group (n=107) experienced the same degree of improvement in pain (p=0.74) and disability (p=0.76) as non-thrust manipulation group (n=36), but had fewer number of therapy sessions (p=0.04), a shorter length of stay (p=0.02) and lower costs (p=0.03).	All participants positive on CPR, therefore unable to assess rule performance.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
					Validation Hallegraeff et al. (2009) RCT	n=64, patients with LBP positive on CPR and attending physical therapy, mean age 40 years (SD = 10), 45% female, 31% had symptoms < 1 week.	1. Experimental group (n=31), 4 sessions over 2.5 weeks of thrust manipulation and strengthening and stretching exercises; 2. Control group (n=33), 4 sessions over 2.5 weeks of physical therapy without manipulation	1. disability (ODQ score) (continuous); 2. pain (VAS[51]) (continuous); 3. mobility (sit and reach test) (continuous); 4. patient perceived improvement after the 4th session (dichotomous)	No significant between-group difference in pain (p=0.26), disability (p=0.38) or mobility (p=0.14) at the fourth treatment. In multivariate ANOVA[52], the experimental group improved greater than the control group for the disability outcome (p=0.001, effect size = 0.21).	32% and 31% in the experimental and control groups respectively, reported to be improved by the fourth treatment. All participants positive on CPR, therefore unable to assess rule performance.
	5.	Predicting 6 month disability outcome for patients participating in a specific exercise program.	6 month disability = 4.4 + 0.24*(baseline ODQ score) + 0.34*(baseline FABQ work subscale score) - 10*(1 if centralization present, otherwise 0)	Linear regression equation	Derivation George et al. (2005) Single cohort secondary analysis	n=28, patients with acute/subacute LBP classified to receive specific exercise using a treatment based classification system, mean age 39 years (SD 10), 61% female, mean 21 days (SD 16) of symptoms.	Specific exercise consistent with Treatment-Based Classification x 4 weeks.	ODQ score at 6 months (continuous)	R^2 = 0.49.	

[51] visual analogue scale
[52] analysis of variance

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	6.	Identifying patients participating in a stabilization exercise program who are unlikely to experience improvement.	1. negative prone instability test; 2. aberrant movements absent; 3. FABQ – physical activity subscale < 9; 4. no hypermobility with lumbar spring testing	Count of predictors	Derivation G. E. Hicks et al. (2005) Single cohort	n=54, patients with LBP referred to physical therapy, mean age 42 years (SD 13), 57% female, mean 41 days of symptoms (SD 44), 28% prevalence of dependent outcome.	Supervised lumbopelvic stabilization exercise program, 16 sessions over 8 weeks and daily home exercises.	< 50% improvement and < 6 point improvement on ODQ at 8 weeks (dichotomous)	For 3 or more predictors present, sens = 0.87 (0.62-0.96); spec = 0.85 (0.70-0.93); +LR = 5.6 (2.6-12.1); -LR = 0.16 (0.04-0.58).	For 3 or more predictors present = 68%. Data as reported in study has been inverted such that increasing positive status on CPR is associated with higher likelihood of dependent outcome. CPR for success reported in this study was not eligible for inclusion in this review as predictor variables not selected via multivariable statistical procedures.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	7.	Predicting activity limitation at 9 weeks in patients with subacute or chronic musculoskeletal pain participating in an exercise-based physiotherapy program	9 week Functional Rating Index = 0.72*baseline Functional Rating Index score(0-100) - 8.93*interpreter required(0 if required, otherwise 1) + 2.31*duration of previous intervention (natural log (months of previous treatment +0.125)) - 4.15*baseline work status (1 = working, otherwise 0) + 13.66	Linear regression equation	Derivation Hewitt et al. (2007) Single cohort with split sample	n=720 (n=360 in training set, n=360 in validation set), patients with subacute or chronic musculoskeletal attending physiotherapy (approx 56% LBP), mean age 40 years (SD 11), 34% female, mean 11 months (SD 15) duration of symptoms.	Exercise-based physiotherapy program consisting of 6-9 weeks of 1 hour gym sessions (3/week) and daily home program.	Activity limitation at 9 weeks measured using a modified version of the Functional Rating Index (continuous)	R^2 = 0.69.	Study also reports on the development of 2 prediction tools that did not meet this review's eligibility criteria.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	8.	Identifying patients being treated with the McKenzie method who are likely to experience improvement	1. duration of symptoms less than 12 weeks; 2. back pain (not neck pain); 3. centralization or abolition of symptoms	Algorithm leading to 8 stratified risk groups.	Derivation May et al. (2008) Secondary analysis of single treatment arm of RCT	n=102, patients with back or neck pain referred by GPs to physiotherapy, secondary analysis of subgroup of 102/161 patients randomized to receive McKenzie treatment in previous RCT, demographics of those in secondary analysis not reported, 21% prevalence of dependent outcome.	McKenzie based treatment delivered by physiotherapists	50% reduction in RMDQ or NPQ[53] from baseline to 6 weeks that is retained at 6 or 12 months ("liberal" definition of success) (dichotomous)		Predicted probability provided in study for each of the 8 stratified groups ranging from 3% (duration ≥ 12 weeks, neck pain, no centralization or abolition) to 69% (duration < 12 weeks, back pain, centralization or abolition).

[53] Northwick Park neck pain questionnaire

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	9.	Identifying patients being treated with the McKenzie method who are likely to experience improvement	1. duration of symptoms less than 12 weeks; 2. back pain (not neck pain)	Algorithm leading to 4 stratified risk groups.	Derivation May et al. (2008) Secondary analysis of single treatment arm of RCT	n=102, patients with back or neck pain referred by GPs to physiotherapy, secondary analysis of subgroup of 102/161 patients randomized to receive McKenzie treatment in previous RCT, demographics of those in secondary analysis not reported, 16% prevalence of dependent outcome.	McKenzie based treatment delivered by physiotherapists	50% reduction in RMDQ or NPQ from baseline to 6 weeks that is retained at 6 and 12 months ("strict" definition of success) (dichotomous)		Predicted probability provided in study for each of the 4 stratified groups ranging from 1% (duration ≥ 12 weeks, neck pain) to 49% (duration < 12 weeks, back pain).

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	10.	Identifying patients with Ankylosing Spondylitis participating in a specific exercise program who are likely to experience improvement	1. SF-36 physical role score > 37; 2. SF-36 bodily pain score > 27; 3. Bath Ankylosing Spondylitis Disease Activity Index score > 31	Count of predictors	Derivation Alonso-Blanco et al. (2009) Single cohort	n=35, patients with Ankylosing Spondylitis referred to a university physical therapy clinic, mean age 46 years (SD 9), 20% female, mean 10 years of symptoms (SD 3), 46% prevalence of dependent outcome.	Specific exercise program delivered in a 1 hour group format in 8 sessions over 8 weeks.	≥20% reduction in Bath Ankylosing Spondylitis Disease Activity Index score and GROC[54] score ≥ +5 at 1 month follow-up (dichotomous)	For 2 or more predictors present, sens = 0.75 (0.51-0.90); spec = 0.93 (0.66-0.99); +LR = 11.2 (1.7-76.0).	For 2 or more predictors present = 91%. CrI not calculated as contingency table unable to accurately derived from study data.
	11.	Identifying patients receiving mechanical lumbar traction who will experience improvement	1. FABQ work subscale score < 21; 2. no neurological deficit, 3. age > 30; 4. non-manual work.	Count of predictors	Derivation Cai et al. (2009) Single cohort	n=129, patients with LBP referred to physiotherapy from orthopaedic outpatient clinic, mean age 31 years (SD 12), 16% female, mean 40 weeks duration of symptoms (SD 82), 19% prevalence of dependent outcome.	3 sessions of mechanical lumbar traction within 9 days, at 30-40% of patient's weight, intermittent (30sec on, 10sec off) x 15 minutes.	> 50% improvement on ODQ by third treatment session (9 days following initial) (dichotomous)	For all 4 predictors present, sens = 0.36 (0.19-0.57); spec = 0.96 (0.90-0.99); +LR = 9.36 (3.13-28).	For all 4 predictors present = 69%. 95% CrI calculated to be 41.1% - 87.0%.

[54] global rating of change

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	12.	Identifying postpartum women receiving lumbopelvic manipulation who will experience improvement	1. positive seated flexion test; 2. positive prone knee bend test; 3. posterior superior iliac spine symmetrical in sitting; 4. pain not extending below the knee.	Count of predictors	Derivation Al-Sayegh et al. (2010) Single cohort	n=69, female patients presenting within 1 year of giving birth with LBP and/or buttock pain, mean age 31 years (SD 6), 100% female, mean 29 weeks duration of symptoms (SD 17), 80% prevalence of dependent outcome.	Up to 2 treatments of high velocity thrust lumbopelvic manipulation on most symptomatic side (up to 2 attempts), 10 reps hand-heel rock range of motion exercise, and advice to remain as active as possible.	> 50% improvement on ODQ by third treatment session (up to 8 days following initial) (dichotomous)	For 2 or more predictors present, sens = 0.65 (0.51-0.77); spec = 0.79 (0.49-0.94); +LR = 3.1 (1.1-8.5).	For 2 or more predictors present = 92%. 95%CrI calculated to be 80.2% - 97.4%.
	13.	Identifying postpartum women receiving lumbopelvic manipulation who will not experience improvement	1. age>35 years; 2.VAS-best>3; 3. negative prone knee bend test.	Count of predictors	Derivation Al-Sayegh et al. (2010) Single cohort	n=69, female patients presenting within 1 year of giving birth with LBP and/or buttock pain, mean age 31 years (SD 6), 100% female, mean 29 weeks duration of symptoms (SD 17), 20% prevalence of dependent outcome.	Up to 2 treatments of high velocity thrust lumbopelvic manipulation on most symptomatic side (up to 2 attempts), 10 reps hand-heel rock range of motion exercise, and advice to remain as active as possible.	≤50% improvement on ODQ by third treatment session (up to 8 days following initial) (dichotomous)	For 2 or more predictors present, sens = 0.43 (0.19-0.7); spec = 0.96 (0.86-0.99); +LR = 11.8 (2.7-52.2).	For 2 or more predictors present = 75%. 95%CrI calculated to be 38.7% - 92.2%.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	14.	Identifying patients who will experience a clinically relevant improvement in disability	1. no evidence of disc degeneration on imaging; 2. no previous surgery; 3. receiving muscle relaxants; 4. not receiving major opioids; 5. having been treated with neuroreflexotherapy; 6. higher baseline RMDQ score; 7. lower baseline LBP severity (VAS); 8. lower baseline leg pain severity (VAS); 9. shorter duration of symptoms (acute (<14 days) / subacute (14-90 days) / chronic (>90 days))	Nomogram	Derivation Kovacs et al. (2012) Single cohort	n= 4220, seeking care for LBP in primary care or at a speciality centre (rheumatology, rehabilitation, neuroreflexotherapy, orthopaedic surgery), mean age 54 years (SD 15), 64% female, median 180 days of pain (IQR 90-365), 74% prevalence of dependent outcome.	95% received neuroreflexotherapy, 59% received analgesics, 15% received physical therapy, 1% underwent surgery	Improvement of ≥ 3 points on Spanish version of RMDQ (0-24) at 3 months (dichotomous)	Calibration - Hosmer-Lemeshow test p = 0.18. Discrimination - Area under receiver operating characteristic curve = 0.64.	Point estimate of outcome probability available for each 'score' on nomogram.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	15.	Identifying patients participating in a Pilates-based exercise program who will experience improvement	1. no leg symptoms in the last week; 2. BMI[55] ≥ 25; 3. total trunk flexion ≤70°; 4. at least 1 hip with average internal and external rotation of ≥25°; 5. duration of symptoms ≤6 months	Count of predictors	Derivation Stolze et al. (2012) Single cohort	n=95, referred or presenting to physical therapy, mean age 56 years (SD 11), 81% female, 68% symptoms > 6 months, 54% prevalence of dependent outcome.	Standardized Pilates-based exercise program using a Reformer, 16 sessions over 8 weeks.	≥ 50% improvement on ODQ at 8 weeks (dichotomous)	For 3 or more predictors present, sens = 0.73 (0.58-0.84); spec = 0.93 (0.81-0.99); +LR = 10.6 (3.5-32.1).	For 3 or more predictors present = 93% (81% - 97%). 95%CrI calculated to be 80.1% - 97.3%.
Pain	16.	Predicting 6 month pain outcome for patients participating in a specific exercise program.	6 month pain intensity = 0.97 + 0.27*(baseline pain score) - 1.6*(1 if centralization present, otherwise 0)	Linear regression equation	Derivation George et al. (2005) Single cohort secondary analysis	n=28, patients with acute/subacute LBP classified to receive specific exercise, mean age 39 years (SD 10), 61% female, mean 21 days (SD 16) of symptoms.	Specific exercise consistent with Treatment-Based Classification x 4 weeks.	NRS (11-point) pain score at 6 months (continuous)	R^2 = 0.29.	

[55] body mass index

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	17.	Predicting pain intensity at 9 weeks in patients with subacute or chronic musculoskeletal pain participating in an exercise-based physiotherapy program	Pain intensity at 9 weeks = 0.41*baseline pain intensity(0-10) + 0.04*baseline activity limitation (0-100 Functional Rating Index) - 0.94*non-English speaking background(1 if English, otherwise 0) + 0.27*duration of previous intervention(natural log(months of previous intervention + 0.123)) + 0.41	Linear regression equation	Derivation Hewitt et al. (2007) Single cohort with split sample	n=720 (n=360 in training set, n=360 in validation set), patients with subacute or chronic musculoskeletal attending physiotherapy (approx 56% LBP), mean age 40 years (SD 11), 34% female, mean 11 months (SD 15) duration of symptoms.	Exercise-based physiotherapy program consisting of 6-9 weeks of 1 hour gym sessions (3/week) and daily home program	Pain intensity at 9 weeks measured using a 10cm VAS (continuous)	R^2 = 0.67.	Study also reports on the development of 2 prediction tools that did not meet this review's eligibility criteria.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	18.	Identifying patients who will experience a clinically relevant improvement in pain in the lower back	1. having been treated with neuroreflexotherapy; 2. no previous surgery; 3. lower baseline RMDQ score; 4. higher baseline LBP severity (VAS); 5. lower baseline leg pain severity (VAS); 6. shorter duration of symptoms (acute (<14 days) / subacute (14-90 days) / chronic (>90 days))	Nomogram	Derivation Kovacs et al. (2012) Single cohort	n= 4406, seeking care for LBP in primary care or at a speciality centre (rheumatology, rehabilitation, neuroreflexotherapy, orthopaedic surgery), mean age 54 years (SD 15), 64% female, median 180 days of pain (IQR 90-365), 79% prevalence of dependent outcome.	95% received neuroreflexotherapy, 59% received analgesics, 15% received physical therapy, 1% underwent surgery	Improvement of ≥ 1.5 points on 10cm VAS for severity of LBP at 3 months (dichotomous)	Calibration - Hosmer-Lemeshow test p = 0.20. Discrimination - Area under receiver operating characteristic curve = 0.65.	Point estimate of outcome probability available for each 'score' on nomogram.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	19.	Identifying patients who will experience a clinically relevant improvement in leg pain	1. having been treated with neuroreflexotherapy; 2. no previous surgery; 3. lower baseline RDDQ score; 4. not receiving an EMG[56]; 5. lower baseline LBP severity (VAS); 6. higher baseline leg pain severity (VAS)	Nomogram	Derivation Kovacs et al. (2012) Single cohort	n= 3359, seeking care for LBP in primary care or at a speciality centre (rheumatology, rehabilitation, neuroreflexotherapy, orthopaedic surgery), mean age 55 years (SD 15), 66% female, median 180 days of pain (IQR 90-365), 75% prevalence of dependent outcome.	95% received neuroreflexotherapy, 59% received analgesics, 15% received physical therapy, 1% underwent surgery	Improvement of ≥ 1.5 points on 10cm VAS for severity of leg pain at 3 months (dichotomous)	Calibration - Hosmer-Lemeshow test p = 0.16. Discrimination - Area under receiver operating characteristic curve = 0.66.	Point estimate of outcome probability available for each 'score' on nomogram.
Recovery	20.	Identifying patients presenting in general practice with a short duration episode of care	1. duration of pain <1 week; 2. SLR ≥ 60°	Count of predictors	Derivation Roland et al. (1983) Single cohort	n=230 episodes of LBP (from 215 patients), 212 episodes with follow-up data at 4 weeks, mean age 41 years (dispersion not reported), 53% female, duration of symptoms not reported, 81% prevalence of dependent outcome.	Not specified	Time from first to last consultation ≤ 15 days (dichotomous)	Not reported, but accuracy of both variables present calculated to be, sens = 0.37 (0.30-0.44); spec = 0.90 (0.77-0.96); +LR = 3.8 (1.5-9.8); -LR = 0.70 (0.60-0.82).	2 predictors present = 94%, 1 predictor present = 79%, 0 predictors present = 58%. 95%CrI calculated for both variables present to be 85.8% - 97.6%.

[56] electromyography

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	21.	Identifying patients receiving chiropractic who will be better by the fourth session	1. absence of leg pain; 2. improved at 2nd visit (either improved pain when turning in bed, sleeping, putting on socks/shoes, walking, or getting up from sitting); 3. not overweight or obese	Count of predictors	Derivation Malmqvist et al. (2008) Single cohort	n=984, patients with LBP receiving chiropractic, 60% between the ages of 21-50, 48% female, 37% had a duration of symptoms of less than 2 weeks, 66% prevalence of dependent outcome.	Chiropractic management as decided by the treating chiropractor.	Definitely better on global assessment by the fourth treatment session (dichotomous)		0 predictors present = 34%; 1 predictor present = 60%; 2 predictors present = 75%; 3 predictors present= 84%. Reported data does not permit calculation of 95%CrI.
	22.	Identifying patients with acute low back pain who are likely to recover at different rates.	1. baseline pain ≤ 7/10; 2. duration of current episode ≤ 5 days; 3. ≤ 1 previous episodes of LBP	Count of predictors	Derivation Hancock, Maher, et al. (2009) Pooled results from RCT	n=239, patients with LBP < 6 weeks duration presenting to a GP, mean age 41 years (SD 16), 44% female, mean 9 days of symptoms (SD 9).	1. Detuned ultrasound and placebo diclofenac; 2. Detuned ultrasound and active diclofenac; 3. Spinal manipulative therapy and placebo diclofenac; 4. Spinal manipulative therapy and active diclofenac.	Number of days from the baseline assessment until recovery from pain (≤1 on 0-10 NRS) (continuous)	Median days to recovery for 0 predictors = 22 days (11-33); 1 predictor = 22 days (19-24); 2 predictors = 15 days (12-18); 3 predictors = 6 days (4-8). Hazard ratios (reference category 0 predictors) 1 predictor = 1.3 (0.7-2.3); 2 predictors = 2.0 (1.2-3.6); 3 predictors = 3.5 (1.8-7.0).	Proportion recovered at 1 week with 0 predictors = 15%; 1 predictor = 13%; 2 predictors = 23%, 3 predictors = 60%. Proportion recovered at 12 weeks with 0 predictors = 70%; 1 predictor = 85%; 2 predictors = 95%; 3 predictors = 95%.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
Surgical intervention	23.	Predicting need for surgical intervention due to non-response to conservative treatment in patients with herniated nucleus pulposus	1. pain intensity; 2. duration of symptoms; 3. crossed straight leg raise test; 4. muscle power grade; 5. number of dermatome deficits	Score chart	Derivation Buranapanitkit et al. (2003) Retrospective single cohort	n=251, patients admitted to hospital with a diagnosis with lumbar herniated nucleus pulposus, mean age 38 years (range 15-60), 60% female, 47% had symptoms less than 3 months, 67% prevalence of dependent outcome.	6 weeks of rest, analgesia, anti-inflammatory medication, and physical therapy.	Requiring surgical intervention as no improvement from 6 weeks of conservative treatment (dichotomous)		Scores < 45 = 15%; scores 45-64 = 53%; scores ≥ 65 = 96%.
Symptom persistence	24.	Identifying patients with acute LBP who are likely to develop persistent symptoms	1. characteristics of current episode (more points for exacerbation of chronic LBP and sciatica); 2. difficulty in walking a short distance or climbing stairs; 3. difficulty rising from bed or chair; 4. duration of certificate to remain off work > 8 days; 5. taking part in a sport	Score chart	Derivation Valat et al. (2000) Single cohort	n=2487, employed patients with acute LBP (< 8 days) presenting to GPs or Rheumatologists, mean age 41 years (SD 9), 43% female, all had symptoms < 8 days, 6% prevalence of dependent outcome.	98% received medication, 52% strict bed rest for < 3 days.	Persistence of unchanged or worsened LBP at week 7 following the initial consultation (dichotomous)		Scores ≤3 = 2.9%; scores 4-6 = 7.9%; scores > 6 = 19.1%

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	25.	Identifying patients who are likely to develop long-term persistent symptoms	1. female; 2. dissatisfaction with employment situation; 3. history of LBP; 4. radiating leg pain; 5. widespread pain; 6. two or more restrictions in spinal movement	Count of predictors	Derivation Thomas et al. (1999) Single cohort	n=180 (167 in multivariable analysis), patients presenting in general practice with new episode of LBP, 59% female, 66% aged between 30-59 years, 75% symptoms < 4 weeks, 34% prevalence of dependent outcome.	Not specified.	≥2 / 10 pain on VAS and Hanover score < 75% at 1 week and 3 and 12 months (dichotomous)	Not reported, but accuracy of 5 or more predictors present calculated to be sens = 0.41 (0.29 - 0.55); spec = 0.92 (0.86 - 0.96); +LR = 5.3 (2.6 - 10.8); -LR = 0.64 (0.50 - 0.81).	0 - 2 predictors present = 6%; 3 predictor present = 27%; 4 predictors present = 35%; 5-6 predictors present= 70%
Work	26.	Predicting probable work outcome following outpatient rehabilitation	1. MMPI[57] depression score; 2. age; 3. duration of problem; 4. duration of time off work; 5. gender	Classification functions	Derivation Cairns et al. (1984) Single cohort	n=100, patients with LBP attending outpatient rehabilitation, mean age 43 years (dispersion data not reported), 50% female, mean 3.5 years duration of symptoms (no dispersion data reported), 52% prevalence of return to work.	Outpatient rehabilitation program 3 hour per day, 5 days a week x 4 weeks, consisting of conditioning exercises, stress management, nutrition advice, medication reduction and biofeedback.	Work status at 1 year post-discharge - working, ready for or in vocational rehabilitation, not working (trichotomous)	67% classification accuracy.	

[57] Minnesota multiphasic personality inventory

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	27.	Predicting probable work outcome following inpatient rehabilitation	1. MMPI hysteria score; 2. age; 3. duration of problem; 4. income source (no disability income, worker's compensation, social security disability, worker's compensation and social security disability); 5. gender	Classification functions	Derivation Cairns et al. (1984) Single cohort	n=100, patients with LBP attending inpatient rehabilitation, mean age 46 years (dispersion data not reported), 67%%, female, mean 8 years duration of symptoms (no dispersion data reported), 15% prevalence of return to work.	Multidisciplinary inpatient treatment program.	Work status at 1 year post-discharge - working, ready for or in vocational rehabilitation, not working (trichotomous)	73% classification accuracy.	

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	28.	Identifying patients consulting in primary care who are at risk of an adverse occupational outcome	1. patient's recovery expectations; 2. radiating pain; 3. previous back surgery; 4. pain intensity; 5. frequent change in position because of back pain; 6. irritability and bad temper; 7. difficulty sleeping.	Algorithm	Derivation Dionne et al. (2005) Single cohort with split sample	Training set n=354, validation set n=506, adults absent from work due to LBP consulting in primary care, full sample demographics - mean age 39 years (SD 11), 42% female, 78% with recurrent or persistent back pain, 17% failed to return to work in good health at 2 years.	Not specified	Return to work in good health at 2 years (success / partial success / failure) (categorical)	For failure vs partial success / success: Training set - sens = 0.79; spec = 0.64. LRs not reported, but calculated from study data to be +LR = 2.2 (1.8 – 2.7); -LR = 0.33 (0.20 – 0.55).	For algorithm predicted failure to return to work in good health = 31% (training set). 95%CrI calculated to be 23.8% - 38.7%.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	29.	Identifying patients off work with chronic LBP and participating in a functional restoration program who are likely to return to work by 6 months	1. duration of complaints (months); 2. functional disability (ODQ score); 3. presence of disc herniation with associated radiculopathy (MRI[58] confirmed herniation or extrusion, unilateral pain, unilateral paraesthesia or pain below the knee in 1 leg, and a SLR discrepancy of at least 15° between legs); 4. fear avoidance beliefs (FABQ)	Nomogram	Derivation Heymans et al. (2007) Retrospective single cohort	n=194, patients attending a physiotherapy functional restoration program and on sick leave due to LBP, mean age 42 years (SD 10), 67% female, mean 21 months (SD 40) of symptoms, 70% prevalence of dependent outcome.	Physiotherapy functional restoration program 3 times per week for 4-8 weeks, consisting of progressive aerobic and resistance exercises delivered using a cognitive-behavioural approach.	Return to work (including modified duties) 6 months following completion of functional restoration program (dichotomous)	Calibration – Slope index = 0.91. Discrimination - Area under receiver operating characteristic curve = 0.76. For a threshold of the nomogram of ≥50% predicted probability of return to work, sens = 62%; spec = 78%. LRs not reported, but approximated from study data to be +LR = 2.8 (1.7 – 4.5); -LR = 0.49 (0.38 – 0.64).	For ≥50% threshold of predicted probability of return to work, post-test probability = 87%. 95%CrI approximated to be 78.5% - 92.0%.

[58] magnetic resonance imaging

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	30.	Identifying patients on sick leave due to LBP who are at risk of more than 6 months of sick leave.	1. job satisfaction (good, reasonable, moderate or poor); 2. fear avoidance beliefs (FABQ); 3. pain intensity (VAS 0-10); 4. duration of complaints (weeks); 5. gender	Score chart	Derivation Heymans et al. (2009) Pooled results from 3 RCTs	n=628, patients on sick leave due to LBP < 8 weeks, mean age 41 years (SD 10), 29% female, median 6 months duration of complaint (IQR 13.3), 19% prevalence of dependent outcome.	RCT 1 (n=134) used behaviourally orientated graded activity program vs usual care; RCT 2 (n=195) used a workplace intervention and graded activity vs usual care; RCT 3 (n=299) used high and low intensity back schools vs usual care.	Prolonged sick leave > 6 months (dichotomous)	For scores ≥10, sens = 0.32; spec = 0.89. LRs not reported, but approximated from study data to be +LR = 2.8 (2.0 – 4.0); -LR = 0.77 (0.68 – 0.87).	For scores ≥10, post-test probability = 41%. 95%CrI approximated to be 31.2% - 50.5%.

Table 9.6 (p. 361) and Table 9.7 (p. 363) summarize the methodological quality of included studies appraised using the QUIPS and the PEDro scales respectively. Quality scores on the PEDro scale ranged from 5 to 9. Quality appraisal of the included derivation studies (Table 9.8, p. 364) identified the following items as the most frequent sources of potential bias: lack of blinded outcome assessment; no assessment of collinearity of predictor variables; no justification for the number of study participants; no reporting of uncertainty intervals for posterior probability estimates; insufficient reporting on the reliability of predictor variables; insufficient number of outcome events per candidate predictor variable; and lack of justification for the selection of candidate predictor variables. Table 9.9 (p. 366) summarises the quality appraisal of included validation studies. The most common sources of potential bias were: lack of assessment of the inter-observer reliability of the CPR; no assessment/reporting of the accuracy of the tool; not having complete follow-up; and no reporting of uncertainty intervals for posterior probability estimates.

Table 9.6 Risk of potential bias of included studies as appraised using QUIPS

Study	CPR reference number	Stage of development	Study participation	Study attrition	Prognostic factor measurement	Outcome measurement	Confounding measurement and account	Analysis
Alonso-Blanco et al. (2009)	10	Derivation	Low	Low	Moderate	Low	N/A	Moderate
Al-Sayegh et al. (2010)	12	Derivation	Low	Low	Moderate	Low	N/A	Moderate
Al-Sayegh et al. (2010)	13	Derivation	Low	Low	Moderate	Low	N/A	Moderate
Buranapanitkit et al. (2003)	23	Derivation	Moderate	Low	Moderate	High	N/A	Moderate
Cai et al. (2009)	11	Derivation	Low	Low	Moderate	Low	N/A	Moderate
Cairns et al. (1984)	26	Derivation	Moderate	Low	Moderate	Moderate	N/A	Moderate
Cairns et al. (1984)	27	Derivation	Moderate	Low	Moderate	Moderate	N/A	Moderate
Cleland et al. (2006)	3	Validation	Moderate	Low	Low	Low	High	Low
C. Cook et al. (2013)	3	Validation	Low	Moderate	Low	Low	Low	Low
Dionne (2005)	2	Validation	Low	Moderate	Low	Low	High	Moderate
Dionne et al. (1997)	2	Derivation	Low	High	Moderate	Low	N/A	Moderate
Dionne et al. (2005)	28	Derivation	Low	Moderate	Moderate	Low	N/A	Low
Dionne et al. (2011)	2	Validation	Low	Moderate	Low	Low	High	Low
Flynn et al. (2002)	3	Derivation	Low	Low	Moderate	Low	N/A	Moderate
Fritz, Childs, et al. (2005)	4	Derivation	Low	Low	Low	Low	N/A	Low
Fritz, Brennan, and Leaman (2006)	4	Validation	Moderate	Low	Low	Low	Moderate	Moderate
George et al. (2005)	5	Derivation	Low	Low	Low	Low	N/A	Low

Study	CPR reference number	Stage of development	Study participation	Study attrition	Prognostic factor measurement	Outcome measurement	Confounding measurement and account	Analysis
George et al. (2005)	16	Derivation	Low	Low	Low	Low	N/A	Low
Hancock, Maher, et al. (2009)	22	Derivation	Low	Low	Moderate	Moderate	N/A	Moderate
Hewitt et al. (2007)	7	Derivation	Low	Low	Moderate	Low	N/A	Low
Hewitt et al. (2007)	17	Derivation	Low	Low	Moderate	Moderate	N/A	Low
Heymans et al. (2007)	29	Derivation	Low	Low	Moderate	Moderate	N/A	Low
Heymans et al. (2009)	30	Derivation	Low	Low	Moderate	Moderate	N/A	Low
G. E. Hicks et al. (2005)	6	Derivation	Low	Low	Moderate	Low	N/A	Low
Kovacs et al. (2012)	14	Derivation	Low	Low	Moderate	Low	N/A	Moderate
Kovacs et al. (2012)	18	Derivation	Low	Low	Moderate	Low	N/A	Moderate
Kovacs et al. (2012)	19	Derivation	Low	Low	Moderate	Low	N/A	Moderate
Malmqvist et al. (2008)	21	Derivation	Low	Moderate	Moderate	Low	N/A	Moderate
May et al. (2008)	8	Derivation	Moderate	Low	Low	Low	N/A	Moderate
May et al. (2008)	9	Derivation	Moderate	Low	Low	Low	N/A	Moderate
Roland et al. (1983)	1	Derivation	Moderate	Low	Moderate	Low	N/A	Moderate
Roland et al. (1983)	20	Derivation	Moderate	Low	Moderate	Moderate	N/A	Moderate
Schwind et al. (2013)	3	Validation	Low	Low	Low	Low	Low	Low
Stolze et al. (2012)	15	Derivation	Low	Low	Low	Low	N/A	Moderate
Thomas et al. (1999)	25	Derivation	Moderate	Moderate	Moderate	Low	N/A	Moderate
Valat et al. (2000)	24	Derivation	Low	Low	High	High	N/A	Moderate

Table 9.7 Risk of potential bias of included studies as appraised using PEDro

Study	CPR reference number	Stage of development	Eligibility criteria	Random allocation	Concealed allocation	Baseline comparability	Blind subjects	Blind therapists	Blind assessors	Adequate follow-up	Intention to treat analysis	Between-group comparisons	Point estimates and variability	Score
Childs et al. (2004)	3	Validation	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	8
Childs et al. (2006)	3	Validation	No	Yes	No	No	No	No	No	Yes	Yes	Yes	Yes	5
Cleland et al. (2009)	3	Validation	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	8
Hallegraeff et al. (2009)	4	Validation	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	7
Hancock, Maher, Latimer, et al. (2008)	3	Validation	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	9
Schenk et al. (2012)	3	Validation	Yes	Yes	No	Yes	No	No	Yes	No	No	Yes	Yes	5
Sutlive et al. (2009)	3	Validation	Yes	Yes	No	Yes	No	No	No	Yes	Yes	Yes	Yes	6

Table 9.8 Methodological appraisal of included derivation studies

Study	CPR reference number	Prospective design	Study site described	Justification for number of participants	Representative sample	Important patient characteristics described	Candidate predictor variables justified	Blinded predictor assessment	Predictor variables have demonstrated reliability	Outcome measure valid and reliable	Blinded outcome assessment	Mathematical techniques described	Reporting and handling of missing data	10 outcome events per variable in final model	10 outcome events per candidate variable	Collinearity of predictor variables assessed	Predictor variables kept continuous	Uncertainty in CPR accuracy described	Uncertainty in posttest probability described	Non-paradoxical performance
Alonso-Blanco et al. (2009)	10	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Partly	No	No	No	No	Yes	No	No
Al-Sayegh et al. (2010)	12	Yes	Partly	Yes	Partly	Yes	No	Yes	No	Yes	No	Yes	No	No	No	No	N/A	Yes	No	Yes
Al-Sayegh et al. (2010)	13	Yes	Partly	Yes	Partly	Yes	No	Yes	No	Yes	No	Yes	No	No	No	No	No	Yes	No	Yes
Buranapanitkit et al. (2003)	23	No	Yes	No	No	Yes	No	No	No	No	No	Yes	No	Yes	No	No	N/A	N/A	No	Yes
Cai et al. (2009)	11	Yes	Yes	No	Yes	Yes	Partly	Yes	Partly	Yes	No	Yes	No	No	No	No	No	Yes	No	Yes
Cairns et al. (1984)	26	No	Yes	No	Yes	Partly	No	No	No	No	No	Yes	No	N/A	N/A	No	Yes	No	N/A	N/A
Cairns et al. (1984)	27	No	Yes	No	No	Partly	No	No	No	No	No	Yes	No	N/A	N/A	No	Yes	No	N/A	N/A
Dionne et al. (1997)	2	Yes	Yes	No	Partly	Yes	Partly	Yes	No	Partly	No	Yes	Partly	N/A	N/A	N/A	N/A	N/A	No	Yes
Dionne et al. (2005)	28	Yes	Yes	No	Yes	Yes	Yes	Yes	Partly	Partly	No	Yes	Partly	N/A	N/A	N/A	N/A	No	Yes	Yes
Flynn et al. (2002)	3	Yes	Yes	No	Yes	Yes	No	Yes	No	Partly	No	Yes	Partly	No	No	No	No	Yes	No	Yes
Fritz, Childs, et al. (2005)	4	Yes	Yes	No	Yes	Yes	Yes	Yes	Partly	Yes	No	Yes	No	No	No	No	N/A	Yes	No	Yes
George et al. (2005)	5	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No	No	No	No	Yes	No	N/A	Yes
George et al. (2005)	16	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	No	No	Yes	No	N/A	Yes

Study	CPR reference number	Prospective design	Study site described	Justification for number of participants	Representative sample	Important patient characteristics described	Candidate predictor variables justified	Blinded predictor assessment	Predictor variables have demonstrated reliability	Outcome measure valid and reliable	Blinded outcome assessment	Mathematical techniques described	Reporting and handling of missing data	10 outcome events per variable in final model	10 outcome events per candidate variable	Collinearity of predictor variables assessed	Predictor variables kept continuous	Uncertainty in CPR accuracy described	Uncertainty in posttest probability described	Non-paradoxical performance
Hancock, Maher, et al. (2009)	22	Yes	Yes	No	Yes	Yes	No	Yes	Partly	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes
Hewitt et al. (2007)	7	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Partly	Yes	Yes	No	Yes	No	N/A	Yes
Hewitt et al. (2007)	17	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	Yes	Partly	Yes	Yes	No	Yes	No	N/A	Yes
Heymans et al. (2007)	29	No	No	No	Partly	Yes	Yes	No	No	No	No	Yes	Yes	Yes	No	No	Yes	No	No	No
Heymans et al. (2009)	30	Yes	Partly	No	Partly	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	No	No	Yes	No	No	Yes
G. E. Hicks et al. (2005)	6	Yes	Yes	No	Yes	Yes	Partly	Yes	Yes	Yes	Yes	Yes	Partly	No	No	No	No	Yes	No	No
Kovacs et al. (2012)	14	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	No	Yes	Yes	Yes	Yes	No	Yes	N/A	No	Yes
Kovacs et al. (2012)	18	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Partly	No	Yes	Yes	Yes	Yes	No	Yes	N/A	No	Yes
Kovacs et al. (2012)	19	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Partly	No	Yes	Yes	Yes	Yes	No	Yes	N/A	No	Yes
Malmqvist et al. (2008)	21	Yes	Yes	No	Partly	Yes	Yes	Yes	No	Partly	No	Yes	Yes	Yes	Yes	No	N/A	N/A	No	Yes
May et al. (2008)	8	Yes	Yes	No	Yes	No	Partly	Yes	Partly	Partly	No	Yes	Yes	No	No	No	N/A	N/A	No	Yes
May et al. (2008)	9	Yes	Yes	No	Yes	No	Partly	Yes	Partly	Partly	No	Yes	Yes	No	No	No	N/A	N/A	No	Yes
Roland et al. (1983)	1	Yes	Yes	No	Partly	Partly	No	Yes	No	Partly	No	Yes	Partly	Yes	No	Partly	No	N/A	No	Yes
Roland et al. (1983)	20	Yes	Yes	No	Partly	Partly	No	Yes	No	No	No	Yes	Partly	Yes	No	Partly	No	N/A	No	Yes
Stolze et al. (2012)	15	Yes	Yes	Yes	Yes	Yes	Partly	Yes	Partly	Partly	No	Yes	Yes	No	No	No	Yes	Yes	Yes	No
Thomas et al. (1999)	25	Yes	Yes	No	No	Yes	Partly	Yes	No	Partly	No	Yes	Partly	No	No	No	No	N/A	No	Yes
Valat et al. (2000)	24	Yes	Partly	No	Partly	Yes	No	Yes	No	No	No	Yes	Partly	Yes	No	No	No	N/A	No	Yes

Table 9.9 Methodological appraisal of included validation studies

Study	CPR reference number	Prospective validation in new sample	Different clinical setting	Representative sample	The rule is applied accurately	Reliability of the rule is assessed	Complete follow-up	Reporting and handling of missing data	Accuracy of the rule described	Accuracy uncertainty described	Posttest probability uncertainty described
Childs et al. (2004)	3	Yes	Partly	Yes	Yes	No	No	Yes	Yes	Yes	No
Childs et al. (2006)	3	Yes	Partly	Yes	Yes	No	No	Yes	No	N/A	No
Cleland et al. (2006)	3	Yes	Yes	Partly	Yes	No	Yes	Yes	No	N/A	No
Cleland et al. (2009)	3	Yes	Yes	No	Yes	No	No	Yes	No	N/A	No
C. Cook et al. (2013)	3	Yes	Yes	Yes	Yes	No	Partly	Partly	Yes	Yes	N/A
Dionne (2005)	2	Yes	Yes	Yes	Yes	No	No	Yes	Yes	No	No
Dionne et al. (2011)	2	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Fritz, Brennan, and Leaman (2006)	4	No	Yes	Partly	Yes	No	Yes	Yes	No	N/A	N/A
Hallegraeff et al. (2009)	4	Yes	Yes	No	Yes	No	N/A	No	No	N/A	No
Hancock, Maher, Latimer, et al. (2008)	3	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	N/A
Schenk et al. (2012)	3	Yes	Yes	Partly	Yes	No	No	Yes	No	N/A	No
Schwind et al. (2013)	3	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	N/A
Sutlive et al. (2009)	3	Yes	No	No	Yes	No	No	Yes	No	N/A	N/A

9.5 Discussion

9.5.1 Characteristics of included studies

This systematic review identified 35 studies reporting on the development of non-surgical prognostic/prescriptive forms of LBP CPRs. It builds upon the existing body of literature with the inclusion of 13 studies that have not been previously reported on in earlier reviews (Beneciuk et al., 2009; Haskins et al., 2012; Lubetzky-Vilnai et al., 2014; May & Rosedale, 2009; Patel et al., 2013; Stanton et al., 2010; van Oort et al., 2012) on this topic. Thirty CPRs were identified with three of these tools known to have progressed to validation - the 'Cassandra rule' for predicting which patients with LBP are more likely to develop long-term significant functional limitations (Dionne, 2005; Dionne et al., 1997; Dionne et al., 2011), and the five-item (Childs et al., 2006; Childs et al., 2004; Cleland et al., 2009; Cleland et al., 2006; C. Cook et al., 2013; Flynn et al., 2002; Hancock, Maher, Latimer, et al., 2008; Schenk et al., 2012; Schwind et al., 2013; Sutlive et al., 2009) and two-item (Fritz, Brennan, & Leaman, 2006; Fritz, Childs, et al., 2005; Hallegraeff et al., 2009) Flynn manipulation CPRs designed to predict which patients being treated with lumbopelvic manipulation are more likely to experience a favourable prognosis.

Functional outcomes were the most common form of dependent variable used in the derivation of the identified CPRs. In contrast, pain and symptom resolution outcomes were much less frequently used. Given that the

performance of prognostic factors may not necessarily generalise across different types of dependent outcomes (C. Cook et al., 2013), it is important that the outcomes used in CPR derivation studies are selected based on the clinical problems they aim to address. However, very little research has been conducted to date on the types of clinical problems for which LBP CPRs should be developed (Haskins, Osmotherly, Southgate, et al., 2015). It is not known if the range of CPRs included in this review would be considered useful by clinicians who treat patients with LBP, although preliminary evidence suggests that the 5-item Flynn manipulation CPR has already been adopted by some practising clinicians (Learman et al., 2012; Sparks et al., 2010).

9.5.2 The Cassandra rule

The 'Cassandra rule' was derived in a population of patients with back pain presenting to primary care physicians and aims to identify individuals with differing degrees of risk of developing long-term significant functional limitations (Dionne et al., 1997). The CPR uses a measure of depression and a measure of somatization from selected items of the Symptoms Checklist 90 Revised questionnaire (Derogatis, 1977) to stratify patients by their degree of risk of having 50% or greater disability on the Roland-Morris Disability questionnaire (Roland & Morris, 1983) at two years. The 'Cassandra rule' was statistically validated using a split sample in the derivation study, and subsequently validated in a prospective cohort of 860 patients absent from work due to LBP consulting in primary care (Dionne, 2005). It was further validated in a prospective cohort of 1262 patients presenting to an

emergency department with non-specific back pain (Dionne et al., 2011). The prevalence of the dependent outcome was very similar across all three studies, ranging from just 16% to 19%. For the dichotomized groups of high/moderate vs low risk of the dependent outcome, the point-estimate of the +LR and –LR ranged from 1.3 to 2.0 and 0.25 to 0.40 respectively across the three studies. The relative consistency in the observed accuracy of this CPR to date provides preliminary confidence that it may perform similarly in other comparable clinical settings particularly in those settings with a similar prevalence of the dependent outcome. When comparing the 'Cassandra rule' to physician prediction, the CPR was found to be more sensitive (82% vs 37%) but less specific (45% vs 85%) in identifying those at risk of a poorer functional outcome (Dionne et al., 2011). This finding, in addition to the relative magnitudes of the +LR and –LR, suggests that 'negative' status on the prediction tool may be more informative to a clinician's prognostic judgement. No studies were identified that examined the utility of this CPR in improving clinical outcomes or resource efficiency. A more parsimonious version of the Cassandra rule comprising just 5 items, including one new question, has been developed and demonstrated to perform similarly to the original tool (Dionne et al., 2011). However, the development of this updated tool did not meet the criteria for inclusion in the present review.

9.5.3 Five-item Flynn manipulation CPR

Ten studies were included in this review regarding the development of a 5-item CPR for identifying patients receiving manipulation who are more likely to experience a favourable functional outcome. Flynn et al. (2002) derived

the tool in a population of patients with LBP presenting to military outpatient physical therapy facilities who were treated with lumbopelvic thrust manipulation (in a supine-lying position), range of motion exercises, and advice to maintain usual activity within the limits of their pain. A successful outcome was defined as more than a 50% improvement on the modified Oswestry Disability Questionnaire (Fritz & Irrgang, 2001) by the third treatment session, which occurred up to 8 days following the initial treatment. Nine validation studies were identified that investigated the generalizability of this CPR to other interventions and dependent outcomes, and explored whether the CPR is a treatment effect modifier.

In patients receiving thrust and non-thrust lumbopelvic manipulation, there is evidence to support positive baseline status on the Flynn manipulation CPR (most commonly defined as ≥ 4 predictors present) as a predictor of reduced disability (Childs et al., 2004; C. Cook et al., 2013; Flynn et al., 2002; Hancock, Maher, Latimer, et al., 2008; Schwind et al., 2013), improved pain (C. Cook et al., 2013; Hancock, Maher, Latimer, et al., 2008), greater patient-perceived extent of recovery (C. Cook et al., 2013), and less treatment sessions (C. Cook et al., 2013). The predictive value of a patient's baseline CPR status has however, been demonstrated to be sensitive to the threshold used to define a successful outcome. For example, Schwind and colleagues found that a patient's baseline status on the Flynn CPR was a predictor of successful outcome when using a definition of success of at least 50% improvement in the Oswestry Disability Questionnaire, but was no longer a significant independent predictor of success when the definition was changed

to at least 30% improvement on the same questionnaire (Schwind et al., 2013). One of the five variables in the Flynn manipulation CPR, no pain distal to the knee, was also identified as an independent predictor of reduced disability in the derivation of a prognostic CPR involving postpartum women with lumbopelvic pain receiving thrust manipulation (Al-Sayegh et al., 2010).

Four validation studies limited their study sample to participants assessed as being positive on the Flynn manipulation CPR in order to investigate the clinical outcomes achieved from interventions different to that used in the original derivation study. Cleland et al. (2006) found that 11 of 12 participants treated with a side-lying thrust manipulation technique achieved 50% or greater improvement in disability at 1 week. A subsequent RCT found similar improvements in pain and function in patients treated with supine-lying or side-lying thrust manipulation, but inferior outcomes in those treated with non-thrust manipulation (Cleland et al., 2009). Two further RCTs identified no significant differences in outcomes between those treated with supine-lying thrust manipulation and neutral gap thrust manipulation (Sutlive et al., 2009) or Mechanical Diagnosis and Therapy (Schenk et al., 2012). These findings suggest that in patients positive on the Flynn manipulation CPR, similar clinical outcomes may be achieved with some alternative interventions.

Two high quality validation studies included in this review investigated whether a patient's baseline status on the Flynn manipulation CPR is a treatment effect modifier. Childs et al. (2004) identified CPR status as a significant effect modifier of thrust manipulation compared to an exercise-

based intervention for the outcomes of pain and disability. In contrast, Hancock, Maher, Latimer, et al. (2008) did not identify CPR status as a significant treatment effect modifier of spinal manipulation compared to sham ultrasound for the same outcomes. There are several differences between these two studies that may account for their conflicting findings. In particular, all patients in the study of Childs et al received thrust manipulation compared to just 5% in the study of Hancock et al. The comparison interventions were also notably different and this may plausibly contribute to the observed differences in findings. As such, the findings of this review provide limited evidence supporting baseline status on the Flynn manipulation CPR as an effect modifier of thrust lumbopelvic manipulation compared to an exercise-based intervention for the outcomes of pain and disability.

9.5.4 Two-item Flynn manipulation CPR

A 2-item variation of the Flynn manipulation CPR was proposed by Fritz, Childs, et al. (2005) using the variables related to duration and distribution of symptoms. Using pooled data from patients with LBP receiving thrust manipulation in two previous studies (Childs et al., 2004; Flynn et al., 2002), the 2-item CPR was found to classify patients the same as the 5-item rule in 84% of cases. The +LR for the 2-item CPR for the dichotomized outcome of more than 50% reduction in disability by the third treatment (7.2) indicates that it would have a ‘moderate’ (Jaeschke et al., 1994) influence on shifting the pre-test probability. Based on the reported data, similar patients receiving this intervention who have both criteria present would have an 85% (95%CrI 71% - 93%) probability of achieving this outcome. Two validation studies of

the two-item CPR were included in this review (Fritz, Brennan, & Leaman, 2006; Hallegraeff et al., 2009), however both studies were conducted using only patients assessed as being positive on the rule and therefore provide limited evidence concerning the rule's accuracy in identifying patients with differing likelihoods of experiencing improvement. Using a retrospective database review, Fritz, Brennan, and Leaman (2006) identified that patients with both criteria present and treated with thrust or non-thrust manipulation experienced greater clinical improvements compared to those treated without manipulation. Further, those treated with thrust manipulation achieved similar clinical outcomes more efficiently than those treated with non-thrust manipulation. Hallegraeff et al. (2009) reported similar findings, with patients with both criteria present and randomized to receive thrust manipulation experiencing small but statistically significant greater improvements in disability compared to those randomized to receive non-manipulative physical therapy care.

9.5.5 Methodological considerations

No studies in this review selected predictor variables for inclusion in a CPR based on their identified function as a treatment effect modifier. This however, would be an important methodological consideration in the development of a CPR designed to identify likely responders to a given intervention (Hancock, Herbert, et al., 2009; Stanton et al., 2010). Although the majority (n=20) of studies included in this review sampled patients receiving a specific treatment program, it is not known if the predictor variables included in the derived CPRs reflect predictors of response to

treatment, or simply non-specific predictors of an outcome independent of the treatment (J. C. Hill & Fritz, 2011). For example, the evidence summarised in this review suggests that patients found positive on the 5-item Flynn manipulation CPR who receive lumbopelvic manipulation are more likely to experience functional improvement compared to those found negative on the rule. However, with the exception of one study (Childs et al., 2004), it is not yet clear if these patients are more likely to improve irrespective of the treatment provided.

The methodological appraisal of included studies identified several opportunities to reduce potential sources of bias in future CPR development studies. The selection of candidate predictor variables needs to be logically justified and considered within the context of probable predictive performance, psychometric properties, and practicality (C. Cook et al., 2010; Haskins, Osmotherly, Southgate, et al., 2015; Lubetzky-Vilnai et al., 2014; Seel et al., 2012). The study sample size should also be justified and sufficiently large to ensure at least 10 outcome events per candidate predictor variable (Bouwmeester et al., 2012; van Oort et al., 2012). The common practice of univariate screening to cull the number of predictors entered into a multivariable model is not effective in reducing the risk of overfitting and may subsequently lead to an increased chance of spurious findings (Babyak, 2004). Predictor variables should also be assessed for collinearity to reduce the likelihood of paradoxical CPR performance whereby the probability of an outcome decreases with increasing positive status on the rule (Lubetzky-Vilnai et al., 2014). This paradoxical performance was

observed in four studies included in this review with none of these studies reporting on whether collinearity was assessed (Table 9.8).

When validating a CPR, researchers should seek to include patients across the full range of possible categories of the tool. By only including patients that represent one particular status on the CPR (e.g., only those considered 'positive'), the performance of the tool in discriminating between patients with differing likelihoods of achieving the dependent outcome is unable to be evaluated. The inter-observer reliability of a patient's status on a CPR is a potential threat to its validity (Laupacis et al., 1997; Stiell & Wells, 1999) and ideally should be evaluated and reported similar to accepted standards for single-item tests (Bossuyt, Reitsma, Bruns, Gatsonis, Glasziou, Irwig, Moher, et al., 2003). Finally, during all phases of a CPR's development the reporting of uncertainty intervals for outcome prevalence, CPR accuracy (e.g., sensitivity, specificity, +LR, -LR), and posterior probabilities would enable a more meaningful interpretation of a study's findings (C. Cook et al., 2010; Haskins, Osmotherly, Tuyl, et al., 2014; Lubetzky-Vilnai et al., 2014).

9.5.6 Limitations

There are limitations of this study that need to be acknowledged. An operational definition of a LBP CPR was developed for this study by the research team to facilitate reproducibility and to transparently detail the types of studies under review. In particular, we aimed to differentiate studies that clearly presented a prediction tool that could be reasonably applied by a clinician for an individual patient from other forms of statistical prediction

models. It is anticipated however that others may have differing views on the sorts of tools that should be considered CPRs. Our definition lead to the exclusion of 11 studies (Axen et al., 2005; Brennan et al., 2006; Childs et al., 2003; Du Bois & Donceel, 2008; Fritz et al., 2007; Fritz et al., 2004; Jellema et al., 2006; Leboeuf-Yde et al., 2004; Skargren, Carlsson, & Oberg, 1998; Teyhen et al., 2007; van der Hulst, Vollenbroek-Hutten, Groothuis-Oudshoorn, & Hermens, 2008) that were included in earlier related reviews on this subject (Beneciuk et al., 2009; Haskins et al., 2012; May & Rosedale, 2009; Patel et al., 2013; van Oort et al., 2012). Our definition of a CPR required the use of multivariable statistics to derive predictor variables. This lead to the exclusion of some tools which have been previously called CPRs. For example, a tool developed by G. E. Hicks et al. (2005) to identify patients participating in a stabilization exercise program who are likely to experience improvement was excluded as the predictors in this tool were selected based on their univariate statistical association with the dependent outcome. Our definition of a CPR also excluded predictive tools such as the STarT Back Screening Tool (J. C. Hill et al., 2008) which sought to only include potentially modifiable prognostic factors.

A highly sensitive search strategy across multiple databases was used to identify potentially eligible studies. However the nomenclature used to describe CPRs is varied and it is therefore possible that some studies were inadvertently omitted. Our supplementary search strategies identified an additional 25 studies that were considered for eligibility, with four of these studies advancing to inclusion in the review. This highlights the importance of

supplementary search strategies (such as hand-searching and citation tracking) in the identification of CPR development studies.

The criteria used in this review to appraise the methodological quality of derivation and validation CPR development studies have not been validated. They were based on standards commonly reported in well-cited CPR methodological texts (Beattie & Nelson, 2006; Childs & Cleland, 2006; Laupacis et al., 1997; McGinn et al., 2000; Stiell & Wells, 1999) and methodological items identified in recent publications (Bouwmeester et al., 2012; C. Cook et al., 2010; Haskins, Osmotherly, Tuyl, et al., 2014; Lubetzky-Vilnai et al., 2014; Seel et al., 2012; van Oort et al., 2012). Many of the criteria used in the present review have however, been used in earlier previous systematic reviews on CPRs. The QUIPS and PEDro scales were used in this review to complement the methodological appraisal of included studies.

9.5.7 Conclusion

In conclusion, this systematic review identified 30 prognostic/prescriptive CPRs relevant to the non-surgical management of adults with LBP. Most have not yet undergone validation and therefore cannot be recommended for use in clinical practice at this time. Clinicians may however, consider using knowledge of the identified individual predictors that comprise these tools to cautiously inform their prognostic clinical judgements (McGinn et al., 2008). Three CPRs included in this review have been identified to have undergone validation. The ‘Cassandra rule’ has been validated in two large prospective

studies and may be applied in comparable clinical settings with similar patient populations with some confidence in its modest prognostic accuracy. It is not yet known if the clinical application of the ‘Cassandra rule’ results in improved patient outcomes or improvements in resource efficiency. Positive status on the 5-item Flynn manipulation CPR has been demonstrated in several studies to be a predictor of reduced disability in patients receiving thrust and non-thrust forms of lumbopelvic manipulation. It is not yet clear however, if a patient’s status on this CPR predicts a more favourable prognostic outcome irrespective of the treatment provided. No evidence was found that addressed whether the clinical application of the Flynn manipulation CPR results in improved patient outcomes or more efficient care. A 2-item variation of the Flynn manipulation CPR has undergone validation in two studies. However in both studies patients who were deemed negative on the rule were excluded, thereby precluding an evaluation of the rule’s predictive accuracy in identifying patients with differing likelihoods of experiencing improvement. More research seeking to validate the derived CPRs identified in this review is warranted. Research evaluating the clinical impact of the application of the ‘Cassandra rule’ and the 5-item Flynn manipulation CPR is also needed.

CHAPTER 10

SUMMARY OF FINDINGS, CONCLUSIONS AND RECOMMENDATIONS

In this concluding chapter, the overall key findings and recommendations that have arisen from this program of research will be summarised. The implications of these findings will be considered from clinical and research methodology perspectives. The limitations of this thesis will be discussed and opportunities for further research will be examined. Finally, a concluding summation of the thesis will be provided.

10.1 Summary of findings

The overall objective of this research program was to facilitate the development of CPRs with the greatest potential to positively influence the physiotherapy management of LBP. This was achieved through a series of five studies and a Clinical Commentary that together sought to address the following three aims:

1. Identify and assess the degree to which CPRs for LBP may be confidently applied in clinical practice using a hierarchical framework for CPR development and an appraisal and synthesis of the existing evidence base.
2. Explore the range of factors that may influence the implementation of CPRs for LBP within Australian physiotherapy practice.

3. Examine the areas of perceived need for LBP CPRs and the range of characteristics such tools need to encompass to be considered clinically meaningful and useful within Australian physiotherapy practice.

10.1.1 Research aim 1

Three systematic reviews were conducted to address the first research aim. At the time when the first systematic review (Chapter 4) was initiated, only one review had been previously published on the topic of CPRs relevant to musculoskeletal physiotherapy practice, although this was limited to CPRs for physical therapy interventions in the derivation phase of development (Beneciuk et al., 2009). The first study in this program of research therefore sought to identify all forms of CPRs (diagnostic, prescriptive and prognostic) developed within the discipline of physiotherapy and relevant to the management of patients with LBP, at any stage of CPR development. A broad operational definition of a CPR was employed, reflecting the limited knowledge regarding the number, type and quality of tools that had been developed to that point. The electronic search was conducted across five medical databases from 1990 to January 2010 and was supplemented with citation tracking and hand-searching of relevant journals. Of the 7,453 unique records screened for eligibility, 23 studies (15 derivation, 8 validation, and 0 impact analysis) were included describing the development of 25 CPRs.

The majority of physiotherapy CPRs under development were diagnostic in function (n=15), with relatively fewer prescriptive (n=7) and prognostic (n=3)

tools. Most tools had been developed in the USA, which may have important implications regarding the awareness of these tools amongst physiotherapists in other regions. The intended function of the derived CPRs was notably divergent, which may reflect lack of knowledge regarding the perceived needs of the target clinical consumers of these tools.

The key finding of the systematic review detailed in Chapter 4 was that 23 of the 25 derived CPRs for LBP developed within the physiotherapy profession had been derived, but not yet subjected to validation or impact analysis. Using a well-accepted hierarchical framework for CPR development (McGinn et al., 2000), these tools could not be recommended for use in clinical practice (see section 3.7, p. 115). This is because the variables identified to have a predictive relationship with the dependent outcome in these studies may simply reflect chance associations, or be specific to the study sample or setting in which the tools were derived (McGinn et al., 2008). Two CPRs were identified to had undergone validation - a five item tool (Flynn et al., 2002) and two item abbreviated variant (Fritz, Childs, et al., 2005) designed to predict a more favourable functional outcome in patients receiving lumbopelvic manipulation. Neither of these tools had however undergone impact analysis, and therefore could not be recommended to be used in practice with confidence that their application would likely result in improved patient outcomes.

The two subsequent systematic reviews (Studies 4 and 5) conducted within this program of research aimed to update and extend the knowledge base

regarding the development of CPRs relevant to the physiotherapy assessment and management of LBP. These studies differed in their scope and methodology to that used in Study 1 in several important ways:

1. Studies 4 and 5 were entirely focused upon either diagnostic (Chapter 8) or prognostic (Chapter 9) forms of LBP CPRs. The findings of these studies were thus explored with greater specificity and depth in regard to the type of tool under investigation.
2. Studies 1 (Chapter 4) and 2 (Chapter 5) highlighted that the use of the term 'clinical prediction rule' is subject to interpretation in both research and clinical contexts. To facilitate greater transparency, reproducibility and to more precisely describe the type of tools under investigation, a clear operational definition of a CPR inclusive of basic methodological standards was applied in Studies 4 and 5.
3. Study 1 focused upon tools developed within the physiotherapy profession. However, in Studies 4 and 5 CPR development studies were included irrespective of the health discipline(s) involved in their development. The broadened scope of Studies 4 and 5 was anticipated to elicit more comprehensive findings that would be of greater significance to physiotherapists in their management of LBP.
4. Studies 4 and 5 included the large volume of more recent research in their respective fields. Further, these reviews were not restricted to CPRs developed after a given date, thus facilitating a broader search of relevant LBP CPRs.
5. A more sensitive electronic search strategy was employed in Studies 4 and 5 based upon the incorporation of a newly developed sensitive

search string designed to identify prediction model studies (Geersing et al., 2012). Further, a greater number of electronic databases were searched allowing the identification of a greater number of potentially eligible studies.

6. The findings of Study 3 Chapter 6 (p. 210), indicated that the predictive precision of a LBP CPR may have important implications regarding its perceived usefulness and likelihood to be implemented in a clinical context. Thus, using the methodology described in the Clinical Commentary (Chapter 7), uncertainty intervals were calculated or approximated for posterior probability estimates in instances where reported data permitted.
7. The quality appraisal in Studies 4 and 5 incorporated recent methodological considerations pertinent to the development of CPRs. The two updated reviews additionally appraised the methodological quality of included studies based upon their underlying study design. Greater opportunities to improve the methodological rigour of LBP CPR development studies were consequently able to be identified.

Study 4 (Chapter 8) was a systematic review of diagnostic forms of LBP CPRs. 'Diagnosis' was not restricted to a pathoanatomic source (see discussion in sections 2.4 and 2.5), but rather more broadly defined as relating to the present status or classification of an individual. This operational definition also served to delineate these tools from their prognostic counterparts, which were investigated in Study 5 (Chapter 9). Using a highly sensitive search strategy, 15 publications reporting on 18

studies were included in the review. Notably, 14 of these publications were additional to those included in Study 1, reflecting both the broader scope and more sensitive search strategy employed in this updated systematic review. Further, six publications (Henschke et al., 2009; Laslett, Aprill, et al., 2006; Laslett, Aprill, et al., 2005; Laslett, McDonald, et al., 2006; Laslett et al., 2003; van der Wurff et al., 2006) that were included in study 1 were excluded in Study 4 as a result of the use of an updated and standardised operational definition of a CPR incorporative of basic methodological standards.

Study 4 identified that 13 diagnostic CPRs for LBP had been derived (see Table 8.5, p. 277). Notably, the clinical presentations that the CPRs aimed to assist to in identifying were quite diverse and included: spinal stenosis (n=3), vertebral fracture (n=2), inflammatory back pain (n=2), spondyloarthritis (n=1), zygapophyseal joint mediated pain (n=1), radicular LBP (n=1), radiographic instability (n=1), spondylolysis (n=1), and psychological disturbance (n=1).Ten of the 13 identified diagnostic LBP CPRs had not been developed beyond their initial derivation, and are consequently not recommended for use in clinical practice at this time (McGinn et al., 2000). Three diagnostic LBP CPRs were identified to have undergone validation - The Japanese Society for Spine Surgery and Related Research CPR to assist in the identification of patients with spinal stenosis; and the 'Berlin criteria' and 'IBP according to experts criteria', which are both designed to assist in the identification of patients presenting with inflammatory back pain (see Figure 8.2, p. 295).

The Japanese Society for Spine Surgery and Related Research CPR (Konno et al., 2007) (see section 8.5.1, p. 294) underwent broad validation in a single prospective study (Kato et al., 2009) that included patients with lower extremity symptoms presenting to hospital orthopaedic outpatient clinics. Such validation increases confidence with regard to the known predictive performance of the CPR in those clinical contexts. The results of that study suggest that a patient's positive status on that tool may have only limited predictive performance in identifying patients with lumbar spinal stenosis (+LR = 1.6, 95%CI 1.3, 2.0). However, negative status on the CPR may be informative in identifying those without the condition (-LR = 0.13, 95%CI 0.04, 0.41). Applying the pre-test probability of lumbar spinal stenosis identified in the validation study (49%) and using the methodology outlined in the Clinical Commentary (Chapter 7), the post-test probability of a patient having lumbar spinal stenosis given their positive status on this tool is 60%, with a 95% credible interval of 50% - 70%. A patient's negative status gives a post-test probability of 11% with a 95% credible interval of 4% - 28%. As no impact analysis studies were identified for this CPR, it is unknown if the application of this tool is likely to result in beneficial outcomes such as improving the efficiency of the diagnostic process, or improving patient outcomes.

The 'Berlin criteria' CPR (Rudwaleit et al., 2006) (see Figure 8.2, p. 295) functions to identify patients with LBP with probable inflammatory causes. Such patients are considered likely to benefit from specialist referral and further tests to investigate for the presence of spondyloarthritis (Golder & Schachna, 2013; Kain et al., 2008). Identifying patients with spondyloarthritis

early in the disease process is considered challenging but important to the successful management of that condition (Rudwaleit, van der Heijde, et al., 2004). The Berlin criteria were identified to have undergone broad validation in three studies, the largest of which involved 648 patients with chronic LBP presenting to a rheumatologist (Sieper, van der Heijde, et al., 2009). The results of that study indicate that a patient positive on the 'Berlin criteria' presenting to a rheumatologist with chronic LBP would have a post-test probability of 88% (95% credible interval 84% - 91%) of having inflammatory back pain. The two smaller validation studies in patients with suspected axial spondyloarthritis (Sieper, van der Heijde, et al., 2009) and anterior uveitis (Chan et al., 2012) provide similar point-estimates for the +LR to that of the larger validation study and range from 2.2 – 3.8. The findings of the validation studies provide increased confidence regarding the known predictive performance of the 'Berlin criteria' in these settings.

'IBP according to experts' (also known as the Assessment of SpondyloArthritis international Society expert criteria, see Figure 8.2, p. 295) (Sieper, van der Heijde, et al., 2009) is a five item CPR that shares the same function as the 'Berlin criteria' in screening for patients with LBP with probable inflammatory back pain. This CPR underwent broad validation concurrently with the 'Berlin criteria' in the multinational study conducted by Sieper, van der Heijde, et al. (2009), and was identified to be more sensitive (80% versus 70%) but less specific (72% versus 81%) than that tool. Using data from the multinational validation study that included patients with chronic LBP presenting to a rheumatologist, a patient's positive status on the 'IBP

according to experts' CPR would shift the pre-test probability of inflammatory back pain from 66% to 85% (95%CrI 81% - 88%). Using a hierarchical framework for CPR development, the evidence considered within Study 4 indicates that clinicians may consider applying the 'IBP according experts' CPR in patients with chronic LBP presenting to a rheumatologist with some confidence in the known predictive accuracy of that tool (McGinn et al., 2000).

Importantly, the evidence considered within Study 4 highlighted that the predictive performance of the 'Berlin criteria' and the 'IBP according to experts' CPRs have not been investigated in a patient population presenting to primary care. Consequently, the accuracy of these tools in the primary care setting is currently unknown. The lack of validation of these instruments in this setting is a significant consideration for their clinical application (McGinn et al., 2000). As the prevalence of inflammatory back pain in patients with LBP presenting in primary care is considered to be much lower than in those presenting to a rheumatologist (Hamilton et al., 2014; Underwood & Dawes, 1995), the predictive performance of both CPRs may differ meaningfully across these settings (Leeflang et al., 2009). Given the lack of impact analysis studies to date, it is also not known whether the application of these tools benefits clinical practice, such as leading to improved patient outcomes or the more accurate/efficient targeting of patients for further investigations and referrals. A further consideration is that additional non-CPR tools and referral strategies have been developed to help aid the earlier diagnosis of spondyloarthritis in primary care, however there

are limited data concerning their comparative predictive performance (Brandt et al., 2007; Hermann et al., 2009; Poddubnyy et al., 2011; Sieper et al., 2013). Consequently, it is not known if the 'Berlin criteria' or the 'IBP according to experts' CPRs are the optimal tools for this purpose.

Study 5 (Chapter 9) focused on prognostic CPRs that function to predict future outcomes. This included prescriptive CPRs that function to predict future relative treatment effects. Of 10,005 records that were screened for eligibility, 35 studies reporting on the development of 30 CPRs were included in the review. Twenty of the publications included in this review were not included in Study 1, and 13 publications had not been included in any previous review on this subject (Beneciuk et al., 2009; Haskins et al., 2012; Lubetzky-Vilnai et al., 2014; May & Rosedale, 2009; Patel et al., 2013; Stanton et al., 2010; van Oort et al., 2012). This highlights both the broader search strategy used in Study 5, as well as growth in this area of research in recent years. In contrast to the diversity observed in Study 4 on diagnostic CPRs, half of the 30 identified prognostic CPRs shared a similar dependent outcome – patient function.

The evidence considered within Study 5 indicates that 27 prognostic CPRs for LBP are in their derivation phase of development and consequently are not able to be recommended for direct use in clinical practice at this time (McGinn et al., 2000; McGinn et al., 2008). Three prognostic CPRs ('Cassandra rule', five item Flynn manipulation CPR, and two item Flynn manipulation CPR) were found to have undergone validation, but none of

these tools have been assessed for their ability to positively influence clinical practice.

The ‘Cassandra rule’ functions to delineate patients with LBP with differing degrees of risk of developing long-term significant functional limitation, defined as 50% or greater disability on the Roland-Morris Disability Questionnaire at two years (Dionne, 2005; Dionne et al., 1997; Dionne et al., 2011). The CPR was initially derived in patients with LBP presenting in primary care in the US, and has undergone broad validation in primary care patients absent from work due to LBP, as well as patients presenting to an emergency department in Canada. The point-estimate of the +LR and –LR ranges from 1.3 to 2.0 and 0.25 to 0.40 for the dichotomized groups of high/moderate vs low risk across the development studies. The evidence evaluated in Study 5 supports the clinical application of the ‘Cassandra rule’ in comparable clinical settings with preliminary confidence in its known, but limited, predictive performance (McGinn et al., 2000). The ‘Cassandra rule’ has been identified to be more sensitive (82% vs 37%) but less specific (45% vs 85%) in identifying those at risk of a poorer functional outcome compared to unassisted physician prediction. Coupled with knowledge of the relative magnitudes of the +LR and –LR, this finding suggests that ‘negative’ status on the prediction tool may be more informative to a clinician’s prognostic judgement in determining the risk of a poor long-term functional outcome.

Ten studies (1 derivation, 9 validation, 0 impact analysis) included in Study 5 concerned the development of the five item Flynn manipulation CPR. This

CPR functions to identify which patients with LBP receiving lumbopelvic manipulation are more likely to experience a favourable functional outcome. This body of evidence provides support that a patient's positive status on the CPR is a predictor of reduced disability, improved pain, greater patient-perceived extent of recovery and fewer treatment sessions amongst those patients with LBP receiving thrust or non-thrust lumbopelvic manipulation, relative to those who are negative on the rule (Childs et al., 2004; C. Cook et al., 2013; Flynn et al., 2002; Hancock, Maher, Latimer, et al., 2008; Schwind et al., 2013). Notably, the predictive performance of this CPR has been demonstrated to be sensitive to the definition of the dependent outcome. Schwind et al. (2013) identified that a patient's baseline status on the five item Flynn manipulation CPR was a predictor of successful outcome when using a definition of success of at least 50% improvement in the Oswestry Disability Questionnaire, but was no longer a significant independent predictor of success when the definition was changed to at least 30% improvement on the same questionnaire.

Two high quality RCTs (Childs et al., 2004; Hancock, Maher, Latimer, et al., 2008) provide conflicting evidence regarding whether the five item Flynn manipulation CPR is a treatment-effect modifier for lumbopelvic manipulation compared to an alternative intervention. Several important differences between these two studies may plausibly explain the differences in findings, most notably large differences in both the active and control interventions. Thus, while the study of Childs et al. (2004) provides preliminary evidence that status on the CPR may moderate the effect-size of disability

improvement from thrust lumbopelvic manipulation compared to an exercise-based intervention, the findings of Hancock, Maher, Latimer, et al. (2008) indicate that the CPR's moderating effect may not generalise to non-thrust manipulative treatment and/or when compared to a sham electrotherapy modality. Consequently, while there is consistent evidence supporting the performance of the five item Flynn manipulation CPR as a prognostic tool amongst those receiving lumbopelvic manipulation, the evidence is less clear regarding whether a patient's status on the tool should be used to preferentially select which patients should be offered this treatment.

Four other validation studies included in Study 5 demonstrate that patients who are positive on the five item Flynn manipulation may achieve similar clinical outcomes when provided with different interventions to that provided in the original derivation study (Cleland et al., 2009; Cleland et al., 2006; Schenk et al., 2012; Sutlive et al., 2009). These studies did not include patients who were negative on the rule, and thus the performance of the CPR in differentiating patients with differing prognoses was not able to be evaluated. Nevertheless, the finding that patients who are positive on the CPR may achieve similar clinical outcomes with differing interventions further challenges any presumption that the tool may be used to preferentially select patients to receive thrust lumbopelvic manipulation.

A two item variation of the Flynn manipulation CPR comprised of just the symptomatic duration and distribution variables was derived by Fritz, Childs, et al. (2005), and subsequently underwent validation in the studies of Fritz,

Brennan, and Leaman (2006) and Hallegraeff et al. (2009). In both validation studies, clinical outcomes were contrasted between patients who were 'positive' on the two item CPR and received lumbopelvic manipulation, and patients who were 'positive' on the CPR and received a non-manipulative intervention. The findings of both studies demonstrate greater disability improvements in those receiving manipulation. However, the design of both studies involved omitting those patients who were 'negative' on the CPR, and thus prevents analysis regarding the predictive performance of the CPR. This is analogous to a study that only recruits participants of one sex not being able to evaluate the relationship between gender and the outcome of interest.

10.1.2 Research aim 2

Study 2 (Chapter 5) sought to address research aim 2 by using qualitative research methodology to explore the range of factors that may influence the implementation of CPRs for LBP within Australian physiotherapy practice. A knowledge, attitudes and practices/behaviour framework (Cabana et al., 1999; Legare et al., 2008) was applied due to its extensive use in previous research to investigate the barriers to the adoption of other clinical innovations (Barlow et al., 2008; Dennison et al., 2007; Larson, 2004; Pogorzelska & Larson, 2008; Rubinson et al., 2005; Schouten et al., 2007), and its explicit recommendation as an appropriate framework to consider the barriers and facilitators to the use of CPRs (Abboud & Cabana, 2001). The key themes that were identified in this study are summarised in Table 5.4 (p. 196), and shared similarities with the identified barriers to the adoption of

clinical practice guidelines, outcome measures and evidence-based practice in physiotherapy (Abrams et al., 2006; Côté et al., 2009; Jette et al., 2003).

Some of the participants in Study 2 had not previously encountered the term or concept of CPRs. Also, there was limited familiarity with the specific content and function of LBP CPRs currently being developed. Given that Studies 1, 4 and 5 indicate that no single LBP CPR has been developed sufficiently at the present time to be able to be applied in clinical practice with confidence that its use will lead to positive benefits, it may not be critical to immediately address these knowledge gaps. However, these findings do highlight that addressing knowledge gaps in the future in regards to both the awareness of CPRs and familiarity with their specific content and function, represents an important step in facilitating the eventual translation of LBP CPR research evidence into practice.

Study 2 identified that some physiotherapists conceptualise CPRs as the formalisation of pattern recognition, and some view CPRs as paralleling existing well-accepted subgrouping mechanisms like Mechanical Diagnosis and Therapy (McKenzie & May, 2003). These insights may be informative to future strategies to facilitate the adoption of well-developed CPRs into practice, and to help understand how physiotherapists will seek to incorporate CPRs into their overall assessment and management of patients with LBP.

A range of both facilitative and inhibitive attitudes toward the use of CPRs for LBP were identified in Study 2. Facilitative attitudes included views that:

1. CPRs are evidence-based practice.
2. CPRs enable greater confidence in making predictions.
3. CPRs may help inform decision-making.
4. CPRs may help novice clinicians.
5. CPRs may positively challenge traditional reasoning strategies.
6. Numeric data may be helpful.

Inhibitive attitudes included views that:

1. CPRs are complicated.
2. CPRs are or could become fads.
3. CPRs could cause intellectual laziness.
4. CPRs have limited generalizability.
5. CPRs may challenge clinicians' autonomy.
6. CPRs may not work because treatment techniques are too varied.
7. CPRs oversimplify the complexities of a clinical presentation.
8. Dislike of the word 'rule'.
9. Existing CPRs are not yet ready to be applied.
10. LBP is too complicated for CPRs.
11. No personal need for a CPR.

While the focus of Study 2 was on CPRs relevant to the assessment and management of LBP, many of the identified facilitative and inhibitive attitudes may plausibly generalise to CPRs developed for other clinical problems. Future LBP CPR development studies and evidence translation strategies

may benefit from ensuring that such tools are constructed and marketed to practising physiotherapists with consideration of these identified clinician attitudes. Examples of such strategies are discussed further in the following subsection concerning implications arising from this research.

Amongst the sample of physiotherapist participants in Study 2, very few expressed that they have previously used a CPR in the management of a patient with LBP. Fewer again acknowledged that they would have used a CPR in the assessment and management of a fictitious LBP patient scenario (Figure 5.1, p. 190). It was felt by participants, however, that when CPRs are used in clinical practice they should not be used in isolation, but instead considered within the broader suite of clinical reasoning processes physiotherapists typically employ, and perhaps best viewed as ‘safety nets’ or ‘second opinions’. The findings of Learman et al. (2012) suggest that these views may be common. In that study, clinicians who reported using a CPR were identified to be no more likely to perform manipulation in the presence of contraindications compared to those who reported not to use a CPR. That is, the CPR was not blindly followed, but instead ‘overruled’ by the clinician as appropriately indicated by the relevant additional information not considered within that tool.

Finally, Study 2 also identified that some physiotherapists perceive that CPRs may negatively impact clinical practice if patient care and clinician autonomy become restricted by third third-party payers inappropriately utilising CPRs. Addressing issues concerning the appropriate application of

CPR evidence would therefore be an important component of any strategy designed to improve the adoption of well-developed LBP CPRs into clinical practice.

10.1.3 Research aim 3

Research aim 3 was addressed in Study 3 (Chapter 6), which used qualitative research methodology to explore the types of CPRs for LBP that Australian physiotherapists wish to see developed and the characteristics of LBP CPRs that physiotherapists believe are important.

Physiotherapy participants in this study wanted LBP CPRs that would facilitate the early and accurate identification of serious spinal pathology (e.g. fracture, cancer, infection). In contrast, there was limited perceived need for CPRs that function to sub-classify patients with LBP by pathoanatomic diagnosis (e.g. intervertebral disc, zygapophyseal joint) as these were seldom considered meaningful to physiotherapy management decisions. Of the 15 diagnostic CPRs identified to be currently under development in Study 4 (Chapter 8), only two function to identify a serious cause of LBP (Delitto et al., 2012)(see Table 2.7, p. 48), and both of these tools are in the derivation phase of development (Roman et al., 2010; Roux et al., 2007). By contrast, and using a definition of a CPR not inclusive of minimum methodological standards, 14 of the 15 diagnostic LBP CPRs developed within physiotherapy identified in Study 1 (Chapter 4), relate to the pathoanatomic diagnosis of LBP. These findings may suggest that the function of diagnostic

forms of LBP CPRs currently under development, may not match the perceived needs of some Australian physiotherapists.

Study 3 also highlighted that Australian physiotherapists may value CPRs that can accurately identify likely responders to an intervention. However, Studies 1 and 5 identified that while many LBP CPRs have been developed with patient populations receiving a specific program of therapy, none have selected variables for inclusion in a CPR based on the identification of treatment modifiers (Hancock, Herbert, et al., 2009; Sun et al., 2010). Consequently, the variables that comprise these CPRs may simply predict those with a favourable prognosis irrespective of the treatment provided, thereby limiting their capacity to inform treatment selection decisions. As such, there may currently be an unmet need for CPRs that function to identify likely responders to LBP treatment modalities. Participants in Study 3 further expressed a desire for LBP CPRs that may predict non-success, worsening or no need for intervention. Studies 1 and 5 highlight that there are currently few tools that may function to fulfil this perceived clinical need.

Physiotherapist participants in Study 3 also welcomed the development of prognostic forms of LBP CPRs. Study 5 identified that 30 prognostic CPRs for LBP have been derived, but only three have undergone validation and none have undergone impact analysis. Two-thirds of the identified CPRs were derived using patient samples receiving a specific treatment program, with most using a measure of patient function as the dependent outcome. Physiotherapists in Study 3 thought meaningful outcomes to predict include

recovery from a presentation, time to return to work, time to return to normal physical activity, likelihood of persisting symptoms, likelihood of requiring surgery and the likelihood of experiencing a recurrence. Approximately one third of the prognostic CPRs for LBP identified in Study 5 function to predict these outcomes.

A number of potentially modifiable characteristics of LBP CPRs were identified in Study 3, that were considered by physiotherapists to influence the clinical meaningfulness of these tools (see Figure 6.2, p. 219). Clinicians thought that LBP CPRs need to be simple, practical and able to be easily applied. Further, they need to be sufficiently evaluated such that physiotherapists may have confidence that their application will benefit clinical practice. In contrast, CPRs that may be perceived as containing a large number of variables, require the use of complicated statistics or comprise variables that have no clear logical relationship with the dependent outcome may be considered less meaningful. An example of the latter discussed by participants in Study 3 was the inclusion of a hip rotation variable in the five item Flynn manipulation CPR (Flynn et al., 2002). This was considered by some to perhaps reflect a statistical artefact and to be no more meaningful than a variable concerning a patient's hair colour. Thus, providing a clear, biologically-plausible and evidence-based explanation for the variables comprising a CPR, in addition to the demonstrated effectiveness of that tool, may have important implications for its acceptance in clinical practice.

Participants in Study 3 further reported that LBP CPRs need to be compatible with traditional clinical reasoning and decision-making strategies. This parallels a finding from Study 2 in that physiotherapists believed that CPRs applied in clinical practice should not be used in isolation, but instead considered within the broader suite of clinical reasoning processes physiotherapists typically employ. It is therefore plausible that developing LBP CPRs that encompass a broad range of assessment and management approaches, and are not overtly biased to a particular management paradigm may help gain greater overall clinician acceptance of CPRs. Strategies aiming to increase the clinical application of well-developed LBP CPRs may benefit from considering and addressing how a CPR may be practically applied in the clinical setting such that it is viewed to complement and not supplant clinical reasoning.

The precision and accuracy of a CPR was also identified as an important characteristic of tools designed to assist in the management of LBP. Analogous to the use of confidence intervals for diagnostic test accuracy and treatment effect sizes (Stratford, 2010), posterior probability uncertainty intervals may inform the application of CPR research evidence by considering their position relative to a threshold level of certainty required by a clinician to make a decision (Fritz & Wainner, 2001). Studies 1, 4 and 5, highlight that the precision of posterior probability estimates are seldom reported in CPR development studies. To help address this gap, a clinical and academic resource (Clinical Commentary - Chapter 7) was developed to generate awareness of this issue, and to provide practical support for the

formal calculation and approximation of uncertainty intervals for posterior probability estimates. The primary method outlined in this text (objective Bayesian method using Monte Carlo simulation) was subsequently used to calculate or approximate uncertainty intervals for posterior probability estimates for CPR development studies included in Studies 4 and 5. Consideration of the precision of posterior probability estimates has important implications for the design and reporting of LBP CPR development studies, and may plausibly impact the perceived usefulness of a CPR in clinical practice.

10.2 Implications

The clinical and research methodology implications arising from each study within this program of research have been detailed in their respective chapters and are briefly summarised below.

10.2.1 Clinical implications

Studies 1, 4 and 5 involved the synthesis of the available body of evidence concerning the development of LBP CPRs to facilitate an assessment of their readiness to be applied in clinical practice. The following recommendations are based on this evidence:

1. With the few exceptions noted below, almost all LBP CPRs are in the derivation stage of development and cannot be recommended to be applied in clinical practice at this time. Clinicians may wish to consider which variables were and were not identified to have a significant

predictive relationship to the target outcome or diagnosis within a derivation study to cautiously inform their clinical practice. Clinicians need to be wary, however, that such relationships may simply reflect chance associations, or may be specific to the unique characteristics of the derivation study's patient sample, clinicians or setting (McGinn et al., 2000; McGinn et al., 2008).

2. The Japanese Society for Spine Surgery and Related Research CPR to assist in the identification of lumbar spinal stenosis (Kato et al., 2009; Konno et al., 2007) (Figure 8.2, p. 295) may be applied in patients with lower extremity symptoms presenting to hospital orthopaedic outpatient clinics and medical centres with some confidence in its known, but limited, predictive performance (McGinn et al., 2000). Based on the validation study data and using a cut-off score of ≥7, the +LR is 1.6 (95%CI 1.3, 2.0) and the -LR is 0.13 (95%CI 0.04, 0.41) for the clinical diagnosis of lumbar spinal stenosis. There is currently no evidence to indicate that application of this CPR produces beneficial clinical consequences, such as improvements in patient outcomes or resource efficiencies.
3. The Berlin criteria comprise a 4-item CPR to assist in the identification of inflammatory back pain (Chan et al., 2012; Rudwaleit et al., 2006; Sieper, van der Heijde, et al., 2009) (Figure 8.2, p. 295). This CPR may be applied in patients with chronic LBP presenting to a rheumatologist with some confidence in its known predictive accuracy for the clinical diagnosis of inflammatory back pain (McGinn et al., 2000). Using data from a multinational validation study (Sieper, van

der Heijde, et al., 2009) and using a cut-off point of ≥2 predictors present, the +LR is approximately 3.8 (95%CI 2.8, 5.0) and the -LR is approximately 0.37 (95%CI 0.31, 0.43). The Berlin criteria have not been validated in patients presenting in primary care, and consequently the predictive performance of this CPR in that setting is not known. There is currently no evidence to indicate that application of this CPR produces beneficial clinical consequences such as improvements in patient outcomes or resource efficiencies.

4. The 'IBP according to experts' CPR assists in the identification of inflammatory back pain (Sieper, van der Heijde, et al., 2009)(Figure 8.2, p. 288). This CPR may be applied in patients with chronic LBP presenting to a rheumatologist with some confidence in its known predictive accuracy for the clinical diagnosis of inflammatory back pain (McGinn et al., 2000). Using data from a multinational validation study (Sieper, van der Heijde, et al., 2009) and using a cut-off point of ≥4 predictors present, the +LR is approximately 2.9 (95%CI 2.3, 3.6) and the -LR is 0.28 (95%CI 0.23, 0.35). The 'IBP according to experts' CPR has not been validated in patients presenting in primary care, and consequently the predictive performance of this CPR in that setting is not known. There is currently no evidence to indicate that application of this CPR produces beneficial clinical consequences, such as improvements in patient outcomes or resource efficiencies.
5. The 'Cassandra rule' (Dionne, 2005; Dionne et al., 1997; Dionne et al., 2011) may be applied in populations of patients with LBP presenting in primary care and emergency departments with some confidence in its

limited prognostic accuracy in identifying patients with differing degrees of risk of developing a poor long-term functional outcome, defined as ≥50% on the Roland Morris Disability Questionnaire at two years post-baseline (McGinn et al., 2000). Using data from the largest and most recent validation study (Dionne et al., 2011), the +LR is 1.50 (95%CI 1.38, 1.64) and the -LR is 0.40 (95%CI 0.25, 0.47) for identifying high/moderate vs low risk patients. There is limited evidence the 'Cassandra rule' is more sensitive, but less specific than physician prediction (Dionne et al., 2011). There is currently no evidence to indicate that application of this CPR produces beneficial clinical consequences, such as improvements in patient outcomes or resource efficiencies.

6. The five item Flynn manipulation CPR (Flynn et al., 2002) may be applied in patients with LBP receiving thrust and non-thrust forms of lumbopelvic manipulation with confidence in its predictive performance in delineating patients with differing degrees of risk of achieving a rapid favourable functional outcome (McGinn et al., 2000). Using data from a high quality validation study involving patients receiving thrust lumbopelvic manipulation (Childs et al., 2004) and using a cut-off point of ≥4 predictors present, the +LR is 13.2 (95%CI 3.4, 52.1) for identifying those patients who will improve by ≥50% on the Oswestry Disability Questionnaire by one week post-baseline. The current body of evidence is limited as to whether the CPR can be applied to preferentially select which patients with LBP should be offered lumbopelvic manipulation. There is currently no evidence to indicate

that application of this CPR produces beneficial clinical consequences such as improvements in patient outcomes or resource efficiencies.

The Clinical Commentary (Chapter 7) provides a resource for clinicians to facilitate the approximation of uncertainty intervals for posterior probabilities in instances where this information has not been provided in the original text. Two methods are advocated in instances where it is appropriate to source prevalence, sensitivity and specific data from a single study:

1. Using Microsoft Excel (2010, Microsoft Corporation, WA, USA), the lower boundary (=beta.inv(0.025, true positive +1, false positive +1)) and upper boundary (=beta.inv(0.975, true positive +1, false positive +1)) of the 95% uncertainty interval may be approximated from data within a 2 x 2 contingency table.
2. Using the statistical freeware R (http://www.r-project.org/), the 95% uncertainty interval may be approximated (binom.test(true positives, true positives + false positives)) from data within a 2 x 2 contingency table.

10.2.2 Research methodology implications

Findings from Studies 2 and 3 enable research methodology recommendations to facilitate the development of CPRs for LBP with the greatest capacity to positively influence physiotherapy practice:

1. Soliciting input from practising clinicians throughout all stages of the development of a CPR may be beneficial to their eventual acceptance and usefulness in clinical practice. This has been previously

recommended in methodological texts (Reilly & Evans, 2006), however, the findings of Studies 1, 4 and 5 suggest that this has seldom been employed in the development of CPRs for LBP.

2. It is likely to be beneficial for researchers to initially assess the perceived need for a CPR to ensure it functions to address a clinically meaningful problem as perceived by the intended clinical consumers. Examples of this include; investigating the magnitude of the clinical problem, measuring clinicians' predictions of the outcome, measuring clinicians' degree of comfort with adopting an alternate practice behaviour (Stiell, Wells, McDowell, et al., 1995), explicitly asking clinicians if they would consider using a proposed CPR (Perry et al., 2009), and asking clinicians to rank the perceived usefulness of a range of hypothetical CPRs (Eagles, Stiell, Clement, Brehaut, Kelly, et al., 2008).
3. From the outset of a CPR's development, it is also likely to be advantageous to explicitly investigate the specific characteristics that the proposed CPR must incorporate in order to be considered meaningful. An example of this is assessing the required predictive precision of a CPR (Perry et al., 2009), which will also help inform the required sample size of the derivation study (Buderer, 1996; S. Jones et al., 2003).
4. Candidate predictor variables should ideally be selected to encompass common clinical assessment techniques and a broad range of management paradigms. They must also have a clear logical relationship to the dependent outcome, which may improve clinician

acceptability of the tool. Further, candidate predictors should ideally be able to be obtained in a timely fashion without the need for sophisticated equipment.

5. The complexity of the clinical interface of the CPR may negatively impact its perceived usefulness in the clinical setting. Thus, consideration must be given to restricting the number of predictor variables, using a clear and easy to apply CPR format (predictor count, algorithm, nomogram etc.), restricting the need for calculations (e.g. variable weightings, regression equations), using graphical aids where appropriate (Björk et al., 2012), and presenting the tool in a manner to optimise its capacity to be memorised (e.g. the use of an acronym as used in the $ABCD^2$ rule (Johnston et al., 2007))
6. It may be preferable to substitute the term 'rule' with a less authoritarian term such as 'tool' or 'guideline' to improve clinician acceptance. Ensuring that CPRs are not perceived as threatening clinician autonomy may be critical to their acceptance in clinical practice, and terms other than 'rule' may highlight the intended function of a CPR to *inform* a clinical decision and not to *form* a clinical decision (Swets et al., 2000a).

Studies 1, 4 and 5 identified opportunities to optimise the methodological quality of CPR development studies. The following recommendations are based on the evidence considered within these studies:

1. The intended function of a CPR should inform the study design, analysis and selection of predictor variables. In the case of

prescriptive CPRs for LBP which are intended to function to help inform treatment selection decisions, CPRs should be derived in RCTs, use tests of interaction and comprise variables identified to be treatment effect modifiers (Hancock, Herbert, et al., 2009; J. C. Hill & Fritz, 2011; Stanton et al., 2010).

2. The sample size of CPR derivation studies needs to be sufficiently large to reduce the risk of overfitting, which has a deleterious effect on the predictive performance of the tool outside of the derivation study data. The common practice of univariate screening of candidate predictor variables for inclusion in a multivariable model may not necessarily reduce the risk of overfitting (Babyak, 2004). It is generally recommended that the study sample should be sufficiently large to ensure at least 10 outcome events per candidate predictor variable (Bouwmeester et al., 2012; Concato et al., 1993; Lewis, 2007; van Oort et al., 2012). In the case of a prescriptive CPR in which interaction effects are assessed, the sample size needs to be approximately four times that required to detect an overall treatment effect of the same magnitude (Brookes et al., 2004).
3. Reference standards need to be measurably valid and reliable. Diagnostic CPRs for sub-presentations of LBP often lack a definitive gold standard, and consequently reference standards need to be reproducibly identifiable and reflective of the broad consensus of a presentation's classification.
4. Candidate predictor variables should be confined to those that have demonstrated reliability (C. Cook et al., 2010). They should further be

logically justifiable and considered within the context of probable predictive performance and practicality (C. Cook et al., 2010; Lubetzky-Vilnai et al., 2014; Seel et al., 2012).

5. Collinearity of predictor variables in regression models requires assessment as correlated variables may give invalid findings and increase the likelihood of developing tools with paradoxical performance, whereby increased positive status on the tool is associated with decreased risk of a target diagnosis or outcome (Lubetzky-Vilnai et al., 2014).
6. Continuous predictor variables should be kept as such, until at least after the multivariable analysis, as their transformation into categorical variables results in poorer-performing models and influences which variables are identified as significantly related to the reference standard (Lubetzky-Vilnai et al., 2014; Schellingerhout et al., 2009).
7. Patient populations in CPR validation studies need to reflect the full range of possible categories for that tool. Studies that seek to include only those patients considered ‘positive’ on a CPR will be unable to evaluate the performance of the tool in delineating patients with differing likelihoods of achieving a target diagnosis or outcome.
8. Validation studies should seek to investigate the inter-observer reliability of a CPR’s application and interpretation, at least within a subset of the study population, as this is a potential threat to its validity (Laupacis et al., 1997; Stiell & Wells, 1999).
9. Throughout all phases of a CPR’s development uncertainty intervals should be reported for outcome prevalence, CPR accuracy (eg.

sensitivity, specificity, likelihood ratios), and for posterior probabilities to facilitate the more meaningful interpretation of a study's findings (C. Cook et al., 2010; Lubetzky-Vilnai et al., 2014).

10.3 Limitations

This subsection will summarise the collective limitations of the studies that comprise this thesis. Each individual study's limitations have been previously detailed in their respective chapters.

The definition of the types of clinical tools that may be classified as CPRs is open to interpretation. This was highlighted in Studies 1 and 2 with evidence of inconsistent use of the term CPR to describe a variety of different tools, within both the academic literature and amongst clinicians. Consequently, a transparent, reproducible and more specific definition of a CPR was developed and applied in Studies 4 and 5 (Table 8.2, p. 267; and Table 9.1, p. 311). The basis of this applied definition was focused upon encompassing the most common use of the term 'CPR' encountered within the literature, and to explicitly differentiate tools that could be reasonably applied by a clinician for an individual patient from other forms of statistical prediction tools and models. Importantly, this definition may not reflect the broader consensus of which tools should be considered CPRs, and has not undergone any form of validation. Contrasting Studies 1, 4, 5 and previous related systematic reviews (Beneciuk et al., 2009; Lubetzky-Vilnai et al., 2014; May & Rosedale, 2009; Patel et al., 2013; Stanton et al., 2010; van

Oort et al., 2012), highlights the significance of variations in the definition of a CPR to the evidence considered within a review. Accordingly, an important limitation of this program of research is that the findings are specific to how a CPR has been operationally defined, which may not necessarily be consistent with how others may choose to define these tools.

A related limitation concerns issues associated with identifying CPR development studies within the medical literature. In the absence of a Medical Subject Heading (MeSH) specific to CPRs, and the varying nomenclature used to described these tools, their identification in the medical literature is complex (Geersing et al., 2012; Holland et al., 2005; Ingui & Rogers, 2001; Keogh et al., 2011; Wong et al., 2003). Highly sensitive electronic search strategies across multiple medical databases were applied in Studies 1, 4 and 5, coupled with hand-searching and citation tracking. The number of studies subsequently included via these supplementary search strategies highlighted the current insufficiency of sole reliance on an electronic search strategy to identify CPR development studies. Therefore, while every reasonable attempt was made to ensure that all relevant studies were included in Studies 1, 4 and 5, it is plausible that some studies may have been inadvertently omitted. This may possibly have implications for the recommendations arising from this research that relate to the appropriateness of CPRs to be applied in clinical practice at this time.

There is currently no validated tool designed to methodologically appraise diagnostic, prescriptive and prognostic CPR development studies at each

stage of their development. Thus, criteria were selected for use in Studies 1, 4 and 5 based on standards commonly reported in well-cited CPR methodological texts (Beattie & Nelson, 2006; Childs & Cleland, 2006; Laupacis et al., 1997; McGinn et al., 2000; Stiell & Wells, 1999) and methodological items identified in recent publications (Bouwmeester et al., 2012; C. Cook et al., 2010; Lubetzky-Vilnai et al., 2014; Seel et al., 2012; van Oort et al., 2012). Many of the criteria have been used in earlier previous systematic reviews of CPRs. Nevertheless, the tools applied in Studies 1, 4 and 5 have not been validated and it would be inappropriate to calculate a sum score for each study or to otherwise quantitatively synthesize the findings of the quality appraisal analysis.

The qualitative research design applied in Studies 2 and 3 was not intended to provide generalizable findings, but to instead provide an in-depth exploration of a range of selected issues concerning CPRs for LBP (Krueger & Casey, 2009). It is not known if the findings arising from Studies 2 and 3 are representative of the clinical population from which participants were selected, or if these findings are generalizable to other clinician populations. Additionally, the sampling strategy of studies 2 and 3 aimed to include the intended clinical consumers of LBP CPRs, and accordingly participants were selected for inclusion irrespective of their baseline knowledge of the topic. It is plausible that participants with greater familiarity with the subject matter may have held differing views to those less knowledgeable.

Finally, a focus group schedule of questions and activities was developed for Studies 2 and 3 which was informed by prior research in this topic. This schedule was not however formally evaluated prior to its implementation. A possible consequence of this limitation was the need to develop and provide study participants in focus groups 3 and 4 with a one-page summary of CPRs with examples (Figure 5.2, p. 191). This was implemented in response to findings from focus groups 1 and 2 that participant knowledge about LBP CPRs was diverse. It was believed that discussion may be better facilitated in focus groups 3 and 4 by amending the focus group schedule to include the provision of a brief standardised summary of LBP CPRs to participants at a set point within the process. Qualitative research is often characterised by the simultaneous collection and analysis of data, thereby enabling researchers to adjust their avenue of investigation to build greater knowledge where opportunities are identified (Krueger & Casey, 2009; Sandelowski, 2000).

10.4 Further research

A number of further research opportunities arising from this program of research are identifiable and are summarised below:

1. Further research in the field of CPRs may benefit from a validated definition that is representative of the broad consensus of which types of tools constitute CPRs. Such a definition may be optimally utilised within a broader classification approach to clinical statistical prediction models/tools that provides clearly defined, mutually exclusive and

collectively exhaustive classifications for each type of tool and model. A validated definition of a CPR may inform the development of a Medical Subject Heading (MeSH) which would greatly facilitate their indexing and retrieval within the medical literature.

2. Validation of the derived CPRs for LBP identified in Studies 1, 4 and 5 is indicated. Validation studies may benefit from focusing on CPRs that function to fulfil an identified unmet clinical need, and whose derivation study is of sufficient methodological quality such that the derived tool may reasonably be expected to generalise outside of the original study data.
3. Validation of the Japanese Society for Spine Surgery and Related Research CPR for lumbar spinal stenosis is indicated in populations of patients with LBP and concomitant lower extremity symptoms presenting to a physiotherapist.
4. Impact analysis of the Japanese Society for Spine Surgery and Related Research CPR for lumbar spinal stenosis is indicated.
5. Validation of the 'Berlin criteria' and 'IBP according to experts' CPRs is indicated in populations of patients presenting in primary care, including those patients with LBP presenting to a physiotherapist.
6. Impact analysis of the 'Berlin criteria' and the 'IBP according to experts' CPRs is indicated.
7. Validation of the 'Cassandra rule' is indicated in patient populations with LBP presenting to a physiotherapist.
8. Impact analysis of the 'Cassandra rule' is indicated.

9. More research is required to investigate the performance of the five item Flynn manipulation CPR as a treatment-effect modifier. The findings of this research will have important implications in regards to the use of the tool to preferentially select patients to receive lumbopelvic manipulation. The predictive performance of the CPR should be further investigated using variations of the active intervention, control intervention and the dependent outcome.
10. Impact analysis of the five item Flynn manipulation is indicated.
11. Validation of the two item Flynn manipulation CPR is indicated using study designs incorporative of patients who are negative and positive on that tool.
12. Development and validation of a standardised tool to appraise the methodological quality of all forms of CPR development studies, at each stage of development, is indicated. The development of such a tool would have important implications regarding the ability to critically appraise the existing body of literature, and to identify opportunities to improve the methodological rigour of future studies.
13. The findings of Study 3 may be used to inform the development of a questionnaire designed to explore physiotherapists' priorities for the development of CPRs for LBP. Generalizable findings arising this research may have important implications regarding identifying which clinical problems CPRs should be aiming to help address, and the modifiable characteristics such tools require to be considered clinically meaningful. Potentially important characteristics may include the

required predictive precision of CPRs for varying functions, and the dependent outcomes considered to have the greatest clinical utility.

14. Research investigating whether the degree of clinician experience or expertise influences the magnitude of benefit derived from the application of a well-developed CPR is indicated. Other characteristics affecting the degree of benefit resulting from the application of a well-developed CPR also require investigation, such that the implementation of such tools may be optimally targeted as scientifically indicated.
15. Research investigating the methods by which well-developed CPRs may be optimally integrated within a clinician's clinical reasoning is indicated. Considerations concerning how, when, and for what and whom CPRs are applied within the clinical encounter may be beneficial.

10.5 Summary of thesis

The identification of meaningful subgroups of patients with LBP is a research priority and CPRs are one of several evidence-based mechanisms proposed to support such sub-classification. The primary objective of this program of research was to undertake a series of studies that may facilitate the development of CPRs with the greatest potential to positively influence the physiotherapy management of LBP. Identifying and assessing the degree to which CPRs for LBP may be confidently applied in clinical practice was considered a primary research aim and three systematic reviews (Studies 1,

4 and 5) were undertaken to address it. The second primary aim concerned exploring the range of factors that may influence the implementation of CPRs for LBP within Australian physiotherapy practice and this was addressed in Study 2. The third primary research aim was addressed in Study 3 and concerned an examination of the areas of perceived need for LBP CPRs in addition to the range of characteristics such tools need to encompass to be considered clinically meaningful and useful within Australian physiotherapy practice. Finally, a Clinical Commentary on the topic of uncertainty intervals for posterior probabilities was produced as an academic and clinical resource.

The evidence synthesised within Studies 1, 4 and 5 identified that a large and growing number of LBP CPRs are under development, however the majority of these tools have not undergone validation and therefore cannot be recommended for direct use in clinical practice at this time. The current lack of impact analysis studies also prevents the assessment of whether the application of LBP CPRs in clinical practice results in beneficial effects on patient outcomes or resource efficiencies. A small number of LBP CPRs have undergone validation, such that clinicians may have some confidence in the predictive accuracy of these tools when applied in similar patient populations and settings. Several opportunities to improve the methodological rigour of future CPR development studies have been identified.

Study 2 identified that physiotherapists' knowledge of LBP CPRs may be quite varied and few participants in that study reported ever using them to inform their clinical decision-making. Potential barriers to the use of LBP CPRs identified in Study 2 included a negative connotation associated with the term 'rule', a perception that CPRs are overly-complex and infrequently applicable, clinical experience obviating the need for such tools, and the potential threat to clinical autonomy and for misuse by third-party payers. Study participants felt that LBP CPRs were best used within the suite of clinical reasoning processes physiotherapists typically employ and considered as second opinions or safety nets that were able to be overruled by the clinician.

The findings of Study 3 highlighted that prognostic forms of CPRs for LBP that function to predict future meaningful outcomes may be welcomed by practising physiotherapists. CPRs that identify likely responders to interventions are likely to be considered useful, as well as diagnostic forms of CPRs that function to identify serious causes of LBP such as fracture and cancer. CPRs that identify which patients are more likely to experience an adverse outcome or to not require physiotherapy intervention may also be welcomed by clinicians. Participants in Study 3 thought that LBP CPRs should be uncomplicated, easy to remember, easy to apply, accurate and precise, and well-supported by research evidence. It was believed that LBP CPRs should not contain an excessive number of variables, use complicated statistics, or contain variables that have no clear logical relationship to the dependent outcome. It was further considered by participants in Study 3 that

LBP CPRs need to be compatible with traditional clinical reasoning and decision-making processes, and sufficiently inclusive of a broad range of management approaches and common clinical assessment techniques.

The findings of this thesis have informed direct clinical recommendations concerning the evidence-based application of LBP CPRs, in addition to recommendations to improve the methodological quality and reporting of future LBP CPR development studies. Further, consideration of the range of views identified in Studies 2 and 3 may inform strategies and future research projects aimed at optimising the development of LBP CPRs with the greatest potential to positively influence physiotherapy practice and implementation strategies that will facilitate the translation of CPR research findings into practice.

APPENDIX 1 Statements of collaboration from co-authors

Statement from Darren A. Rivett relating to papers published with Robin Haskins

I, Darren A. Rivett, attest that Research Higher Degree candidate, Robin Haskins, contributed to the listed publications by contributing to the study conception and design, data acquisition, data analysis and interpretation, and manuscript preparation and revision for publication.

1. Haskins, R., Rivett, D. A., & Osmotherly, P. G. (2012). Clinical prediction rules in the physiotherapy management of low back pain: a systematic review. *Manual Therapy, 17*(1), 9-21.
2. Haskins, R., Osmotherly, P. G., Southgate, E., & Rivett, D. A. (2014). Physiotherapists' knowledge, attitudes and practices regarding clinical prediction rules for low back pain. *Manual Therapy, 19*(2), 142-151.
3. Haskins, R., Osmotherly, P. G., Southgate, E., & Rivett, D. A. (2015). Australian physiotherapists' priorities for the development of clinical prediction rules for low back pain: a qualitative study. *Physiotherapy, 101*(1), 44-49.
4. Haskins, R., Osmotherly, P. G., Tuyl, F., & Rivett, D. A. (2014). Uncertainty in Clinical Prediction Rules: The Value of Credible Intervals. *Journal of Orthopaedic & Sports Physical Therapy, 44*(2), 85-91.
5. Haskins, R., Osmotherly, P. G., & Rivett, D. A. (2015). Diagnostic clinical prediction rules for specific subtypes of low back pain: A

systematic review. *Journal of Orthopaedic & Sports Physical Therapy, 45*(2), 61-76.

6. Haskins, R., Osmotherly, P. G., & Rivett, D. A. (2015). Validation and impact analysis of prognostic clinical prediction rules for low back pain is needed: a systematic review. *Journal of Clinical Epidemiology,* accepted 10th February 2015.

…… Date: 16/02/15

Professor Darren A. Rivett

…… Date: 16/02/15

Robin Haskins

…… Date: 18/02/15

Professor Robert Callister

Assistant Dean Research Training, Faculty of Health and Medicine

Statement from Peter G. Osmotherly relating to papers published with Robin Haskins

I, Peter G. Osmotherly, attest that Research Higher Degree candidate, Robin Haskins, contributed to the listed publications by contributing to the study conception and design, data acquisition, data analysis and interpretation, and manuscript preparation and revision for publication.

1. Haskins, R., Rivett, D. A., & Osmotherly, P. G. (2012). Clinical prediction rules in the physiotherapy management of low back pain: a systematic review. *Manual Therapy, 17*(1), 9-21.
2. Haskins, R., Osmotherly, P. G., Southgate, E., & Rivett, D. A. (2014). Physiotherapists' knowledge, attitudes and practices regarding clinical prediction rules for low back pain. *Manual Therapy, 19*(2), 142-151.
3. Haskins, R., Osmotherly, P. G., Southgate, E., & Rivett, D. A. (2015). Australian physiotherapists' priorities for the development of clinical prediction rules for low back pain: a qualitative study. *Physiotherapy, 101*(1), 44-49.
4. Haskins, R., Osmotherly, P. G., Tuyl, F., & Rivett, D. A. (2014). Uncertainty in Clinical Prediction Rules: The Value of Credible Intervals. *Journal of Orthopaedic & Sports Physical Therapy, 44*(2), 85-91.
5. Haskins, R., Osmotherly, P. G., & Rivett, D. A. (2015). Diagnostic clinical prediction rules for specific subtypes of low back pain: A

systematic review. *Journal of Orthopaedic & Sports Physical Therapy, 45*(2), 61-76.

6. Haskins, R., Osmotherly, P. G., & Rivett, D. A. (2015). Validation and impact analysis of prognostic clinical prediction rules for low back pain is needed: a systematic review. *Journal of Clinical Epidemiology,* accepted 10th February 2015.

..... Date:

Dr Peter G. Osmotherly

..... Date:

Robin Haskins

..... Date:

Professor Robert Callister

Assistant Dean Research Training, Faculty of Health and Medicine

Statement from Erica Southgate relating to papers published with Robin Haskins

I, Erica Southgate, attest that Research Higher Degree candidate, Robin Haskins, contributed to the listed publications by contributing to the study conception and design, data acquisition, data analysis and interpretation, and manuscript preparation and revision for publication.

1. Haskins, R., Osmotherly, P. G., Southgate, E., & Rivett, D. A. (2014). Physiotherapists' knowledge, attitudes and practices regarding clinical prediction rules for low back pain. *Manual Therapy, 19*(2), 142-151.
2. Haskins, R., Osmotherly, P. G., Southgate, E., & Rivett, D. A. (2015). Australian physiotherapists' priorities for the development of clinical prediction rules for low back pain: a qualitative study. *Physiotherapy, 101*(1), 44-49.

..... Date:

Dr Erica Southgate

..... Date:

Robin Haskins

..... Date:

Professor Robert Callister

Assistant Dean Research Training, Faculty of Health and Medicine

Statement from Frank Tuyl relating to papers published with Robin Haskins

I, Frank Tuyl, attest that Research Higher Degree candidate, Robin Haskins, contributed to the listed publication by contributing to the clinical commentary conception and design, and manuscript preparation and revision for publication.

1. Haskins, R., Osmotherly, P. G., Tuyl, F., & Rivett, D. A. (2014). Uncertainty in Clinical Prediction Rules: The Value of Credible Intervals. *Journal of Orthopaedic & Sports Physical Therapy, 44*(2), 85-91.

.... Date:

Dr Frank Tuyl

..... Date:

Robin Haskins

..... Date:

Professor Robert Callister

Assistant Dean Research Training, Faculty of Health and Medicine

APPENDIX 2 Ethical approval for studies presented in CHAPTER 5 and CHAPTER 6

HUMAN RESEARCH ETHICS COMMITTEE

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Notification of Expedited Approval

To Chief Investigator or Project Supervisor:	**Professor Darren Rivett**
Cc Co-investigators / Research Students:	**Mr Peter Osmotherly** **Doctor Erica Southgate** **Mr Robin Haskins**
Re Protocol:	**Physiotherapists' knowledge, attitudes, beliefs and practices toward low back pain clinical prediction rules**
Date:	**15-Mar-2011**
Reference No:	**H-2011-0032**
Date of Initial Approval:	**15-Mar-2011**

Thank you for your **Response to Conditional Approval (minor amendments)** submission to the Human Research Ethics Committee (HREC) seeking approval in relation to the above protocol.

Your submission was considered under **Expedited** review by the Ethics Administrator.

I am pleased to advise that the decision on your submission is **Approved** effective **15-Mar-2011**.

In approving this protocol, the Human Research Ethics Committee (HREC) is of the opinion that the project complies with the provisions contained in the National Statement on Ethical Conduct in Human Research, 2007, and the requirements within this University relating to human research.

Approval will remain valid subject to the submission, and satisfactory assessment, of annual progress reports. *If the approval of an External HREC has been "noted" the approval period is as determined by that HREC.*

The full Committee will be asked to ratify this decision at its next scheduled meeting. A formal *Certificate of Approval* will be available upon request. Your approval number is **H-2011-0032**.

If the research requires the use of an Information Statement, ensure this number is inserted at the relevant point in the Complaints paragraph prior to

distribution to potential participants You may then proceed with the research.

Conditions of Approval

This approval has been granted subject to you complying with the requirements for *Monitoring of Progress*, *Reporting of Adverse Events*, and *Variations to the Approved Protocol* as detailed below.

PLEASE NOTE:
In the case where the HREC has "noted" the approval of an External HREC, progress reports and reports of adverse events are to be submitted to the External HREC only. In the case of Variations to the approved protocol, or a Renewal of approval, you will apply to the External HREC for approval in the first instance and then Register that approval with the University's HREC.

- ***Monitoring of Progress***

Other than above, the University is obliged to monitor the progress of research projects involving human participants to ensure that they are conducted according to the protocol as approved by the HREC. A progress report is required on an annual basis. Continuation of your HREC approval for this project is conditional upon receipt, and satisfactory assessment, of annual progress reports. You will be advised when a report is due.

- ***Reporting of Adverse Events***

1. It is the responsibility of the person **first named on this Approval Advice** to report adverse events.
2. Adverse events, however minor, must be recorded by the investigator as observed by the investigator or as volunteered by a participant in the research. Full details are to be documented, whether or not the investigator, or his/her deputies, consider the event to be related to the research substance or procedure.
3. Serious or unforeseen adverse events that occur during the research or within six (6) months of completion of the research, must be reported by the person first named on the Approval Advice to the (HREC) by way of the Adverse Event Report form within 72 hours of the occurrence of the event or the investigator receiving advice of the event.
4. Serious adverse events are defined as:
    - Causing death, life threatening or serious disability.
    - Causing or prolonging hospitalisation.
    - Overdoses, cancers, congenital abnormalities, tissue damage, whether or not they are judged to be caused by the investigational agent or procedure.
    - Causing psycho-social and/or financial harm. This covers everything from perceived invasion of privacy, breach of confidentiality, or the diminution of social reputation, to the creation of psychological fears and trauma.

- Any other event which might affect the continued ethical acceptability of the project.

5. Reports of adverse events must include:
   - Participant's study identification number;
   - date of birth;
   - date of entry into the study;
   - treatment arm (if applicable);
   - date of event;
   - details of event;
   - the investigator's opinion as to whether the event is related to the research procedures; and
   - action taken in response to the event.

6. Adverse events which do not fall within the definition of serious or unexpected, including those reported from other sites involved in the research, are to be reported in detail at the time of the annual progress report to the HREC.

- ***Variations to approved protocol***

If you wish to change, or deviate from, the approved protocol, you will need to submit an *Application for Variation to Approved Human Research*. Variations may include, but are not limited to, changes or additions to investigators, study design, study population, number of participants, methods of recruitment, or participant information/consent documentation. **Variations must be approved by the (HREC) before they are implemented** except when Registering an approval of a variation from an external HREC which has been designated the lead HREC, in which case you may proceed as soon as you receive an acknowledgement of your Registration.

Linkage of ethics approval to a new Grant

HREC approvals cannot be assigned to a new grant or award (ie those that were not identified on the application for ethics approval) without confirmation of the approval from the Human Research Ethics Officer on behalf of the HREC.

Best wishes for a successful project.

Professor Alison Ferguson
Chair, Human Research Ethics Committee

For communications and enquiries:
Human Research Ethics Administration

Research Services
Research Integrity Unit
HA148, Hunter Building
The University of Newcastle
Callaghan NSW 2308
T +61 2 492 18999
F +61 2 492 17164
Human-Ethics@newcastle.edu.au

HUMAN RESEARCH ETHICS COMMITTEE

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Notification of Expedited Approval

To Chief Investigator or Project Supervisor:	**Professor Darren Rivett**
Cc Co-investigators / Research Students:	**Mr Peter Osmotherly** **Doctor Erica Southgate** **Mr Robin Haskins**
Re Protocol:	**Physiotherapists' knowledge, attitudes, beliefs and practices toward low back pain clinical prediction rules**
Date:	**01-Feb-2012**
Reference No:	**H-2011-0032**

Thank you for your **Variation** submission to the Human Research Ethics Committee (HREC) seeking approval in relation to a variation to the above protocol.

Variation to:

1. Extend the period for data collection to 30 June 2012.

2. Amend the eligibility criteria *from* "currently working within the Newcastle or Central Coast regions of NSW" *to* "currently working within the Hunter, Central Coast, New England or Mid North Coast regions of NSW".

3. Introduce a new recruitment method. This will involve the inclusion of a recruitment notice in a professional email bulletin sent to members of the Australian Physiotherapy Association.

- Information Statement (v3, dated 15/12/2011)
- Initial Contact Cover Letter (v3, dated 15/12/2011)
- Recruitment Notice (v1, dated 15/12/2011)

Your submission was considered under **Expedited** review by the Chair/Deputy Chair.

I am pleased to advise that the decision on your submission is **Approved** effective **30-Jan-2012**.

The full Committee will be asked to ratify this decision at its next scheduled meeting. A formal *Certificate of Approval* will be available upon request.

Dr Jean Harkins
Acting Chair, Human Research Ethics Committee

For communications and enquiries:
Human Research Ethics Administration

Research Services
Research Integrity Unit
HA148, Hunter Building
The University of Newcastle
Callaghan NSW 2308
T +61 2 492 18999
F +61 2 492 17164
Human-Ethics@newcastle.edu.au

APPENDIX 3 Information statement and consent form for studies presented in CHAPTER 5 and CHAPTER 6

FACULTY OF HEALTH

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Professor Darren A. Rivett
Head, School of Health Sciences
The University of Newcastle
Phone: (02) 4921 7220
E-mail: Darren.Rivett@newcastle.edu.au

Information Statement for the Research Project:
Physiotherapists' knowledge, attitudes, beliefs and practices toward low back pain clinical prediction rules.
Document Version 2: dated 3/3/11

You are invited to participate in the research project identified above which is being conducted by Mr Robin Haskins as part of his PhD studies under the supervision of Professor Darren Rivett and Mr Peter Osmotherly from The University of Newcastle.

Why is the research being done?
The purpose of the research is to evaluate the knowledge, attitudes, beliefs and practices of Physiotherapists regarding clinical prediction rules for low back pain. The results will provide direction in the development of clinical prediction rules that have the greatest potential to positively influence clinical practice.

Who can participate in the research?
We are seeking musculoskeletal Physiotherapists in the Hunter and Central Coast regions who manage patients with low back pain and have an awareness of clinical prediction rules.

What choice do you have?
Participation in this research is entirely your choice. Only those people who give their informed consent will be included in the project. A follow up telephone call in one week will be made to confirm your participation. Whether or not you decide to participate, your decision will not disadvantage you.

What would you be asked to do?
If you agree to participate, you will be asked to attend a focus group which will last between 1.5 and 2 hours. Focus groups will be held on The University of Newcastle's campuses in Newcastle and on the Central Coast. Participants will be invited to share their knowledge, attitudes, beliefs and practices regarding low back pain clinical prediction rules. Within the focus group session, participants will also be asked to provide comments upon a fictitious case scenario. The focus group will be tape recorded and the discussion transcribed, however your confidentiality will be protected by the research team as all identifying details including names, places and organisations will be replaced by pseudonyms in the final

transcript and in publications. If you agree to participate, you will have the right to withdraw from the study at any time including during and after the focus group.

How much time will it take?
The focus group will last between 1.5 and 2 hours.

What are the risks and benefits of participating?
There is no identifiable benefit or risk associated with your decision to participate.

How will your privacy be protected?
If you decide to participate, your confidentiality will be protected by the research team. All potentially identifying details including names, places and organisations will be replaced by pseudonyms in the final transcript and in any resultant publications. Only the research team will have access to any data that identifies you. All audio and written transcripts from the focus groups will be stored for 5 years in a locked cabinet at The University of Newcastle. Electronic data will be stored in a password-protected electronic database.

How will the information collected be used?
The information collected in this study will be presented in a thesis to be submitted for Mr Haskins' PhD degree, in addition to published papers in scientific journals and at professional conferences. Individual participants will not be identified in any reports arising from the project. Participants will be offered a summary of the results upon completion of the study.

What do you need to do to participate?
Please read this Information Statement and be sure you understand its contents before you consent to participate. If there is anything you do not understand, or you have questions, please contact a member of the research team.
A member of the research team will contact you by telephone in one week. Whether or not you choose to participate, you may notify the researchers to inform them of your preference not to receive this follow-up phone call.

Further information
If you would like further information please contact Professor Darren Rivett at The University of Newcastle on (02) 49217220 or Darren.Rivett@newcastle.edu.au.

Thank you for considering this invitation.

FACULTY OF HEALTH

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Professor Darren A. Rivett
Chief Investigator
Head, School of Health Sciences
The University of Newcastle

Mr Peter G. Osmotherly
Lecturer in Physiotherapy
School of Health Sciences
The University of Newcastle

Mr Robin Haskins
Student Researcher
School of Health Sciences
The University of Newcastle

Dr Erica Southgate
Lecturer in Education
School of Education
The University of Newcastle

Complaints about this research

This project has been approved by the University's Human Research Ethics Committee, Approval No. H-2011-0032.

Should you have concerns about your rights as a participant in this research, or you have a complaint about the manner in which the research is conducted, it may be given to the researcher, or, if an independent person is preferred, to the Human Research Ethics Officer, Research Office, The Chancellery, The University of Newcastle, University Drive, Callaghan NSW 2308, Australia, telephone (02) 49216333, email Human-Ethics@newcastle.edu.au.

FACULTY OF HEALTH

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Professor Darren A. Rivett
Head, School of Health Sciences
The University of Newcastle
Phone: (02) 4921 7220
E-mail: Darren.Rivett@newcastle.edu.au

Consent Form for the Research Project:
Physiotherapists' knowledge, attitudes, beliefs and practices toward low back pain clinical prediction rules.
version 2: dated 3/03/11

I agree to participate in the above research project and give my consent freely.

I understand that the project will be conducted as described in the Information Statement, a copy of which I have retained.

I understand I can withdraw from the project at any time and do not have to give any reason for withdrawing.

I consent to

- participate in a focus group
- providing comments upon a fictitious case scenario
- being recorded within the focus group for the purposes of transcription

I understand that my personal information will remain confidential to the researchers.

I have had the opportunity to have questions answered to my satisfaction.

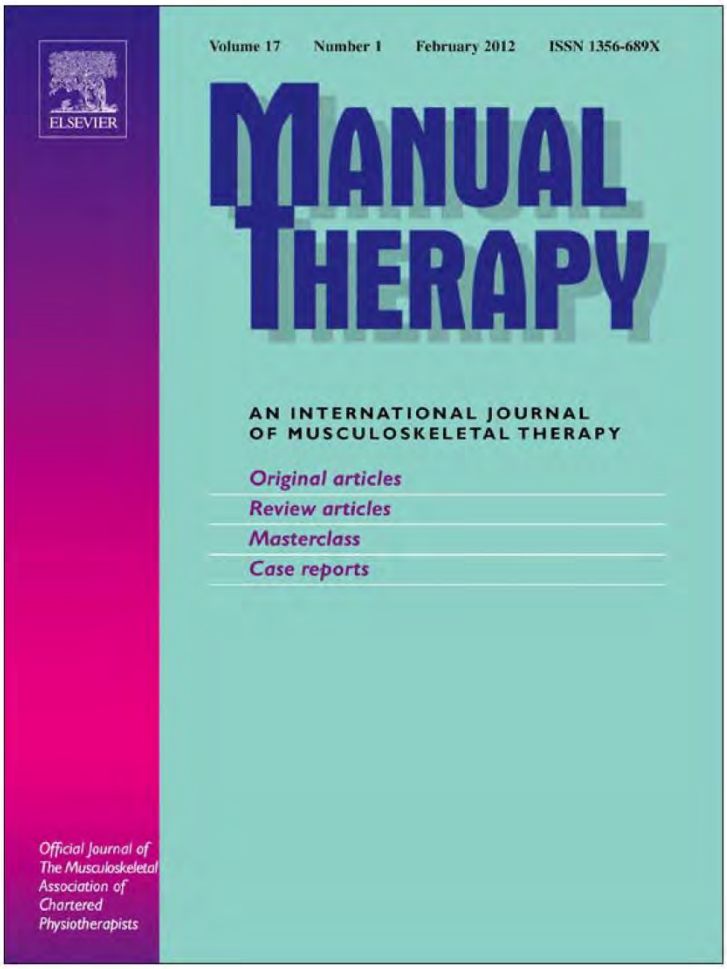
Print Name: _____

Signature: _____ **Date:** _____

APPENDIX 4 Published papers

Study 1 - Chapter 4 (p.127)

Haskins, R., Rivett, D. A., & Osmotherly, P. G. (2012). Clinical prediction rules in the physiotherapy management of low back pain: a systematic review. Manual Therapy, 17(1), 9-21.



Manual Therapy 17 (2012) 9–21

Contents lists available at ScienceDirect

Manual Therapy

journal homepage: www.elsevier.com/math

Systematic review

Clinical prediction rules in the physiotherapy management of low back pain: A systematic review

Robin Haskins*, Darren A. Rivett, Peter G. Osmotherly

School of Health Sciences, The University of Newcastle, NSW 2308, Australia

ARTICLE INFO

Article history:
Received 16 January 2011
Received in revised form
28 April 2011
Accepted 9 May 2011

Keywords:
Low back pain
Physical Therapy (Specialty)
Decision making
Probability

ABSTRACT

Objective: To identify, appraise and determine the clinical readiness of diagnostic, prescriptive and prognostic Clinical Prediction Rules (CPRs) in the physiotherapy management of Low Back Pain (LBP).
Data sources: MEDLINE, EMBASE, CINAHL, AMED and the Cochrane Database of Systematic Reviews were searched from 1990 to January 2010 using sensitive search strategies for identifying CPR and LBP studies. Citation tracking and hand-searching of relevant journals were used as supplemental strategies.
Study selection: Two independent reviewers used a two-phase selection procedure to identify studies that explicitly aimed to develop one or more CPRs involving the physiotherapy management of LBP. Diagnostic, prescriptive and prognostic studies investigating CPRs at any stage of their development, derivation, validation, or impact-analysis, were considered for inclusion using a priori criteria. 7453 unique records were screened with 23 studies composing the final included sample.
Data extraction: Two reviewers independently extracted relevant data into evidence tables using a standardised instrument.
Data synthesis: Identified studies were qualitatively synthesized. No attempt was made to statistically pool the results of individual studies. The 23 scientifically admissible studies described the development of 25 unique CPRs, including 15 diagnostic, 7 prescriptive and 3 prognostic rules. The majority (65%) of studies described the initial derivation of one or more CPRs. No studies investigating the impact phase of rule development were identified.
Conclusions: The current body of evidence does not enable confident direct clinical application of any of the identified CPRs. Further validation studies utilizing appropriate research designs and rigorous methodology are required to determine the performance and generalizability of the derived CPRs to other patient populations, clinicians and clinical settings.

1. Introduction

A Clinical Prediction Rule (CPR) is "a clinical tool that quantifies the individual contributions that various components of the history, physical examination and basic laboratory results make towards the diagnosis, prognosis, or likely response to treatment in an individual patient" (McGinn et al., 2008). These tools aim to facilitate clinical decision-making in the assessment and treatment of individual patients (Beattie and Nelson, 2006) and are thought to be of greatest potential when they are developed and utilised for clinical conditions that involve complex clinical decision making.

Low Back Pain (LBP) is a common and costly complaint (Riihimaki, 1996; Andersson, 1998; Walker, 1999) that has been specifically identified as an ideal target for CPRs due to its heterogeneous population and numerous treatment alternatives (Fritz, 2009). Clinical trials (Fritz et al., 2003; Long et al., 2004; Brennan et al., 2006) have highlighted the benefits of LBP classification systems that aim to 'match' interventions according to the particular sub-group of patients. Concordantly, there has been a surge in the number of publications that discuss the development and application of CPRs that are relevant to the assessment and treatment of LBP (Beneciuk et al., 2009; May and Rosedale, 2009; Stanton et al., 2010). However, before a CPR can be confidently incorporated into clinical practice, it must undergo a process of development that investigates the rule's performance, generalizability, and influence upon clinical outcomes and/or resource consumption.

Numerous publications have discussed the common methodological standards that should apply to the development of CPRs (Wasson et al., 1985; Laupacis et al., 1997; Randolph et al., 1998; Stiell and Wells, 1999; McGinn et al., 2000, 2008; Beattie and

Abbreviations: CPR, clinical prediction rule; LBP, low back pain; QUADCPR, quality checklist for prescriptive derivation-based clinical prediction rules.

* Corresponding author. Tel.: +61 2 4921 7069; fax: +61 2 4921 7053.
E-mail address: Robin.Haskins@newcastle.edu.au (R. Haskins).

1356-689X/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.math.2011.05.001

Nelson, 2006; Childs and Cleland, 2006; Cook, 2008), although the specific criteria often differ between studies. It is, however, commonly accepted that a hierarchical process of rule development is utilised (McGinn et al., 2000), initially commencing with derivation of the rule, and then progressing to a process of validation and then subsequent investigation of its clinical impact.

CPRs that have been derived, but not yet validated are not considered ready for clinical use (McGinn et al., 2000, 2008; Reilly and Evans, 2006). Even rigorously derived rules may reflect chance associations between variables and the target condition or outcome, or they may be unique to the studied population or other characteristics of that clinical setting (McGinn et al., 2008). This is reflected in the finding that most CPRs perform less accurately in subsequent studies involving different patients (Toll et al., 2008). Despite these limitations, it has been suggested that derived CPRs may inform clinical practice by providing clinicians with an understanding of some of the most important predictors of a given target condition or outcome (McGinn et al., 2008).

The process of validation investigates a rule's performance and generalizability to other patient populations, clinicians and clinical settings. Importantly, the validation of a CPR cannot be accomplished by a single study, but requires a process involving a series of studies that test the internal and external validity of the rule across a broad range of clinical environments (Hancock et al., 2009a). Narrow validation of a CPR involves investigating the performance of the rule in a similar patient population and similar clinical setting to the derivation study. A CPR that has been demonstrated to perform well in such a setting is considered to be ready for cautious clinical application to patients that are representative of the studied population (McGinn et al., 2000, 2008).

Confidence in the rule's accuracy improves as it is progressively investigated in various other settings comprising different clinicians and patients with differing prevalence of disease or injury and with differing responsiveness to treatment. CPRs that demonstrate consistent and strong performance in this process of broad validation are considered ready to be applied in clinical practice with confidence in their accuracy (McGinn et al., 2000).

It is not appropriate, however, to assume that the clinical application of a rigorously-validated rule will result in improved clinical care. Impact-analysis is the process of CPR development that involves testing a rule's ability to positively influence clinical outcomes and/or resource consumption, and change clinicians' behaviour (McGinn et al., 2008). Ideally, this involves a direct comparison to usual clinical care or judgement (Toll et al., 2008). Rules that are demonstrated to be highly accurate and perform well across multiple clinical environments may actually be no more accurate, or even worse, than unassisted clinician judgement. Rigorously-validated CPRs that have been demonstrated to produce beneficial clinical consequences via impact-analysis can be confidently incorporated into clinical practice (McGinn et al., 2000, 2008; Reilly and Evans, 2006).

Before clinicians can consider incorporating the growing number of CPRs into their practice, a determination of their readiness for clinical application is required. Previous systematic reviews of CPRs relevant to physiotherapy (Beneciuk et al., 2009; May and Rosedale, 2009; Stanton et al., 2010) have focused upon the identification of prescriptive rules that facilitate treatment decision-making by identifying variables that moderate the magnitude of the treatment-effect. These reviews have specifically excluded studies concerning diagnosis and prognosis, thereby preventing a complete assessment of the available CPRs a physiotherapist may consider in their clinical management of LBP. Further, the quality appraisal systems used in these reviews have not been reflective of the consensus of the common methodological standards for CPR development.

As no universally-accepted standardised tool currently exists for the methodological appraisal of studies of CPRs (Fritz, 2009), previous systematic reviews have used a variety of means to evaluate the quality of included studies. Some reviews have utilised standardised tools that were developed to appraise prognostic (Beneciuk et al., 2009) and diagnostic studies (Bachmann et al., 2004; Hess et al., 2008). Criticism in this approach has focused upon recognising that methodological standards for the development of CPRs differ to that of other types of studies (Stanton et al., 2009). Other reviews (Wisnivesky et al., 2005; Dahri and Loewen, 2007; May and Rosedale, 2009; Stanton et al., 2010) have developed checklists based upon previously proposed methodological standards. A potential problem with this approach is that the proposed methodological standards differ between texts, leading to the possible inclusion of extraneous criteria or the possible exclusion of important criteria dependent upon the text(s) selected. For example, although Stiell and Wells (1999) highlight the importance of a representative sample in the derivation phase of a rule's development, this criterion is omitted from other well-cited texts (Laupacis et al., 1997; McGinn et al., 2000).

The aim of the present review was to identify, appraise and determine the clinical readiness of CPRs in the physiotherapy management of LBP.

2. Methods

2.1. Data sources and searches

A systematic literature search of MEDLINE, EMBASE, CINAHL, AMED and the Cochrane Database of Systematic Reviews from 1990 to January 2010 limited to articles available in English was conducted. A sensitive search strategy for CPRs (Ingui and Rogers, 2001) that has been used in previous systematic reviews (Dahri and Loewen, 2007; Beneciuk et al., 2009; May and Rosedale, 2009) was employed in combination with the search strategy recommended by the Cochrane Back Group (2009) for identifying articles relevant to LBP (Appendix 1). Citation tracking and hand-searching of relevant journals were used as supplemental search strategies.

2.2. Study selection

For a study describing the development of a CPR to be included in the review it had to meet the following criteria:

1. Studies needed to explicitly aim to develop one or more CPRs involving the physiotherapy management of LBP. The operational definition of a CPR for this study was that defined by McGinn et al. (2008). Although it has been suggested that there should be a minimum of three variables in a CPR (Laupacis et al., 1997; Stiell and Wells, 1999), previous systematic reviews (Tamariz et al., 2004; Wisnivesky et al., 2005) have included studies with two or more predictor variables. To ensure all relevant studies were identified, this review used the more liberal definition of a CPR as that containing two or more predictor variables.
2. Substantial practice variation between low back pain treatment providers (Kent and Keating, 2005; Werner et al., 2005) including marked differences in the methods chosen to assess this condition (Kent et al., 2009) makes it arguably inappropriate to assume that the selection and assessment of potential predictor variables will generalise across disciplines. Thus, it was determined a priori that for a study to be included, the assessment of potential predictor variables was required to be performed by a physiotherapist to ensure their direct relevance to the primary research aim.

3. Consistent with the definition of a CPR employed in this review (McGinn et al., 2008), predictor variables were required to be independently meaningful.
4. Diagnostic, prescriptive and prognostic studies investigating CPRs at any stage of their development (McGinn et al., 2000), derivation, validation, or impact-analysis, were included.

No restriction was placed upon the type of potential predictor variables (eg. history items, imaging modalities, physical examination items, psychological variables etc) under investigation in the studies considered for inclusion. Further, no restriction was placed upon the clinical setting or the type of patients with LBP under investigation in studies considered for eligibility in this review.

Identified studies were downloaded into an electronic reference management system (EndNote, version X2.0.1[1]) and duplicates were removed.

Two reviewers performed the first-stage screening of titles and abstracts based upon the stated eligibility criteria. Any study denoted eligible by either reviewer was progressed to the second-stage of eligibility screening. Additionally, studies identified by citation tracking and hand-searching of relevant journals were progressed to the second-stage. The full-text of included studies was obtained and examined by two reviewers. During this second-stage of screening, concordance between reviewers determined inclusion, with disagreements resolved by consensus, or if needed by a third reviewer.

2.3. Data extraction and quality assessment

A standardised instrument was used for data extraction. Information collected from each study included the country of origin, the number of rules developed, study design, stated objective, and details of the patient population. The reviewers also investigated whether included studies specifically used the term "clinical prediction rule". The hierarchy of evidence for CPRs (McGinn et al., 2000) was initially employed to determine which stage of CPR development an article was describing. Studies were subsequently defined as derivation, validation or impact-analysis.

Consistent with the aim of the present review, the quality of the included studies were evaluated against the well-cited methodological standards that are employed by researchers in the development of all forms of CPRs. These criteria reflect the necessary methodological requirements to develop any form of a CPR and should be considered as an extension to the various methodological requisites that are specific to the underlying study design. In the absence of an appropriate standardised tool and to avoid the limitations of unsystematically selecting criteria from previous reports, we initially identified the key texts describing the methodological standards common to the development of all forms of CPRs, including those used in previous systematic reviews. From these texts, five (Laupacis et al., 1997; Stiell and Wells, 1999; McGinn et al., 2000; Beattie and Nelson, 2006; Childs and Cleland, 2006) were selected based upon their inclusion in previous reviews, their number of citations in MEDLINE and EMBASE and their relevance to the research aim. Criteria that were represented in two or more of the five selected texts were included in the methodological appraisal of the included studies. This review employed definitions of the accepted CPR quality criteria that have been previously published (Laupacis et al., 1997; Stiell and Wells, 1999; McGinn et al., 2000; Beattie and Nelson, 2006; Childs and Cleland, 2006). A checklist was subsequently developed for each of the three phases of rule development. The research designs of the included studies were anticipated to be extensively heterogeneous ranging from randomised controlled and observational intervention studies, to cross-sectional diagnostic investigations and longitudinal prognostic studies. Consequently, no attempt was made to appraise and contrast the included studies against the methodological standards that are specific to their unique underlying research design.

Two reviewers independently appraised the methodological quality of the included studies. Each criterion was evaluated independently with concordance between examiners determining the appropriate outcome. Disagreement was resolved by consensus and if needed, by a third reviewer. For a criterion to be marked as being met, studies must have entirely fulfilled the requirements of that criterion with no occasions of disparity. For example, in studies that aimed to develop two or more CPRs, all rules within the study must have achieved the requirements of that criterion for it to be considered met. Criteria marked as 'unclear' or 'not met' were consolidated to enable the dichotomisation of each criterion as 'met' or 'not met'.

The research design of studies investigating predictors of responsiveness to intervention were specifically evaluated for their ability to identify treatment-effect modifiers. These variables, also known as 'moderators', are the baseline characteristics that identifies subgroups of patients with differing treatment effect-sizes for a given intervention (Kraemer et al., 2002, 2006; Turner et al., 2007; MacKinnon and Luecken, 2008; Kraemer and Gibbons, 2009). Recent commentary in the rehabilitation literature (Hancock et al., 2009a) has highlighted the inadequacy of single-arm research designs in identifying the variables that influence a patient's responsiveness to an intervention. Controlled trials are required in all stages of prescriptive CPR development to discriminate between the non-specific prognostic factors associated with clinical outcome, and the specific treatment-effect modifying variables that help further guide clinical decision making. The distinction between single-arm prescriptive CPR studies and prognostic CPR studies was determined by the stated clinical aim of the CPR in each study.

2.4. Data synthesis and analysis

Due to the anticipated heterogeneity of the included studies, no attempt was made to statistically pool the results of individual studies.

Between-rater agreement was evaluated for each stage of the screening process and for the methodological appraisal of the included studies. The absolute and chance-corrected degrees of agreement (κ) with 95% confidence intervals were calculated for both stages of the screening procedure. Between group comparisons were analysed following exploratory data analysis and relevant parametric or non-parametric tests were applied. All statistical analyses were conducted using Stata 11.0[2].

3. Results

3.1. Study selection

The database search strategy yielded 10 202 studies. Another twelve studies were identified via hand-searching of relevant journals and citation-tracking of included studies. Following the removal of duplicate records, 7453 records were screened via title

[1] EndNote version X2.0.1, Thomson Reuters, 2141 Palomar Airport Road, Suite 350, Carlsbad, CA 92011, USA.

[2] Stata version 11.0, StataCorp LP, 4905 Lakeway Drive, College Station, Texas 77845, USA.

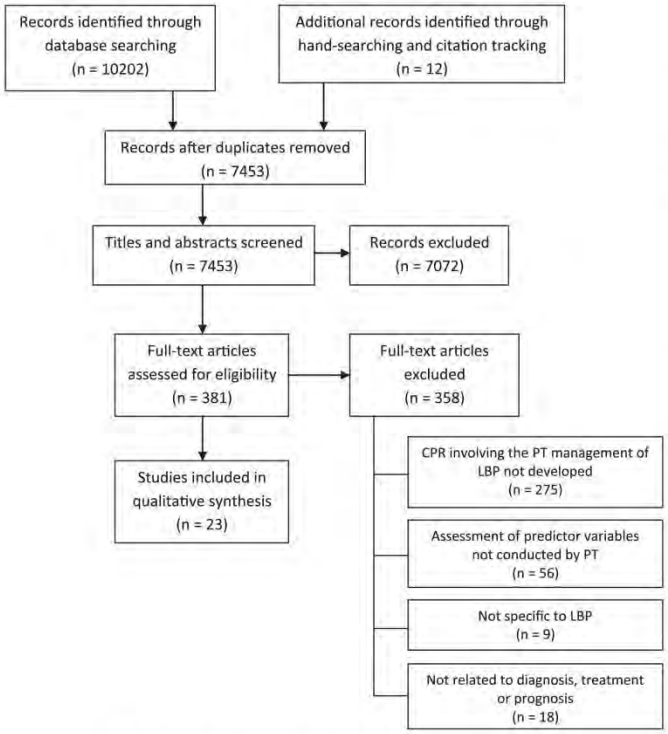


Fig. 1. Flow chart of search strategy and study selection.

and abstract with 381 records progressing to the second stage of screening. The full-text copies of these studies were located and reviewed with 23 studies composing the final included sample. The reasons for exclusion are highlighted in Fig. 1.

The absolute agreement between raters for the first and second-round screening procedures was 96.6% and 94% respectively. The chance-corrected degree of agreement was observed to be "moderate" (Sackett et al., 1991) for both procedures with $\kappa = 0.49$ (95% CI 0.43–0.55) for the screening by titles and abstracts, and $\kappa = 0.53$ (95% CI 0.35–0.72) for the screening by full-text. All but one episode of disagreement between raters was resolved by consensus, with the remaining study ruled to be included by the third reviewer.

3.2. Characteristics of included studies

The majority of included studies ($n = 15$) originated from the USA. Three studies were conducted in Australia and two in The Netherlands. The remaining three studies were conducted in Singapore, Spain and the United Kingdom. Although the search strategy enabled the inclusion of studies from 1990, the earliest year of publication of the included sample was 2002. The majority of included studies developed just one CPR, although some studies investigated up to five rules in one publication.

Fifteen derivation and eight validation studies compose the included sample. No studies investigating the impact phase of rule development were identified. Fourteen studies describe CPRs used to influence treatment decision-making. Ten (43%) of the included studies relate to the prediction of clinical outcome with the use of spinal manipulation. Seven studies concern diagnosis and only two prognostic studies were included. Across the 23 included publications, 25 unique CPRs are described including 15 diagnostic, 7 prescriptive and 3 prognostic rules. Appendices 2a, 2b and 2c detail the identified CPRs and the relevant studies that have contributed to their development.

3.3. Qualitative appraisal of included studies

Quality scoring for the derivation and validation studies is provided in Tables 1 and 2 respectively. "Substantial" (Sackett et al., 1991) between-rater agreement was observed for the quality scoring with an absolute degree of agreement of 88.7% ($\kappa = 0.74$, 95% CI 0.66–0.81). Three episodes of disagreement required resolution by a third reviewer, with the remaining disagreements being resolved by consensus.

Five of the 14 publications (36%) concerning prescriptive CPRs used a randomised controlled-study design that would permit the identification of treatment-effect modifiers.

Although all included studies satisfied the operational definition of a CPR, not all articles specifically used the term. Of the 23 included studies, only 15 (65%) explicitly used the term "clinical prediction rule" when describing the clinical tool being developed. It was more common for prescriptive studies to use the term "clinical prediction rule", compared to diagnostic and prognostic studies ($p < 0.001$).

4. Discussion

There has been a rapid growth in the number of studies reporting upon the development of CPRs in the physiotherapy literature. This trend mirrors that seen in Medicine, particularly in the fields of Emergency and Intensive Care and may be reflective of a progressive move towards models of clinical decision-making that are increasingly data-driven and firmly founded upon the process of scientific enquiry. The quest to identify meaningful sub-groups of patients will have important implications for clinical practice, particularly for presentations, such as LBP, which are confounded by their degree of heterogeneity and numerous treatment alternatives.

To our knowledge, the present review is the first to systematically locate, appraise and determine the clinical readiness of diagnostic, prescriptive and prognostic CPRs involving the physiotherapy management of LBP in all phases of their development. Twenty-five unique CPRs were identified encompassing a diverse range of factors. While the growth in this research is arguably important for LBP treatment providers, this observed large variation in CPR themes may reflect the current lack of understanding of clinicians' priorities for CPRs. Investigation of the areas of perceived clinical need for CPRs would facilitate the development of rules with the greatest potential to positively influence clinical practice (Eagles et al., 2008).

Previous systematic reviews of CPRs in the physical rehabilitation literature (Beneciuk et al., 2009; May and Rosedale, 2009; Stanton et al., 2010) have included four studies involving the physiotherapy management of LBP which were excluded in the present review. Two studies (Fritz et al., 2004, 2007) included in earlier reviews have investigated the characteristics that are associated with treatment outcomes. However, as both studies did not develop a clinical tool that may be applied to an individual patient they did not meet the present review's eligibility criteria. One excluded study (Brennan et al., 2006) was determined to have investigated a classification system while the other excluded study (Teyhen et al., 2007) was limited to describing the arthrokinematic characteristics of a sub-group that were positive on a previously derived CPR.

4.1. Summary of evidence

Based upon the findings of the present review, the available evidence does not support the direct clinical application of any of the identified CPRs for LBP at this time. Of the 25 unique CPRs

Table 1
Derivation study quality appraisal.

	Diagnostic									Prescriptive				Prognostic	
	Fritz et al., 2005b	Henschke et al., 2009	Laslett et al., 2006a	Laslett et al., 2005	Laslett et al., 2006b	Laslett et al., 2003	van der Wurff et al., 2006	Alonso-Blanco et al., 2009	Cai et al., 2009	Flynn et al., 2002	Fritz et al., 2005a	Hicks et al., 2005	May et al., 2008	George et al., 2005	Hancock et al., 2009b
Prospective design	✓	✓	✓	✓	✓[a]	✓	✓	✓	✓	✓	✓[a]	✓	✓[a]	✓[a]	✓
Outcomes defined	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Outcome clinical important	✓	✓	✓	No[b]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Blinded outcome assessment	✓	No	No	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	No
All important predictors included	No	✓	No	No	No	No	No	No	No	✓	No	No	No	✓	✓
Predictive variables clearly defined	✓	No	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Blinded predictor assessment	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Assessment of the reliability of the predictive variables	✓	No	No	✓	No	✓	✓	✓	✓	✓	✓	✓	✓	✓	No
Important patient characteristics described	✓	✓	✓	✓	✓	No	✓	✓	✓	✓	✓	✓	No	✓	✓
Representative sample	No	✓	No	No	No	No	No	✓	✓	✓	✓	✓	✓	No	✓
Study site described	No	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Justification for the number of study subjects	No	✓	No	No	No	No	No	No	No	No	No	No	No	No	No
≥10 outcome events per independent variable in the rule	✓	No	✓	No[c]	No	No[c]	No[c]	No	No	No	No[c]	No[c]	No	No[d]	✓
Mathematical techniques described	✓	No	✓	✓	No	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Multivariable analysis	✓	✓	✓	No	✓	No	No	✓	✓	✓	No	No[c]	✓	✓	✓
Results of the rule described	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Clinically sensible/ reasonable	No	✓	No	✓	✓	✓	✓	No	No	✓	✓	✓	No	✓	✓
Easy to use	✓	✓	No[d]	✓	No[d]	No[d]	✓	✓	✓	✓	✓	✓	✓	✓	✓
Probability of diagnosis or outcome described	✓	✓	No	No	✓	No	✓	✓	✓	✓	✓	✓	✓	No	✓
Course of action described	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
Specifically uses the term "Clinical prediction rule"	No	No	No	No	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes

[a] Secondary analysis of prospectively derived data.
[b] Single diagnostic injection not consistent with current SIJ diagnostic criterion standard (Szadek et al., 2009).
[c] Multivariable regression not performed for all prediction rules.
[d] More than one rule presented. Not all rules satisfy criterion.

Table 2
Validation study quality appraisal.

	Prescriptive							
	Childs et al., 2003	Childs et al., 2004	Childs et al., 2006	Cleland et al., 2006	Cleland et al., 2009	Fritz et al., 2006	Hancock et al., 2008b	Hallegraeff et al., 2009
Prospective validation in new patient population	No[a]	✓	✓	✓	✓	No	✓	✓
Different clinical setting to derivation study	No[a]	No	No	✓	✓	✓	✓	✓
Different clinicians to derivation study	No[a]	No[a]	No[a]	✓	✓	✓	✓	✓
Representative sample	No	✓	✓	No	✓	No	✓	No
The rule is applied accurately	No	✓	✓	✓	✓	✓	✓	No
Complete follow-up	✓	✓	✓	✓	✓	✓	✓	✓
Accuracy of the rule in the validation study sample described	No	✓	No[b]	No	No	No	✓	No
Assessment of the inter-observer reliability of the rule	No	No	No	No	No	No	No[a]	No
Specifically uses the term "Clinical Prediction Rule"	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes

[a] Unclear. Insufficient information.
[b] Absence of target outcome in sub-group preventing appropriate statistical analysis.

identified, only two have progressed to the process of validation and no rule has been investigated for its ability to positively influence clinical outcomes and/or resource consumption.

The 5-item spinal manipulation CPR derived by Flynn et al. (2002) in a single-arm study design is one of the CPRs that has been further investigated in a series of validation studies. Recent commentary in the literature (Allison, 2009; Hancock et al., 2009a; Cook et al., 2010) and in two Physical Therapy podcasts (Fritz et al., 2009a, 2009b) have discussed the limitations of single-arm study designs in the development of prescriptive CPRs. The lack of a control group enables the identification of non-specific prognostic variables but is unable to investigate the moderators of treatment-effect. Controlled-study designs utilizing tests of interactions are required to identify on whom and under what circumstances treatments produce different outcomes (Kraemer et al., 2002; Hancock et al., 2009a). Accordingly, it has been suggested that the subsequent study undertaken by Childs et al. (2004) is most appropriately considered a derivation study and not a validation study. This is because it was the first controlled-study that enabled the investigation of the CPR as a treatment response modifier, in contrast to a non-specific prognostic factor (Hancock et al., 2009a).

Of the remaining validation studies that have aimed to develop the 5-item spinal manipulation CPR in new cohorts of patient populations, only two (Hancock et al., 2008b; Cleland et al., 2009) have used a controlled-study design. Cleland et al. (2009) aimed to examine the generalizability of the CPR to different thrust and non-thrust manipulative techniques. The generalizability of a CPR to other procedures is most appropriately determined by controlled-study designs that investigate if a patient's status on the rule significantly moderates the effect-size of an intervention (Assmann et al., 2000; Kraemer et al., 2002, 2006; Turner et al., 2007; MacKinnon and Luecken, 2008). However, as the patient population in this study were all positive on the spinal manipulation CPR, the performance of the rule in identifying those with a difference in treatment responsiveness remained untested. Finally, in the well-designed validation study by Hancock et al. (2008b), the spinal manipulation CPR was found to perform no better than chance in identifying patients likely to respond to this intervention. Positive status on the rule, however, was found to be a non-specific prognostic factor. One of the many possible explanations for the observed findings noted by these researchers (Hancock et al., 2008a, 2008b) and others (Hebert and Perle, 2008) is the difference in treatment provided in this study compared to the original derivation studies (Flynn et al., 2002; Childs et al., 2004), with high-velocity thrust manipulative techniques only being used on a very small proportion of the patients in this study.

The 2-item pragmatic spinal manipulation CPR derived by Fritz et al. (2005a) was based upon the collated results of two previous studies (Flynn et al., 2002; Childs et al., 2004) used to develop the 5-item rule. This abbreviated form of the spinal manipulation CPR was found to strongly identify those patients with a good outcome following treatment. However, as no control group was included in the derivation, the variables may represent prognostic factors that may have no specific relationship with the intervention provided. Two subsequent studies (Fritz et al., 2006; Hallegraeff et al., 2009) attempting to validate this rule restricted their patient populations to only those that were positive on the pragmatic spinal manipulation CPR. As previously noted, without the inclusion of patients that are also negative on the rule, a prescriptive CPR's performance is unable to be rigorously investigated. Consequently, the body of evidence does not yet enable confidence in the direct clinical application of either the 5-item or 2-item spinal manipulation CPRs in identifying subgroups of patients with differences in responsiveness to this intervention.

The 23 rules that have been derived, but not yet proceeded to validation may inform clinical practice by providing clinicians with an understanding of some of the most important predictors of a given target condition or outcome (McGinn et al., 2008). However, even in this limited application clinicians must exercise due caution as predictor variables may simply reflect chance associations or unique characteristics of the studied population or setting. Further, prescriptive predictor variables identified through single-arm study designs may not identify the relevant features that modify the effect of a given intervention, but instead reflect non-specific prognostic factors (Hancock et al., 2009a).

It has been argued that the biologic plausibility of predictor variables be carefully considered throughout the derivation of a CPR to minimise the likelihood of including factors that reflect chance associations with the target outcome (Childs and Cleland, 2006; Fritz et al., 2009b; Raney et al., 2009). However, the primary function of a CPR is to accurately predict a target outcome and not to identify the determinants of that outcome. The composite of factors that together accurately predict a given outcome are of most value, regardless of whether this relationship is confounded by other variables (Katz, 2006). To illustrate this point, consider that although carrying a cigarette lighter will not cause lung cancer, it may accurately predict a greater likelihood of developing the disease (Katz, 2006). Excluding predictive variables that are not believed at the time to be causally related to the target outcome may result in the development of CPRs with inferior predictive accuracy. Consequently, the process of rigorous validation of derived CPRs is the most suitable method to identify and

exclude those variables that previously reflected chance associations with the target outcome (McGinn et al., 2008).

4.2. Methodological quality

Substantial variation was observed in the methodological quality of the fifteen included derivation studies. In addition to the previously mentioned research-design limitations of many prescriptive CPR studies, other common methodological shortcomings included the omission of important predictor variables, not providing a justification for the sample size and not including an appropriate number of outcome events per independent predictor when performing multivariable regression analysis.

Including the most probable predictor variables in the investigation aims to ensure that important relevant factors are not omitted (Laupacis et al., 1997). However, this needs to be balanced with restricting the analysis to a pre-determined small number of variables, ideally for only one outcome, to reduce the likelihood of eliciting findings that are due to chance and random error (Assmann et al., 2000). Researchers should consider examining the results of secondary-analyses of randomised controlled trials and the findings of single-arm treatment studies to help guide the selection of variables (Fritz et al., 2009a).

Only one of the included derivation studies explicitly justified the size of the studied population. Larger sample sizes enable more precise estimates of a rule's predictive power, which in turn enhances confidence in its clinical application (Childs and Cleland, 2006; McGinn et al., 2008). A further consideration is that the investigation of treatment-effect modifiers in prescriptive CPRs requires much larger sample sizes in comparison to identifying main effects between treatment groups. Simulation studies have demonstrated that a study with an 80% power of detecting a given overall effect would require four times the number of subjects to maintain this power in detecting an interaction effect of the same magnitude (Brookes et al., 2004).

Researchers developing CPRs need to carefully consider the prevalence of the target outcome or condition when determining the sample size to ensure that there is a sufficient number of outcome events to satisfy the assumptions implicit to the statistical analysis. Seventy percent of the included derivation studies that used multivariable regression analysis did not have an adequate number of outcome events per independent variable in the model. Guidelines for the development of multivariable logistic regression and Cox proportional hazard models advocate a minimum of ten outcome events per independent variable to reduce the likelihood of identifying erroneous associations and to improve the precision of the findings (Concato et al., 1993). For multiple linear regression, it is recommended that there should be at least ten patients for every variable selected (Lewis, 2007).

Similar to the variance observed in the derivation studies, the methodological quality for the eight included validation studies varied substantially. No validation study included in this review investigated the inter-observer reliability of the CPR. Guidelines on the validation of CPRs have recommended that researchers examine the inter-observer reliability of the rule, at least within a subset of the study population, to ensure consistency in the interpretation of a patient's status on the rule (Laupacis et al., 1997; Stiell and Wells, 1999).

4.3. Study limitations

The search strategy employed in this review has been demonstrated to have high sensitivity for the detection of CPR studies (Ingui and Rogers, 2001) and has been used in other systematic reviews (Dahri and Loewen, 2007; Beneciuk et al., 2009; May and Rosedale, 2009). However, due to inconsistent nomenclature used to describe these clinical tools, it is plausible that not all potentially eligible studies were identified.

The primary aim of this review was the identification and appraisal of CPRs in the physiotherapy management of LBP. Due to substantial between-discipline practice differences in the assessment of LBP (Kent et al., 2009), it was determined a priori that for a study to be included, the assessment of potential predictor variables was required to be performed by a physiotherapist. This eligibility criterion resulted in the exclusion of studies that had developed CPRs using other LBP treatment providers for the assessment of predictor variables. While outside the scope of the present review, the value and validity of such CPRs for physiotherapy practice arguably merits investigation.

The sensitive operational definition of a CPR used in this review enabled the inclusion of studies that may not have explicitly used the term "clinical prediction rule". Consequently, the methodological standards that would be considered by researchers explicitly aiming to develop a CPR may not have been considered in the design of these other studies. As the quality appraisal tool used in this review reflects these well-cited standards for CPR development, it is perhaps not surprising that a large variation of quality was observed between those studies that did and did not explicitly use the term "clinical prediction rule".

The methodological appraisal tool used in this review was developed via a systematic process that aimed to minimise bias in the selection of appropriate quality criteria. While we believe this approach represents an improvement upon that used in previous systematic reviews of CPRs, our checklist has not been formally validated, and consequently the results need to interpreted with caution. The degree of between-rater agreement was high for the majority of the quality criteria, however, it is clear that some variables particularly those relating to the appraisal of validation studies would benefit from measures to further improve rater concordance. An important consideration is that the quality criteria used in this review reflects the well-cited methodological standards that are common to diagnostic, prescriptive and prognostic forms of CPRs. Although this approach appropriately reflects the primary aim of this review and enables a qualitative comparison of the included studies, it is acknowledged that the omission of appraisal criteria that are specific to the development of each particular form of CPR may represent a potential limitation of the present study. Recently, a quality checklist for prescriptive derivation-based CPRs (the QUADCPR) has been developed using Delphi methods (Cook et al., 2010). While this checklist will require further investigation of its reliability and validity, and is not advocated for the retrospective appraisal of CPR studies, it constitutes an important contribution in providing clear methodological guidelines for developing future studies aiming to derive prescriptive rules.

5. Conclusions

This review is the first to systematically locate, appraise and determine the clinical readiness of diagnostic, prescriptive and prognostic CPRs involving the physiotherapy management of LBP in all phases of their development. Twenty-five unique rules were identified across fifteen derivation and eight validation studies. No impact studies were located. The current body of evidence does not enable confident direct clinical application of any of the identified CPRs. Further validation studies utilizing appropriate research designs and rigorous methodology are required to determine the performance and generalizability of the derived CPRs to other patient populations, clinicians and clinical settings.

Acknowledgements

We thank Dane P Fehlberg for his assistance with screening and appraising the literature.

Appendix 1. Database search strategies

MEDLINE, EMBASE and the Cochrane Database of Systematic Reviews.

1	Validat$.mp. or Predict$.ti. or Rule$.mp.
2	(Predict$ and (Outcome$ or risk$ or model$)).mp.
3	((History or Variable$ or Criteria or Scor$ or Characteristic$ or Finding$ or Factor$) and (Predict$ or Model$ or Decision$ or Identif$ or Prognos$)).mp.
4	Decision$.mp. and ((Model$ or Clinical$).mp. or Logistic Models/)
5	(Prognostic and (History or Variable$ or Criteria or Scor$ or Characteristic$ or Finding$ or Factor$ or Model$)).mp.
6	4 or 1 or 3 or 2 or 5
7	dorsalgia.ti,ab.
8	exp Back Pain/
9	backache.ti,ab.
10	exp Low Back Pain/
11	(lumbar adj pain).ti,ab.
12	coccyx.ti,ab.
13	coccydynia.ti,ab.
14	sciatica.ti,ab.
15	sciatica/
16	spondylosis.ti,ab.
17	lumbago.ti,ab.
18	11 or 7 or 9 or 17 or 12 or 15 or 14 or 8 or 16 or 10 or 13
19	6 and 18
20	("1990" or "1991" or "1992" or "1993" or "1994" or "1995" or "1996" or "1997" or "1998" or "1999" or "2000" or "2001" or "2002" or "2003" or "2004" or "2005" or "2006" or "2007" or "2008" or "2009").yr.
21	19 and 20

CINAHL, AMED.

S22	S13 and S21
S21	S12 and S20
S20	S15 or S16 or S17 or S18 or S19
S19	prognostic and (history or variable* or criteria or scor* or characteristic* or finding* or factor* or model*)
S18	decision* and (model* or clinical* or mh Logistic Models)
S17	(history or variable* or criteria or scor* or characteristic* or finding* or factor*) and (predict* or model* or decision* or identif* or prognos*)
S16	predict* and (outcome* or risk* or model*)
S15	Validat* or ti Predict* or Rule*
S14	S12 and S13
S13	yr 1990 or yr 1991 or yr 1992 or yr 1993 or yr 1994 or yr 1995 or yr 1996 or yr 1997 or yr 1998 or yr 1999 or yr 2000 or yr 2001 or yr 2002 or yr 2003 or yr 2004 or yr 2005 or yr 2006 or yr 2007 or yr 2008 or yr 2009
S12	S1 or S2 or S3 or S4 or S5 or S6 or S7 or S8 or S9 or S10 or S11
S11	ti lumbago or ab lumbago
S10	ti spondylosis or ab spondylosis
S9	mh sciatica
S8	ti sciatica or ab sciatica
S7	ti coccydynia or ab coccydynia
S6	ti coccyx or ab coccyx
S5	ti (lumbar n0 pain) or ab (lumbar n0 pain)
S4	mh Low Back Pain+
S3	ti backache or ab backache
S2	mh Back Pain+
S1	ti dorsalgia or ab dorsalgia

Appendix 2a. Diagnostic clinical prediction rules included in qualitative synthesis.

CPR	Variables	Publication	Stage of rule development	Sample	Results/outcome
Radiographic instability	Lumbar Flexion > 53°, lack of hypomobility with intervertebral motion testing (2 variables)	Fritz et al., 2005b	Derivation	n = 49, LBP +/− leg pain, referred for imaging on suspicion of instability, mean 39.2 years old, 57% female, median 78 days of symptoms, 57% prevalence of target condition.	If 2 variables positive, $+LR^a$ = 12.8 (95% CI 0.79–211.6). If 1 variable positive, +LR = 4.3 (95% CI 1.8–10.6).
Diskogenic pain CPR1	CP[b], PPE[c], VABLE[d], Ext Losse (4 variables)	Laslett et al., 2006a	Derivation	n = 216, LBP +/− leg pain, referred to specialist diagnostic centre, mean 44.2 years old, 43% female, mean 158 weeks of symptoms, 35% prevalence of target condition. Only 107 patients received reference standard.	If 1 or more variables positive, then +LR = 1.9 (95% CI 1.1–3.2) and $-LR^f$ = 0.37 (95% CI 0.21–0.65). If 2 variables positive, then +LR = 6.7 (95% CI 0.95–50) and −LR = 0.73(0.61–0.97)
Diskogenic pain CPR2	No CP, PPE, VABLE, Ext Loss (4 variables)	Laslett et al., 2006a	Derivation	n = 216, LBP +/− leg pain, referred to specialist diagnostic centre, mean 44.2 years old, 43% female, mean 158 weeks of symptoms, 35% prevalence of target condition. Only 107 patients received reference standard.	If 2 variables positive, then sensitivity = 37% (95% CI 24–50) and specificity = 100% (95% CI 82–100). LR's not calculated due to 100% specificity.
Diskogenic pain CPR3	PPE, VABLE, Ext Loss (3 variables)	Laslett et al., 2006a	Derivation	n = 216, LBP +/− leg pain, referred to specialist diagnostic centre, mean 44.2 years old, 43% female, mean 158 weeks of symptoms, 35% prevalence of target condition. Only 107 patients received reference standard.	If 2 variables positive, then +LR = 6.5 (95% CI 0.9–46.3) and −LR = 0.77 (95% CI 0.66–0.9).
SIJ mediated pain CPR1	Distraction, Compression, Thigh thrust, Gaenslen's (right), Gaenslen's (left), Sacral Thrust (6 variables)	Laslett et al., 2005	Derivation	n = 48, buttock pain +/− LBP +/− leg pain, referred to specialist diagnostic centre with suspicion of SIJ pain, mean 42.1 years old, 67% female, mean 32 months of symptoms, 33% prevalence of target condition.	If 3 or more variables positive, then +LR = 4.29 (95% CI 2.34–8.58) and −LR = 0.8 (95% CI 0.14–0.37)
SIJ mediated pain CPR2	Distraction, Thigh Thrust, Compression, Sacral Thrust (4 variables)	Laslett et al., 2005	Derivation	n = 48, buttock pain +/− LBP +/− leg pain, referred to specialist diagnostic centre with suspicion of SIJ pain, mean 42.1 years old, 67% female, mean 32 months of symptoms, 33% prevalence of target condition.	If 2 positives, then +LR = 4 (95% CI 2.13–8.08) and −LR = 0.16 (95% CI 0.04–0.47)

SIJ mediated pain CPR3	Distraction, Thigh Thrust, Gaenslen's test, Compression, Sacral Thrust (5 variables)	Laslett et al., 2003	Derivation	$n = 43$ (subset of patients from Laslett et al., 2005 using different reference standard), buttock pain +/− LBP +/− leg pain, referred to specialist diagnostic centre with suspicion of SIJ pain, insufficient data to report precise demographic details, 26% prevalence of target condition.	If 3 or more positives, then +LR = 4.16 (95% CI 2.16−8.39) and −LR = 0.12 (95% CI 0.02−0.49).
SIJ mediated pain CPR4	No CP/peripheralisation, Distraction, Thigh Thrust, Gaenslen's test, Compression, Sacral Thrust (6 variables)	Laslett et al., 2003	Derivation	$n = 34$ (subset of patients from Laslett et al., 2005 using different reference standard), buttock pain +/− LBP +/− leg pain, referred to specialist diagnostic centre with suspicion of SIJ pain, insufficient data to report precise demographic details, 32% prevalence of target condition.	If no CP/periphalisation and if 3 or more positives of remaining variables, then +LR = 6.97 (95% CI 2.7−20.27) and −LR = 0.11 (95% CI 0.02−0.44)
SIJ mediated pain CPR5	Distraction, Compression, Thigh Thrust, Patrick sign, Gaenslen's test (5 variables)	van der Wurff et al., 2006	Derivation	$n = 60$, buttock pain +/− leg pain, referred for invasive procedures, mean 51 years old, 78% female, mean 98 months of symptoms, 45% prevalence of target condition.	If 3 or more positives, then +LR = 4.02 (95% CI 2.04−7.89) and −LR 0.19 (95% CI 0.07−0.47)
Z-jt mediated pain CPR1	Age ≥ 50, symptoms best walking, symptoms best sitting, onset pain is paraspinal, MSPQ[g] > 13, ext/rot test[h], no CP (7 variables)	Laslett et al., 2006b	Derivation	$n = 120$, LBP +/− leg pain, referred to specialist diagnostic centre with suspicion of z-jt pain, mean 43 years old, 46% female, mean 158 weeks of symptoms, 11% prevalence of target condition.	If 4 or more positives, then +LR = 7.6 (95% CI 4.5−13.7) and −LR = 0.0 (95% CI 0.0−0.35)
Z-jt mediated pain CPR2	Age ≥ 50, symptoms best walking, symptoms best sitting, onset pain is paraspinal, MSPQ > 13, ext/rot test (6 variables)	Laslett et al., 2006b	Derivation	$n = 120$, LBP +/− leg pain, referred to specialist diagnostic centre with suspicion of z-jt pain, mean 43 years old, 46% female, mean 158 weeks of symptoms, 11% prevalence of target condition.	If 2 or more positives, then +LR = 1.6 (95% CI 1.5−1.8) and −LR = 0.0 (95% CI 0.0−0.69).
Z-jt mediated pain CPR3	Age ≥ 50, symptoms best walking, symptoms best sitting, onset pain is paraspinal, MSPQ > 13 (5 variables)	Laslett et al., 2006b	Derivation	$n = 120$, LBP +/− leg pain, referred to specialist diagnostic centre with suspicion of z-jt pain, mean 43 years old, 46% female, mean 158 weeks of symptoms, 11% prevalence of target condition.	If 1 or more positives, then +LR = 1.4 (95% CI 1.3−1.5) and −LR = 0.0 (95% CI 0.0−0.95).
Z-jt mediated pain CPR4	Age ≥ 50, symptoms best walking, symptoms best sitting, onset pain is paraspinal, ext/rot test (5 variables)	Laslett et al., 2006b	Derivation	$n = 120$, LBP +/− leg pain, referred to specialist diagnostic centre with suspicion of z-jt pain, mean 43 years old, 46% female, mean 158 weeks of symptoms, 11% prevalence of target condition.	If 2 or more positives, then +LR = 2.0 (95% CI 1.8−2.5) and −LR = 0.0 (95% CI 0.0−0.49).
Z-jt mediated pain CPR5	Age ≥ 50, symptoms best walking, symptoms best sitting, onset pain is paraspinal, ext/rot test (5 variables)	Laslett et al., 2006b	Derivation	$n = 120$, LBP +/− leg pain, referred to specialist diagnostic centre with suspicion of z-jt pain, mean 43 years old, 46% female, mean 158 weeks of symptoms, 11% prevalence of target condition.	If 3 or more positives, then +LR = 9.7 (95% CI 5.0−18.8) and −LR = 0.17 (95% CI 0.05−0.6).
Vertebral fracture	Female sex, age > 70, significant trauma, prolonged use of corticosteroids (4 variables)[i]	Henschke et al., 2009	Derivation	$n = 1172$, acute LBP +/− leg pain patients presenting to a primary care provider, mean 44 years old, 47% female, 59% had duration of less than one week, 0.7% prevalence of target condition.	If 2 or more positives, then +LR = 15.5 (95% CI 7.2−24.6). If 3 or more positives, then +LR = 218.3(95% CI 45.6−953.8).

[a] +LR = positive likelihood ratio.
[b] CP = centralization phenomenon.
[c] PPE = persistent low back pain between episodes of acute low back pain.
[d] VABLE = subjective report of 'vulnerability' when in the semi-stooped position or when performing twisting actions.
[e] Ext Loss = visual estimation of moderate or major loss of lumbar extension range of movement.
[f] −LR − negative likelihood ratio.
[g] MSPQ = Modified somatic perception questionnaire.
[h] Ext/Rot test = Extension/Rotation test.
[i] Predictor variables not exclusively assessed by physiotherapists. Physiotherapists = 72.6%, general practitioners = 22.8%, chiropractors = 4.6%.

Appendix 2b. Prescriptive clinical prediction rules included in qualitative synthesis.

CPR	Variables	Publication	Stage of rule development	Sample	Results/outcome	Methodological notes
Spinal manipulation	Duration of symptoms < 16 days, FABQ-W[a] < 19, at least 1 hip with >35° IR ROM[b], hypomobility with lumbar spring testing, no symptoms distal to knee (5 variables)	Flynn et al., 2002	Derivation	$n = 71$, LBP +/− leg pain, baseline ODQ[c] score ≥ 30%, referred to physiotherapy, mean 37.6 years old, 41% female, mean 42 days of symptoms, 45% prevalence of target outcome.	If 4 or more positives, then +LRd = 24.38 (95% CI 4.63 −139.41)	Single-arm design. Therefore unable to identify treatment-effect modifiers.
		Childs et al., 2003	Validation	$n = 2$ (case reports), 54 and 26 year old males, LBP and buttock pain respectively. One patient met 5 CPR criteria, the other patient met just 1 (or 2) criteria.	Only the patient with all 5 criteria positive experienced dramatic improvement in pain and disability following manipulation.	Research design prevents identification of treatment-effect modifiers.
		Childs et al., 2004	Validation	$n = 131$ (RCT), LBP +/− leg pain, baseline ODQ score ≥ 30%, referred to physiotherapy, mean 33.9 years old, 42% female, median 27 days of symptoms, 29% prevalence of target outcome at 1/52 and 50% at 4/52.	Significant 3 way-interaction between CPR status (≥4/5 = positive), Rx-group and time for pain and disability. For dichotomized outcome (success/failure) the interaction between CPR status and Rx-group strongly predicted success. For patients receiving manipulation, CPR positive status had +LR = 13.2 (95% CI 3.4−52.1). For patients CPR positive the NNT with manipulation = 1.3 (95% CI 1.1−1.9)	RCT. Therefore treatment-effect modifiers able to be identified.
		Childs et al., 2006	Validation	$n = 131$ (RCT), LBP +/− leg pain, baseline ODQ score ≥ 30%, referred to physiotherapy, mean 33.9 years old, 42% female, median 27 days of symptoms.	Aimed to investigate if CPR status is predictive of a worsening in disability. No patient that was CPR positive and received manipulation worsened, preventing appropriate statistical analysis.	Secondary analysis of 2004 RCT. Therefore treatment-effect modifiers able to be identified.
		Cleland et al., 2006	Validation	$n = 12$ (case series), LBP, ODQ score ≥ 30%, referred to physiotherapy, all CPR positive (≥4/5 = positive), mean 39 years old, 42% female, median 19 days of symptoms.	Aimed to investigate generalizability of CPR status to another high-velocity thrust manipulation procedure. 11 out of 12 patients (92%) achieved the target outcome of 'success' at 1/52 following intervention.	All patients CPR positive, therefore unable to determine rule performance. Research design prevents identification of treatment-effect modifiers.
		Hancock et al., 2008b	Validation	$n = 239$ (RCT), LBP < 6/52 duration, presenting to general practitioner, mean 40.7 years old, 44% female, mean 9 days of symptoms.	Non-significant 3-way interaction between Rx-group, CPR status (≥4/5 = positive) and time for pain ($p = 0.805$) and disability ($p = 0.6$). Patients that were CPR positive had better pain and disability outcomes independent of treatment group.	RCT. Therefore treatment-effect modifiers able to be identified. Spinal manipulative technique differed to derivation study. Only 5% of sample received high-velocity thrust manipulation.
		Cleland et al., 2009	Validation	$n = 112$ (RCT), LBP +/− leg pain, attending an outpatient physiotherapy clinic, modified ODQ baseline score >25%, all CPR positive (≥4/5 = positive), mean 40.3 years old, 52% female, median 45 days of symptoms.	Aimed to investigate the generalizability of CPR to another high-velocity thrust manipulation procedure and a non-thrust manipulative technique. No difference between the 2 high-velocity thrust procedures in pain and disability at any time point. Outcomes poorer in the non-thrust group.	All patients CPR positive, therefore unable to determine rule performance. RCT. Therefore treatment-effect modifiers able to be identified.

Spinal manipulation – pragmatic rule	Duration of symptoms < 16 days, no symptoms distal to knee (2 variables)	Fritz et al., 2005a	Derivation	$n = 141$ (data from 2 previous studies (Flynn et al., 2002; Childs et al., 2004)), LBP +/− leg pain, baseline ODQ score ≥ 30%, referred to physiotherapy, mean 35.5 years old, 49% female, median 22 days of symptoms, 45% prevalence of target outcome.	If both criteria positive, then +LR = 7.2 (95% CI 3.2–16.1).	Single-arm design. Therefore unable to identify treatment-effect modifiers.
		Fritz et al., 2006	Validation	$n = 215$ (retrospective review of clinical database), occupational LBP, receiving Rx in outpatient physiotherapy clinic, all CPR positive (2/2 = positive), mean 35.9 years old, 32% female, mean 5.3 days of symptoms.	66.5% received manipulation (49.8% thrust and 16.7% non-thrust). Patients receiving manipulation experienced greater reductions in pain and disability with treatment, compared to those not receiving manipulation.	Research design prevents the identification of treatment-effect modifiers. All patients CPR positive, therefore unable to determine rule performance.
		Hallegraeff et al., 2009	Validation	$n = 64$ (RCT), acute LBP, all CPR positive (2/2 = positive), mean 39 years old, 45% female, 31% had symptoms less than 1/52.	Significant interaction for disability at 2.5 weeks between CPR status (including the additional criterion of age > 35 years) and Rx-group. No significant interactions for pain or lumbar spinal mobility.	All patients CPR positive (by derivation study criteria), therefore unable to determine rule performance. RCT. Therefore treatment-effect modifiers able to be identified. Analysis performed with the additional CPR criterion of age >35 years.
Lumbar traction	FABQ-W < 21, no neurological deficit, age > 30, non-manual work job status (4 variables)	Cai et al., 2009	Derivation	$n = 129$, diagnosis related to the lumbosacral spine +/− leg pain, referred from orthopaedics to physiotherapy, mean 30.9 years old, 16% female, mean 40 weeks of symptoms, 19% prevalence of target outcome.	If 3 or more positives, then +LR = 3.04 (95% CI 2.04–4.53). If all 4 positive, then +LR = 9.36 (95% CI 3.13–28.0).	Single-arm design. Therefore unable to identify treatment-effect modifiers.
Stabilisation exercise – success	Age < 40 years, average SLR[e] > 91°, aberrant movement present, positive prone instability test (4 variables)	Hicks et al., 2005	Derivation	$n = 54$, LBP +/− leg pain, referred to outpatient physiotherapy clinics, mean 42.4 years old, 57% female, mean 41 days of symptoms, 33% prevalence of target outcome (success).	If 3 or more positives, then +LR = 4.0 (95% CI 1.6–10.0). If 2 or more positives, then +LR = 1.9 (95% CI 1.2–2.9).	Single-arm design. Therefore unable to identify treatment-effect modifiers.
Stabilisation exercise – failure	Prone instability test, aberrant movement, hypermobility, FABQ physical activity subscale > 8 (4 variables)	Hicks et al., 2005	Derivation	$n = 54$, LBP +/− leg pain, referred to outpatient physiotherapy clinics, mean 42.4 years old, 57% female, mean 41 days of symptoms, 72% prevalence of target outcome (not failure).	In the absence of 2 or more positives (ie. 1 or 0 positives), then −LR[f] = 0.18 (95% CI 0.08–0.38).	Single-arm design. Therefore unable to identify treatment-effect modifiers.
McKenzie approach (MDT[g])	<12/52 duration, centralization or abolition of symptoms with MDT loading strategies (2 variables)	May et al., 2008	Derivation	$n = 102$ (secondary analysis of single-arm of RCT), back and neck pain patients referred by GP's to Physiotherapy, study sample demographics not provided.	For those patients with back pain, the presence of both predictor variables gave a probability of success ('liberal' definition provided in study) of 68.9%. The absence of both variables gave a probability of success of 10%.	Single-arm design. Therefore unable to identify treatment-effect modifiers.
Specific exercise program for Ankylosing Spondylitis	SF-36 Physical Role > 37, SF-36 Bodily Pain > 27, BASDAI[h] > 31 (3 variables)	Alonso-Blanco et al., 2009	Derivation	$n = 35$, patients with AS referred to physiotherapy clinic, mean 45.7 years old, 20% female, mean 9.7 years of symptoms, 46% prevalence of target outcome.	If 2 or more positives, then +LR = 11.2 (95% CI 1.7–76.0). If 3 or more positives, then +LR = 2.6 (95% CI 1.6–4.0).	Single-arm design. Therefore unable to identify treatment-effect modifiers.

[a] FABQ-W = Fear avoidance beliefs questionnaire work subscale.
[b] IR ROM = internal rotation range of movement.
[c] ODQ = Oswestry disability questionnaire.
[d] +LR = positive likelihood ratio.
[e] SLR = straight leg raise.
[f] −LR = negative likelihood ratio.
[g] MDT = mechanical diagnosis and therapy.
[h] BASDAI = Bath ankylosing spondylitis disease activity index.

Appendix 2c. Prognostic clinical prediction rules included in qualitative synthesis.

CPR	Variables	Publication	Stage of rule development	Sample	Results/outcome	Notes
6 month pain outcome for acute/subacute LBP	Baseline pain intensity (0–10 NRS[a]), CP[b] (present = 1, absent = 0) (2 variables)	George et al., 2005	Derivation	$n = 28$ (secondary analysis of sub-group in earlier clinical trial), LBP < 60 days duration, aged 18–55 years, demographic details of this sub-group not reported.	6 month pain intensity (0–10 NRS) = 0.97 + 0.27(Pain 0–10 NRS) − 1.6 (CP).	Analysis limited to only those patients that were classified for 'specific exercise'.
6 month disability outcome for acute/subacute LBP	Baseline disability (ODQc), FABQ-W[d], CP (present = 1, absent = 0) (3 variables)	George et al., 2005	Derivation	$n = 28$ (secondary analysis of sub-group in earlier clinical trial), LBP < 60 days duration, aged 18–55 years, demographic details of this sub-group not reported.	6 month disability (ODQ) = 4.4 + 0.24 (ODQ) + 0.34(FABQ-W) − 10 (CP).	Analysis limited to only those patients that were classified for 'specific exercise'.
Time to recovery from acute LBP	Baseline pain ≤ 7/10, duration of current episode ≤5 days, and ≤1 previous episodes (3 variables)	Hancock et al., 2009b	Derivation	$n = 239$ (RCT). LBP +/− leg pain <6/52, presenting to GPs, mean age 40.7 years, 44% female, mean 9 days of symptoms.	If 3 variables positive, then median days to recovery (from baseline assessment) = 6 (95% CI 4–8). If no variables are positive, then median days to recovery = 22 (95% CI 11–33).	All arms of study included in analysis.

[a] NRS = numerical rating scale.
[b] CP = centralisation phenomenon.
[c] ODQ = Oswesty Disability Questionnaire.
[d] FABQ-W = Fear Avoidance Beliefs Questionnaire Work Subscale.

References

Allison SC. On "A guide to interpretation of studies". Hancock M, et al. Physical Therapy. 2009;89:698–704. Physical Therapy 2009;89(10):1098–9.

Alonso-Blanco C, Fernandez-de-las-Penas C, Cleland JA. Preliminary clinical prediction rule for identifying patients with ankylosing spondylitis who are likely to respond to an exercise program: a pilot study. American Journal of Physical Medicine & Rehabilitation 2009;88(6):445–54.

Andersson GB. Epidemiology of low back pain. Acta Orthopaedica Scandinavica Supplementum 1998;281:28–31.

Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. Lancet 2000;355(9209):1064–9.

Bachmann LM, Haberzeth S, Steurer J, ter Riet G. The accuracy of the Ottawa knee rule to rule out knee fractures: a systematic review. Annals of Internal Medicine 2004;140(2):121–4.

Beattie P, Nelson R. Clinical prediction rules: what are they and what do they tell us? Australian Journal of Physiotherapy 2006;52(3):157–63.

Beneciuk JM, Bishop MD, George SZ. Clinical prediction rules for physical therapy interventions: a systematic review. Physical Therapy 2009;89(2):114–24.

Brennan GP, Fritz JM, Hunter SJ, Thackeray A, Delitto A, Erhard RE. Identifying subgroups of patients with acute/subacute "nonspecific" low back pain: results of a randomized clinical trial. Spine 2006;31(6):623–31.

Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. Journal of Clinical Epidemiology 2004;57(3):229–36.

Cai C, Pua YH, Lim KC. A clinical prediction rule for classifying patients with low back pain who demonstrate short-term improvement with mechanical lumbar traction. European Spine Journal 2009;18(4):554–61.

Childs J, Fritz J, Piva S, Erhard R. Clinical decision making in the identification of patients likely to benefit from spinal manipulation: a traditional versus an evidence-based approach. Journal of Orthopaedic and Sports Physical Therapy 2003;33(5):259–72.

Childs JD, Cleland JA. Development and application of clinical prediction rules to improve decision making in physical therapist practice. Physical Therapy 2006; 86(1):122–31.

Childs JD, Flynn TW, Fritz JM. A perspective for considering the risks and benefits of spinal manipulation in patients with low back pain. Manual Therapy 2006; 11(4):316–20.

Childs JD, Fritz JM, Flynn TW, Irrgang JJ, Johnson KK, Majkowski GR, et al. A clinical prediction rule to identify patients with low back pain most likely to benefit from spinal manipulation: a validation study. Annals of Internal Medicine 2004; 141(12):920–8.

Cleland J, Fritz J, Kulig K, Davenport T, Eberhart S, Magel J, et al. Comparison of the effectiveness of three manual physical therapy techniques in a subgroup of patients with low back pain who satisfy a clinical prediction rule: a randomized clinical trial. Spine 2009;34(25):2720–9.

Cleland JA, Fritz JM, Whitman JM, Childs JD, Palmer JA. The use of a lumbar spine manipulation technique by physical therapists in patients who satisfy a clinical prediction rule: a case series. Journal of Orthopaedic & Sports Physical Therapy 2006;36(4):209–14.

Cochrane Back Group. Search strategies for back group specialized registry, http://www.mrw.interscience.wiley.com/cochrane/clabout/articles/BACK/frame.html; 2009 [accessed January 2009].

Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariable models. Annals of Internal Medicine 1993;118(3):201–10.

Cook C. Potential pitfalls of clinical prediction rules. Journal of Manual and Manipulative Therapy 2008;16(2):69–71.

Cook C, Brismee JM, Pietrobon R, Sizer Jr P, Hegedus E, Riddle DL. Development of a quality checklist using Delphi methods for prescriptive clinical prediction rules: the QUADCPR. Journal of Manipulative and Physiological Therapeutics 2010;33(1):29–41.

Dahri K, Loewen P. The risk of bleeding with warfarin: a systematic review and performance analysis of clinical prediction rules. Thrombosis & Haemostasis 2007;98(5):980–7.

Eagles D, Stiell IG, Clement CM, Brehaut J, Kelly AM, Mason S, et al. International survey of emergency physicians' priorities for clinical decision rules. Academic Emergency Medicine 2008;15(2):177–82.

Flynn T, Fritz J, Whitman J, Wainner R, Magel J, Rendeiro D, et al. A clinical prediction rule for classifying patients with low back pain who demonstrate short-term improvement with spinal manipulation. Spine 2002;27(24):2835–43.

Fritz J, Hancock M, Herbert R, Riddle D. July 2009 debate: Clinical prediction rules – part 1, http://www.scienceaudio.net/ptj/200907/ptj_200907_debate_1.mp3; 2009a [accessed 27.02.10].

Fritz J, Hancock M, Herbert R, Riddle D. July 2009 debate: Clinical prediction rules – part 2, http://www.scienceaudio.net/ptj/200907/ptj_200907_debate_2.mp3; 2009b [accessed 27.02.10].

Fritz JM. Clinical prediction rules in physical therapy: coming of age? Journal of Orthopaedic & Sports Physical Therapy 2009;39(3):159–61.

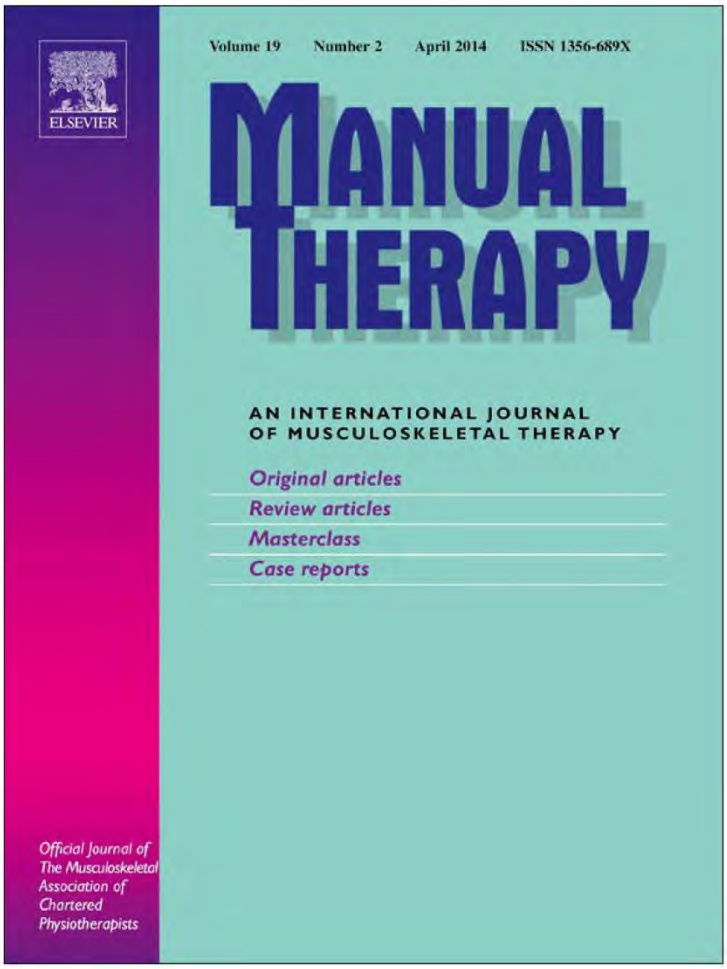
Fritz JM, Brennan GP, Leaman H. Does the evidence for spinal manipulation translate into better outcomes in routine clinical care for patients with occupational low back pain? A case-control study. Spine Journal: Official Journal of the North American Spine Society 2006;6(3):289–95.

Fritz JM, Childs JD, Flynn TW. Pragmatic application of a clinical prediction rule in primary care to identify patients with low back pain with a good

prognosis following a brief spinal manipulation intervention. BMC Family Practice 2005a;6(29).
Fritz JM, Delitto A, Erhard RE. Comparison of classification-based physical therapy with therapy based on clinical practice guidelines for patients with acute low back pain: a randomized clinical trial. Spine 2003;28(13):1363–71.
Fritz JM, Lindsay W, Matheson JW, Brennan GP, Hunter SJ, Moffit SD, et al. Is there a subgroup of patients with low back pain likely to benefit from mechanical traction? Results of a randomized clinical trial and subgrouping analysis. Spine 2007;32(26):E793–800.
Fritz JM, Piva SR, Childs JD. Accuracy of the clinical examination to predict radiographic instability of the lumbar spine. European Spine Journal 2005b;14(8):743–50.
Fritz JM, Whitman JM, Flynn TW, Wainner RS, Childs JD. Factors related to the inability of individuals with low back pain to improve with a spinal manipulation. Physical Therapy 2004;84(2):173–90.
George SZ, Bialosky JE, Donald DA. The centralization phenomenon and fear-avoidance beliefs as prognostic factors for acute low back pain: a preliminary investigation involving patients classified for specific exercise. Journal of Orthopaedic & Sports Physical Therapy 2005;35(9):580–8.
Hallegraeff JM, de Greef M, Winters JC, Lucas C. Manipulative therapy and clinical prediction criteria in treatment of acute nonspecific low back pain. Perceptual & Motor Skills 2009;108(1):196–208.
Hancock M, Herbert RD, Maher CG. A guide to interpretation of studies investigating subgroups of responders to physical therapy interventions. Physical Therapy 2009a;89(7):698–704.
Hancock MJ, Maher CG, Herbert RD. Answer to the letter to the editor of J. Hebert et al. concerning "Hancock MJ, Maher CG, Latimer J, Herbert RD, McAuley JH (2008) Independent evaluation of a clinical prediction rule for spinal manipulative therapy: a randomised controlled trial. European Spine Journal 2008a;17(10):1403–4.
Hancock MJ, Maher CG, Latimer J, Herbert RD, McAuley JH. Independent evaluation of a clinical prediction rule for spinal manipulative therapy: a randomised controlled trial. European Spine Journal 2008b;17(7):936–43.
Hancock MJ, Maher CG, Latimer J, Herbert RD, McAuley JH. Can rate of recovery be predicted in patients with acute low back pain? Development of a clinical prediction rule. European Journal of Pain 2009b;13(1):51–5.
Hebert JJ, Perle SM. Letter to the editor concerning "Independent evaluation of a clinical prediction rule for spinal manipulative therapy: a randomised controlled trial" (M. Hancock et al.). European Spine Journal 2008;17(10):1401–2.
Henschke N, Maher CG, Refshauge KM, Herbert RD, Cumming RG, Bleasel J, et al. Prevalence of and screening for serious spinal pathology in patients presenting to primary care settings with acute low back pain. Arthritis & Rheumatism 2009;60(10):3072–80.
Hess EP, Thiruganasambandamoorthy V, Wells GA, Erwin P, Jaffe AS, Hollander JE, et al. Diagnostic accuracy of clinical prediction rules to exclude acute coronary syndrome in the emergency department setting: a systematic review. CJEM Canadian Journal of Emergency Medical Care 2008;10(4):373–82.
Hicks GE, Fritz JM, Delitto A, McGill SM. Preliminary development of a clinical prediction rule for determining which patients with low back pain will respond to a stabilization exercise program. Archives of Physical Medicine and Rehabilitation 2005;86(9):1753–62.
Ingui BJ, Rogers MA. Searching for clinical prediction rules in MEDLINE. Journal of the American Medical Informatics Association 2001;8(4):391–7.
Katz MH. Multivariable analysis. A practical guide for clinicians. 2nd ed. New York: Cambridge University Press; 2006.
Kent P, Keating JL. Classification in nonspecific low back pain: what methods do primary care clinicians currently use? Spine 2005;30(12):1433–40.
Kent PM, Keating JL, Taylor NF. Primary care clinicians use variable methods to assess acute nonspecific low back pain and usually focus on impairments. Manual Therapy 2009;14(1):88–100.
Kraemer HC, Frank E, Kupfer DJ. Moderators of treatment outcomes: clinical, research, and policy importance. Journal of the American Medical Association 2006;296(10):1286–9.
Kraemer HC, Gibbons RD. Why does the randomized clinical trial methodology so often mislead clinical decision making? Focus on moderators and mediators of treatment. Psychiatric Annals 2009;39(7):736–45.
Kraemer HC, Wilson GT, Fairburn CG, Agras WS. Mediators and moderators of treatment effects in randomized clinical trials. Archives of General Psychiatry 2002;59(10):877–83.
Laslett M, Aprill CN, McDonald B, Oberg B. Clinical predictors of lumbar provocation discography: a study of clinical predictors of lumbar provocation discography. European Spine Journal 2006a;15(10):1473–84.
Laslett M, Aprill CN, McDonald B, Young SB. Diagnosis of sacroiliac joint pain: validity of individual provocation tests and composites of tests. Manual Therapy 2005;10(3):207–18.
Laslett M, McDonald B, Aprill CN, Tropp H, Oberg B. Clinical predictors of screening lumbar zygapophyseal joint blocks: development of clinical prediction rules. Spine Journal 2006b;6(4):370–9.
Laslett M, Young SB, Aprill CN, McDonald B. Diagnosing painful sacroiliac joints: a validity study of a McKenzie evaluation and sacroiliac provocation tests. Australian Journal of Physiotherapy 2003;49(2):89–97.
Laupacis A, Sekar N, Stiell IG. Clinical prediction rules: a review and suggested modifications of methodological standards. Journal of the American Medical Association 1997;277(6):488–94.
Lewis S. Regression analysis. Practical Neurology 2007;7(4):259–64.
Long A, Donelson R, Fung T. Does it matter which exercise? A randomized control trial of exercise for low back pain. Spine 2004;29(23):2593–602.
MacKinnon DP, Luecken LJ. How and for whom? Mediation and moderation in health psychology. Health Psychology 2008;27(Suppl. 2):S99–100.
May S, Gardiner E, Young S, Klaber-Moffett J. Predictor variables for a positive long-term functional outcome in patients with acute and chronic neck and back pain treated with a McKenzie approach: a secondary analysis. Journal of Manual and Manipulative Therapy 2008;16(3):155–60.
May S, Rosedale R. Prescriptive clinical prediction rules in back pain research: a systematic review. Journal of Manual and Manipulative Therapy 2009;17(1):36–45.
McGinn T, Wyer P, Wisnivesky J, Devereauz PJ, Stiell I, Richardson S, et al. Advanced topics in diagnosis: clinical prediction rules. In: Guyatt G, Rennie D, Meade MO, Cook DJ, editors. Users' guides to the medical literature. A manual for evidence-based clinical practice. 2nd ed. USA: McGraw Hill Medical; 2008. p. 491–505.
McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS. Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-based medicine working group. JAMA 2000;284(1):79–84.
Randolph AG, Guyatt GH, Calvin JE, Doig G, Richardson WS, Cook DJ. Understanding articles describing clinical prediction tools. Critical Care Medicine 1998;26(9):1603–12.
Raney NH, Petersen EJ, Smith TA, Cowan JE, Rendeiro DG, Deyle GD, et al. Development of a clinical prediction rule to identify patients with neck pain likely to benefit from cervical traction and exercise. European Spine Journal 2009;18(3):382–91.
Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. Annals of Internal Medicine 2006;144(3):201–9.
Riihimaki H. Epidemiology and pathogenesis of non-specific low back pain: what does the epidemiology tell us? Bulletin of the Hospital for Joint Diseases 1996;55(4):197–8.
Sackett D, Hayes R, Guyatt G, Tugwell P. Clinical epidemiology: a basic science for clinical medicine. 2nd ed. Boston, USA: Brown & Co; 1991.
Stanton TR, Hancock MJ, Maher CG, Koes BW. Critical appraisal of clinical prediction rules that aim to optimize treatment selection for musculoskeletal conditions. Physical Therapy 2010;90(6):843–54.
Stanton TR, Maher CG, Hancock M. On clinical prediction rules for physical therapy interventions. Physical Therapy 2009;89(4):394.
Stiell IG, Wells GA. Methodologic standards for the development of clinical decision rules in emergency medicine. Annals of Emergency Medicine 1999;33(4):437–47.
Szadek KM, van der Wurff P, van Tulder MW, Zuurmond WW, Perez RS. Diagnostic validity of criteria for sacroiliac joint pain: a systematic review. Journal of Pain 2009;10(4):354–68.
Tamariz LJ, Eng J, Segal JB, Krishnan JA, Bolger DT, Streiff MB, et al. Usefulness of clinical prediction rules for the diagnosis of venous thromboembolism: a systematic review. American Journal of Medicine 2004;117(9):676–84.
Teyhen DS, Flynn TW, Childs JD, Abraham LD. Arthrokinematics in a subgroup of patients likely to benefit from a lumbar stabilization exercise program. Physical Therapy 2007;87(3):313–25.
Toll DB, Janssen KJM, Vergouwe Y, Moons KGM. Validation, updating and impact of clinical prediction rules: a review. Journal of Clinical Epidemiology 2008;61(11):1085–94.
Turner JA, Holtzman S, Mancl L. Mediators, moderators, and predictors of therapeutic change in cognitive-behavioral therapy for chronic pain. Pain 2007;127(3):276–86.
van der Wurff P, Buijs EJ, Groen GJ. A multitest regimen of pain provocation tests as an aid to reduce unnecessary minimally invasive sacroiliac joint procedures. Archives of Physical Medicine & Rehabilitation 2006;87(1):10–4.
Walker BF. The prevalence of low back pain in Australian adults. A systematic review of the literature from 1966–1998. Asia-Pacific Journal of Public Health 1999;11(1):45–51.
Wasson JH, Sox HC, Neff RK, Goldman L. Clinical prediction rules. Applications and methodological standards. New England Journal of Medicine 1985;313(13):793–9.
Werner EL, Indahl A, Werner EL, Indahl A. Knowledge, practice and attitudes to back pain among doctors, physiotherapists and chiropractors. Tidsskrift for Den Norske Laegeforening 2005;125(13):1794–7.
Wisnivesky JP, Serebrisky D, Moore C, Sacks HS, Iannuzzi MC, McGinn T. Validity of clinical prediction rules for isolating inpatients with suspected tuberculosis. A systematic review. Journal of General Internal Medicine 2005;20(10):947–52.

Study 2 - Chapter 5 (p.174)

Haskins, R., Osmotherly, P. G., Southgate, E., & Rivett, D. A. (2014). Physiotherapists' knowledge, attitudes and practices regarding clinical prediction rules for low back pain. Manual Therapy, 19(2), 142-151.



Manual Therapy 19 (2014) 142–151

Contents lists available at ScienceDirect

Manual Therapy

journal homepage: www.elsevier.com/math

Original article

Physiotherapists' knowledge, attitudes and practices regarding clinical prediction rules for low back pain

Robin Haskins*, Peter G. Osmotherly, Erica Southgate, Darren A. Rivett

The University of Newcastle, Australia

ARTICLE INFO

Article history:
Received 7 April 2013
Received in revised form
23 August 2013
Accepted 23 September 2013

Keywords:
Low back pain
Clinical prediction rules
Physiotherapy

ABSTRACT

Clinical Prediction Rules (CPRs) have been developed to assist in the physiotherapy management of low back pain (LBP) although little is known about the factors that may influence their implementation in clinical practice. This study used qualitative research methodology to explore the knowledge, attitudes and practices/behaviours of physiotherapists in relation to these tools. Four semi-structured focus groups involving 26 musculoskeletal physiotherapists were conducted across three Australian geographic regions. A fictitious LBP case scenario was developed and used to facilitate group discussion. Participant knowledge of CPRs was found to be mixed, with some clinicians never having previously encountered the term or concept. LBP CPRs were often conceptualised as a formalisation of pattern recognition. Attitudes towards CPRs expressed by study participants were wide-ranging with several facilitating and inhibiting views identified. It was felt that more experienced clinicians had limited need of such tools. Only a small number of participants expressed that they had ever used LBP CPRs in clinical practice. To optimise the successful adoption of an LBP CPR, researchers should consider avoiding the use of the term 'rule' and ensure that the tool and its interface are uncomplicated and easy to use. Understanding potential barriers, the needs of clinicians and the context in which CPRs will be implemented will help facilitate the development of tools with the highest potential to positively influence physiotherapy practice.

1. Introduction

The identification of meaningful sub-groups of patients with low back pain is a priority area for LBP research and is believed to have the potential to lead to substantial improvements in patient care (Borkan and Cherkin, 1996; Henschke et al, 2007; Foster et al., 2009; Costa et al., 2013). Although the idea of sub-grouping patients with LBP is not new (Riddle, 1998; McCarthy et al., 2004), more recently greater emphasis has been placed upon the use of statistical procedures to identify the factors that delineate patients with LBP with differing prognoses and degrees of responsiveness to certain interventions. One such sub-grouping mechanism is the clinical prediction rule (CPR).

A CPR is a clinical tool that is used to inform decision-making by quantifying the probability of a given outcome, diagnosis or treatment response using a parsimonious set of factors from the history, physical examination and other investigations (McGinn et al., 2008). In recent years a growing number of CPRs relevant to physiotherapy have been derived for LBP presentations for a wide variety of diagnostic, prognostic and prescriptive functions (Beneciuk et al., 2009; May and Rosedale, 2009; Stanton et al., 2010; Haskins et al., 2012). At this time however, it is not clear if these tools are consistent with the perceived needs of physiotherapists or will be accepted by them.

Limited evidence suggests that LBP CPRs may be accepted and used by some US physical therapists. A recent US study found that 40% of surveyed physical therapists who routinely employ lumbar thrust manipulation report using a CPR (Learman et al., 2012). Outside of a US context, however, there is no discernible research data on physiotherapists' awareness or use of LBP CPRs. Awareness of Emergency Medicine CPRs has been demonstrated to vary internationally and to be highest in the countries in which the tools have been developed (Graham et al., 2001; Eagles et al., 2008). As most LBP CPRs relevant to physiotherapy practice have been developed in the US (Haskins et al., 2012), it is likely that awareness and use of these tools in other countries may be much lower.

In addition to limited awareness, previous research has highlighted that once CPRs have been validated and demonstrated to positively impact clinical practice, there are a number of individual and system level barriers that may impede their successful adoption (Graham et al., 1998, 2001; Brehaut et al., 2005; Brehaut et al., 2006; Stiell et al., 2006; Eagles et al., 2008; Beutel et al., 2012). Table 1 provides an overview of the literature-informed potential barriers to the adoption of LBP CPRs in physiotherapy practice

* Corresponding author. School of Health Sciences, The University of Newcastle, Callaghan, NSW 2308, Australia. Tel.: +61 2 4922 3079; fax: +61 2 4921 7053.
E-mail address: Robin.Haskins@newcastle.edu.au (R. Haskins).

1356-689X/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.math.2013.09.005

Table 1
Literature-informed potential barriers to the adoption of LBP CPRs in physiotherapy practice.

Theme	Subtheme	Potential barrier	Description
Knowledge	Awareness	Lack of awareness	Unaware of the existence of LBP CPRs
	Familiarity	Lack of familiarity	Insufficient knowledge of the content of LBP CPRs to enable their application
	Forgetting	Forgetting	Inadvertently omitting to implement LBP CPRs
Attitudes	Agreement in general	Too 'cookbook'	Perception that LBP CPRs oversimplify the complexities of the clinical encounter
		Dislike of the term 'rule'	Aversion to using LBP CPRs due to the term 'rule' implying an authoritative influence on decision-making
		Challenge to autonomy	Perception that LBP CPRs are a threat to professional autonomy
		Biased synthesis	Perception that the development of the tool was biased
		Not practical	Perception that LBP CPRs are unclear or impractical to follow
		Unspecified overall lack of agreement with using the tool	Lack of agreement with LBP CPRs in general
	Expectancy	No perceived benefit to patient outcomes	Perception that using LBP CPRs will not lead to improved patient outcomes
		No perceived benefit to health care processes	Perception that using LBP CPRs will not lead to improved health care processes
	Self-efficacy	Lack of self-efficacy	Belief that one cannot use LBP CPRs
	Motivation	Lack of motivation/Inertia of current practice	Lack of motivation to use LBP CPRs or to change one's habits
Practices/Behaviours	Patient factors	Lack of consistency with patient preferences	Perceived inability to reconcile patient preferences with the use of LBP CPRs
	Factors associated with LBP CPRs as an innovation	Lack of triability	Perception that LBP CPRs cannot be tried or experimented with
		Lack of compatibility	Perception that LBP CPRs are not consistent with one's own approach
		High complexity	Perception that LBP CPRs are difficult to understand and use
		Lack of observability	Lack of the visibility of the results of using LBP CPRs
		Not communicable	Perception that it is not possible to communicate with colleagues about LBP CPRs to reach a mutual understanding
		Increased uncertainty	Perception that the use of LBP CPRs will increase uncertainty
		Not modifiable	Lack of flexibility to modify or adapt LBP CPRs
	Environmental factors	Lack of time	Insufficient time to use LBP CPRs
		Lack of resources	Insufficient resources to use LBP CPRs
		Organisational constraints	Insufficient support from the organisation to use LBP CPRs
		Lack of reimbursement	Insufficient reimbursement for using LBP CPRs
		Increased medicolegal liability	Perceived increased risk of legal actions arising from using LBP CPRs

based on the current body of evidence using a framework of knowledge, attitudes and practices/behaviours (Cabana et al., 1999; Legare et al., 2008). This framework has been used in previous research to help identify the barriers to the adoption of other clinical innovations, such as clinical practice guidelines (Larson, 2004; Schouten et al, 2007; Pogorzelska and Larson, 2008) and clinical protocols (Rubinson et al., 2005; Dennison et al., 2007; Barlow et al., 2008), and has been recommended as an appropriate framework to investigate the barriers to the use of CPRs (Abboud and Cabana, 2001). Recognition of the facilitators and barriers to the use of LBP CPRs will enable the development of tailored strategies that may assist the adoption of these tools into practice (Bero et al., 1998; Cabana et al., 2002; Grol and Wensing, 2004; Mehta, 2004; National Institute of Clinical Studies, 2006).

Although considerable work has been invested in the development of LBP CPRs for physiotherapy practice, very little is known about how they will be integrated within the complex thinking and decision-making processes of clinical reasoning (Edwards et al., 2004). Limited evidence suggests that clinicians using LBP CPRs may not necessarily use them in isolation but rather consider them within the context of all other available information to inform their decision-making (Learman et al., 2012). Understanding the ways in which physiotherapists apply LBP CPRs in the clinical setting will also be informative to designing strategies to optimise their use.

What physiotherapists know about LBP CPRs, as well as their attitudes and practices in relation to these tools remains largely unknown but will underpin their successful adoption into clinical practice (National Institute of Clinical Studies, 2006). Qualitative research methodology seeks to construct meaning and knowledge through the understanding of human experience (Petty et al., 2012a) and provides an appropriate avenue to gain deep understanding and greater insight into the factors that influence LBP CPR implementation in physiotherapy. The generation of such knowledge is anticipated to inform strategies that may optimise the development of LBP CPRs with the greatest potential to positively impact physiotherapy practice.

2. Methods

2.1. Design

Qualitative Descriptive design is intended to provide a clear description of a specific phenomenon or experience from the perspective of research participants (Magilvy and Thomas, 2009). It is an approach that seeks to identify and explore rich straight description on particular topics using language reflective of that used by participants and with minimal interpretative meaning inferred by the researcher (Sandelowski, 2000; Neergaard et al., 2009; Sandelowski, 2010). Qualitative Descriptive design was deemed an appropriate approach to gain firsthand insight into the knowledge, attitudes and practices/behaviours of physiotherapists in relation to LBP CPRs. The investigation of these domains is a well-recognised approach used to examine the barriers to the adoption of evidence in practice (Lang et al., 2007).

2.2. Participants

Purposive sampling (Greenwood and Parsons, 2000) is a sampling technique that involves the selective recruitment of participants who may provide the best insight into the research questions. This sampling technique was used in this study to recruit physiotherapists of varying degrees of experience who manage patients with low back pain, in both private and public sectors, across

Table 2
Schedule of activities for focus groups.

Section	Foci	Activities	Prompts	Approximate time
Introduction	Participant backgrounds, clinical experience, work setting, experience managing patients with LBP.	Participants introduce themselves to the group.	Please tell the group a little about yourself? Where do you work? How long have you been working as a physiotherapist? What proportion of your caseload are LBP patients?	10 min
Case study	Knowledge, attitudes and practices regarding CPRs for LBP. Clinical decision making in the assessment and management of LBP.	Participants are asked to read a fictitious LBP case study and discuss their perspectives on the assessment and management of that patient.	What are your thoughts regarding this patient's diagnosis/prognosis/management? What information is important in helping you make these decisions? Is anyone aware of any CPRs that could be used? Would anyone consider using a CPR? Which one(s) and why?	30 min
CPRs for LBP	Knowledge, attitudes and practices regarding CPRs for LBP.	Group discussion on CPRs for LBP. Participants in focus groups 3 and 4 received one-page summary about CPRs following discussion about knowledge of the topic (Appendix 2).	What do you understand about the term 'CPR'? Which CPRs have you heard of? Can you describe any CPRs? What are your experiences with using LBP CPRs? How would you incorporate LBP CPRs into your clinical reasoning? How do you feel about CPRs for LBP? What are the barriers to using LBP CPRs?	40 min
Priorities for LBP CPR development	Participant priorities for the development of LBP CPRs.	Group discussion on participants' priorities for LBP CPR development.	In the management of LBP, are there any areas of your practice that may benefit from a CPR? Which types of LBP CPRs would be most useful? What characteristics do they need to have to be useful? What advice would you give to researchers who are developing LBP CPRs?	30 min

metropolitan and regional areas of New South Wales, Australia. It was considered by the research team that this sample is likely to be a target consumer group of LBP CPRs.

Potential participants were identified and recruited using phonebook listings and an online professional search tool. Additionally, an advertisement for the study was included within a professional email bulletin sent to all members of the Australian Physiotherapy Association.

2.3. Data collection

Four focus groups each lasting between 1.5 and 2 h were conducted. Each group consisted of between 5 and 11 participants and was moderated by a member of the research team. A focus group schedule (Krueger and Casey, 2009) of activities and questions was developed and informed by research exploring the knowledge, attitudes and practices/behaviours of emergency physicians in relation to CPRs (Graham et al., 1998, 2001; Brehaut et al., 2005; Brehaut et al., 2006; Stiell et al., 2006; Eagles et al., 2008). A questioning route was developed centred upon addressing the following key research questions:

1. What is the knowledge of musculoskeletal physiotherapists in regard to CPRs?
2. What are the attitudes of musculoskeletal physiotherapists toward CPRs for LBP?
3. What are the self-reported practices/behaviours of musculoskeletal physiotherapists in relation to LBP CPRs?

A semi-structured format for each focus group was used. Table 2 details the foci, activities, prompts and approximate time spent on each section of the focus group.

A fictitious LBP case scenario (Appendix 1) was developed based upon a previously published case study (Glynn and Weisbach, 2010) and adapted to include the predictor variables of 25 CPRs for LBP identified in a recent systematic review (Haskins et al., 2012). The credibility and consistency of the case scenario was checked by four specialist musculoskeletal physiotherapists and Fellows of the Australian College of Physiotherapists (Australian Physiotherapy Association, 2013) before being used in the focus groups.

After conducting the first two focus groups, the research team decided to provide participants in the remaining focus groups with a one-page summary (Appendix 2) detailing a common definition and example of each type of CPR. This was instigated in response to the research team's recognition that participant knowledge about LBP CPRs was diverse and it was believed that discussion concerning attitudes and practices/behaviours may be facilitated by providing a brief standardised summary to all participants. Qualitative research is often characterised by the simultaneous collection and analysis of data, thereby enabling researchers to adjust their avenue of investigation to build greater knowledge where opportunities are identified (Sandelowski, 2000; Krueger and Casey, 2009).

2.4. Data analysis

The audio file from each digitally recorded focus group was transcribed and analysed using Thematic Networks (Attride-Stirling, 2001). Focus group transcriptions were uploaded to NVivo[1] and pseudonyms were substituted for participant names and places. Two processes were used to code the data. Transcripts were read several times by the first author and then segments of text were coded based upon the identification of recurrent themes (Morse et al., 2008). All recurrent themes identified by the first author were coded during this process independent of personal beliefs concerning their relationship with the study's research

[1] NVivo version 9, QSR International Pty Ltd, 651 Doncaster Road, Doncaster, Vic 3108, Australia.

Table 3
Participant characteristics.

Participant pseudonym	Gender	Clinical experience (years)	Place of entry-level qualification	Current work setting	Focus group
David	Male	3	ACT, Australia	Public hospital	1
Colleen	Female	31	NSW, Australia	Private practice	1
Jason	Male	24	NSW, Australia	Private practice	1
Mark	Male	14	NSW, Australia	Private practice	1
Kevin	Male	6	NSW, Australia	Public hospital	1
Courtney	Female	34	NSW, Australia	Private practice	2
Donald	Male	29	NSW, Australia	Private practice	2
Logan	Male	19	Germany	Private practice	2
Corey	Male	23	NSW, Australia	Private practice	2
Lachlan	Male	27	NSW, Australia	Private practice	2
Bill	Male	23	NSW, Australia	Private practice	3
Rupert	Male	8	NSW, Australia	Public hospital	3
Henry	Male	17	NSW, Australia	Public hospital	3
Fred	Male	7	NSW, Australia	Private practice	3
Cameron	Male	18	United Kingdom	Public hospital	3
Christian	Male	3	NSW, Australia	Private practice	4
Neil	Male	3	NSW, Australia	Private practice	4
Clayton	Male	1	NSW, Australia	Private practice	4
Kathleen	Female	1	NSW, Australia	Private practice	4
Tyrone	Male	6	NSW, Australia	Private practice	4
Erik	Male	28	NSW, Australia	Private practice	4
Melanie	Female	1	NSW, Australia	Private practice	4
Hugh	Male	25	Victoria, Australia	Private practice	4
Harold	Male	21	NSW, Australia	Private practice	4
Charlene	Female	missing	NSW, Australia	Private practice	4
Yvette	Female	missing	United Kingdom	Private practice	4

questions. This was done to enhance the trustworthiness of the data analysis process by ensuring that the full research team were involved in the selection of themes that best related to the study's research questions. Following this, the first author coded data according to a list of potential barriers to the adoption of CPRs identified in the literature and presented in Table 1.

Themes identified from both rounds of coding were examined by the research team and those that best related to the study's research questions were included. Themes that were considered to be overlapping were combined and re-coded to produce a smaller set of mutually exclusive themes. Clusters of themes with commonality were arranged into organising themes (Attride-Stirling, 2001). Organising themes were grouped together based on the research question they addressed.

Member checking was used to enhance the trustworthiness of the study's findings (Krefting, 1991; Petty et al., 2012b). This process involved e-mailing a one-page collated summary of the research team's interpretation of key themes from the focus groups to participants and inviting their feedback on any and all aspects of the summary (Mays and Pope, 2000). The feedback provided by participants was considered by all members of the research team in regards to whether it confirmed or challenged the research team's interpretation of the findings. The trustworthiness of the study's findings was also improved with the use of peer debriefing, whereby the selection and organisation of included themes, and the consideration of participant feedback, were discussed and agreed upon by all members of the research team (Creswell and Miller, 2000; Petty et al., 2012b).

3. Findings

Four focus groups involving a total of 26 participants were conducted. Participant characteristics are summarised in Table 3. The participant sample were predominantly male (77%), worked in a private setting (81%) and had an average of 15.5 years (SD 11) of clinical experience. Three new graduate physiotherapists participated in the study. Nine participants were previously known to the first author through various professional networks. The use of pseudonyms and peer debriefing throughout data analysis minimised any risk of bias that could result from existing researcher–participant relationships.

The first round of coding led to the identification of 62 recurrent themes. Sixteen (62%) of the literature-informed themes (Table 1) were identified within the transcribed text. Integrating the two coding processes led to the development of 27 non-overlapping themes relevant to the study's research questions which were arranged into 7 organising themes (Table 4). Participant feedback on the summarized themes was primarily confirmatory and did not lead to substantial modifications.

3.1. Knowledge

Two organising themes related to the research question regarding physiotherapists' knowledge of CPRs were identified.

3.1.1. Awareness of and familiarity with LBP CPRs

Participants reported mixed awareness of CPRs with some participants ($n = 5$) not having previously encountered the term or concept. A CPR developed to identify patients with LBP who are more likely to respond favourably to spinal manipulation (Flynn et al., 2002) was the most commonly recognised CPR, although many of the criteria that constituted the tool were not commonly identifiable by participants.

> *Fred: I have a vague recollection (of CPRs) from uni but I must admit I'm pretty ignorant of them.*

> *Kevin: I know there's lots of studies on clinical prediction rules with low back pain and there's even some meta-analyses of those studies.*

3.1.2. Conceptualisation of LBP CPRs

Parallels were identified between CPRs and patient management paradigms such as Mechanical Diagnosis and Therapy (McKenzie & May, 2003), with the sub-classification of patients into smaller and more homogenous groups. Most believed that CPRs were simply the

Table 4
Summary of themes.

Themes	Organising themes	Research questions
Awareness of CPRs is varied	Awareness and familiarity	Knowledge about CPRs
Familiarity with CPRs is varied		
Conceptualisation of CPRs is varied	Conceptualisation	
CPRs are the formalisation of existing reasoning processes		
CPRs are evidence-based practice	Facilitative attitudes	Attitudes toward CPRs
CPRs enable greater confidence in making predictions		
CPRs may help inform decision-making		
CPRs may help novice clinicians		
CPRs may positively challenge traditional reasoning strategies		
Numeric data may be helpful		
CPRs are complicated	Inhibitive attitudes	
CPRs are or could become fads		
CPRs could cause intellectual laziness		
CPRs have limited generalisability		
CPRs may challenge clinicians' autonomy		
CPRs may not work because treatment techniques are too varied		
CPRs oversimplify the complexities of a clinical presentation		
Dislike of the word 'rule'		
Existing CPRs are not yet ready to be applied		
LBP is too complicated for CPRs		
No personal need for a CPR	Current practices/behaviours	Practices/behaviours and implementation issues
Some CPRs are used without knowledge that they are CPRs		
Use of CPRs is varied		
CPRs may function as second opinions or as a safety net	CPRs within the clinical reasoning process	
CPRs should not be used in isolation		
CPRs should only be applied to patients for which they have been developed		
Third party payers may use CPRs	Third party payer issues	

formalisation of clinical reasoning strategies like pattern recognition that physiotherapists commonly use. For example:

> *Jason: It's just really formalising and detailing something that we do all the time…. It's what your experience is developing and that's developed for you to be able to look at something and say, "Look, I think this is what this is and I know if I go down this path with it I'm going to be likely to get a good outcome"…It's just an informal thing, they're part of our art.*

3.2. Attitudes

A wide range of attitudes toward LBP CPRs were expressed. The identified themes were clustered into two organising themes based on their facilitative or inhibitive influence on the implementation of CPRs in clinical practice.

3.2.1. Facilitative attitudes towards LBP CPRs

CPRs were viewed positively by some participants as consistent with evidence-based practice. The conscientious use of numerical data and probabilities to inform decision-making was welcomed by some participants, although it was acknowledged that CPRs could be applied clinically without reference to numbers. Statistically derived data were considered by some to be valuable in challenging existing models and assumptions. It was considered that CPRs may be helpful in informing clinical decision-making for LBP, in particular by enabling clinicians to have greater confidence in their predictions.

> *Christian: I do think that it (using CPRs) is evidence based practice…One little thing that we pick up from a clinical prediction rule might inform something that we might change in our practice.*

> *Jason: I think that it's (developing and using CPRs is) a big step in the right direction for us as clinicians…We'd probably have more confidence in being able to say to people "Look, if we do this (treatment) for people with your sorts of signs and symptoms, we get a good outcome…".*

The value of CPRs for LBP was considered to be clinician-dependent, with most participants expressing a view that novice clinicians with limited experience may benefit the most.

> *Tyrone: It (CPRs) certainly would have helped as a new grad six years ago… If I had those (CPRs), it would have made it a lot easier…*

> *Kathleen: I think for a new grad these clinical prediction rules are excellent…*

3.2.2. Inhibitive attitudes towards LBP CPRs

Some participants viewed CPRs as overly complicated and seldom generalisable to the patients that they treat.

> *Corey: …my scant reading of them (CPRs) is that they're too complicated and not trustworthy enough..*

> *Cameron: You know I've seen a few of them come off the market now, the manip(ulation), the stabilisation, even the directional preference one which I find a more common utility, but even that, I'm still finding unfortunately my patients don't fit any of these.*

The term 'rule' was viewed negatively by a number of participants and there was a perception by some that CPRs oversimplified the complexities of a clinical presentation and the clinical reasoning process. Variability in patients, as well as the way in which treatments such as manipulation are applied, was considered by some to adversely affect the utility of CPRs for LBP presentations. Clinical experience was believed by some to obviate the need for LBP CPRs.

> *Rupert: You've spent twenty-three years developing your own algorithm and you go, that would be a thousand times more complicated than anything that gets put into this thing and it's not about this tool trying to replace that but then you also think oh but hang on, I can sort of do this automatically almost, why do I change to that.*

Some physiotherapists believed that CPRs for LBP were not sufficiently developed at this time to enable confident application

in the clinical setting and that their premature adoption could have negative implications for the profession.

Cameron: … I'm just a bit worried about some of this (CPR research) being the next big thing and being part of the vernacular in every day clinical practice before it's been tested in multiple populations.

David: My thoughts are that they're not absolute at this stage… I like the overall idea but I think there' a long way to go (before CPRs can be used).

There was a perception that CPRs could become the next professional 'fad' and be viewed by some clinicians as a sort of magical panacea, despite the current lack of evidence of a positive impact.

Bill: It reminds me of a patient with arthritis searching for a cure and if someone is proposing that this particular thing is fantastic they'll jump on the bandwagon and do it because really there's not really any great answer for arthritis, arthritis pain, there's not really any great answer for low back pain and how physios are treating low back pain. We want it, we're looking for it, it's not there.

Rupert: I don't want to feel like I've just drunk the Kool Aid… It's almost too easy.

Many clinicians expressed a concern regarding the potential of CPRs to cause 'intellectual laziness', as well as negatively impacting upon the autonomy of the clinician.

Fred: You could start to get intellectually sloppy, you know, 'clinical reasoning sloppiness'.

Cameron: What about a society that we're working in in 30 years' time where these prediction models say you must go this direction and you can't have that treatment and you have to have this treatment… "I'm sorry doctor or physio you don't have that latitude and that freedom" …

3.3. Practices/behaviours and implementation issues

Three organising themes related to the research question concerning the practices/behaviours of physiotherapists in relation to LBP CPRs.

3.3.1. Current practices/behaviours

A small minority ($n = 3$) of participants expressed that they would have used a CPR within their assessment and management of the fictitious LBP case scenario. Of those previously familiar with CPRs for LBP, only a small number ($n = 7$) of participants acknowledged that they had ever used them to inform their decision-making in clinical practice. A greater number expressed that they had used CPRs for non-LBP presentations, with the most commonly cited rules being the Ottawa ankle (Stiell et al., 1992) and knee (Stiell et al., 1995) rules, and Wells et al.'s CPR for deep vein thrombosis (Wells et al., 1995). A CPR for the diagnosis of sacroiliac joint mediated pain (Laslett et al., 2003) was reported to be used by several participants both for the fictitious case scenario and within routine clinical practice, although it was not always recognised by participants as a type of CPR.

Melanie: Well, at uni they don't even teach you for example that SIJ (sacroiliac joint), that that's a clinical prediction rule. That's just how you assess an SIJ. That's not a clinical prediction rule.

3.3.2. CPRs within the clinical reasoning process

Clinicians believed that when CPRs are used in clinical practice they should not be used in isolation, but rather used within the suite of clinical reasoning processes physiotherapists typically employ including the consideration of patient expectations and preferences.

Kathleen: I think we've still got to use your own clinical judgement and your own intuition as well. I think these (CPRs) need to be used to complement all of that. I don't think we can just rely solely on clinical prediction rules.

Some participants considered that CPRs best serve clinical practice as second opinions or as 'safety nets', and are able to be overruled by the clinician.

Colleen: They're (CPRs) confirming things that you might be a little bit unsure about.

Henry: The rule might say one thing but we've already decided that we may not necessarily do that.

Many stressed the importance of restricting the use of CPRs to the patient populations for which they were intended.

Donald: …if those rules are built up around a certain type of patient and someone expects us to apply them to every sort of patient, you're asking for chaos, it's not going to work.

3.3.3. Third party payer issues

A recurring theme identified across the focus groups was that CPRs may be used by third party payers, such as insurance companies and government funded health services. Some perceived this as beneficial and thought this may help minimize overservicing and the use of ineffective treatment modalities. The majority of participants however, considered the use of CPRs by third party payers as predominantly negative and believed that this would restrict clinician autonomy and preclude the incorporation of patient preferences into decision-making.

Hugh: I just don't want to get painted into a corner where if I don't treat according to these clinical prediction rules WorkCover (government insurance) might say…"Well, that's not evidence based. That's not gold standard treatment. Why are we paying you to treat this person when it's not following your clinical prediction rule?"

Kevin: WorkCover (government insurance) would use a rule to cut people off from funding… They'd use it in the worst possible way.

4. Discussion

The findings of this study have highlighted that a range of factors related to clinician knowledge, attitudes and practices/behaviours may influence the adoption of LBP CPRs into physiotherapy clinical practice. Many of these factors share similarities with the identified barriers to the adoption of other innovations in physiotherapy, including the use of clinical practice guidelines (Côté et al., 2009), outcome measures (Abrams et al., 2006) and the application of evidence-based practice (Jette et al., 2003).

Knowledge of CPRs, in terms of awareness and familiarity, was quite mixed among the participants in this study, with some not having previously encountered this term or concept. This might suggest that as an innovation, CPRs have not as yet permeated into the mainstream conversation of practising clinicians at least within parts of NSW, Australia. Previous research suggests that awareness of CPRs may be highest in the countries in which the tools have been developed (in this case predominantly the US) (Graham et al., 2001; Eagles et al., 2008) and subsequently the knowledge about LBP CPRs of participants in this study may plausibly contrast to that of clinicians in those regions. Addressing knowledge gaps about LBP CPRs may be an important first step in any strategy designed to enhance the adoption of these tools.

CPRs were seen by many as the formalisation of a traditional reasoning process used by experienced clinicians and this view may have influenced some of the attitudes expressed by participants in this study. Attitudes were notably diverse and encompassed positions that may be seen as both facilitating and inhibitive to the implementation of LBP CPRs. A key belief that emerged across the

focus groups was that the benefit of using these tools was experience-dependent. That is, novice physiotherapists may benefit from the use of LBP CPRs, however more experienced clinicians had limited need of such tools. This belief may have substantial implications for the adoption of LBP CPRs in physiotherapy practice and warrants timely investigation.

For an LBP CPR to be successfully incorporated into physiotherapy practice, the findings of this study suggest there are modifiable characteristics that may enhance its acceptability. While participants in this study expressed they would not blindly adhere to a CPR independent of and naïve to other clinical information and decision-making processes, the word 'rule' had negative connotations and was considered by some to be an implementation barrier. Avoiding the term 'rule' may therefore be a simple but important strategy in improving the use of CPRs. Similar to this study's findings, previous research in the field of Emergency Medicine has found that less than 10% of physicians prefer the term 'rule' when describing these tools (Graham et al., 2001). Less authoritarian terms like 'tool' or 'guideline' may be more palatable and perhaps more consistent with the intended function of CPRs – that is, to help *inform* decision-making, not dictate decision-making (Swets et al., 2000). The findings of this study also highlight the importance of ensuring that CPRs appear uncomplicated and easy to use. This may include making sure that all aspects of the tool are clear and unambiguous (Brehaut et al., 2010), using graphical aids where appropriate (Björk et al., 2012) and soliciting input from practicing clinicians throughout their development (Reilly and Evans, 2006).

Participants in this study infrequently incorporated CPRs into their assessment or management of LBP. This stands in contrast to a recent US study that found that 40% of surveyed physical therapists who routinely employ thrust manipulation report using a CPR (Learman et al., 2012). Learman et al. further identified that clinicians who reported to use a CPR were no more likely to perform manipulation in the presence of contraindications than those who do not use the tool. That is, clinicians did not blindly 'obey' a CPR but rather considered it within the context of all other presenting information. Physiotherapists in the present study reported a similar attitude toward the use of CPRs and believed that the optimal use of these tools was nested within the suite of clinical reasoning strategies clinicians typically employ. Further, physiotherapists felt strongly that third-party payers, naïve to all of the available information, should be prevented from using CPRs to direct the clinician to provide particular forms of therapy.

A limitation of the current study is that the findings represent the thoughts and opinions of study participants and may not be generalisable to other populations. Readers should carefully consider the methods and analytic strategies used in this research when considering the degree to which the findings may be transferable to their own setting (Krueger and Casey, 2009).

5. Conclusions

This is the first study outside of a US context to explore the knowledge, attitudes and practices of physiotherapists in regards to LBP CPRs. Most of the participants in this study reported to be aware of LBP CPRs however very few reported to have ever used them to inform their decision-making. Barriers to the use of LBP CPRs identified in this study included a negative connotation associated with the term 'rule', a perception that CPRs are overly-complex and infrequently applicable, clinical experience obviating the need for such tools, and the potential threat to clinical autonomy and for misuse by third-party payers. Study participants felt that LBP CPRs were best used within the suite of clinical reasoning processes physiotherapists typically employ and considered as second opinions or safety nets that were be able to be overruled by the clinician. Consideration of these views may inform strategies that will optimise the development of LBP CPRs with the highest potential to positively influence physiotherapy practice and implementation strategies that will optimise their adoption into clinical practice.

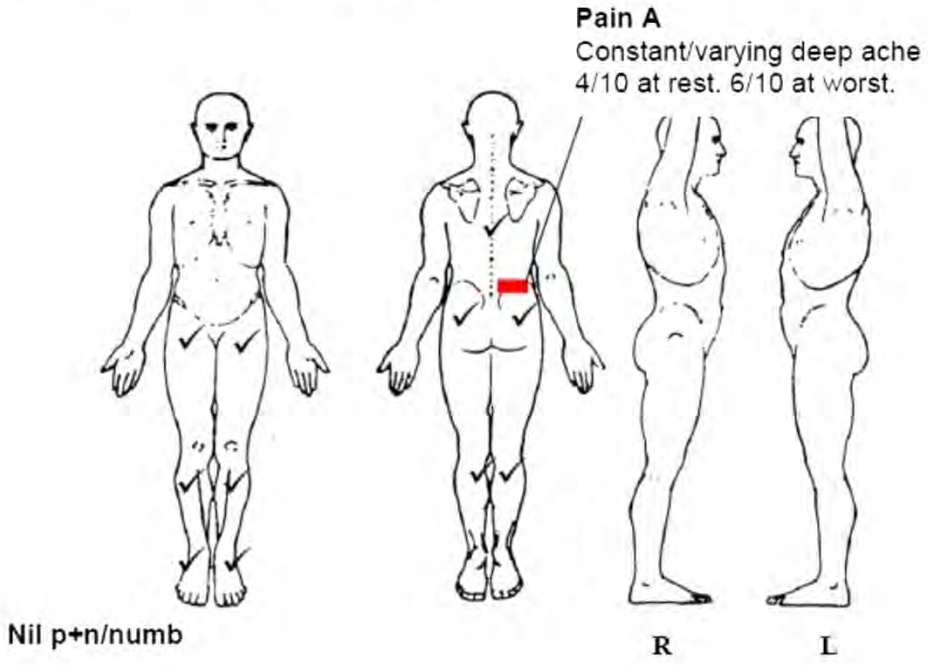
Appendix 1. Case scenario

Summary:

32 year old male computer programmer complaining of a 10 day history of right sided low back pain.

Subjective examination

Body chart



Behaviour

Aggravating factors
- Forward bending
- Sitting for greater than 15 min

Easing factors
- Rest
- Best when lying on back with knees bent (crook-lying)

Cauda equina screening
- Nil change to bladder or bowel function
- Nil paraesthesia/anaesthesia in saddle region

Self report outcome measures

Outcome measure	Score	Notes
Modified Oswestry disability index	34%	Higher score = greater disability
Numeric pain rating scale: current	4/10	Higher score = greater current pain
Numeric pain rating scale: worst in past 24/24	6/10	Higher score = greater pain in past 24/24
Numeric pain rating scale: best in past 24/24	4/10	Higher score = greater pain in past 24/24
Fear avoidance belief questionnaire: work subscale	4/42	Higher score = greater fear and avoidance of work-related activities
Fear avoidance belief questionnaire: physical activity subscale	7/24	Higher score = greater fear and avoidance of physical activities
Modified somatic perception questionnaire	5/39	Higher score = heightened somatic awareness

24-Hour pattern
- Nil diurnal variation. Activity dependent.

Irritability
- Bending forward immediately causes 6/10 pain which eases to 4/10 within a few seconds after returning to a neutral standing posture

History

Current history
- Pain first commenced 10 days ago
- Pain spontaneously occurred a few hours after gardening (involved heavy lifting and bending)
- Nil change in presentation since first onset

Past history
- 3 previous episodes of mild low back pain over the past 5 years
- Each previous episode has resolved spontaneously after 2–3 days without treatment. Nil persisting symptoms between episodes.
- Otherwise, nil significant

Social history
- Lives with wife. No kids.
- Nil regular sporting activities.
- Has not taken any time off work due to low back pain.

Treatment history
- Nil to date

Special questions

General health
- Nil significant. Nil surgery.

Medications
- Panadol. Up to 8 tablets daily since onset, resulting in "mild relief".
- Nil steroids.

Investigations
- Nil to date.

Weight
- Stable.

Spinal cord screening
- Nil paraethesia/anaesthesia
- Nil gait abnormalities

Physical examination

Observation

Standing: Slight forward head posture with protracted scapulae. Decreased kyphosis at mid-thoracic spine. Increased lumbar lordosis. Nil lateral shift. Nil other significant findings.

Sitting: Slight forward head posture. Nil other significant findings.

Physiological movements

Lumbar spine.

Flexion: mild restriction of movement (43°). Pain 6/10. No aberrant movement pattern. Patient reports to feel 'vulnerable' at the beginning of the movement.

Extension: moderate restriction of movement (12°). Nil change in pain.

(R) Lateral flexion: Mild restriction of movement (25°). Pain 5/10.

(L) Lateral flexion: moderate restriction of movement (18°). Pain 6/10.

(R) Side glide: Mild restriction of movement. Pain 5/10.

(L) Side glide: Moderate restriction of movement. Pain 6/10.

Rotation: L = R. Nil restriction of movement. Nil change in pain.

Extension/rotation: L = R. Nil change in pain.

Hip.

Internal rotation: L = R. 40°. Nil change in pain.

Repeated movements

Flexion in standing x 10: Nil effect on symptoms or mechanical presentation.

Extension in standing x 10: Nil effect on symptoms or mechanical presentation.

Flexion in lying x 10: Nil effect on symptoms or mechanical presentation.

Extension in lying x 10: Nil effect on symptoms or mechanical presentation.

Neurological testing

Motor: nil abnormalities detected.

Sensation: nil abnormalities detected.

Reflexes (KJ & AJ): Nil abnormalities detected.

Neurodynamic testing

Straight leg raise: L = R. 70° of hip flexion. Nil change in symptoms.

Slump test: L = R. Nil restriction of movement. Nil change in symptoms.

Sacroiliac joint pain provocation tests

Distraction: nil change in pain.
Thigh thrust: L = R. nil change in pain.
Gaenslen's test: L = R. nil change in pain.
Compression: nil change in pain.
Sacral thrust: nil change in pain.
Patrick sign: L = R. nil change in pain.

Muscular/motor assessment

Transversus abdominis – Prone test with pressure biofeedback unit

- Pressure reduced by 3 mmHg (70 mmHg–67 mmHg) × 10 s with normal respiratory pattern.
- Nil spinal movement during test and nil palpable abdo bulging.
- Consistent test response across 10 repeated tests.

Segmental lumbar multifidus – Prone palpation test

- Nil palpable deficits

Special tests

Prone instability test: negative for pain reduction.

Palpation

Passive accessory intervertebral movement testing

- Generally, hypomobile throughout thoracic and lumbar spine
- Familiar pain reproduced with central and right unilateral PA testing over L_4 and L_5 to a similar degree
- Nil evidence of hypermobility at any spinal segment

Appendix 2. Clinical prediction rules for low back pain

Diagnostic decision making *Diagnostic CPRs* Possible functions: 1. To help clinicians quantify the likelihood of a particular diagnosis given the presence or absence of certain signs and symptoms. 2. To help identify which patients do not require further testing for a particular diagnosis given the presence or absence of certain signs and symptoms.	Example: Diagnosis of SIJ mediated pain[a] Positive on CPR if 3 or more of the following tests reproduce the patient's complaint: 1. Distraction 2. Thigh thrust 3. Gaenslen's test 4. Compression 5. Sacral thrust. +LR = 4.16 (95%CI 2.16–8.39) −LR = 0.12 (95%CI 0.02–0.49)
Treatment decision making *Prescriptive CPRs* Possible function: 1. To help clinicians identify which patients have a higher likelihood of success for a given intervention in comparison to an alternate intervention	Example: Success with lumbopelvic manipulation compared with aerobic and strengthening exercises[b] Positive on CPR if 4 or more of the following tests reproduce the patient's complaint: 1. Duration of current episode<16 days 2. No symptoms below the knee 3. Fear-avoidance belief questionnaire work subscale<19 4. At least 1 hypomobile lumbar segment 5. At least 1 hip with more than 35° of internal rotation ROM. For those positive on CPR, NNT = 1.3 (95%CI 1.1–1.9)
Prognostication *Prognostic CPRs* Possible functions: 1. To help clinicians quantify the likely clinical outcome for an individual given the presence or absence of certain signs and symptoms 2. To help clinicians identify which patients may not require intervention	Example: Time to recovery from acute low back pain[c] Positive on CPR if all 3 of the following are present: 1. Baseline pain≤7/10 2. Duration of current episode≤5 days 3. ≤ 1 previous episodes. For those positive on CPR, median days to recovery = 6 days (95%CI 4–8)

CPR = clinical prediction rule, CI = confidence interval, NNT = number needed to treat.

[a] Laslett M, Young SB, Aprill CN, McDonald B. Diagnosing painful sacroiliac joints: a validity study of a McKenzie evaluation and sacroiliac provocation tests. Australian Journal of Physiotherapy 2003; 49(2):89–97.

[b] Childs JD, Fritz JM, Flynn TW, Irrgang JJ, Johnson KK, Majkowski GR et al. A clinical prediction rule to identify patients with low back pain most likely to benefit from spinal manipulation: a validation study. Annals of Internal Medicine 2004; 141(12):920–8.

[c] Hancock MJ, Maher CG, Latimer J, Herbert RD, McAuley JH. Can rate of recovery be predicted in patients with acute low back pain? Development of a clinical prediction rule. European Journal of Pain 2009; 13(1):51–5.

References

Abboud P, Cabana MD. Understanding barriers to the adoption of clinical decision rules. Ann Emerg Med 2001;38(6):703–4.

Abrams D, Davidson M, Harrick J, Harcourt P, Zylinski M, Clancy J. Monitoring the change: current trends in outcome measure usage in physiotherapy. Man Ther 2006;11(1):46–53.

Attride-Stirling J. Thematic networks: an analytic tool for qualitative research. Qual Res 2001;1(3):385–405.

Australian Physiotherapy Association. 2013.

Barlow G, Nathwani D, Myers E, Sullivan F, Stevens N, Duffy R, et al. Identifying barriers to the rapid administration of appropriate antibiotics in community-acquired pneumonia. J Antimicrob Chemother 2008;61(2):442–51.

Beneciuk JM, Bishop MD, George SZ. Clinical prediction rules for physical therapy interventions: a systematic review. Phys Ther 2009;89(2):114–24.

Bero LA, Grilli R, Grimshaw JM, Harvey E, Oxman AD, Thomson MA. Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. Br Med J 1998;317(7156):465–8.

Beutel BG, Trehan SK, Shalvoy RM, Mello MJ. The Ottawa knee rule: examining use in an academic emergency department. West J Emerg Med 2012;13(4):366–72.

Björk J, Ekelund U, Ohlsson M. Risk predictions for individual patients from logistic regression were visualized with bar–line charts. J Clin Epidemiol 2012;65:335–42.

Borkan JMMDP, Cherkin DCP. An agenda for primary care research on low back pain. Spine 1996;21(24):2880–4.

Brehaut JC, Graham ID, Wood TJ, Taljaard M, Eagles D, Lott A, et al. Measuring acceptability of clinical decision rules: validation of the Ottawa acceptability of decision rules instrument (OADRI) in four countries. Med Decis Making 2010;30(3):398–408.

Brehaut JC, Stiell IG, Graham ID. Will a new clinical decision rule be widely used? The case of the Canadian C-spine rule. Acad Emerg Med 2006;13(4):413–20.

Brehaut JC, Stiell IG, Visentin L, Graham ID. Clinical decision rules "in the real world": how a widely disseminated rule is used in everyday practice. Acad Emerg Med 2005;12(10):948–56.

Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. J. Am. Med. Assoc. 1999;282(15):1458–65.

Cabana MD, Rushton JL, Rush AJ. Implementing practice guidelines for depression: applying a new framework to an old problem. Gen Hosp Psychiatry 2002;24(1):35–42.

Costa LD, Koes BW, Pransky G, Borkan J, Maher CG, Smeets RJ. Primary care research priorities in low back pain: an update. Spine 2013;38(2):148–56.

Côté AM, Durand MJ, Tousignant M, Poitras S. Physiotherapists and use of low back pain guidelines: a qualitative study of the barriers and facilitators. J Occup Rehabil 2009;19(1):94–105.

Creswell JW, Miller DL. Determining validity in qualitative inquiry. Theor Into Pract 2000;39(3):124–30.

Dennison CR, Mendez-Tellez PA, Wang W, Pronovost PJ, Needham DM. Barriers to low tidal volume ventilation in acute respiratory distress syndrome: survey development, validation, and results. Crit Care Med 2007;35(12):2747–54.

Eagles D, Stiell IG, Clement CM, Brehaut J, Taljaard M, Kelly AM, et al. International survey of emergency physicians' awareness and use of the Canadian cervical-spine rule and the Canadian computed tomography head rule. Acad Emerg Med 2008;15(12):1256–61.

Edwards I, Jones M, Carr J, Braunack-Mayer A, Jensen GM. Clinical reasoning strategies in physical therapy. Phys Ther 2004;84(4):312–30.

Flynn T, Fritz J, Whitman J, Wainner R, Magel J, Rendeiro D, et al. A clinical prediction rule for classifying patients with low back pain who demonstrate short-term improvement with spinal manipulation. Spine 2002;27(24):2835–43.

Foster NE, Dziedzic KS, Windt DAWM, Fritz JM, Hay EM. Research priorities for non-pharmacological therapies for common musculoskeletal problems: nationally and internationally agreed recommendations. BMC Musculoskelet Disord 2009;10(1):3.

Glynn PE, Weisbach PD. Case studies., clinical prediction rules. A physical therapy reference manual [Chapter 9]. Sudbury, Massachusetts: Jones and Barlett Publishers; 2010. p. 226–32.

Graham ID, Stiell IG, Laupacis A, McAuley L, Howell M, Clancy M, et al. Awareness and use of the Ottawa ankle and knee rules in 5 countries: can publication alone be enough to change practice? Ann Emerg Med 2001;37(3):259–66.

Graham ID, Stiell IG, Laupacis A, O'Connor AM, Wells GA. Emergency physicians' attitudes toward and use of clinical decision rules for radiography. Acad Emerg Med 1998;5(2):134–40.

Greenwood J, Parsons M. A guide to the use of focus groups in health care research: part 2. Contemp Nurse 2000;9(2):181–91.

Grol R, Wensing M. What drives change? Barriers to and incentives for achieving evidence-based practice. Med J Aust 2004;180(6):S57–60.

Haskins R, Rivett DA, Osmotherly PG. Clinical prediction rules in the physiotherapy management of low back pain: a systematic review. Man Ther 2012;17(1):9–21.

Henschke N, Maher CG, Refshauge KM, Das A, McAuley JH. Low back pain research priorities: a survey of primary care practitioners. BMC Fam Pract 2007;8:40.

Jette DU, Bacon K, Batty C, Carlson M, Ferland A, Hemingway RD, et al. Evidence-based practice: beliefs, attitudes, knowledge, and behaviors of physical therapists. Phys Ther 2003;83(9):786–805.

Krefting L. Rigor in qualitative research: the assessment of trustworthiness. Am J Occup Ther 1991;45(3):214–22.

Krueger RA, Casey MA. Focus groups. A practical guide for applied research. 4th ed. Thousand Oaks, California: Sage Publications, Inc.; 2009.

Lang ES, Wyer PC, Haynes RB. Knowledge translation: closing the evidence-to-practice gap. Ann Emerg Med 2007;49(3):355–63.

Larson E. A tool to assess barriers to adherence to hand hygiene guideline. Am J Infect Control 2004;32(1):48–51.

Laslett M, Young SB, Aprill CN, McDonald B. Diagnosing painful sacroiliac joints: a validity study of a McKenzie evaluation and sacroiliac provocation tests. Aust J Physiother 2003;49(2):89–97.

Learman K, Showalter C, Cook C. Does the use of a prescriptive clinical prediction rule increase the likelihood of applying inappropriate treatments? A survey using clinical vignettes. Man Ther 2012;17(6):538–43.

Legare F, Ratte S, Gravel K, Graham ID. Barriers and facilitators to implementing shared decision-making in clinical practice: update of a systematic review of health professionals' perceptions. Patient Educ Couns 2008;73(3):526.

Magilvy JK, Thomas E. A first qualitative project: qualitative descriptive design for novice researchers. J Specialists Pediatr Nurs 2009;14(4):298–300.

May S, Rosedale R. Prescriptive clinical prediction rules in back pain research: a systematic review. J Man Manipulative Ther 2009;17(1):36–45.

Mays N, Pope C. Qualitative research in health care: assessing quality in qualitative research. Br Med J 2000;320(7226):50–2.

McCarthy C, Arnall F, Strimpakos N, Freemont A, Oldham J. The biopsychosocial classification of non-specific low back pain: a systematic review. Phys Ther Rev 2004;9(1):17–30.

McGinn T, Wyer P, Wisnivesky J, Devereauz PJ, Stiell I, Richardson S, et al. Advanced topics in diagnosis: clinical prediction rules. In: Guyatt G, Rennie D, Meade MO, Cook DJ, editors. Users' guides to the medical literature. A manual for evidence-based clinical practice. 2nd ed. USA: McGraw Hill Medical; 2008. p. 491–505.

McKenzie RA, May S. The lumbar spine: mechanical diagnosis and therapy. Waikanae, New Zealand; 2003.

Mehta NB. The doctors' challenge: how can we follow guidelines better. Cleveland Clinic J Med 2004;71(2):81–2.

Morse JM, Barrett M, Mayan M, Olson K, Spiers J. Verification strategies for establishing reliability and validity in qualitative research. Int J Qual Methods 2008;1(2):13–22.

National Institute of Clinical Studies. Identifying barriers to evidence uptake. Melbourne, Australia: NICS; 2006.

Neergaard MA, Olesen F, Andersen RS, Sondergaard J. Qualitative description–the poor cousin of health research? BMC Med Res Methodol 2009;9:52.

Petty NJ, Thomson OP, Stew G. Ready for a paradigm shift? Part 1: introducing the philosophy of qualitative research. Man Ther 2012a;17:267–74.

Petty NJ, Thomson OP, Stew G. Ready for a paradigm shift? Part 2: introducing qualitative research methodologies and methods. Man Ther 2012b;17:378–84.

Pogorzelska M, Larson EL. Assessment of attitudes of intensive care unit staff toward clinical practice guidelines. DCCN - Dimensions Crit Care Nurs 2008;27(1):30–8.

Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. Ann Intern Med 2006;144(3):201–9.

Riddle DL. Classification and low back pain: a review of the literature and critical analysis of selected systems. Phys Ther 1998;78(7):708–37.

Rubinson L, Wu AW, Haponik EF, Diette GB. Why is it that internists do not follow guidelines for preventing intravascular catheter infections? Infect Control Hosp Epidemiol 2005;26(6):525–33.

Sandelowski M. Whatever happened to qualitative description? Res Nurs Health 2000;23(4):334–40.

Sandelowski M. What's in a name? Qualitative description revisited. Res Nurs Health 2010;33(1):77–84.

Schouten JA, Hulscher MEJL, Natsch S, Kullberg BJ, Van Der Meer JWM, Grol RPTM. Barriers to optimal antibiotic use for community-acquired pneumonia at hospitals: a qualitative study. Qual Saf Health Care 2007;16(2):143–9.

Stanton TR, Hancock MJ, Maher CG, Koes BW. Critical appraisal of clinical prediction rules that aim to optimize treatment selection for musculoskeletal conditions. Phys Ther 2010;90(6):843–54.

Stiell IG, Brehaut J, Clement C, Grimshaw J, Graham I, Brison R, et al. Perceived barriers to the implementation of the Canadian C-spine rule and the Canadian CT head rule. Can J Emerg Med 2006;8(3):206–7.

Stiell IG, Greenberg GH, McKnight RD, Nair RC, McDowell I, Worthington JR. A study to develop clinical decision rules for the use of radiography in acute ankle injuries. Ann Emerg Med 1992;21(4):384–90.

Stiell IG, Greenberg GH, Wells GA, McKnight RD, Cwinn AA, Cacciotti T, et al. Derivation of a decision rule for the use of radiography in acute knee injuries. Ann Emerg Med 1995;26(4):405–13.

Swets JA, Dawes RM, Monahan J. Better decisions through science. Sci. Am 2000;283(4):82–7.

Wells PS, Hirsh J, Anderson DR, Lensing AWA, Foster G, Kearon C, et al. Accuracy of clinical assessment of deep-vein thrombosis. The Lancet 1995;345(8961):1326–30.

Study 3 - Chapter 6 (p. 210)

Haskins, R., Osmotherly, P. G., Southgate, E., & Rivett, D. A. (2015). Australian physiotherapists' priorities for the development of clinical prediction rules for low back pain: a qualitative study. Physiotherapy, 101(1), 44-49.

ELSEVIER

Physiotherapy

Physiotherapy 101 (2015) 44–49

Australian physiotherapists' priorities for the development of clinical prediction rules for low back pain: A qualitative study

Robin Haskins [a,*], Peter G. Osmotherly [a], Erica Southgate [b], Darren A. Rivett [a]

[a] *School of Health Sciences, The University of Newcastle, University Drive, Callaghan, NSW 2308, Australia*
[b] *School of Education, The University of Newcastle, University Drive, Callaghan, NSW 2308, Australia*

Abstract

Objective To identify the types of clinical prediction rules (CPRs) for low back pain (LBP) that Australian physiotherapists wish to see developed and the characteristics of LBP CPRs that physiotherapists believe are important.
Design Qualitative study using semi-structured focus groups.
Setting Metropolitan and regional areas of New South Wales, Australia.
Participants Twenty-six physiotherapists who manage patients with LBP (77% male, 81% private practice).
Results Participants welcomed the development of prognostic forms of LBP CPRs. Tools that assist in identifying serious spinal pathology, likely responders to interventions, patients who are likely to experience an adverse outcome, and patients not requiring physiotherapy management were also considered useful. Participants thought that LBP CPRs should be uncomplicated, easy to remember, easy to apply, accurate and precise, and well-supported by research evidence. They should not contain an excessive number of variables, use complicated statistics, or contain variables that have no clear logical relationship to the dependent outcome. It was considered by participants that LBP CPRs need to be compatible with traditional clinical reasoning and decision-making processes, and sufficiently inclusive of a broad range of management approaches and common clinical assessment techniques.
Conclusion There were several identified areas of perceived need for LBP CPR development and a range of characteristics such tools need to encompass to be considered clinically meaningful and useful by physiotherapists in this study. Targeting and incorporating the needs and preferences of physiotherapists is likely to result in the development of tools for LBP with the greatest potential to positively impact clinical practice.

Keywords: Decision support techniques; Low back pain; Physical therapy specialty; Clinical prediction rules

Introduction

Clinical prediction rules (CPRs) are an aid to support clinical decision-making [1]. They are generally a simple predictive tool designed to be used with individual patients [2]. Unlike other forms of decision aids, CPRs most commonly provide a clinician with the quantified probability of a patient having a certain diagnosis or achieving a particular prognostic outcome [3]. CPRs come in many different formats and have been developed for a wide range of clinical problems. In some instances, CPRs provide an approach to stratified patient care, enabling treatments to be targeted to particular patient subgroups [4]. Over the past decade a growing number of CPRs have been developed within physiotherapy, with many relating to the management of low back pain (LBP) [5–8]. To date, such tools are remarkably diverse with little consistency in the type of clinical problems they aim to address. While the growth in the development of LBP CPRs is arguably important for the physiotherapy profession, the wide-ranging diversity in these tools may reflect a current lack of awareness about what clinicians actually want or need.

* Corresponding author. Tel.: +61 2 49223079; fax: +61 2 49217053.
E-mail addresses: Robin.Haskins@newcastle.edu.au (R. Haskins), Peter.Osmotherly@newcastle.edu.au (P.G. Osmotherly), Erica.Southgate@newcastle.edu.au (E. Southgate), Darren.Rivett@newcastle.edu.au (D.A. Rivett).

http://dx.doi.org/10.1016/j.physio.2014.04.005

There has been substantial dialogue in the recent physiotherapy literature regarding the appropriate methodology required to derive, validate and assess the impact of CPRs [9–14]. In contrast, there is a lack of literature about the types of problems for which CPRs should be developed or the characteristics and features they need to encompass to be considered useful by physiotherapists.

Given the substantial resources and time required to develop these tools, there is a need right from the preliminary stages of their development to ensure that CPRs will be accepted by clinicians and viewed as useful in addressing an important clinical problem [15]. Investigating and explicitly addressing clinician needs in the preliminary development of a CPR may be an important step in supporting the effective translation of CPR research evidence into clinical practice.

The aim of this study was to explore and describe the types and characteristics of LBP CPRs that are considered important by practicing physiotherapists working in the musculoskeletal field.

Method

Design

A qualitative descriptive method was employed to gain insight into physiotherapists' priorities for CPR development in relation to LBP. This method is intended to provide a clear description of a specific phenomenon or experience from the perspectives of research participants [16–18]. It concentrates on thematic analysis which seeks to identify common threads across participant perspectives in qualitative data [19]. Focus groups are commonly used in the early stages of product development to gain insight into the target consumers' thoughts and feelings about that product [20]. This approach is arguably well suited for exploring the needs and preferences of practising physiotherapists who are target clinical consumers of LBP CPRs. Four semi-structured focus groups, each lasting 1.5 to 2 hours and consisting of between 5 and 11 participants, were conducted across three geographic regions of New South Wales, Australia incorporating both metropolitan and regional areas. The first focus group was moderated by the third author (female, PhD, senior lecturer, experienced moderator) and the following three groups were moderated by the first author (male, B.Phty (Hons), physiotherapist, student of qualitative research methods). Groups were conducted outside of business hours to facilitate recruitment and held on a locally based university campus or in a private function centre. Ethical approval for the study was granted by the University of Newcastle's Human Research Ethics Committee.

Previous research in the field of Emergency Medicine [15,21] informed the development of a focus group schedule of questions. Participants were asked about areas of their practice with patients with LBP they thought may benefit from a CPR and the characteristics such tools require to be useful and meaningful. Clinicians were asked to share their beliefs on how LBP CPRs are most appropriately incorporated within physiotherapy practice and about any advice they would give to researchers who were considering developing LBP CPRs. Each focus group was recorded using a digital voice recorder. The audio file from each group was transcribed and used for data analysis.

Participants

Participants were recruited according to a purposive sampling framework [22] that would reflect the likely clinical consumers of LBP CPRs. Participants were selected according to the following characteristics: registered practising physiotherapist; working in public or private practice; having a caseload inclusive of patients with LBP; and proficiency in English. The study design deliberately included clinicians with a range of clinical experience from recent graduates to those with several decades of practice. Public listings were used to identify and recruit potentially eligible participants, in addition to an advertisement within an electronic bulletin e-mailed to all members of the Australian Physiotherapy Association.

Data analysis

Focus group transcriptions were uploaded to NVivo (Version 9, QSR International Pty Ltd, Doncaster, Victoria, Australia) and pseudonyms were substituted for participant names and places. Transcripts were read several times by the first author and then segments of text were inductively coded. Inductive coding is data-driven and is different from deductive coding in that there is minimal attempt to interpret the data through pre-existing categories derived from the literature [23]. Clusters of basic themes with commonality were arranged into organising themes (Figs. 1 and 2) [24]. Ongoing analysis with inductive coding and the development of these thematic levels [25] occurred over the course of the focus groups. Thematic saturation [23] occurred at the fourth focus group. Organising themes were then related to the study's research questions to develop a smaller number of organising themes for each of the research questions [24]. Ongoing analysis informed a decision to include a one page

Physiotherapists' priorities for LBP CPR development

Organising Themes	Basic Themes
Diagnosis	• Physiotherapists want LBP CPRs that enable the early and accurate identification of serious spinal pathology
	• There is limited desire for the development of CPRs that facilitate the sub-classification of non-specific LBP by pathoanatomic diagnosis
Intervention	• Physiotherapists want LBP CPRs that predict non-success, worsening or no need for intervention
	• Physiotherapists want LBP CPRs that accurately identify likely responders to intervention
Prognosis	• Physiotherapists have strong desire for LBP CPRs that accurately predict a patient's probable prognosis

Fig. 1. Summary of the types of LBP CPRs physiotherapists wish to see developed.

Physiotherapists' requirements of LBP CPRs

Organising Themes	Basic Themes
Application considerations	• LBP CPRs need to be simple, practical and easily applied
	• LBP CPRs should be developed and applied to clearly defined presentations
	• LBP CPRs need to be compatible with traditional clinical reasoning and decision-making strategies
Credibility and meaningfulness	• LBP CPRs need to make sense and should contain predictor variables that have a clear logical relationship with the dependent outcome
	• LBP CPRs need to be meaningful and clinically relevant
Performance expectations	• Physiotherapists require confidence that use of a LBP CPR will lead to improved patient outcomes
	• LBP CPRs need to be accurate to be useful

Fig. 2. Summary of physiotherapists' desired characteristics of LBP CPRs.

explanation describing CPRs for focus groups 3 and 4 to help direct participants to the topic of the focus group, that is CPRs.

To enhance credibility and confirmability, peer debriefing with other members of the research team was used throughout all stages of the data analysis [26,27]. Three members of the research team (RH, PO, DR) have previous research experience and a declared interest in the development of LBP CPRs. Member checking was employed by sending a summary of the research team's interpretation of the key study themes to participants and inviting and integrating their feedback into the study's findings [28]. Dependability was enhanced by confirmation of the accuracy of audio transcriptions [29] and cross-checking with field notes.

Results

Participants

Twenty-six physiotherapists participated in this study including three new-graduates (Table S1). The majority of participants were male (77%) and were working in a private clinical setting (81%). The average amount of clinical experience was 16 years (SD 11). Nine participants were previously known to the first author through various professional networks. No participants were previously known to the third author who moderated the first focus group. The use of pseudonyms and peer debriefing throughout data analysis minimised any risk of bias that could result from existing researcher–participant relationships.

Supplementary Table S1 can be found, in the online version, at http://dx.doi.org/10.1016/j.physio.2014.04.005.

Types of LBP CPRs that physiotherapists wish to see developed

Three organising themes were derived for the research question concerning the types of LBP CPRs physiotherapists wish to see developed (Fig. 1).

Diagnosis

Participants ($n = 9$) expressed a desire for CPRs to be developed that would enable them to accurately diagnose serious spinal pathology, as the following quote illustrates.

"It would be really nice if we could have a clinical prediction rule that would rule out the really heavily nasty stuff that can exist in spines... A little case story that I use to highlight is that I know a girl recently who's had a severe spinal cord injury with a history of having years and years of chiropractic treatment... There's nothing wrong that the chiropractor did in terms of his treatment... Unfortunately it was a tumour and it wasn't a musculoskeletal thing. But they can present as musculoskeletal things and to have a clinical prediction rule that was really good at being able to tell the difference would be nice." (Clive)

There was limited identified need for the development of CPRs that facilitate the sub-classification of non-specific LBP by pathoanatomic diagnosis. A few participants ($n = 5$) welcomed such tools, with similarities drawn with the classification of peripheral joint presentations. Other participants ($n = 13$), however, believed that they would have little impact on clinical practice as pathoanatomic diagnoses are not commonly influential in physiotherapy management decisions, as the following quote illustrates.

"I definitely wouldn't say that I use it (CPRs) in terms of identifying pathology, which is the medical model, which I don't think applies very well to physiotherapy. I think we're better to categorise people in groups of mechanical presentations." (Jeff)

Intervention

Participants ($n = 10$) wanted CPRs that would predict which patients with LBP would worsen from a particular intervention, or who would not benefit from physiotherapy treatment.

"It's also (helpful for CPRs to inform) when should we not treat this patient, and send them to someone, say they're a surgical candidate or whatever. That's how I would use a clinical prediction rule." (Emma)

"Rules that tell us when to lay off them (not treat patients with LBP) are good." (Brian)

LBP CPRs that identify patients who are more likely to achieve a successful outcome from a particular intervention were considered useful by the majority of participants ($n = 16$) and were thought to be helpful in informing treatment decision-making.

"It would be great to have an idea that if you do this treatment on this type of thing that you're going to get a great response. That would be fantastic rather than farting around (aimlessly continuing) with the stuff that you did for 10 years." (Adrian)

"I think if we had a prediction rule for treatment that said that this one generally gets better with this treatment... then that

definitely would prejudice my treatment (decision making) for sure." (Jerry)

Prognosis

Participants ($n = 20$) commonly expressed a view that CPRs which can accurately predict a patient's probable prognosis would be very valuable. Several physiotherapists ($n = 8$) expressed that determining a patient with LBP's likely prognosis was particularly challenging. Therefore, participants wanted CPRs that would predict time to recovery from a presentation, time to return to work, time to return to normal physical activity, likelihood of persisting symptoms, likelihood of requiring surgery, and the likelihood of experiencing a recurrence.

"Rules that talk about when they (patients with LBP) should return to work or not return to work – that would probably be a big (priority)." (Brian)

That's the horrible question that is asked to us. I think it's the hardest question of all, "how long?" (Jeff)

"I think that's where these prognostic models can actually help us." (Tim)

Physiotherapists' desired characteristics of LBP CPRs

Three organising themes were derived for the research question concerning physiotherapists' required characteristics of LBP CPRs to maximise their clinical utility (Fig. 2).

Application considerations

Participants felt that LBP CPRs must be simple and easy to apply for the benefit of both clinicians and patients ($n = 13$). Many considered it was important that tools focused on being practical and not being overly complicated by too many variables or complex statistics. The interface of the prediction tool needed to be compatible with the clinical environment and preferably the CPR could be able to be easily memorised. Some ($n = 5$) also expressed a view that predictor variables should be able to be obtained in a timely fashion using existing routine clinical measures and not require the need for sophisticated equipment.

"I think keeping them simple so practitioners can use them but also patients, in the sense that we're involving them in decision making." (Adam)

"I'd have to be able to use it straight up without looking at the computer and without researching it." (Ethan)

Physiotherapists also expressed a view that LBP CPRs would be more clinically useful if they were developed for very clear, well-defined patient presentations.

"I would say pick a clearly defined presentation, a typical well-defined presentation and work on that particular presentation, a predictor rule for that." (George)

"You have to really have a very specific subgroup and then you can make perhaps a rule that applies for that subgroup only, and that it makes sense." (Francis)

It was commonly reported by participants ($n = 19$) that LBP CPRs need to be compatible with their clinical reasoning and decision-making processes. CPRs were unlikely to be considered useful if they were viewed to be based on discordant management paradigms or insufficiently encompassing of a preferred management philosophy. Participants felt it was important that the developers of CPRs investigate the types of assessment techniques clinicians typically use and believe to be important in decision-making.

"It depends how broad based your clinical predication rules are. Whose models are they based on or which treatment or assessment philosophy are they based on?" (Terry)

"We follow clear clinical reasoning paths and... it has to fit into your clinical reasoning" (Emma)

Credibility and meaningfulness

Participants expressed a view that the selection of predictor variables in LBP CPRs should be based on sound clinical reasons and not solely on statistical procedures. A logical relationship (*e.g.*, biological plausibility) between predictor variables and the dependent outcome was considered by some ($n = 6$) to be very important, and if lacking, a potential threat to a rule's acceptance and implementation.

"To me it doesn't fit a particular model (hip rotation variable in spinal manipulation CPR), it just seems to be a statistical aberration that's popped up and they think oh, that's interesting. It would be like saying well if you've got red hair you're going to respond better to a manipulation than someone with brown hair. So it's got to match in with a clinical reasoning model I think." (George)

It was commonly ($n = 17$) believed that LBP CPRs need to address meaningful and important problems to be considered clinically useful. Examples of such problems identified by participants have been summarised in Fig. 1.

"From a face validity point of view, unless us, as clinicians see them as adding value to our day in day out practice we're unlikely to adopt them at this stage." (Tim)

Performance expectations

Participants commonly ($n = 16$) reported a view that evidence of a positive benefit on patient outcomes was required before a LBP CPR could be confidently used by physiotherapists in clinical practice.

"The idea (of LBP CPRs) I think sounds good if I can guarantee that I'm always getting the best likelihood of doing the best thing for this patient." (Phil)

"... I'm still waiting for some real juicy stuff that's going to be quick and easy to use but have a profound effect on our practice as well, consistently outperform me..." (Tim)

The accuracy and precision of predictions made by LBP CPRs was also considered by participants to be very important. Many participants ($n = 14$) felt that LBP CPRs needed to be highly accurate to be considered useful for clinical practice.

"I'd have to be able to be very, very confident, so really high (likelihood) ratios, to then apply it." (Ethan)

"You could have all the gut experience and thirty years' experience in the world but if this prediction model works... then you'd use it no matter who you think you were... because it can consistently outperform you..." (Tim)

Discussion

The successful translation of new knowledge into practice requires consideration and incorporation of the needs of clinicians [30]. This study is the first to explore the priorities of physiotherapists regarding the development of LBP CPRs. It has highlighted several areas of perceived need and a range of characteristics required of such tools to be useful to practicing physiotherapists.

One of the key findings from this study was that participants very commonly ($n = 20$) believed that prognostic CPRs would be helpful to their clinical practice. However, of the three major types of LBP CPRs (diagnostic, prescriptive and prognostic) that have been derived in physiotherapy to date, prognostic tools number the fewest [5]. In contrast, relatively few participants ($n = 5$) expressed a need for tools that facilitate the pathoanatomic sub-classification of non-specific LBP, but these have been derived in the greatest number [5]. Consistent with the identified needs of physiotherapists in this study, timely validation and impact assessment of existing derived prognostic LBP CPRs [31,32] may be warranted. Additionally, physiotherapists in this study would welcome CPRs for prognostic outcomes, such as the likely time to return to work and the probability of experiencing a recurrence.

The results of this study support the ongoing development of diagnostic LBP CPRs that facilitate the early and accurate identification of serious spinal pathology, such as vertebral fracture [33]. In addition, the ongoing development of prescriptive LBP CPRs that function to identify those patients with a relatively higher likelihood of success from a given intervention was supported by participants. Notably, physiotherapists in this study ($n = 10$) also expressed a need for tools that can identify patients who are likely to worsen from a given intervention, and also those who may not require physiotherapy management. To date, relatively little emphasis in the physiotherapy CPR literature has been placed upon the identification of these latter groups.

Participants in this study identified several modifiable properties of LBP CPRs that may enhance their clinical utility and meaningfulness. Tools that are uncomplicated, easy to remember, easy to apply and well-supported by research evidence were considered most useful. In contrast, negative attributes reported by participants in this study related to having a large number of variables, use of complicated statistics, or the inclusion of variables that have no clear logical relationship to the dependent outcome. Study participants expressed a view that LBP CPRs need to be compatible with traditional clinical reasoning and decision-making processes and sufficiently inclusive of a broad range of management approaches and common clinical assessment techniques. The accuracy and precision of LBP CPRs was also identified to be a consideration for some physiotherapists in this study ($n = 14$).

Some limitations of this study should be acknowledged. The findings represent the thoughts and opinions of study participants who practice physiotherapy in metropolitan and regional areas in New South Wales, Australia, and may not necessarily generalise to other clinician populations [20]. It was not possible to determine the degree to which the study sample is reflective of the wider potential participant population. The sample predominantly included males (77%) and those working within a private setting (81%), and it is not known whether such characteristics are related to the views expressed in this study. Particular care would therefore seem warranted in generalising the study's findings to non-comparable clinician populations. The sampling strategy aimed to include the intended clinical consumers of LBP CPRs and as such, no restrictions were placed on the baseline knowledge of study participants in regards to the study topic. It is plausible that participants who had greater prior awareness and familiarity with LBP CPRs may have held different views on the subject compared to those less knowledgeable. However, the primary aim of this research was to undertake an exploration of the range of views held by practicing physiotherapists and it was therefore determined *a priori* to not screen for inclusion based on subject knowledge. Finally, the question schedule was not formally evaluated prior to its implementation in the focus groups.

It is anticipated that the findings of this study will help researchers to develop LPB CPRs that have the greatest potential to positively influence physiotherapy clinical practice. Given the large time and resource commitment required to develop CPRs, targeting and incorporating the identified needs and preferences of the intended clinical users from the preliminary stages of development merits consideration, and may plausibly enhance the translation of research findings into clinical practice.

Ethical approval: University of Newcastle Human Research Ethics Committee (reference number H-2011-0032).

Funding: None declared.

Conflict of interest: None declared.

References

[1] Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. Ann Intern Med 2006;144:201–9.

[2] Beattie P, Nelson R. Clinical prediction rules: what are they and what do they tell us? Aust J Physiother 2006;52:157–63.

[3] McGinn T, Wyer P, Wisnivesky J, Devereauz PJ, Stiell I, Richardson S, *et al.* Advanced topics in diagnosis: clinical prediction rules. In: Guyatt G, Rennie D, Meade MO, Cook DJ, editors. Users' guides to the medical literature: a manual for evidence-based clinical practice. 2nd ed. USA: McGraw Hill Medical; 2008. p. 491–505.

[4] Foster NE, Hill JC, O'Sullivan P, Hancock M. Stratified models of care. Best Pract Res Clin Rheumatol 2013;27:649–61.

[5] Haskins R, Rivett DA, Osmotherly PG. Clinical prediction rules in the physiotherapy management of low back pain: a systematic review. Man Ther 2012;17:9–21.

[6] May S, Rosedale R. Prescriptive clinical prediction rules in back pain research: a systematic review. J Man Manip Ther 2009;17: 36–45.

[7] Beneciuk JM, Bishop MD, George SZ. Clinical prediction rules for physical therapy interventions: a systematic review. Phys Ther 2009;89:114–24.

[8] Stanton TR, Hancock MJ, Maher CG, Koes BW. Critical appraisal of clinical prediction rules that aim to optimize treatment selection for musculoskeletal conditions. Phys Ther 2010;90:843–54.

[9] Kent P, Hancock M, Petersen DH, Mjøsund HL. Clinimetrics corner: choosing appropriate study designs for particular questions about treatment subgroups. J Man Manip Ther 2010;18:147–52.

[10] Kamper SJ, Maher CG, Hancock MJ, Koes BW, Croft PR, Hay E. Treatment-based subgroups of low back pain: a guide to appraisal of research studies and a summary of current evidence. Best Pract Res Clin Rheumatol 2010;24:181–91.

[11] Kent P, Keating JL, Leboeuf-Yde C. Research methods for subgrouping low back pain. BMC Med Res Methodol 2010;10:62.

[12] Nee RJ, Coppieters MW. Interpreting research on clinical prediction rules for physiotherapy treatments. Man Ther 2011;16:105–8.

[13] Hancock M, Herbert RD, Maher CG. A guide to interpretation of studies investigating subgroups of responders to physical therapy interventions. Phys Ther 2009;89:698–704.

[14] Haskins R, Osmotherly PG, Tuyl F, Rivett DA. Uncertainty in clinical prediction rules: the value of credible intervals. J Orthop Sports Phys Ther 2014;44:85–91.

[15] Eagles D, Stiell IG, Clement CM, Brehaut J, Kelly A-M, Mason S, *et al.* International survey of emergency physicians' priorities for clinical decision rules. Acad Emerg Med 2008;15:177–82.

[16] Neergaard MA, Olesen F, Andersen RS, Sondergaard J. Qualitative description – the poor cousin of health research? BMC Med Res Methodol 2009;9:52.

[17] Sandelowski M. What's in a name? Qualitative description revisited. Res Nurs Health 2010;33:77–84.

[18] Magilvy JK, Thomas E. A first qualitative project: qualitative descriptive design for novice researchers. J Spec Pediatr Nurs 2009;14:298–300.

[19] Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: implications for conducting a qualitative descriptive study. Nurs Health Sci 2013;15:398–405.

[20] Krueger RA, Casey MA. Focus groups: a practical guide for applied research. 4th ed. Thousand Oaks, CA: Sage Publications, Inc.; 2009.

[21] Brehaut JC, Graham ID, Wood TJ, Taljaard M, Eagles D, Lott A, *et al.* Measuring acceptability of clinical decision rules: validation of the Ottawa acceptability of decision rules instrument (OADRI) in four countries. Med Decis Making 2010;30:398–408.

[22] Greenwood J, Parsons M. A guide to the use of focus groups in health care research: Part 2. Contemp Nurse 2000;9:181–91.

[23] Fade S, Swift J. Qualitative research in nutrition and dietetics: data analysis issues. J Hum Nutr Diet 2011;24:106–14.

[24] Attride-Stirling J. Thematic networks: an analytic tool for qualitative research. Qual Res 2001;1:385–405.

[25] Morse JM, Barrett M, Mayan M, Olson K, Spiers J. Verification strategies for establishing reliability and validity in qualitative research. Int J Qual Methods 2008;1:13–22.

[26] Petty NJ, Thomson OP, Stew G. Ready for a paradigm shift? Part 2: Introducing qualitative research methodologies and methods. Man Ther 2012;17:378–84.

[27] Creswell JW, Miller DL. Determining validity in qualitative inquiry. Theory Pract 2000;39:124–30.

[28] Mays N, Pope C. Qualitative research in health care: assessing quality in qualitative research. Br Med J 2000;320:50–2.

[29] Slade SC, Molloy E, Keating JL. People with non-specific chronic low back pain who have participated in exercise programs have preferences about exercise: a qualitative study. Austr J Physiother 2009;55: 115–21.

[30] Graham ID, Logan J, Harrison MB, Straus SE, Tetroe J, Caswell W, *et al.* Lost in knowledge translation: time for a map. J Contin Educ Health Prof 2006;26:13–24.

[31] Hancock MJ, Maher CG, Latimer J, Herbert RD, McAuley JH. Can rate of recovery be predicted in patients with acute low back pain? Development of a clinical prediction rule. Eur J Pain 2009;13: 51–5.

[32] George SZ, Bialosky JE, Donald DA. The centralization phenomenon and fear-avoidance beliefs as prognostic factors for acute low back pain: a preliminary investigation involving patients classified for specific exercise. J Orthop Sports Phys Ther 2005;35:580–8.

[33] Henschke N, Maher CG, Refshauge KM, Herbert RD, Cumming RG, Bleasel J, *et al.* Prevalence of and screening for serious spinal pathology in patients presenting to primary care settings with acute low back pain. Arthritis Rheum 2009;60:3072–80.

Clinical Commentary - Chapter 7 (p. 232)

Haskins, R., Osmotherly, P. G., Tuyl, F., & Rivett, D. A. (2014). Uncertainty in Clinical Prediction Rules: The Value of Credible Intervals. Journal of Orthopaedic & Sports Physical Therapy, 44(2), 85-91.

The published version of this clinical commentary was removed due to copyright restrictions.

Study 4 - Chapter 8 (p. 256)

Haskins, R., Osmotherly, P. G., & Rivett, D. A. (2015). Diagnostic clinical prediction rules for specific subtypes of low back pain: A systematic review. Journal of Orthopaedic & Sports Physical Therapy, 45(2), 61-76.

The published version of this study was removed due to copyright restrictions.

Study 5 - Chapter 9 (p. 304)

Haskins, R., Osmotherly, P. G., & Rivett, D. A. (2015). Validation and impact analysis of prognostic clinical prediction rules for low back pain is needed: a systematic review. Journal of Clinical Epidemiology, 68(7), 821-832.

ELSEVIER

CrossMark

Journal of Clinical Epidemiology 68 (2015) 821–832

Journal of Clinical Epidemiology

Validation and impact analysis of prognostic clinical prediction rules for low back pain is needed: a systematic review

Robin Haskins*, Peter G. Osmotherly, Darren A. Rivett

School of Health Sciences, University of Newcastle, University Drive, Callaghan, New South Wales 2308, Australia

Accepted 9 February 2015; Published online 14 February 2015

Abstract

Objectives: To identify prognostic forms of clinical prediction rules (CPRs) related to the nonsurgical management of adults with low back pain (LBP) and to evaluate their current stage of development.

Study Design and Setting: Systematic review using a sensitive search strategy across seven databases with hand searching and citation tracking.

Results: A total of 10,005 records were screened for eligibility with 35 studies included in the review. The included studies report on the development of 30 prognostic LBP CPRs. Most of the identified CPRs are in their initial phase of development. Three CPRs were found to have undergone validation—the Cassandra rule for predicting long-term significant functional limitations and the five-item and two-item Flynn manipulation CPRs for predicting a favorable functional prognosis in patients being treated with lumbopelvic manipulation. No studies were identified that investigated whether the implementation of a CPR resulted in beneficial patient outcomes or improved resource efficiencies.

Conclusion: Most of the identified prognostic CPRs for LBP are in the initial phase of development and are consequently not recommended for direct application in clinical practice at this time. The body of evidence provides emergent confidence in the limited predictive performance of the Cassandra rule and the five-item Flynn manipulation CPR in comparable clinical settings and patient populations.

Keywords: Clinical prediction rule; Low back pain; Prognosis; Decision support techniques; Systematic review; Derivation; Validation; Impact analysis

1. Introduction

The stratification of patients into meaningful subgroups is a priority area of low back pain (LBP) research [1]. Identifying patients with LBP with differing prognoses and targeting interventions based on the relative likelihood of treatment benefit provides individual and population-level benefits, including improved patient outcomes and efficiencies in resource consumption [2–6]. Clinical prediction rules (CPRs) are one of several overlapping methods proposed to facilitate such stratification [7].

CPRs are simple statistical prediction tools designed to be used with individual patients that comprise a small number of clinical variables that have been identified to be independently predictive of a given diagnosis, outcome, or treatment effect [8]. Prognostic forms of CPRs consist of nonspecific prognostic variables that inform predictions concerning future outcomes such as pain, disability, and return to work. Such tools are therefore well suited for screening and prioritizing patients for interventions and informing advice provided to patients and other parties regarding anticipated prognoses [7,9,10]. Prescriptive CPRs are a special type of prognostic CPR that inform predictions regarding the relative treatment effect a patient may experience from an intervention. The variables that comprise a prescriptive CPR are treatment effect modifiers, which are the baseline variables that differentiate patient subgroups who experience differing magnitudes of treatment effect [11–13]. Thus, prescriptive CPRs function to inform clinical decisions regarding treatment selection [9,14].

The development of a CPR broadly occurs across three main phases, whereby the tool is initially derived, then prospectively validated in new patient cohorts, and finally evaluated for its ability to positively impact clinical practice [15]. The validation of a CPR is important as predictor variables may simply reflect chance statistical associations or

Conflict of interest: None.

Funding: None.

* Corresponding author. Tel.: (61)-2-4922-3079; fax: (61)-2-4922-3995.

E-mail address: Robin.Haskins@uon.edu.au (R. Haskins).

http://dx.doi.org/10.1016/j.jclinepi.2015.02.003

0895-4356/

What is new?

Key findings

- Thirty prognostic clinical prediction rules relevant to the nonsurgical management of adults with low back pain have been derived. Three have also undergone validation, but none have undergone impact evaluation.

What this adds to what was known?

- Most clinical prediction rules for low back pain are in the initial phase of development and cannot be recommended for use in clinical practice at this time.

What is the implication and what should change now?

- The "Cassandra rule" is a clinical prediction rule that may be applied in comparable clinical settings and patient populations with some confidence in its modest prognostic accuracy in identifying patients with differing degrees of risk of developing a poor long-term functional outcome.
- The five-item Flynn manipulation clinical prediction rule may be applied in comparable clinical settings and patient populations to inform prognostic judgments about which patients receiving lumbopelvic manipulation are more likely to experience a favorable functional outcome.

the CPR may be specific to the study sample or setting in which it was derived [16]. CPRs that have been demonstrated to perform consistently across different patient groups and across broad clinical settings may be applied in practice with confidence in their accuracy [17]. Impact analysis is an important final step in the development of a CPR as it evaluates whether the implementation of a validated CPR is likely to have meaningful beneficial consequences [17,18]. Such benefits may include more accurate selection and prioritization of patients requiring intervention, improved patient outcomes, and reduced costs of care [7,9,19].

The limited data concerning the use of CPRs for LBP in clinical practice suggest that many clinicians have an awareness of such tools and consider their application in their clinical decision making [20–23]. Consequently, the identification of the range of existing prognostic CPRs for LBP, and an appraisal of their appropriateness to be applied in clinical practice at this time, is potentially of significant clinical benefit. Previous systematic reviews of CPRs relevant to the nonsurgical management of LBP have limited their scope to tools designed for specific interventions [12,24–27], a particular health profession [24,26–29], or to a particular stage of CPR development [24,26,27]. It is probable that many prognostic CPRs related to the nonsurgical management of LBP have not yet been identified in systematic reviews to date.

Therefore, the aim of this systematic review was to identify prognostic forms of CPRs related to the nonsurgical management of adults with LBP and to evaluate their current stage of development. It is anticipated that the evidence identified in this review will be informative to clinicians managing patients with LBP and to researchers involved in the development of LBP CPRs.

2. Methods

This systematic review sought to include studies reporting on the derivation, validation, or impact analysis of one or more prognostic or prescriptive CPRs related to the nonsurgical management of adults with LBP. A CPR was operationally defined as "a clinical tool that quantifies the individual contributions that various components of the history, physical examination and basic laboratory results make toward the diagnosis, prognosis, or likely response to treatment in an individual patient"[16]. Eligibility criteria were developed by the research team to address the review's research question and are summarized in Table 1. No restrictions were placed on the year of study publication, stage of CPR development, types of predictor variables under consideration (e.g., physical tests, history items, psychosocial factors, and so forth), types of nonsurgical management interventions, or the professional disciplines involved in the development of a CPR. CPRs were included independent of whether they were developed specifically for patients receiving a particular nonsurgical intervention.

The database search strategy (Appendix A at www.jclinepi.com) incorporated search strings identified to have high sensitivity for prognostic prediction model studies [30–32] and disease-specific filters for back-related disorders [33]. Seven databases were searched from their inception to July 2013: MEDLINE (1946–), EMBASE (1947–), Cochrane Central Register of Controlled Trials (1898–), PsychINFO (1806–), CINAHL (1937–), AMED (1985–), and Index to Chiropractic Literature (1981–). Identified records were downloaded into EndNote (Thomson Reuters), and duplicates were removed. Citation tracking and hand searching were conducted as supplementary search strategies.

Two independent reviewers selected studies for inclusion using a two-step process [34,35]. First, the titles and abstracts of identified records were screened by both reviewers with studies deemed eligible by either reviewer progressing to the second stage of screening. In the second stage, the full text of studies were screened by both reviewers with concordance determining eligibility. Episodes

Table 1. Study eligibility criteria

Inclusion criteria	
1	Reports on the derivation, validation, and/or impact analysis of one or more prognostic or prescriptive CPRs related to the nonsurgical management of adults with LBP
2	The CPR under development contains two or more predictor variables
3	The CPR under development was initiated by a formal derivation process in which a larger pool of candidate predictor variables was refined to a smaller set of variables based on their identified independent predictive value using formal multivariate statistical procedures
4	A tool is clearly presented in sufficient detail that may be applied by a clinician to predict a prognostic outcome or likelihood of treatment response in an individual patient
5	Published in English
Exclusion criteria	
1	Limited to the investigation of modifiable and/or determinant predictor variables
2	CPR not capable of directly contributing to patient care
3	Conference proceedings/abstracts, dissertations, commentaries, reviews, editorials, letters, study protocols, $n = 1$ designs (case reports), books, book chapters, book reviews, practice guidelines

Abbreviations: CPR, clinical prediction rule; LBP, low back pain.

of disagreement were resolved by consensus and, if needed, by a third independent reviewer.

A standardized tool was used for data extraction. Information regarding study design, patient population, CPR function, predictor variables, dependent outcomes, statistical analysis, tool format, tool performance, and the reporting of uncertainty intervals was recorded for each included study. The stage of tool development was defined as derivation, validation, or impact analysis using a well-recognized hierarchical CPR development framework [17]. Contingency tables for dichotomized outcomes were extracted, calculated, or approximated for specified cutoff points of a CPR where reported study data permitted. When not reported, the sensitivity, specificity, positive likelihood ratio (+LR), and negative likelihood ratio (−LR) with 95% confidence interval (CI) were calculated or approximated. Uncertainty intervals for posterior probabilities were calculated where study data permitted using the objective Bayesian method using Monte Carlo simulation [36].

The quality appraisal of included studies was conducted by examining the risk of bias relevant to both study design and the methodological factors specific to CPR development. All studies were initially evaluated using standardized appraisal tools relevant to their specific research designs. Prognostic studies were appraised using the Quality In Prognosis Studies (QUIPS) tool [37], and randomized controlled trials (RCTs) were appraised using the Physiotherapy Evidence Database (PEDro) scale [38]. The QUIPS tool was applied in instances in which the data from an RCT had been pooled across treatment groups for the development of a CPR. There is no standardized tool to appraise the methodological quality of factors specific to the development of all forms of CPRs at each stage of their development. Consequently, an updated version of a quality appraisal tool used in a previous systematic review on this topic was applied [28]. This tool was developed to be inclusive of the commonly represented quality criteria for all forms of CPRs at each stage of their development in well-cited methodological texts [15,17,39–41]. These criteria were updated in this review (Appendix B at www.jclinepi.com) to incorporate additional items identified in recent publications [27,29,36,42–44] and to exclude items that formed part of the eligibility criteria of this review (e.g., use of multivariate analysis). Quality appraisal of the included studies was conducted by two independent reviewers. Concordance between reviewers determined quality criterion status, with disagreement resolved by consensus, or if needed by a third independent reviewer. Negative status for a quality criterion was recorded in instances in which a study was found not to report evidence concerning that criterion.

The absolute and chance-corrected degree of interrater agreement were calculated for each stage of study selection. A sum quality score was calculated and reported for studies appraised with the PEDro tool [45]. Quality appraisals using all other instruments in this review are presented descriptively. The review was not eligible for protocol registration because of the lack of a specific intervention, exposure, or outcome measure under consideration. The article was prepared in accordance with the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines [46].

3. Results

The database search strategy yielded 12,347 records. A further 25 records were identified via hand searching and citation tracking. After the removal of duplicates, 10,005 records were screened via title and abstract for eligibility with 352 records advancing to the second stage of screening. The full texts of these studies were reviewed, and 35 were determined to be eligible [47–81]. As illustrated in Fig. 1, the most common reason for a study's exclusion was not satisfying the study's operational definition of a CPR.

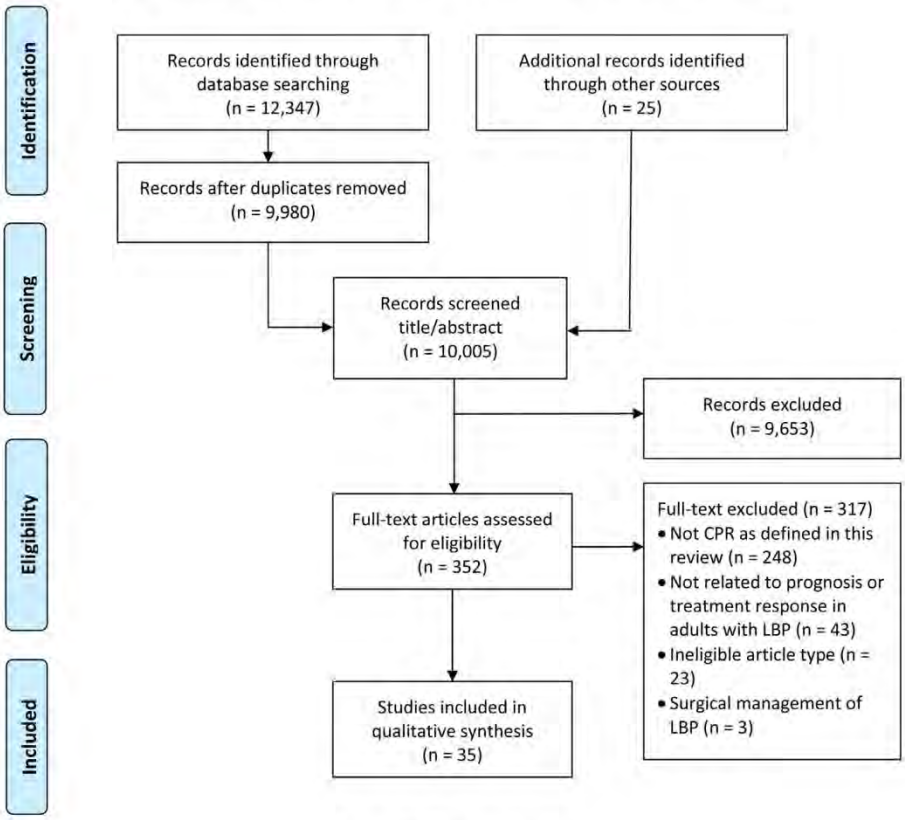


Fig. 1. PRISMA flowchart.

Interrater agreement for the first and second stages of screening was 98% and 93%, respectively. The chance-corrected degree of agreement was "moderate" [82] for both stages with $\kappa = 0.51$ (95% CI: 0.46, 0.57) for the screening by titles and abstracts and $\kappa = 0.59$ (95% CI: 0.45, 0.73) for the screening by full text. Of the 26 episodes of disagreement in the second stage of screening, all but two cases were resolved by consensus, with a third reviewer later including one of these studies.

Thirty CPRs were identified in this review and are summarized in Supplementary Material 1/Appendix at www.jclinepi.com. Most ($n = 20$) of the identified CPRs were derived using populations of patients receiving a specific treatment program. Functional outcomes were modeled as the dependent variable in the derivation of 15 CPRs. Work-related outcomes ($n = 5$), pain intensity ($n = 4$), recovery ($n = 3$), symptom persistence ($n = 2$), and need for surgical intervention ($n = 1$) were used as the dependent outcomes in the derivation of the remaining CPRs. No CPR included in this review selected variables for inclusion in the tool by examining tests of interaction to identify effect modifiers. Three CPRs were identified to have undergone validation in one or more studies. No impact analysis studies were identified.

Appendices C and D at www.jclinepi.com summarize the methodological quality of included studies appraised using the QUIPS and the PEDro scales, respectively. Quality scores on the PEDro scale ranged from 5 to 9. Quality appraisal of the included derivation studies (Table 2) identified the following items as the most frequent sources of potential bias: lack of blinded outcome assessment, no assessment of collinearity of predictor variables, no justification for the number of study participants, no reporting of uncertainty intervals for posterior probability estimates, insufficient reporting on the reliability of predictor variables, insufficient number of outcome events per candidate predictor variable, and lack of justification for the selection of candidate predictor variables. Table 3 lists the quality appraisal of included validation studies. The most common sources of potential bias were lack of assessment of the interobserver reliability of the CPR, no assessment/reporting of the accuracy of the tool, not having complete follow-up, and no reporting of uncertainty intervals for posterior probability estimates.

4. Discussion

4.1. Characteristics of included studies

This systematic review identified 35 studies reporting on the development of nonsurgical prognostic/prescriptive forms of LBP CPRs. It builds on the existing body of

literature with the inclusion of 13 studies that have not been previously reported on in earlier reviews [12,24–29] on this topic. Thirty CPRs were identified with three of these tools known to have progressed to validation—the "Cassandra rule" for predicting which patients with LBP are more likely to develop long-term significant functional limitations [52,73,74], and the five-item [54,68–72,77–80] and two-item [55,75,76] Flynn manipulation CPRs designed to predict which patients being treated with lumbopelvic manipulation are more likely to experience a favorable prognosis.

Functional outcomes were the most common form of dependent variable used in the derivation of the identified CPRs. In contrast, pain and symptom resolution outcomes were much less frequently used. Given that the performance of prognostic factors may not necessarily generalize across different types of dependent outcomes [72], it is important that the outcomes used in CPR derivation studies are selected based on the clinical problems they aim to address. However, very little research has been conducted to date on the types of clinical problems for which LBP CPRs should be developed [83]. It is not known if the range of CPRs included in this review would be considered useful by clinicians who treat patients with LBP, although preliminary evidence suggests that the five-item Flynn manipulation CPR has already been adopted by some practising clinicians [20,23].

4.2. The Cassandra rule

The "Cassandra rule" was derived in a population of patients with back pain presenting to primary care physicians and aims to identify individuals with differing degrees of risk of developing long-term significant functional limitations [52]. The CPR uses a measure of depression and a measure of somatization from selected items of the Symptoms Checklist 90 Revised questionnaire [84] to stratify patients by their degree of risk of having 50% or greater disability on the Roland-Morris Disability questionnaire [85] at 2 years. The "Cassandra rule" was statistically validated using a split sample in the derivation study and subsequently validated in a prospective cohort of 860 patients absent from work due to LBP consulting in primary care [73]. It was further validated in a prospective cohort of 1,262 patients presenting to an emergency department with nonspecific back pain [74]. The prevalence of the dependent outcome was very similar across all three studies, ranging from just 16% to 19%. For the dichotomized groups of high/moderate vs. low risk of the dependent outcome, the point estimate of the +LR and −LR ranged from 1.3 to 2.0 and 0.25 to 0.40, respectively, across the three studies. The relative consistency in the observed accuracy of this CPR to date provides preliminary confidence that it may perform similarly in other comparable clinical settings particularly in those settings with a similar prevalence of the dependent outcome. When comparing the "Cassandra rule" to physician prediction, the CPR was found to be more sensitive (82% vs. 37%) but less specific (45% vs. 85%) in identifying those at risk of a poorer functional outcome [74]. This finding, in addition to the relative magnitudes of the +LR and −LR, suggests that "negative" status on the prediction tool may be more informative to a clinician's prognostic judgment. No studies were identified that examined the utility of this CPR in improving clinical outcomes or resource efficiency. A more parsimonious version of the Cassandra rule comprising just five items, including one new question, has been developed and demonstrated to perform similarly to the original tool [74]. However, the development of this updated tool did not meet the criteria for inclusion in the present review.

4.3. Five-item Flynn manipulation CPR

Ten studies were included in this review regarding the development of a five-item CPR for identifying patients receiving manipulation who are more likely to experience a favorable functional outcome. Flynn et al. [54] derived the tool in a population of patients with LBP presenting to military outpatient physical therapy facilities who were treated with lumbopelvic thrust manipulation (in a supine-lying position), range of motion exercises, and advice to maintain usual activity within the limits of their pain. A successful outcome was defined as more than a 50% improvement on the modified Oswestry Disability Questionnaire [86] by the third treatment session, which occurred up to 8 days after the initial treatment. Nine validation studies were identified that investigated the generalizability of this CPR to other interventions and dependent outcomes and explored whether the CPR is a treatment effect modifier.

In patients receiving thrust and nonthrust lumbopelvic manipulation, there is evidence to support positive baseline status on the Flynn manipulation CPR (most commonly defined as $\geq$4 predictors present) as a predictor of reduced disability [54,68,72,77,79], improved pain [72,77], greater patient-perceived extent of recovery [72], and less treatment sessions [72]. The predictive value of a patient's baseline CPR status has, however, been demonstrated to be sensitive to the threshold used to define a successful outcome. For example, Schwind et al. [79] found that a patient's baseline status on the Flynn CPR was a predictor of successful outcome when using a definition of success of at least 50% improvement in the Oswestry Disability Questionnaire but was no longer a significant independent predictor of success when the definition was changed to at least 30% improvement on the same questionnaire. One of the five variables in the Flynn manipulation CPR, no pain distal to the knee, was also identified as an independent predictor of reduced disability in the derivation of a prognostic CPR involving postpartum women with lumbopelvic pain receiving thrust manipulation [48].

Four validation studies limited their study sample to participants assessed as being positive on the Flynn

Table 2. Methodological appraisal of included derivation studies

Study	CPR reference number	Prospective design	Study site described	Justification for number of participants	Representative sample	Important patient characteristics described	Candidate predictor variables justified	Blinded predictor assessment	Predictor variables have demonstrated reliability
Alonso-Blanco et al. 2009 [47]	10	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Al-Sayegh et al. 2010 [48]	12	Yes	Partly	Yes	Partly	Yes	No	Yes	No
Al-Sayegh et al. 2010 [48]	13	Yes	Partly	Yes	Partly	Yes	No	Yes	No
Buranapanitkit et al. 2003 [49]	23	No	Yes	No	No	Yes	No	No	No
Cai et al. 2009 [50]	11	Yes	Yes	No	Yes	Yes	Partly	Yes	Partly
Cairns et al. 1984 [51]	26	No	Yes	No	Yes	Partly	No	No	No
Cairns et al. 1984 [51]	27	No	Yes	No	No	Partly	No	No	No
Dionne et al. 1997 [52]	2	Yes	Yes	No	Partly	Yes	Partly	Yes	No
Dionne et al. 2005 [53]	28	Yes	Yes	No	Yes	Yes	Yes	Yes	Partly
Flynn et al. 2002 [54]	3	Yes	Yes	No	Yes	Yes	No	Yes	No
Fritz et al. 2005 [55]	4	Yes	Yes	No	Yes	Yes	Yes	Yes	Partly
George et al. 2005 [56]	5	Yes	Yes	No	No	Yes	Yes	Yes	Yes
George et al. 2005 [56]	16	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Hancock et al. 2009 [57]	22	Yes	Yes	No	Yes	Yes	No	Yes	Partly
Hewitt et al. 2007 [58]	7	Yes	Yes	No	Yes	Yes	Yes	Yes	No
Hewitt et al. 2007 [58]	17	Yes	Yes	No	Yes	Yes	Yes	Yes	No
Heymans et al. 2007 [59]	29	No	No	No	Partly	Yes	Yes	No	No
Heymans et al. 2009 [60]	30	Yes	Partly	No	Partly	Yes	Yes	Yes	No
Hicks et al. 2005 [61]	6	Yes	Yes	No	Yes	Yes	Partly	Yes	Yes
Kovacs et al. 2012 [62]	14	Yes	Yes	Yes	Yes	Yes	No	Yes	No
Kovacs et al. 2012 [62]	18	Yes	Yes	Yes	Yes	Yes	No	Yes	No
Kovacs et al. 2012 [62]	19	Yes	Yes	Yes	Yes	Yes	No	Yes	No
Malmqvist et al. 2008 [63]	21	Yes	Yes	No	Partly	Yes	Yes	Yes	No
May et al. 2008 [64]	8	Yes	Yes	No	Yes	No	Partly	Yes	Partly
May et al. 2008 [64]	9	Yes	Yes	No	Yes	No	Partly	Yes	Partly
Roland et al. 1983 [65]	1	Yes	Yes	No	Partly	Partly	No	Yes	No
Roland et al. 1983 [65]	20	Yes	Yes	No	Partly	Partly	No	Yes	No
Stolze et al. 2012 [66]	15	Yes	Yes	Yes	Yes	Yes	Partly	Yes	Partly
Thomas et al. 1999 [81]	25	Yes	Yes	No	No	Yes	Partly	Yes	No
Valat et al. 2000 [67]	24	Yes	Partly	No	Partly	Yes	No	Yes	No

Abbreviation: CPR, clinical prediction rule.

manipulation CPR to investigate the clinical outcomes achieved from interventions different to that used in the original derivation study. Cleland et al. [70] found that 11 of 12 participants treated with a side-lying thrust manipulation technique achieved 50% or greater improvement in disability at 1 week. A subsequent RCT found similar improvements in pain and function in patients treated with supine-lying or side-lying thrust manipulation but inferior outcomes in those treated with nonthrust manipulation [71]. Two further RCTs identified no significant differences in outcomes between those treated with supine-lying thrust manipulation and neutral gap thrust manipulation [80] or mechanical diagnosis and therapy [78]. These findings suggest that in patients positive on the Flynn manipulation CPR, similar clinical outcomes may be achieved with some alternative interventions.

Two high-quality validation studies included in this review investigated whether a patient's baseline status on the Flynn manipulation CPR is a treatment effect modifier. Childs et al. [68] identified CPR status as a significant effect modifier of thrust manipulation compared with an exercise-based intervention for the outcomes of pain and disability. In contrast, Hancock et al. [77] did not identify CPR status as a significant treatment effect modifier of spinal manipulation compared with sham ultrasound for the same outcomes. There are several differences between these two studies that may account for their conflicting findings. In particular, all patients in the study of Childs et al. received thrust manipulation compared with just 5% in the study of Hancock et al. The comparison interventions were also notably different, and this may plausibly contribute to the observed differences in findings. As such, the findings of this review provide limited evidence supporting baseline status on the Flynn manipulation CPR as an effect modifier of thrust lumbopelvic manipulation compared with an exercise-based intervention for the outcomes of pain and disability.

4.4. Two-item Flynn manipulation CPR

A two-item variation of the Flynn manipulation CPR was proposed by Fritz et al. [55] using the variables related to duration and distribution of symptoms. Using pooled data from patients with LBP receiving thrust manipulation in two previous studies [54,68], the two-item CPR was found to classify patients the same as the five-item rule in 84%

Outcome measure valid and reliable	Blinded outcome assessment	Mathematical techniques described	Reporting and handling of missing data	10 Outcome events per variable in final model	10 Outcome events per candidate variable	Collinearity of predictor variables assessed	Predictor variables kept continuous	Uncertainty in CPR accuracy described	Uncertainty in posttest probability described	Nonparadoxical performance
Yes	No	Yes	Partly	No	No	No	No	Yes	No	No
Yes	No	Yes	No	No	No	No	N/A	Yes	No	Yes
Yes	No	Yes	No	No	No	No	No	Yes	No	Yes
No	No	Yes	No	Yes	No	No	N/A	N/A	No	Yes
Yes	No	Yes	No	No	No	No	No	Yes	No	Yes
No	No	Yes	No	N/A	N/A	No	Yes	No	N/A	N/A
No	No	Yes	No	N/A	N/A	No	Yes	No	N/A	N/A
Partly	No	Yes	Partly	N/A	N/A	N/A	N/A	N/A	No	Yes
Partly	No	Yes	Partly	N/A	N/A	N/A	N/A	No	Yes	Yes
Partly	No	Yes	Partly	No	No	No	No	Yes	No	Yes
Yes	No	Yes	No	No	No	No	N/A	Yes	No	Yes
Yes	No	Yes	No	No	No	No	Yes	No	N/A	Yes
Yes	No	Yes	No	Yes	No	No	Yes	No	N/A	Yes
No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes
Yes	No	Yes	Partly	Yes	Yes	No	Yes	No	N/A	Yes
No	No	Yes	Partly	Yes	Yes	No	Yes	No	N/A	Yes
No	No	Yes	Yes	Yes	No	No	Yes	No	No	No
No	No	Yes	Yes	Yes	No	No	Yes	No	No	Yes
Yes	Yes	Yes	Partly	No	No	No	No	Yes	No	No
Yes	No	Yes	Yes	Yes	Yes	No	Yes	N/A	No	Yes
Partly	No	Yes	Yes	Yes	Yes	No	Yes	N/A	No	Yes
Partly	No	Yes	Yes	Yes	Yes	No	Yes	N/A	No	Yes
Partly	No	Yes	Yes	Yes	Yes	No	N/A	N/A	No	Yes
Partly	No	Yes	Yes	No	No	No	N/A	N/A	No	Yes
Partly	No	Yes	Yes	No	No	No	N/A	N/A	No	Yes
Partly	No	Yes	Partly	Yes	No	Partly	No	N/A	No	Yes
No	No	Yes	Partly	Yes	No	Partly	No	N/A	No	Yes
Partly	No	Yes	Yes	No	No	No	Yes	Yes	Yes	No
Partly	No	Yes	Partly	No	No	No	No	N/A	No	Yes
No	No	Yes	Partly	Yes	No	No	No	N/A	No	Yes

of cases. The +LR for the two-item CPR for the dichotomized outcome of more than 50% reduction in disability by the third treatment (7.2) indicates that it would have a "moderate" [87] influence on shifting the pretest probability. Based on the reported data, similar patients receiving this intervention who have both criteria present would have an 85% (95% credible interval: 71%, 93%) probability of achieving this outcome. Two validation studies of the two-item CPR were included in this review [75,76]; however, both studies were conducted using only patients assessed as being positive on the rule and therefore provide limited evidence concerning the rule's accuracy in identifying patients with differing likelihoods of experiencing improvement. Using a retrospective database review, Fritz et al. [75] identified that patients with both criteria present and treated with thrust or nonthrust manipulation experienced greater clinical improvements compared with those treated without manipulation. Furthermore, those treated with thrust manipulation achieved similar clinical outcomes more efficiently than those treated with nonthrust manipulation. Hallegraef et al. [76] reported similar findings with patients with both criteria present and randomized to receive thrust manipulation experiencing small but statistically significant greater improvements in disability compared with those randomized to receive nonmanipulative physical therapy care.

4.5. Methodological considerations

No studies in this review selected predictor variables for inclusion in a CPR based on their identified function as a treatment effect modifier. This, however, would be an important methodological consideration in the development of a CPR designed to identify likely responders to a given intervention [11,12]. Although most studies (n = 20) included in this review sampled patients receiving a specific treatment program, it is not known if the predictor variables included in the derived CPRs reflect predictors of response to treatment or simply nonspecific predictors of an outcome independent of the treatment [9]. For example, the evidence summarized in this review suggests that patients found positive on the five-item Flynn manipulation CPR who receive lumbopelvic manipulation are more likely to experience functional improvement compared with those found negative on the rule. However, with the exception of one study [68], it is not yet clear if these patients are more likely to improve irrespective of the treatment provided.

Table 3. Methodological appraisal of included validation studies

Study	CPR reference number	Prospective validation in new sample	Different clinical setting	Representative sample	The rule is applied accurately	Reliability of the rule is assessed	Complete follow-up	Reporting and handling of missing data	Accuracy of the rule described	Accuracy uncertainty described	Posttest probability uncertainty described
Childs et al. 2004 [68]	3	Yes	Partly	Yes	Yes	No	No	Yes	Yes	Yes	No
Childs et al. 2006 [69]	3	Yes	Partly	Yes	Yes	No	No	Yes	No	N/A	No
Cleland et al. 2006 [70]	3	Yes	Yes	Partly	Yes	No	Yes	Yes	No	N/A	No
Cleland et al. 2009 [71]	3	Yes	Yes	No	Yes	No	No	Yes	No	N/A	No
Cook et al. 2013 [72]	3	Yes	Yes	Yes	Yes	No	Partly	Partly	Yes	Yes	N/A
Dionne 2005 [73]	2	Yes	Yes	Yes	Yes	No	No	Yes	Yes	No	No
Dionne et al. 2011 [74]	2	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Fritz et al. 2006 [75]	4	No	Yes	Partly	Yes	No	Yes	Yes	No	N/A	N/A
Hallegraeff et al. 2009 [76]	4	Yes	Yes	No	Yes	No	N/A	No	No	N/A	No
Hancock et al. 2008 [77]	3	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	N/A
Schenk et al. 2012 [78]	3	Yes	Yes	Partly	Yes	No	No	Yes	No	N/A	No
Schwind et al. 2013 [79]	3	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	N/A
Sutlive et al. 2009 [80]	3	Yes	No	No	Yes	No	No	Yes	No	N/A	N/A

Abbreviation: CPR, clinical prediction rule.

The methodological appraisal of included studies identified several opportunities to reduce potential sources of bias in future CPR development studies. The selection of candidate predictor variables needs to be logically justified and considered within the context of probable predictive performance, psychometric properties, and practicality [27,42,44,83]. The study sample size should also be justified and sufficiently large to ensure at least 10 outcome events per candidate predictor variable [29,43]. The common practice of univariate screening to cull the number of predictors entered into a multivariate model is not effective in reducing the risk of overfitting and may subsequently lead to an increased chance of spurious findings [88]. Predictor variables should also be assessed for collinearity to reduce the likelihood of paradoxical CPR performance whereby the probability of an outcome decreases with increasing positive status on the rule [27]. This paradoxical performance was observed in four studies included in this review with none of these studies reporting on whether collinearity was assessed (Table 2).

When validating a CPR, researchers should seek to include patients across the full range of possible categories of the tool. By only including patients who represent one particular status on the CPR (e.g., only those considered "positive"), the performance of the tool in discriminating between patients with differing likelihoods of achieving the dependent outcome is unable to be evaluated. The interobserver reliability of a patient's status on a CPR is a potential threat to its validity [39,40] and ideally should be evaluated and reported similar to accepted standards for single-item tests [89]. Finally, during all phases of a CPR's development, the reporting of uncertainty intervals for outcome prevalence, CPR accuracy (e.g., sensitivity, specificity, +LR, −LR), and posterior probabilities would enable a more meaningful interpretation of a study's findings [27,36,42].

4.6. Limitations

There are limitations of this study that need to be acknowledged. An operational definition of an LBP CPR was developed for this study by the research team to facilitate reproducibility and to transparently detail the types of studies under review. In particular, we aimed to differentiate studies that clearly presented a prediction tool that could be reasonably applied by a clinician for an individual patient from other forms of statistical prediction models. It is anticipated, however, that others may have differing views on the sorts of tools that should be considered CPRs. Our definition lead to the exclusion of 11 studies [4,90–99] that were included in earlier related reviews on this subject [24–26,28,29]. Our definition of a CPR required the use of multivariate statistics to derive predictor variables. This lead to the exclusion of some tools that have been

previously called CPRs. For example, a tool developed by Hicks et al. [61] to identify patients participating in a stabilization exercise program who are likely to experience improvement was excluded as the predictors in this tool were selected based on their univariate statistical association with the dependent outcome. Our definition of a CPR also excluded predictive tools such as the STarT Back Screening Tool [100], which sought to only include potentially modifiable prognostic factors.

A highly sensitive search strategy across multiple databases was used to identify potentially eligible studies. However, the nomenclature used to describe CPRs is varied, and it is therefore possible that some studies were inadvertently omitted. Our supplementary search strategies identified an additional 25 studies that were considered for eligibility, with four of these studies advancing to inclusion in the review. This highlights the importance of supplementary search strategies (such as hand searching and citation tracking) in the identification of CPR development studies.

The criteria used in this review to appraise the methodological quality of derivation and validation CPR development studies have not been validated. They were based on standards commonly reported in well-cited CPR methodological texts [15,17,39–41] and methodological items identified in recent publications [27,29,36,42–44]. Many of the criteria used in the present review have, however, been used in earlier previous systematic reviews on CPRs. The QUIPS and PEDro scales were used in this review to complement the methodological appraisal of included studies.

4.7. Conclusion

In conclusion, this systematic review identified 30 prognostic/prescriptive CPRs relevant to the nonsurgical management of adults with LBP. Most have not yet undergone validation and therefore cannot be recommended for use in clinical practice at this time. Clinicians may, however, consider using knowledge of the identified individual predictors that comprise these tools to cautiously inform their prognostic clinical judgments [16]. Three CPRs included in this review have been identified to have undergone validation. The "Cassandra rule" has been validated in two large prospective studies and may be applied in comparable clinical settings with similar patient populations with some confidence in its modest prognostic accuracy. It is not yet known if the clinical application of the "Cassandra rule" results in improved patient outcomes or improvements in resource efficiency. Positive status on the five-item Flynn manipulation CPR has been demonstrated in several studies to be a predictor of reduced disability in patients receiving thrust and nonthrust forms of lumbopelvic manipulation. It is not yet clear, however, if a patient's status on this CPR predicts a more favorable prognostic outcome irrespective of the treatment provided. No evidence was found that addressed whether the clinical application of the Flynn manipulation CPR results in improved patient outcomes or more efficient care. A two-item variation of the Flynn manipulation CPR has undergone validation in two studies. However, in both studies, patients who were deemed negative on the rule were excluded, thereby precluding an evaluation of the rule's predictive accuracy in identifying patients with differing likelihoods of experiencing improvement. More research seeking to validate the derived CPRs identified in this review is warranted. Research evaluating the clinical impact of the application of the "Cassandra rule" and the five-item Flynn manipulation CPR is also needed.

Acknowledgments

The authors thank Dane P Fehlberg for his assistance with screening and appraising the literature.

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jclinepi.2015.02.003.

References

[1] Costa Lda C, Koes BW, Pransky G, Borkan J, Maher CG, Smeets RJ. Primary care research priorities in low back pain: an update. Spine 2013;38:148–56.
[2] Foster NE, Mullis R, Hill JC, Lewis M, Whitehurst DGT, Doyle C, et al. Effect of stratified care for low back pain in family practice (IMPaCT back): a prospective population-based sequential comparison. Ann Fam Med 2014;12:102–11.
[3] Hill JC, Whitehurst DGT, Lewis M, Bryan S, Dunn KM, Foster NE, et al. Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. Lancet 2011;378:1560–71.
[4] Brennan GP, Fritz JM, Hunter SJ, Thackeray A, Delitto A, Erhard RE. Identifying subgroups of patients with acute/subacute "nonspecific" low back pain: results of a randomized clinical trial. Spine 2006;31:623–31.
[5] Long A, Donelson R, Fung T. Does it matter which exercise?: a randomized control trial of exercise for low back pain. Spine 2004;29:2593–602.
[6] Whitehurst DG, Bryan S, Lewis M, Hill J, Hay EM. Exploring the cost–utility of stratified primary care management for low back pain compared with current best practice within risk-defined subgroups. Ann Rheum Dis 2012;71:1796–802.
[7] Foster NE, Hill JC, O'Sullivan P, Hancock M. Stratified models of care. Best Pract Res Clin Rheumatol 2013;27:649–61.
[8] Randolph AG, Guyatt GH, Calvin JE, Doig G, Richardson WS. Understanding articles describing clinical prediction tools. Crit Care Med 1998;26:1603–12.
[9] Hill JC, Fritz JM. Psychosocial influences on low back pain, disability, and response to treatment. Phys Ther 2011;91:712–21.
[10] Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? BMJ 2009;338:b375.
[11] Hancock M, Herbert RD, Maher CG. A guide to interpretation of studies investigating subgroups of responders to physical therapy interventions. Phys Ther 2009;89:698–704.
[12] Stanton TR, Hancock MJ, Maher CG, Koes BW. Critical appraisal of clinical prediction rules that aim to optimize

treatment selection for musculoskeletal conditions. Phys Ther 2010;90:843–54.
[13] Kraemer HC, Frank E, Kupfer DJ. Moderators of treatment outcomes: clinical, research, and policy importance. JAMA 2006; 296:1286–9.
[14] Cook C. Potential pitfalls of clinical prediction rules. J Man Manip Ther 2008;16:69–71.
[15] Childs JD, Cleland JA. Development and application of clinical prediction rules to improve decision making in physical therapist practice. Phys Ther 2006;86:122–31.
[16] McGinn T, Wyer P, Wisnivesky J, Devereauz PJ, Stiell I, Richardson S, et al. Advanced topics in diagnosis: clinical prediction rules. In: Guyatt G, Rennie D, Meade MO, Cook DJ, editors. Users' guides to the medical literature a manual for evidence-based clinical practice second edition. USA: McGraw Hill Medical; 2008:491–505.
[17] McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS. Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-Based Medicine Working Group. JAMA 2000;284:79–84.
[18] Toll D, Janssen K, Vergouwe Y, Moons K. Validation, updating and impact of clinical prediction rules: a review. J Clin Epidemiol 2008; 61:1085–94.
[19] Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. Ann Intern Med 2006;144:201–9.
[20] Learman K, Showalter C, Cook C. Does the use of a prescriptive clinical prediction rule increase the likelihood of applying inappropriate treatments? A survey using clinical vignettes. Man Ther 2012; 17:538–43.
[21] Willett G, Johnson G, Jones K. The effect of a hybrid continuing education course on outpatient physical therapy for individuals with low back pain. Internet J Allied Health Sci Pract 2011;9:1–11.
[22] Haskins R, Osmotherly PG, Southgate E, Rivett DA. Physiotherapists' knowledge, attitudes and practices regarding clinical prediction rules for low back pain. Man Ther 2014;19:142–51.
[23] Sparks CL, McGehee WF, Buettner MA, Scott KM. Use of a clinical prediction rule for low back pain by outpatient physical therapists. J Man Manip Ther 2010;18:237.
[24] Beneciuk JM, Bishop MD, George SZ. Clinical prediction rules for physical therapy interventions: a systematic review. Phys Ther 2009; 89:114–24.
[25] May S, Rosedale R. Prescriptive clinical prediction rules in back pain research: a systematic review. J Man Manip Ther 2009;17: 36–45.
[26] Patel S, Friede T, Froud R, Evans DW, Underwood M. Systematic review of randomized controlled trials of clinical prediction rules for physical therapy in low back pain. Spine 2013;38:762–9.
[27] Lubetzky-Vilnai A, Ciol M, McCoy SW. Statistical analysis of clinical prediction rules for rehabilitation interventions: current state of the literature. Arch Phys Med Rehabil 2014;95:188–96.
[28] Haskins R, Rivett DA, Osmotherly PG. Clinical prediction rules in the physiotherapy management of low back pain: a systematic review. Man Ther 2012;17:9–21.
[29] van Oort L, van den Berg T, Koes BW, de Vet RHCW, Anema HJR, Heymans MW, et al. Preliminary state of development of prediction models for primary care physical therapy: a systematic review. J Clin Epidemiol 2012;65:1257–66.
[30] Geersing G-J, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons K. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. PLoS One 2012;7:e32844.
[31] Ingui BJ, Rogers MA. Searching for clinical prediction rules in MEDLINE. J Am Med Inform Assoc 2001;8:391–7.
[32] Holland JL, Wilczynski NL, Haynes RB, The Hedges Team. Optimal search strategies for identifying sound clinical prediction studies in EMBASE. BMC Med Inform Decis Mak 2005;5:11.
[33] Bombardier C, van Tulder M, Bronfort G, Chou R, Corbin T, Deyo R, et al. Cochrane Back Group. About the Cochrane Collaboration (Cochrane Review Groups (CRGs)). 2014;Issue 1. Art. No.: BACK.
[34] Edwards P, Clarke M, DiGuiseppi C, Pratap S, Roberts I, Wentz R. Identification of randomized controlled trials in systematic reviews: accuracy and reliability of screening records. Stat Med 2002;21: 1635–40.
[35] Higgins J, Green S. Cochrane handbook for systematic reviews of interventions. The Cochrane Collaboration; 2011: Available at www.cochrane-handbook.org Accessed March, 2011.
[36] Haskins R, Osmotherly PG, Tuyl F, Rivett DA. Uncertainty in clinical prediction rules: the value of credible intervals. J Orthop Sports Phys Ther 2014;44:85–91.
[37] Hayden JA, Côté P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. Ann Intern Med 2006; 144:427–37.
[38] Maher CG, Sherrington C, Herbert RD, Moseley AM, Elkins M. Reliability of the PEDro scale for rating quality of randomized controlled trials. Phys Ther 2003;83:713–21.
[39] Laupacis A, Sekar N, Stiell IG. Clinical prediction rules: a review and suggested modifications of methodological standards. JAMA 1997;277:488–94.
[40] Stiell IG, Wells GA. Methodologic standards for the development of clinical decision rules in emergency medicine. Ann Emerg Med 1999;33:437–47.
[41] Beattie P, Nelson R. Clinical prediction rules: what are they and what do they tell us? Aust J Physiother 2006;52:157–63.
[42] Cook C, Brismée J-M, Pietrobon R, Sizer P Jr, Hegedus E, Riddle DL. Development of a quality checklist using Delphi methods for prescriptive clinical prediction rules: the QUADCPR. J Manipulative Physiol Ther 2010;33:29–41.
[43] Bouwmeester W, Zuithoff NP, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, et al. Reporting and methods in clinical prediction research: a systematic review. PLoS Med 2012;9: e1001221.
[44] Seel RT, Steyerberg EW, Malec JF, Sherer M, Macciocchi SN. Developing and evaluating prediction models in rehabilitation populations. Arch Phys Med Rehabil 2012;93:S138–53.
[45] de Morton NA. The PEDro scale is a valid measure of the methodological quality of clinical trials: a demographic study. Aust J Physiother 2009;55:129–33.
[46] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Ann Intern Med 2009;151: W65–94.
[47] Alonso-Blanco C, Fernández-de-las-Penas C, Cleland JA. Preliminary clinical prediction rule for identifying patients with ankylosing spondylitis who are likely to respond to an exercise program: a pilot study. Am J Phys Med Rehabil 2009;88:445–54.
[48] Al-Sayegh NA, George SE, Boninger ML, Rogers JC, Whitney SL, Delitto A. Spinal mobilization of postpartum low back and pelvic girdle pain: an evidence-based clinical rule for predicting responders and nonresponders. PM R 2010;2:995–1005.
[49] Buranapanitkit B, Tautakul K, Lim A, Geater A, Chomchan U. Clinical predictors for outcome of treatment of lumbar herniated nucleus palposus. J Musculoskelet Res 2003;7:61–9.
[50] Cai C, Pua YH, Lim KC. A clinical prediction rule for classifying patients with low back pain who demonstrate short-term improvement with mechanical lumbar traction. Eur Spine J 2009;18:554–61.
[51] Cairns D, Mooney V, Crane P. Spinal pain rehabilitation: inpatient and outpatient treatment results and development of predictors for outcome. Spine 1984;9:91–5.
[52] Dionne CE, Koepsell TD, Von Korff M, Deyo RA, Barlow WE, Checkoway H. Predicting long-term functional limitations among back pain patients in primary care settings. J Clin Epidemiol 1997;50:31–43.

[53] Dionne CE, Bourbonnais R, Fremont P, Rossignol M, Stock SR, Larocque I. A clinical return-to-work rule for patients with back pain. CMAJ 2005;172:1559–67.
[54] Flynn T, Fritz J, Whitman J, Wainner R, Magel J, Rendeiro D, et al. A clinical prediction rule for classifying patients with low back pain who demonstrate short-term improvement with spinal manipulation. Spine 2002;27:2835–43.
[55] Fritz JM, Childs JD, Flynn TW. Pragmatic application of a clinical prediction rule in primary care to identify patients with low back pain with a good prognosis following a brief spinal manipulation intervention. BMC Fam Pract 2005;6:29.
[56] George SZ, Bialosky JE, Donald DA. The centralization phenomenon and fear-avoidance beliefs as prognostic factors for acute low back pain: a preliminary investigation involving patients classified for specific exercise. J Orthop Sports Phys Ther 2005;35:580–8.
[57] Hancock MJ, Maher CG, Latimer J, Herbert RD, McAuley JH. Can rate of recovery be predicted in patients with acute low back pain? Development of a clinical prediction rule. Eur J Pain 2009;13:51–5.
[58] Hewitt JA, Hush JM, Martin MH, Herbert RD, Latimer J. Clinical prediction rules can be derived and validated for injured Australian workers with persistent musculoskeletal pain: an observational study. Aust J Physiother 2007;53:269–76.
[59] Heymans M, Ford J, McMeeken J, Chan A, de Vet HC, van Mechelen W. Exploring the contribution of patient-reported and clinician based variables for the prediction of low back work status. J Occup Rehabil 2007;17:383–97.
[60] Heymans MW, Anema JR, van Buuren S, Knol DL, van Mechelen W, de Vet HC. Return to work in a cohort of low back pain patients: development and validation of a clinical prediction rule. J Occup Rehabil 2009;19:155–65.
[61] Hicks GE, Fritz JM, Delitto A, McGill SM. Preliminary development of a clinical prediction rule for determining which patients with low back pain will respond to a stabilization exercise program. Arch Phys Med Rehabil 2005;86:1753–62.
[62] Kovacs FM, Seco J, Royuela A, Corcoll Reixach J, Abraira V, Spanish Back Pain Research N. Predicting the evolution of low back pain patients in routine clinical practice: results from a registry within the Spanish National Health Service. Spine J 2012;12:1008–20.
[63] Malmqvist S, Leboeuf-Yde C, Ahola T, Andersson O, Ekström K, Pekkarinen H, et al. The nordic back pain subpopulation program: predicting outcome among chiropractic patients in Finland. Chiropr Osteopat 2008;16:13. Online access only 12 p.
[64] May S, Gardiner E, Young S, Klaber-Moffett J. Predictor variables for a positive long-term functional outcome in patients with acute and chronic neck and back pain treated with a McKenzie approach: a secondary analysis. J Man Manip Ther 2008;16:155–60.
[65] Roland M, Morrell D, Morris R. Can general practitioners predict the outcome of episodes of back pain? Br Med J 1983;286:523–6.
[66] Stolze LR, Allison SC, Childs JD. Derivation of a preliminary clinical prediction rule for identifying a subgroup of patients with low back pain likely to benefit from Pilates-based exercise. J Orthop Sports Phys Ther 2012;42:425–36.
[67] Valat JP, Goupille P, Rozenberg S, Urbinelli R, Allaert F, Spine group of the Societe F. Acute low back pain: predictive index of chronicity from a cohort of 2487 subjects. Spine Group of the Societe Francaise de Rhumatologie. Joint Bone Spine 2000;67:456–61.
[68] Childs JD, Fritz JM, Flynn TW, Irrgang JJ, Johnson KK, Majkowski GR, et al. A clinical prediction rule to identify patients with low back pain most likely to benefit from spinal manipulation: a validation study. [Summary for patients in Ann Intern Med. 2004 Dec 21;141(12):I39; PMID: 15611484]. Ann Intern Med 2004;141:920–8.
[69] Childs JD, Flynn TW, Fritz JM. A perspective for considering the risks and benefits of spinal manipulation in patients with low back pain. Man Ther 2006;11:316–20.
[70] Cleland JA, Fritz JM, Whitman JM, Childs JD, Palmer JA. The use of a lumbar spine manipulation technique by physical therapists in patients who satisfy a clinical prediction rule: a case series. J Orthop Sports Phys Ther 2006;36:209–14.
[71] Cleland J, Fritz J, Kulig K, Davenport T, Eberhart S, Magel J, et al. Comparison of the effectiveness of three manual physical therapy techniques in a subgroup of patients with low back pain who satisfy a clinical prediction rule: a randomized clinical trial. Spine 2009;34:2720–9.
[72] Cook C, Learman K, O'Halloran B, Showalter C, Kabbaz V, Goode A, et al. Which prognostic factors for low back pain are generic predictors of outcome across a range of recovery domains? Phys Ther 2013;93:32–40.
[73] Dionne CE. Psychological distress confirmed as predictor of long-term back-related functional limitations in primary care settings. J Clin Epidemiol 2005;58:714–8.
[74] Dionne CE, Le Sage N, Franche R-L, Dorval M, Bombardier C, Deyo RA. Five questions predicted long-term, severe, back-related functional limitations: evidence from three large prospective studies. J Clin Epidemiol 2011;64:54–66.
[75] Fritz JM, Brennan GP, Leaman H. Does the evidence for spinal manipulation translate into better outcomes in routine clinical care for patients with occupational low back pain? A case-control study. Spine J 2006;6:289–95.
[76] Hallegraeff JM, de Greef M, Winters JC, Lucas C. Manipulative therapy and clinical prediction criteria in treatment of acute nonspecific low back pain. [Erratum appears in Percept Mot Skills. 2009 Jun;108(3):981 Note: Hallegraeff, HJM [corrected to Hallegraeff, J M]]. Percept Mot Skills 2009;108:196–208.
[77] Hancock MJ, Maher CG, Latimer J, Herbert RD, McAuley JH. Independent evaluation of a clinical prediction rule for spinal manipulative therapy: a randomised controlled trial. Eur Spine J 2008;17:936–43.
[78] Schenk R, Dionne C, Simon C, Johnson R. Effectiveness of mechanical diagnosis and therapy in patients with back pain who meet a clinical prediction rule for spinal manipulation. J Man Manip Ther 2012;20:43–9.
[79] Schwind J, Learman K, O'Halloran B, Showalter C, Cook C. Different minimally important clinical difference (MCID) scores lead to different clinical prediction rules for the Oswestry disability index for the same sample of patients. J Man Manip Ther 2013;21:71–8.
[80] Sutlive TG, Mabry LM, Easterling EJ, Durbin JD, Hanson SL, Wainner RS, et al. Comparison of short-term response to two spinal manipulation techniques for patients with low back pain in a military beneficiary population. Mil Med 2009;174:750–6.
[81] Thomas E, Silman AJ, Croft PR, Papageorgiou AC, Jayson IVM, Macfarlane GJ. Predicting who develops chronic low back pain in primary care: a prospective study. BMJ 1999;318:1662–7.
[82] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
[83] Haskins R, Osmotherly PG, Southgate E, Rivett DA. Australian physiotherapists' priorities for the development of clinical prediction rules for low back pain: a qualitative study. Physiotherapy 2015;101:44–9.
[84] Derogatis LR. Symptoms Checklist-90. Administration, scoring and procedures manual for the revised version. Baltimore: Clinical Psychometric Research; 1977.
[85] Roland M, Morris R. A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. Spine 1983;8:141–4.
[86] Fritz JM, Irrgang JJ. A comparison of a modified Oswestry low back pain disability questionnaire and the Quebec back pain disability scale. Phys Ther 2001;81:776–88.
[87] Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. JAMA 1994;271:703–7.

[88] Babyak MA. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. Psychosom Med 2004;66:411−21.

[89] Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Ann Intern Med 2003; 138:W1−12.

[90] Axen I, Jones JJ, Rosenbaum A, Lovgren PW, Halasz L, Larsen K, et al. The nordic back pain subpopulation program: validation and improvement of a predictive model for treatment outcome in patients with low back pain receiving chiropractic treatment. J Manipulative Physiol Ther 2005;28:381−5.

[91] Childs JD, Fritz JM, Piva SR, Erhard RE. Clinical decision making in the identification of patients likely to benefit from spinal manipulation: a traditional versus an evidence-based approach. J Orthop Sports Phys Ther 2003;33:259−72.

[92] Fritz JM, Whitman JM, Flynn TW, Wainner RS, Childs JD. Factors related to the inability of individuals with low back pain to improve with a spinal manipulation. Phys Ther 2004;84:173−90.

[93] Fritz JM, Lindsay W, Matheson JW, Brennan GP, Hunter SJ, Moffit SD, et al. Is there a subgroup of patients with low back pain likely to benefit from mechanical traction? Results of a randomized clinical trial and subgrouping analysis. Spine 2007;32:E793−800.

[94] Jellema P, van der Horst HE, Vlaeyen JWS, Stalman WAB, Bouter LM, van der Windt DA. Predictors of outcome in patients with (sub)acute low back pain differ across treatment groups. Spine 2006;31:1699−705.

[95] Leboeuf-Yde C, Gronstvedt A, Borge JA, Lothe J, Magnesen E, Nilsson O, et al. The nordic back pain subpopulation program: demographic and clinical predictors for outcome in patients receiving chiropractic treatment for persistent low back pain. J Manipulative Physiol Ther 2004;27:493−502.

[96] Skargren EI, Carlsson PG, Oberg BE. One-year follow-up comparison of the cost and effectiveness of chiropractic and physiotherapy as primary management for back pain: subgroup analysis, recurrence, and additional health care utilization. Spine 1998;23: 1875−83.

[97] Teyhen DS, Flynn TW, Childs JD, Abraham LD. Arthrokinematics in a subgroup of patients likely to benefit from a lumbar stabilization exercise program. Phys Ther 2007;87:313−25.

[98] van der Hulst M, Vollenbroek-Hutten MM, Groothuis-Oudshoorn KG, Hermens HJ. Multidisciplinary rehabilitation treatment of patients with chronic low back pain: a prognostic model for its outcome. Clin J Pain 2008;24:421−30.

[99] Du Bois M, Donceel P. A screening questionnaire to predict no return to work within 3 months for low back pain claimants. Eur Spine J 2008;17:380−5.

[100] Hill JC, Dunn KM, Lewis M, Mullis R, Main CJ, Foster NE, et al. A primary care back pain screening tool: identifying patient subgroups for initial treatment. Arthritis Rheum 2008;59:632−41.

Appendix A – Search Strategy

Medline via OVID (1946 - July 2013)	
1	dorsalgia.ti,ab OR exp Back Pain/ OR backache.ti,ab OR exp Low Back Pain/ OR (lumbar adj pain).ti,ab OR coccyx.ti,ab OR coccydynia.ti,ab OR sciatica.ti,ab OR sciatic neuropathy/ OR spondylosis.ti,ab OR lumbago.ti,ab OR back disorder$.ti,ab
2	(Validat$ OR Predict$.ti. OR Rule$) OR (Predict$ AND (Outcome$ OR Risk$ OR Model$)) OR ((History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$) AND (Predict$ OR Model$ OR Decision$ OR Identif$ OR Prognos$)) OR (Decision$ AND (Model$ OR Clinical$ OR Logistic Models/)) OR (Prognostic AND (History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$ OR Model$))
3	Stratification OR ROC Curve/ OR Discrimination OR Discriminate OR c-statistic OR c statistic OR "Area under the curve" OR AUC OR Calibration OR Indices OR Algorithm OR Multivariable
4	1 AND (2 OR 3)
5	limit 4 to english
6	limit 5 to humans

Embase via OVID (1947 - July 2013)	
1	dorsalgia.mp. OR back pain.mp. OR exp LOW BACK PAIN/ OR exp BACKACHE/ OR (lumbar adj pain).mp. OR coccyx.mp. OR coccydynia.mp. OR sciatica.mp. OR exp ISCHIALGIA/ OR spondylosis.mp. OR lumbago.mp. OR back disorder$.ti,ab.
2	predict:.tw. OR exp methodology OR validat:.tw.
3	1 AND 2
4	limit 3 to english
5	limit 4 to humans
6	limit 5 to exclude medline journals

Cochrane Central Register of Controlled Trials via OVID (1898 - July 2013)	
1	exp Back Pain/ OR back ache OR exp Low Back Pain/ OR (lumbar adj pain) OR coccyx OR coccydynia OR sciatica OR spondylosis OR exp Spine/ OR exp Spinal Diseases/ OR lumbago OR discitis OR (disc adj degeneration) OR (disc adj prolapse) OR (disc adj herniation) OR spinal fusion OR spinal neoplasms OR (facet adj joints) OR exp Intervertebral Disk/ or postlaminectomy OR arachnoiditis OR (failed adj back) OR exp Cauda Equina/ OR (lumbar adj vertebra$) OR (spinal adj stenosis) OR (slipped adj (disc$ or disk$)) OR (degenerat$ adj (disc$ or disk$)) OR (stenosis adj (spine or root or spinal)) OR (displace$ adj (disc$ or disk$)) OR (prolap$ adj (disc$ or disk$)) OR exp Sciatic Neuropathy/ OR sciatic$ OR back disorder$ OR (back adj pain)
2	(Validat$ OR Predict$.ti. OR Rule$) OR (Predict$ AND (Outcome$ OR Risk$ OR Model$)) OR ((History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$) AND (Predict$ OR Model$ OR Decision$ OR Identif$ OR Prognos$)) OR (Decision$ AND (Model$ OR Clinical$ OR Logistic Models/)) OR (Prognostic AND (History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$ OR Model$))
3	Stratification OR ROC Curve/ OR Discrimination OR Discriminate OR c-statistic OR c statistic OR "Area under the curve" OR AUC OR Calibration OR Indices OR Algorithm OR Multivariable
4	1 AND (2 OR 3)

5	limit 4 to medline records
6	limit 4 to embase records
7	4 NOT (5 OR 6)

PsychINFO via OVID (1806 - July 2013)	
1	back pain/ OR lumbar spinal cord/ OR (low adj back adj pain).mp OR (back adj pain).mp OR spinal column/ OR (lumbar adj2 vertebra$).mp OR coccyx.mp OR sciatica.mp OR lumbago.mp OR dorsalgia.mp OR back disorder$.mp OR ((disc or disk) adj degenerat$).mp OR ((disc or disk) adj herniat$).mp OR ((disc or disk) adj prolapse$).mp OR (failed adj back).mp
2	(Validat$ OR Predict$.ti. OR Rule$) OR (Predict$ AND (Outcome$ OR Risk$ OR Model$)) OR ((History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$) AND (Predict$ OR Model$ OR Decision$ OR Identif$ OR Prognos$)) OR (Decision$ AND (Model$ OR Clinical$ OR Logistic Models/)) OR (Prognostic AND (History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$ OR Model$))
3	Stratification OR ROC Curve/ OR Discrimination OR Discriminate OR c-statistic OR c statistic OR "Area under the curve" OR AUC OR Calibration OR Indices OR Algorithm OR Multivariable
4	1 AND (2 OR 3)
5	limit 4 to english
6	limit 5 to human

CINAHL via EBSCO (1937 - July 2013)	
1	"dorsalgia" OR (MH "Back Pain+") OR (MH "Low Back Pain") OR "backache" OR (lumbar W1 pain) OR (lumbar N5 pain) OR (MH "Coccyx") OR (MH "Sciatica") OR "sciatica" OR "coccyx" OR "coccydynia" OR "back disorder*" OR (MH "Lumbar Vertebrae") OR (lumbar N2 vertebra) OR (MH "Thoracic Vertebrae") OR (MH "Spondylolisthesis") OR (MH "Spondylolysis") OR "lumbago"
2	(validat* OR ti predict* OR rule*) OR (predict* AND (outcome* OR risk* OR model*)) OR ((history OR variable* OR criteria OR scor* OR characteristic* OR finding* OR factor*) AND (predict* OR model* OR decision* OR identif* OR prognos*)) OR (decision* AND (model* OR clinical* OR MH "logistic regression+")) OR (prognostic AND (history OR variable* OR criteria OR scor* OR characteristic* OR finding* OR factor* OR model*))
3	stratification OR mh "ROC Curve" OR discrimination OR discriminate OR c-statistic OR c statistic OR "Area under the curve" OR AUC OR calibration OR indices OR algorithm OR multivariable
4	S1 AND (S2 OR S3)
5	applied limit to English
6	applied limit to humans
7	applied limit to exclude Medline records

AMED via OVID (1985 - July 2013)	
1	dorsalgia.ti,ab OR exp Back Pain/ OR backache.ti,ab OR exp Low Back Pain/ OR (lumbar adj pain).ti,ab OR coccyx.ti,ab OR coccydynia.ti,ab OR sciatica.ti,ab OR sciatic neuropathy/ OR spondylosis.ti,ab OR lumbago.ti,ab OR back disorder$.ti,ab

2	(Validat$ OR Predict$.ti. OR Rule$) OR (Predict$ AND (Outcome$ OR Risk$ OR Model$)) OR ((History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$) AND (Predict$ OR Model$ OR Decision$ OR Identif$ OR Prognos$)) OR (Decision$ AND (Model$ OR Clinical$ OR Logistic Models/)) OR (Prognostic AND (History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$ OR Model$))
3	Stratification OR ROC Curve/ OR Discrimination OR Discriminate OR c-statistic OR c statistic OR "Area under the curve" OR AUC OR Calibration OR Indices OR Algorithm OR Multivariable
4	1 AND (2 OR 3)
5	limit 4 to english

Index of Chiropractic Literature (- July 2013)	
1	Subject:"Back" OR Subject:"Back Injuries" OR Subject:"Back Pain" OR Subject:"Low Back Pain" OR Subject:"Lumbar" OR Subject:"Lumbosacral Region" OR Subject:"Sciatica" OR All Fields:sciatica OR Subject:"Coccyx" OR Subject:"Sacroiliac Joint" OR Subject:"Sacrum"
2	(Validat* OR Predict* OR Rule*) OR (Predict*AND (Outcome* OR Risk* OR Model*)) OR ((History OR Variable* OR Criteria OR Scor* OR Characteristic* OR Finding* OR Factor*) AND (Predict* OR Model* OR Decision* OR Identif* OR Prognos*)) OR (Decision* AND (Model* OR Clinical* OR "Logistic Model*")) OR (Prognostic AND (History OR Variable* OR Criteria OR Scor* OR Characteristic* OR Finding* OR Factor* OR Model*))
3	Stratification OR "ROC Curve" OR Discrimination OR Discriminate OR c-statistic OR "c statistic" OR "Area under the curve" OR AUC OR Calibration OR Indices OR Algorithm OR Multivariable
4	S2 OR S3
5	S1 AND S4

Appendix B – Quality Appraisal Criteria

Derivation study quality appraisal criteria

Domain	Criteria	Definition
Design	1. Prospective design	The study is conducted forwards in time.
	2. Study site described	The nature of study site is described in sufficient detail to enable comparison to other settings.
Participants	3. Justification for the number of study subjects is reported	Justification is provided for the number of subjects enrolled into the study.
	4. Representative sample	The reported method of patient selection is free of bias so that study subjects encompass a wide clinical and demographic spectrum and are representative of all patients seen at the site with the designated condition.
	5. Important patient characteristics described	The study subjects are well described in terms of inclusion criteria, method of selection and clinical and demographic characteristics.
Predictor variables	6. Selection of candidate predictor variables justified	The selection of candidate predictor variables is justified with appropriate reasoning and may include previous literature, psychometric properties, clinical reasoning, and/or expert opinion.
	7. Blinded predictor assessment	The assessment of the predictor variables is determined without knowledge of the outcome. If the study was prospective and the predictor variables were clearly collected prior to the outcome event, then assessment can be considered to be blind. If the study was retrospective and the authors did not mention blinding, it will be assumed that it was not blinded.
	8. Predictor variables have demonstrated reliability	Predictor tests are reported to be reliable (kappa >=0.60 or ICC >=0.70) either through previous report or through report within the findings of the study.
Outcomes	9. Outcome measure has demonstrated reliability and validity	The outcome measure is reported to have demonstrated reliability and validity. Literature is cited to support the outcome measure and psychometric characteristics of the outcome measure are reported.
	10. Blinded outcome assessment	Interpretation of the outcome assessment is reported to be determined without knowledge of the status of the predictor variables. If a study does not comment upon whether the outcome was categorised without knowledge of the predictor variables, it will be assumed that it was not blinded. If the outcome is self-reported (eg. VAS), blinding is considered to be present if the participant is blinded to their status on the predictor variables.
Analysis	11. Mathematical techniques described	The mathematical techniques employed are adequately described.

	12. Reporting and handling of missing data described	Missing data (eg. values per participant, missing values per predictor, lost to follow-up) and how it was handled (eg. omitted, imputation) are reported.
	13. At least 10 outcome events per independent variable in the final multivariable model	There are at least 10 outcome events per independent variable in the final multivariable model. The number of outcome events is defined in proportional hazards analysis by the count of 'failure' events. In logistic regression the number of outcome events is the smaller number of binary outcomes of the dependent variable. For linear regression models there should be at least 10 patients per variable in the final model.
	14. At least 10 outcome events per candidate predictor variable	As per item 13, except the number of candidate predictor variables replaces the number independent variables in the final model.
	15. Collinearity of predictor variables tested	Collinearity of predictor variables were examined such as testing pairwise correlations or the variance inflation factor.
	16. Continuous predictor variables are kept continuous in the multivariable analysis	Continuous predictor variables were kept as continuous variables in the multivariable analysis.
CPR performance	17. Uncertainty in the accuracy of the CPR is described	Uncertainty intervals are reported for accuracy statistics of the CPR.
	18. Uncertainty in the posttest probability is described	Uncertainty intervals are reported for posttest probabilities.
	19. CPR performance is non-paradoxical	The performance of a CPR behaves logically, such that the probability of a given outcome does not decrease at any point with increasing positive status on that tool.

Validation study quality appraisal criteria

Domain	Criteria	Definition
Design	1. Prospective validation in new patient population	The study is conducted forwards in time in a different population to the derivation study.
	2. Different clinical setting to derivation study	The CPR is tested in a clinical setting that is different to the derivation study.
Participants	3. Representative sample	The method of patient selection is free of bias so that study subjects encompass a wide clinical and demographic spectrum and are representative of all patients seen at the site with the designated condition.
CPR application	4. The rule is applied accurately	The rule is applied exactly as described in the derivation study.

	5. Assessment of the reliability of the rule	The reliability of the interpretation of a rule is explicitly measured using at least a representative subset of the study sample.
Follow-up	6. Complete follow-up	There is complete follow-up. All patients are subjected to the gold or criterion standard to determine their true outcome compared to that predicted by the rule.
	7. Reporting and handling of missing data described	Missing data (eg. values per participant, missing values per predictor, lost to follow-up) and how it was handled (eg. omitted, imputation) is reported.
CPR performance	8. Accuracy of the rule described	The accuracy/performance of the rule is described.
	9. Uncertainty in the accuracy of the CPR is described	Uncertainty intervals are reported for accuracy statistics of the CPR.
	10. Uncertainty in the posttest probability is described	Uncertainty intervals are reported for posttest probabilities.

Appendix C. Risk of potential bias of included studies as appraised using QUIPS

Study	CPR reference number	Stage of development	Study participation	Study attrition	Prognostic factor measurement	Outcome measurement	Confounding measurement and account	Analysis
Alonso-Blanco et al 2009	10	Derivation	Low	Low	Moderate	Low	N/A	Moderate
Al-Sayegh et al 2010	12	Derivation	Low	Low	Moderate	Low	N/A	Moderate
Al-Sayegh et al 2010	13	Derivation	Low	Low	Moderate	Low	N/A	Moderate
Buranapanitkit et al 2003	23	Derivation	Moderate	Low	Moderate	High	N/A	Moderate
Cai et al 2009	11	Derivation	Low	Low	Moderate	Low	N/A	Moderate
Cairns et al 1984	26	Derivation	Moderate	Low	Moderate	Moderate	N/A	Moderate
Cairns et al 1984	27	Derivation	Moderate	Low	Moderate	Moderate	N/A	Moderate
Cleland et al 2006	3	Validation	Moderate	Low	Low	Low	High	Low
Cook et al 2013	3	Validation	Low	Moderate	Low	Low	Low	Low
Dionne 2005	2	Validation	Low	Moderate	Low	Low	High	Moderate
Dionne et al 1997	2	Derivation	Low	High	Moderate	Low	N/A	Moderate
Dionne et al 2005	28	Derivation	Low	Moderate	Moderate	Low	N/A	Low
Dionne et al 2011	2	Validation	Low	Moderate	Low	Low	High	Low
Flynn et al 2002	3	Derivation	Low	Low	Moderate	Low	N/A	Moderate
Fritz et al 2005	4	Derivation	Low	Low	Low	Low	N/A	Low
Fritz et al 2006	4	Validation	Moderate	Low	Low	Low	Moderate	Moderate
George et al 2005	5	Derivation	Low	Low	Low	Low	N/A	Low
George et al 2005	16	Derivation	Low	Low	Low	Low	N/A	Low
Hancock et al 2009	22	Derivation	Low	Low	Moderate	Moderate	N/A	Moderate
Hewitt et al 2007	7	Derivation	Low	Low	Moderate	Low	N/A	Low
Hewitt et al 2007	17	Derivation	Low	Low	Moderate	Moderate	N/A	Low
Heymans et al 2007	29	Derivation	Low	Low	Moderate	Moderate	N/A	Low

Heymans et al 2009	30	Derivation	Low	Low	Moderate	Moderate	N/A	Low
Hicks et al 2005	6	Derivation	Low	Low	Moderate	Low	N/A	Low
Kovacs et al 2012	14	Derivation	Low	Low	Moderate	Low	N/A	Moderate
Kovacs et al 2012	18	Derivation	Low	Low	Moderate	Low	N/A	Moderate
Kovacs et al 2012	19	Derivation	Low	Low	Moderate	Low	N/A	Moderate
Malmqvist et al 2008	21	Derivation	Low	Moderate	Moderate	Low	N/A	Moderate
May et al 2008	8	Derivation	Moderate	Low	Low	Low	N/A	Moderate
May et al 2008	9	Derivation	Moderate	Low	Low	Low	N/A	Moderate
Roland et al 1983	1	Derivation	Moderate	Low	Moderate	Low	N/A	Moderate
Roland et al 1983	20	Derivation	Moderate	Low	Moderate	Moderate	N/A	Moderate
Schwind et al 2013	3	Validation	Low	Low	Low	Low	Low	Low
Stolze et al 2012	15	Derivation	Low	Low	Low	Low	N/A	Moderate
Thomas et al 1999	25	Derivation	Moderate	Moderate	Moderate	Low	N/A	Moderate
Valat et al 2000	24	Derivation	Low	Low	High	High	N/A	Moderate

Appendix D. Risk of potential bias of included studies as appraised using PEDro

Study	CPR reference number	Stage of development	Eligibility criteria	Random allocation	Concealed allocation	Baseline comparability	Blind subjects	Blind therapists	Blind assessors	Adequate follow-up	Intention to treat analysis	Between-group comparisons	Point estimates and variability	Score
Childs et al 2004	3	Validation	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	8
Childs et al 2006	3	Validation	No	Yes	No	No	No	No	No	Yes	Yes	Yes	Yes	5
Cleland et al 2009	3	Validation	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	8
Hallegraeff et al 2009	4	Validation	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	7
Hancock et al 2008	3	Validation	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	9
Schenk et al 2012	3	Validation	Yes	Yes	No	Yes	No	No	Yes	No	No	Yes	Yes	5
Sutlive et al 2009	3	Validation	Yes	Yes	No	Yes	No	No	No	Yes	Yes	Yes	Yes	6

Supplementary Material 1. Prognostic/Prescriptive LBP CPRs

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
Function	1.	Identifying patients presenting in general practice who are likely to have a low level of disability at 4 weeks	1. duration of pain <1 week; 2. SLR* ≥ 60°	Count of predictors	Derivation Roland et al 1983 Single cohort	n=230 episodes of LBP† (from 215 patients attending GP‡ practice), 192 episodes with follow-up data at 4 weeks, mean age 41 years (dispersion not reported), 53% female, duration of symptoms not reported, 84% prevalence of dependent outcome.	Not specified	RMDQ§ (24-items) score of < 14 points at 4 weeks (dichotomous)	Not reported, but accuracy of both variables present calculated to be, sens** = 0.40 (0.33-0.48); spec†† = 0.93 (0.79-0.98); +LR‡‡ = 6.0 (1.6-23.3); -LR§§ = 0.64 (0.55-0.75).	2 predictors present = 97%, 1 predictor present = 81%, 0 predictors present = 61%. 95%CrI*** calculated for both variables present to be 90.0% - 99.1%.
	2. Cassandra rule	Identifying patients at high risk of sustaining long-term significant functional limitations	1. Symptoms Checklist 90 Revised Depression score; 2. Symptoms Checklist 90 Revised Somatization score	Algorithm leading to 4 stratified risk groups. Dichotomized to 2 risk groups in later publications (Dionne et al 2005, 2011)	Derivation Dionne et al 1997 Single cohort with split sample	Consulting primary care physician with back pain. Training set n=569, mean age 47 years (SD††† 14), 52% female, mean 13 years (SD 13) since LBP onset, 95 lost to follow-up, 15% prevalence of dependent outcome. Validation set n=644, mean age 47 years (SD 15), 53% female, mean 12 years (SD 13) since LBP onset, 96 lost to follow-up, 16% prevalence of dependent outcome.	Not specified	Modified RMDQ (16-items) score ≥ 50% at 2 years (dichotomous)	Reported in later publications (Dionne et al 2005, 2011) for dichotomized risk groups (high/moderate vs low) sens = 0.86 (0.77-0.93); spec = 0.57 (0.53-0.62). LR's not reported, but approximated from study data and later publications to be +LR = 2.0 (1.7-2.3); -LR = 0.25 (0.14-0.42).	Validation set - risk of outcome for group 1 (Depression<0.444) = 5%; group 2 (Depression≥0.444 but <1.5 and Somatization < 0.333) = 4%; group 3 (Depression≥0.444 but <1.5 and Somatization ≥ 0.333) =19%; group 4 (Depression ≥1.5) = 36%. Post-test probability not reported for dichotomized risk groups, but high/moderate group risk approximated from study data and later publications to be 27%. 95%CrI for high/moderate risk group approximated to be 22.0% - 32.5%.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
					Validation Dionne 2005 Single cohort	n=860, adults absent from work due LBP consulting in primary care, mean age 39 years (SD 11), 42% female, 78% with recurrent or persistent back pain, 18% prevalence of dependent outcome.	Not specified	RMDQ (24-items) score of ≥ 50% at 2 years (dichotomous)	For dichotomized risk groups (high/moderate vs low) sens = 0.91; spec = 0.29. LRs not reported, but approximated to be +LR = 1.3 (1.2-1.4); -LR = 0.31 (0.19-0.52).	Risk of outcome for group 1 (Depression<0.444) = 6%; group 2 (Depression≥0.444 but <1.5 and Somatization < 0.333) = 7%; group 3 (Depression≥0.444 but <1.5 and Somatization ≥ 0.333) =14%; group 4 (Depression ≥1.5) = 27%. Post-test probability for dichotomized high/moderate group = 22%. 95%CrI for high/moderate risk group approximated to be 18.9% - 25.2%.
					Validation Dionne et al 2011 Single cohort	n=1262, patients presenting to an emergency department with nonspecific back pain, mean age 41 years (SD 14), 48% female, 57% with recurrent or persistent symptoms, 19% prevalence of dependent outcome.	Not specified	RMDQ (24-items) score of ≥ 50% at 2 years (dichotomous)	For dichotomized risk groups (high/moderate vs low) sens = 0.82 (0.76-0.87); spec = 0.45 (0.42-0.49); +LR = 1.50 (1.38-1.64); -LR = 0.40 (0.25-0.47). CPR‡‡‡ more sensitive (82% vs 37%) but less specific (45% vs 85%) than physician prediction.	Post-test probability for dichotomized high/moderate group = 22% (0.22-0.29). 95%CrI for high/moderate risk group approximated to be 22.1% - 28.8%. CPR modified in this study to refine it to 5-items, however, this newly derived tool does not meet the definition of a CPR in this review.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	3.	Identifying patients receiving lumbopelvic manipulation who are likely to experience improvement	1. duration of symptoms < 16 days; 2. FABQ[§§§] work subscale score < 19; 3. at least 1 hip with > 35° of internal rotation ROM[****]; 4. hypomobility in the lumbar spine; 5. no symptoms distal to the knee	Count of predictors	Derivation Flynn et al 2002 Single cohort	n=71, patients with LBP presenting to military outpatient physical therapy facilities, mean age 38 years (SD 11), 41% female, mean 42 days (SD 55) of symptoms, 45% prevalence of dependent outcome.	Up to 2 treatments of high velocity thrust lumbopelvic manipulation (up to 2 attempts on each side), 10 reps supine pelvic tilt ROM exercise, and advice to maintain usual activity level within the limits of pain.	> 50% improvement on ODQ[††††] by third treatment session (up to 8 days following initial) (dichotomous)	For 4 or more predictors present, sens = 0.63 (0.45-0.77); spec = 0.97 (0.87-1.0); +LR = 24.38 (4.63-139.41).	For 4 or more predictors present = 95%. 95%CrI calculated to be 77.1% - 98.9%.
					Validation Childs et al 2004 RCT[‡‡‡‡]	n=131, patients with LBP referred to physiotherapy, mean age 34 years (SD 11), 42% female, median 27 days of symptoms (IQR[§§§§] not reported), 44% prevalence of dependent outcome in manipulation group at 1 week, and 63% at 4 weeks.	1. Manipulation group - high velocity thrust spinal manipulation and ROM exercise on first 2 sessions, advice to maintain usual activity, aerobic and strengthening exercises (5 sessions over 4 weeks); 2. Exercise group - aerobic and strengthening exercises, advice to maintain usual activity (5 sessions over 4 weeks)	1. ≥50% improvement on ODQ at 1 week (dichotomous); 2. ≥50% improvement on ODQ at 4 weeks (dichotomous); 3. ODQ score at 1 week, 4 weeks and 6 months (continuous); 4. pain (0-10 NRS[*****]) at 1 week, 4 weeks and 6 months (continuous)	For 4 or more predictors present, significant 3-way interaction for rule status, treatment group, and time, for ODQ and pain. In manipulation group, having 4 or more predictors present had accuracy of +LR = 13.2 (3.4-52.1) for improvement at 1 week.	For 4 or more predictors present in manipulation group, post-test probability at 1 week = 92%. 95%CrI calculated to be 72.8% - 97.3%.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
					Validation Childs et al 2006 RCT secondary analysis	n=131, patients with LBP referred to physiotherapy, mean age 34 years (SD 11), 42% female, median 27 days of symptoms (IQR not reported), 99% prevalence of dependent outcome in manipulation group at 1 week, and 97% at 4 weeks.	1. Manipulation group - high velocity thrust spinal manipulation and ROM exercise on first 2 sessions, advice to maintain usual activity, aerobic and strengthening exercises (5 sessions over 4 weeks); 2. Exercise group - aerobic and strengthening exercises, advice to maintain usual activity (5 sessions over 4 weeks)	No worsening of disability defined as not having ≥6 point increase in ODQ score (dichotomous) (inverted from study)	No patients with 4 or more predictors present who received manipulation experienced worsening. Accuracy of 4 or more predictors present for prediction of outcome in manipulation group at 1 week not reported, but able to be derived from study data to be sens = 0.33 (0.23-0.45); spec = 1.0 (0.21-1.0); +LR = ∞; -LR = 0.67 (0.56 – 0.79).	For 4 or more predictors present in manipulation group, post-test probability of not being worse at 1 week = 100%. 95%CrI calculated to be 89.0% - 100%.
					Validation Cleland et al 2006 Case series	n=12, patients with LBP attending an outpatient physiotherapy clinic and all CPR positive (≥4 predictors present), mean age 39 years (SD 9), 42% female, median 19 days of symptoms (range 8-148), 92% prevalence of dependent outcome.	Side-lying thrust manipulation and ROM exercise x 2 sessions within 1 week	≥50% improvement on ODQ at 1 week (dichotomous)		11 of 12 (92%) participants improved at 1 week. All participants positive on CPR, therefore unable to assess rule performance.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
					Validation Hancock et al 2008 RCT	n=239, patients with LBP < 6 weeks duration presenting to a GP, mean age 41 years (SD 16), 44% female, mean 9 days of symptoms (SD 9).	1. Spinal manipulative therapy (n=119), 2-3 times per week for a maximum of 12 sessions over 4 weeks, 5% received thrust manipulative techniques; 2. Placebo (n=120), detuned pulsed ultrasound, matched to active treatment group contact (both groups further randomised to receive either placebo or active diclofenac)	1. pain (11 point NRS); 2. disability (RMDQ) measured at 1,2,4 and 12 weeks (all continuous)	For 4 or more predictors present, no significant 3-way interaction between treatment group, CPR status and time for either pain (p=0.805) or disability (p=0.600). Positive rule status predicted improved pain at 2 weeks (p=0.015), and improved disability at 2 (p=0.033) and 12 weeks (0.015) independent of treatment group.	.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
					Validation Cleland et al 2009 RCT	n=112, patients with LBP attending an outpatient physiotherapy clinic and all CPR positive (≥4 predictors present), mean age 40 years (SD 12), 52% female, median 45 days of symptoms (IQR 27-60).	1. Supine thrust manipulation group - supine high velocity thrust spinal manipulation and ROM exercise on first 2 sessions, exercise regime for next 3 sessions (5 sessions over 4 weeks); 2. Side-lying thrust manipulation group - side-lying high velocity thrust spinal manipulation and ROM exercise on first 2 sessions, exercise regime for next 3 sessions (5 sessions over 4 weeks); 3. Non-thrust manipulation group - central lower lumbar non-thrust manipulation (mobilization) and ROM exercise for first 2 sessions, exercise regime for next 3 sessions (5 sessions over 4 weeks).	1. ≥50% improvement on ODQ at 1 week (dichotomous); 2. ≥50% improvement on ODQ at 4 weeks (dichotomous); 3. ≥50% improvement on ODQ at 6 months (dichotomous); 4. ODQ score at 1 week, 4 weeks, and 6 months (continuous); 5. pain (0-10 NRS) at 1 week, 4 weeks and 6 months (continuous)	Significant group by time interaction for ODQ ($p<0.001$) and pain ($p=0.001$).Pair-wise comparisons indicate non-thrust group achieved inferior results to thrust manipulation groups, and no significant difference between thrust manipulation groups. Significant between group difference in proportion achieving a successful outcome at 1 week (between group difference $p < 0.001$); 4 weeks (between group difference $p < 0.001$), and 6 months (between group difference $p = 0.009$).	Success at 1 week, supine thrust manipulation group = 54.1%, side-lying thrust manipulation group = 52.6%, and non-thrust manipulation group = 8.1%; at 4 weeks, supine thrust manipulation group = 86.5%, side-lying thrust manipulation group = 81.6%, and non-thrust manipulation group = 18.9%; at 6 months, supine thrust manipulation group = 91.9%, side-lying thrust manipulation group = 89.5%, and non-thrust manipulation group = 67.6%. All participants positive on CPR, therefore unable to assess rule performance.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
					Validation Sutlive et al 2009 RCT	n=60, military health care beneficiaries with LBP who are CPR positive (3 of 5 criteria present), mean age 26 years (SD 9), 48% female, 62% symptoms < 16 days.	1. lumbopelvic manipulation and pelvic tilt range of motion exercise (1 session); 2.neutral gap manipulation and pelvic tilt range of motion exercise (1 session)	1. pain (11 point NRS) at 48 hours; 2. disability (ODQ score) at 48 hours (both continuous)	No significant between group difference in the degree of improvement in pain (p=0.591) or disability (p=0.668) at 48 hours follow-up.	All participants positive on CPR, therefore unable to assess rule performance. Positive status on CPR was defined as ≥3 predictors present.
					Validation Schenk et al 2012 RCT	n=31 (analysis limited to n=26), patients with LBP referred to physical therapy who were positive on the CPR (analysis restricted to ≥4/5 criteria present), mean age 42 years (calculated from group demographics) (no dispersion reported) 61% female, mean 17 days (no dispersion reported) of symptoms (calculated from group demographics).	1. Mechanical Diagnosis and Therapy using directional preference established at initial session completed as home and clinic exercises; 2. High velocity thrust lumbopelvic manipulation, with 30 and 20 reps of hand-heel rock exercise at first 2 sessions respectively, and hourly home exercises in the patient's directional preference from session 3 until discharge	1. > 50% improvement on ODQ by discharge at week 4 (dichotomous); 2. ODQ score at discharge (continuous); 3. pain NRS at discharge (continuous)	No between group difference at discharge in ODQ score (p=0.31), pain (p=0.08), or the proportion improved by >50% on ODQ (p=0.16).	25% and 56% improved in manipulation group and Mechanical Diagnosis and Therapy group respectively. All participants positive on CPR, therefore unable to assess rule performance.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
					Validation Cook et al 2013 Pooled results from RCT	n=149, patients with LBP attending outpatient physiotherapy, mean age 48 years (SD 15), 53% female, mean 34 weeks of symptoms (SD 99), 71 (49%) positive (≥4/5 predictors present) on CPR.	1. Thrust manipulation for first 2 sessions, then physical therapist directed care (n=76); 2. Non-thrust manipulation for first 2 sessions, then physical therapist directed care (n=73).	1. ≥50% improvement on ODQ at discharge (dichotomous); 2. ≥2.5 points of improvement on 11-point NRS at discharge (dichotomous); 3. self-reported recovery ≥75% at discharge (dichotomous); 4. total visits ≤6 sessions (dichotomous); 5. ODQ change score (continuous); 6. NRS change score (continuous); 7. total visits (continuous); 8. extent of recovery (continuous)	Positive status on CPR (≥4/5 predictors present) was an independent predictor for each of the 4 dichotomous dependent outcomes; 1. OR[††††] =2.9(1.4-6.2); 2. OR=4.8(1.8-10.4); 3. OR=4.0(1.6-9.8); 4. 3.7(1.7-7.6); and for each of the 4 continuous dependent outcomes; 5. β = -4.2 (-7.7 to -0.69); 6. β = -0.98 (-1.5 to -0.47); 7. β = 0.32 (0.19-0.45); 8. β = -10.8 (-18.3 to -3.1).	Data pooled from both treatment groups, with treatment group allocation included as a covariate in each model.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
					Validation Schwind et al 2013 Pooled data from RCT	n=149, patients with LPB attending outpatient physiotherapy, mean age 48 years (SD 15), 53% female, mean 34 weeks of symptoms (SD 99).	1. Thrust manipulation for first 2 sessions, then physical therapist directed care (n=76); 2. Non-thrust manipulation for first 2 sessions, then physical therapist directed care (n=73).	1. ≥ 50% improvement on ODQ by discharge; 2. ≥ 30% improvement on ODQ by discharge; 3. ≥ 17 point improvement on ODQ by discharge; 4. ≥ 10 point improvement on ODQ by discharge; 5. ≥ 5 point improvement on ODQ by discharge; 6. final ODQ score ≤ 20% (all dichotomous)	Predictors retained in each model are different depending on the cut-off point of the dependent outcome. Positive status on the CPR (≥4/5 predictors present) was retained as an independent predictor variable in 3 of the 6 multivariable predictive models. Dependent outcome 1 - positive CPR status p=0.005 (OR 2.9, 95%CI[‡‡‡‡‡] 1.4-6.2); Dependent outcome 3 - positive status on CPR p=0.007 (OR 3.4, 95%CI 1.4-8.2); Dependent outcome 6 - positive status on CPR p=0.029 (OR 3.3 (95%CI 1.1-9.6).	
4.	Identifying patients receiving lumbopelvic manipulation who are likely to experience improvement	1. duration of symptoms < 16 days; 2. no symptoms distal to the knee	Count of predictors	Derivation Fritz et al 2005 Pooled data from 2 studies	n=141, patients with LBP referred to physiotherapy at predominantly military health care facilities, mean age 36 years (SD 11), 49% female, median 22 days of symptoms (range 1-2775), 45% prevalence of dependent outcome.	Up to 2 treatments of high velocity thrust lumbopelvic manipulation (up to 2 attempts on each side), 10 reps supine pelvic tilt ROM exercise, and advice to maintain usual activity level within the limits of pain.	> 50% improvement on ODQ by third treatment session (up to 8 days following initial) (dichotomous)	For both predictors present, sens = 0.56 (0.43-0.67); spec = 0.93 (0.84-0.96); +LR = 7.2 (3.2-16.1). 83.7% classification accuracy compared to 5-item CPR.	For both predictors present = 85%. 95% CrI calculated to be 71.4% - 93.0%.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
					Validation Fritz et al 2006 Retrospective database review	n=215, patients positive on CPR (both predictors present) and received physical therapy for occupational LBP, mean age 36 years (SD 10), 32% female, mean 5.3 days of symptoms (SD 4.7).	107 (50%) received thrust manipulation in first 2 sessions, 36 (17%) received non-thrust manipulation in first 2 sessions, and 72 (33%) received no manipulation in first 2 sessions.	1. change in ODQ score; 2. change in pain score; 3. number of treatment sessions; 4. length of stay; 5.costs (all continuous)	Manipulation group (n=143) improved more than non-manipulation group (n=72) in pain (p=0.008) and disability (p=0.01) and had a shorter length of stay (p=0.02), but there was no difference in number of therapy sessions (p=0.35) or costs (p=0.94). Thrust manipulation group (n=107) experienced the same degree of improvement in pain (p=0.74) and disability (p=0.76) as non-thrust manipulation group (n=36), but had fewer number of therapy sessions (p=0.04), a shorter length of stay (p=0.02) and lower costs (p=0.03).	All participants positive on CPR, therefore unable to assess rule performance.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
					Validation Hallegraeff et al 2009 RCT	n=64, patients with LBP positive on CPR and attending physical therapy, mean age 40 years (SD = 10), 45% female, 31% had symptoms < 1 week.	1. Experimental group (n=31), 4 sessions over 2.5 weeks of thrust manipulation and strengthening and stretching exercises; 2. Control group (n=33), 4 sessions over 2.5 weeks of physical therapy without manipulation	1. disability (ODQ score) (continuous); 2. pain (VAS[§§§§§]) (continuous); 3. mobility (sit and reach test) (continuous); 4. patient perceived improvement after the 4th session (dichotomous)	No significant between-group difference in pain (p=0.26), disability (p=0.38) or mobility (p=0.14) at the fourth treatment. In multivariate ANOVA[******], the experimental group improved greater than the control group for the disability outcome (p=0.001, effect size = 0.21).	32% and 31% in the experimental and control groups respectively, reported to be improved by the fourth treatment. All participants positive on CPR, therefore unable to assess rule performance.
	5.	Predicting 6 month disability outcome for patients participating in a specific exercise program.	6 month disability = 4.4 + 0.24*(baseline ODQ score) + 0.34*(baseline FABQ work subscale score) - 10*(1 if centralization present, otherwise 0)	Linear regression equation	Derivation George et al 2005 Single cohort secondary analysis	n=28, patients with acute/subacute LBP classified to receive specific exercise using a treatment based classification system, mean age 39 years (SD 10), 61% female, mean 21 days (SD 16) of symptoms.	Specific exercise consistent with Treatment-Based Classification x 4 weeks.	ODQ score at 6 months (continuous)	R^2 = 0.49.	

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	6.	Identifying patients participating in a stabilization exercise program who are unlikely to experience improvement.	1. negative prone instability test; 2. aberrant movements absent; 3. FABQ – physical activity subscale < 9; 4. no hypermobility with lumbar spring testing	Count of predictors	Derivation Hicks et al 2005 Single cohort	n=54, patients with LBP referred to physical therapy, mean age 42 years (SD 13), 57% female, mean 41 days of symptoms (SD 44), 28% prevalence of dependent outcome.	Supervised lumbopelvic stabilization exercise program, 16 sessions over 8 weeks and daily home exercises.	< 50% improvement and < 6 point improvement on ODQ at 8 weeks (dichotomous)	For 3 or more predictors present, sens = 0.87 (0.62-0.96); spec = 0.85 (0.70-0.93); +LR = 5.6 (2.6-12.1); -LR = 0.16 (0.04-0.58).	For 3 or more predictors present = 68%. Data as reported in study has been inverted such that increasing positive status on CPR is associated with higher likelihood of dependent outcome. CPR for success reported in this study was not eligible for inclusion in this review as predictor variables not selected via multivariable statistical procedures.
	7.	Predicting activity limitation at 9 weeks in patients with subacute or chronic musculoskeletal pain participating in an exercise-based physiotherapy program	9 week Functional Rating Index = 0.72*baseline Functional Rating Index score(0-100) - 8.93*interpreter required(0 if required, otherwise 1) + 2.31*duration of previous intervention (natural log (months of previous treatment +0.125)) - 4.15*baseline work status (1 = working, otherwise 0) + 13.66	Linear regression equation	Derivation Hewitt et al 2007 Single cohort with split sample	n=720 (n=360 in training set, n=360 in validation set), patients with subacute or chronic musculoskeletal attending physiotherapy (approx 56% LBP), mean age 40 years (SD 11), 34% female, mean 11 months (SD 15) duration of symptoms.	Exercise-based physiotherapy program consisting of 6-9 weeks of 1 hour gym sessions (3/week) and daily home program.	Activity limitation at 9 weeks measured using a modified version of the Functional Rating Index (continuous)	R^2 = 0.69.	Study also reports on the development of 2 prediction tools that did not meet this review's eligibility criteria.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	8.	Identifying patients being treated with the McKenzie method who are likely to experience improvement	1. duration of symptoms less than 12 weeks; 2. back pain (not neck pain); 3. centralization or abolition of symptoms	Algorithm leading to 8 stratified risk groups.	Derivation May et al 2008 Secondary analysis of single treatment arm of RCT	n=102, patients with back or neck pain referred by GPs to physiotherapy, secondary analysis of subgroup of 102/161 patients randomized to receive McKenzie treatment in previous RCT, demographics of those in secondary analysis not reported, 21% prevalence of dependent outcome.	McKenzie based treatment delivered by physiotherapists	50% reduction in RMDQ or NPQ[††††††] from baseline to 6 weeks that is retained at 6 or 12 months ("liberal" definition of success) (dichotomous)		Predicted probability provided in study for each of the 8 stratified groups ranging from 3% (duration ≥ 12 weeks, neck pain, no centralization or abolition) to 69% (duration < 12 weeks, back pain, centralization or abolition).
	9.	Identifying patients being treated with the McKenzie method who are likely to experience improvement	1. duration of symptoms less than 12 weeks; 2. back pain (not neck pain)	Algorithm leading to 4 stratified risk groups.	Derivation May et al 2008 Secondary analysis of single treatment arm of RCT	n=102, patients with back or neck pain referred by GPs to physiotherapy, secondary analysis of subgroup of 102/161 patients randomized to receive McKenzie treatment in previous RCT, demographics of those in secondary analysis not reported, 16% prevalence of dependent outcome.	McKenzie based treatment delivered by physiotherapists	50% reduction in RMDQ or NPQ from baseline to 6 weeks that is retained at 6 and 12 months ("strict" definition of success) (dichotomous)		Predicted probability provided in study for each of the 4 stratified groups ranging from 1% (duration ≥ 12 weeks, neck pain) to 49% (duration < 12 weeks, back pain).
	10.	Identifying patients with Ankylosing Spondylitis participating in a specific exercise program who are likely to experience improvement	1. SF-36 physical role score > 37; 2. SF-36 bodily pain score > 27; 3. Bath Ankylosing Spondylitis Disease Activity Index score > 31	Count of predictors	Derivation Alonso-Blanco et al 2009 Single cohort	n=35, patients with Ankylosing Spondylitis referred to a university physical therapy clinic, mean age 46 years (SD 9), 20% female, mean 10 years of symptoms (SD 3), 46% prevalence of dependent outcome.	Specific exercise program delivered in a 1 hour group format in 8 sessions over 8 weeks.	≥20% reduction in Bath Ankylosing Spondylitis Disease Activity Index score and GROC[‡‡‡‡‡‡] score ≥ +5 at 1 month follow-up (dichotomous)	For 2 or more predictors present, sens = 0.75 (0.51-0.90); spec = 0.93 (0.66-0.99); +LR = 11.2 (1.7-76.0).	For 2 or more predictors present = 91%. CrI not calculated as contingency table unable to accurately derived from study data.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	11.	Identifying patients receiving mechanical lumbar traction who will experience improvement	1. FABQ work subscale score < 21; 2. no neurological deficit, 3. age > 30; 4. non-manual work.	Count of predictors	Derivation Cai et al 2009 Single cohort	n=129, patients with LBP referred to physiotherapy from orthopaedic outpatient clinic, mean age 31 years (SD 12), 16% female, mean 40 weeks duration of symptoms (SD 82), 19% prevalence of dependent outcome.	3 sessions of mechanical lumbar traction within 9 days, at 30-40% of patient's weight, intermittent (30sec on, 10sec off) x 15 minutes.	> 50% improvement on ODQ by third treatment session (9 days following initial) (dichotomous)	For all 4 predictors present, sens = 0.36 (0.19-0.57); spec = 0.96 (0.90-0.99); +LR = 9.36 (3.13-28).	For all 4 predictors present = 69%. 95% CrI calculated to be 41.1% - 87.0%.
	12.	Identifying postpartum women receiving lumbopelvic manipulation who will experience improvement	1. positive seated flexion test; 2. positive prone knee bend test; 3. posterior superior iliac spine symmetrical in sitting; 4. pain not extending below the knee.	Count of predictors	Derivation Al-Sayegh et al 2010 Single cohort	n=69, female patients presenting within 1 year of giving birth with LBP and/or buttock pain, mean age 31 years (SD 6), 100% female, mean 29 weeks duration of symptoms (SD 17), 80% prevalence of dependent outcome.	Up to 2 treatments of high velocity thrust lumbopelvic manipulation on most symptomatic side (up to 2 attempts), 10 reps hand-heel rock range of motion exercise, and advice to remain as active as possible.	> 50% improvement on ODQ by third treatment session (up to 8 days following initial) (dichotomous)	For 2 or more predictors present, sens = 0.65 (0.51-0.77); spec = 0.79 (0.49-0.94); +LR = 3.1 (1.1-8.5).	For 2 or more predictors present = 92%. 95%CrI calculated to be 80.2% - 97.4%.
	13.	Identifying postpartum women receiving lumbopelvic manipulation who will not experience improvement	1. age>35 years; 2.VAS-best>3; 3. negative prone knee bend test.	Count of predictors	Derivation Al-Sayegh et al 2010 Single cohort	n=69, female patients presenting within 1 year of giving birth with LBP and/or buttock pain, mean age 31 years (SD 6), 100% female, mean 29 weeks duration of symptoms (SD 17), 20% prevalence of dependent outcome.	Up to 2 treatments of high velocity thrust lumbopelvic manipulation on most symptomatic side (up to 2 attempts), 10 reps hand-heel rock range of motion exercise, and advice to remain as active as possible.	≤50% improvement on ODQ by third treatment session (up to 8 days following initial) (dichotomous)	For 2 or more predictors present, sens = 0.43 (0.19-0.7); spec = 0.96 (0.86-0.99); +LR = 11.8 (2.7-52.2).	For 2 or more predictors present = 75%. 95%CrI calculated to be 38.7% - 92.2%.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	14.	Identifying patients who will experience a clinically relevant improvement in disability	1. no evidence of disc degeneration on imaging; 2. no previous surgery; 3. receiving muscle relaxants; 4. not receiving major opioids; 5. having been treated with neuroreflexotherapy; 6. higher baseline RMDQ score; 7. lower baseline LBP severity (VAS); 8. lower baseline leg pain severity (VAS); 9. shorter duration of symptoms (acute (<14 days) / subacute (14-90 days) / chronic (>90 days))	Nomogram	Derivation Kovacs et al 2012 Single cohort	n= 4220, seeking care for LBP in primary care or at a speciality centre (rheumatology, rehabilitation, neuroreflexotherapy, orthopaedic surgery), mean age 54 years (SD 15), 64% female, median 180 days of pain (IQR 90-365), 74% prevalence of dependent outcome.	95% received neuroreflexotherapy, 59% received analgesics, 15% received physical therapy, 1% underwent surgery	Improvement of ≥ 3 points on Spanish version of RMDQ (0-24) at 3 months (dichotomous)	Calibration - Hosmer-Lemeshow test p = 0.18. Discrimination - Area under receiver operating characteristic curve = 0.64.	Point estimate of outcome probability available for each 'score' on nomogram.
	15.	Identifying patients participating in a Pilates-based exercise program who will experience improvement	1. no leg symptoms in the last week; 2. BMI[§§§§§§] ≥ 25; 3. total trunk flexion ≤70°; 4. at least 1 hip with average internal and external rotation of ≥25°; 5. duration of symptoms ≤6 months	Count of predictors	Derivation Stolze et al 2012 Single cohort	n=95, referred or presenting to physical therapy, mean age 56 years (SD 11), 81% female, 68% symptoms > 6 months, 54% prevalence of dependent outcome.	Standardized Pilates-based exercise program using a Reformer, 16 sessions over 8 weeks.	≥ 50% improvement on ODQ at 8 weeks (dichotomous)	For 3 or more predictors present, sens = 0.73 (0.58-0.84); spec = 0.93 (0.81-0.99); +LR = 10.6 (3.5-32.1).	For 3 or more predictors present = 93% (81% - 97%). 95%CrI calculated to be 80.1% - 97.3%.
Pain	16.	Predicting 6 month pain outcome for patients participating in a specific exercise program.	6 month pain intensity = 0.97 + 0.27*(baseline pain score) - 1.6*(1 if centralization present, otherwise 0)	Linear regression equation	Derivation George et al 2005 Single cohort secondary analysis	n=28, patients with acute/subacute LBP classified to receive specific exercise, mean age 39 years (SD 10), 61% female, mean 21 days (SD 16) of symptoms.	Specific exercise consistent with Treatment-Based Classification x 4 weeks.	NRS (11-point) pain score at 6 months (continuous)	R^2 = 0.29.	

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	17.	Predicting pain intensity at 9 weeks in patients with subacute or chronic musculoskeletal pain participating in an exercise-based physiotherapy program	Pain intensity at 9 weeks = 0.41*baseline pain intensity(0-10) + 0.04*baseline activity limitation (0-100 Functional Rating Index) - 0.94*non-English speaking background(1 if English, otherwise 0) + 0.27*duration of previous intervention(natural log(months of previous intervention + 0.123)) + 0.41	Linear regression equation	Derivation Hewitt et al 2007 Single cohort with split sample	n=720 (n=360 in training set, n=360 in validation set), patients with subacute or chronic musculoskeletal attending physiotherapy (approx 56% LBP), mean age 40 years (SD 11), 34% female, mean 11 months (SD 15) duration of symptoms.	Exercise-based physiotherapy program consisting of 6-9 weeks of 1 hour gym sessions (3/week) and daily home program	Pain intensity at 9 weeks measured using a 10cm VAS (continuous)	R^2 = 0.67.	Study also reports on the development of 2 prediction tools that did not meet this review's eligibility criteria.
	18.	Identifying patients who will experience a clinically relevant improvement in pain in the lower back	1. having been treated with neuroreflexotherapy; 2. no previous surgery; 3. lower baseline RMDQ score; 4. higher baseline LBP severity (VAS); 5. lower baseline leg pain severity (VAS); 6. shorter duration of symptoms (acute (<14 days) / subacute (14-90 days) / chronic (>90 days))	Nomogram	Derivation Kovacs et al 2012 Single cohort	n= 4406, seeking care for LBP in primary care or at a speciality centre (rheumatology, rehabilitation, neuroreflexotherapy, orthopaedic surgery), mean age 54 years (SD 15), 64% female, median 180 days of pain (IQR 90-365), 79% prevalence of dependent outcome.	95% received neuroreflexotherapy, 59% received analgesics, 15% received physical therapy, 1% underwent surgery	Improvement of ≥ 1.5 points on 10cm VAS for severity of LBP at 3 months (dichotomous)	Calibration - Hosmer-Lemeshow test p = 0.20. Discrimination - Area under receiver operating characteristic curve = 0.65.	Point estimate of outcome probability available for each 'score' on nomogram.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	19.	Identifying patients who will experience a clinically relevant improvement in leg pain	1. having been treated with neuroreflexotherapy; 2. no previous surgery; 3. lower baseline RDDQ score; 4. not receiving an EMG[*******]; 5. lower baseline LBP severity (VAS); 6. higher baseline leg pain severity (VAS)	Nomogram	Derivation Kovacs et al 2012 Single cohort	n= 3359, seeking care for LBP in primary care or at a speciality centre (rheumatology, rehabilitation, neuroreflexotherapy, orthopaedic surgery), mean age 55 years (SD 15), 66% female, median 180 days of pain (IQR 90-365), 75% prevalence of dependent outcome.	95% received neuroreflexotherapy, 59% received analgesics, 15% received physical therapy, 1% underwent surgery	Improvement of ≥ 1.5 points on 10cm VAS for severity of leg pain at 3 months (dichotomous)	Calibration - Hosmer-Lemeshow test p = 0.16. Discrimination - Area under receiver operating characteristic curve = 0.66.	Point estimate of outcome probability available for each 'score' on nomogram.
Recovery	20.	Identifying patients presenting in general practice with a short duration episode of care	1. duration of pain <1 week; 2. SLR ≥ 60°	Count of predictors	Derivation Roland et al 1983 Single cohort	n=230 episodes of LBP (from 215 patients), 212 episodes with follow-up data at 4 weeks, mean age 41 years (dispersion not reported), 53% female, duration of symptoms not reported, 81% prevalence of dependent outcome.	Not specified	Time from first to last consultation ≤ 15 days (dichotomous)	Not reported, but accuracy of both variables present calculated to be, sens = 0.37 (0.30-0.44); spec = 0.90 (0.77-0.96); +LR = 3.8 (1.5-9.8); -LR = 0.70 (0.60-0.82).	2 predictors present = 94%, 1 predictor present = 79%, 0 predictors present = 58%. 95%CrI calculated for both variables present to be 85.8% - 97.6%.
	21.	Identifying patients receiving chiropractic who will be better by the fourth session	1. absence of leg pain; 2. improved at 2nd visit (either improved pain when turning in bed, sleeping, putting on socks/shoes, walking, or getting up from sitting); 3. not overweight or obese	Count of predictors	Derivation Malmqvist et al 2008 Single cohort	n=984, patients with LBP receiving chiropractic, 60% between the ages of 21-50, 48% female, 37% had a duration of symptoms of less than 2 weeks, 66% prevalence of dependent outcome.	Chiropractic management as decided by the treating chiropractor.	Definitely better on global assessment by the fourth treatment session (dichotomous)		0 predictors present = 34%; 1 predictor present = 60%; 2 predictors present = 75%; 3 predictors present= 84%. Reported data does not permit calculation of 95%CrI.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	22.	Identifying patients with acute low back pain who are likely to recover at different rates.	1. baseline pain ≤ 7/10; 2. duration of current episode ≤ 5 days; 3. ≤ 1 previous episodes of LBP	Count of predictors	Derivation Hancock et al 2009 Pooled results from RCT	n=239, patients with LBP < 6 weeks duration presenting to a GP, mean age 41 years (SD 16), 44% female, mean 9 days of symptoms (SD 9).	1. Detuned ultrasound and placebo diclofenac; 2. Detuned ultrasound and active diclofenac; 3. Spinal manipulative therapy and placebo diclofenac; 4. Spinal manipulative therapy and active diclofenac.	Number of days from the baseline assessment until recovery from pain (≤1 on 0-10 NRS) (continuous)	Median days to recovery for 0 predictors = 22 days (11-33); 1 predictor = 22 days (19-24); 2 predictors = 15 days (12-18); 3 predictors = 6 days (4-8). Hazard ratios (reference category 0 predictors) 1 predictor = 1.3 (0.7-2.3); 2 predictors = 2.0 (1.2-3.6); 3 predictors = 3.5 (1.8-7.0).	Proportion recovered at 1 week with 0 predictors = 15%; 1 predictor = 13%; 2 predictors = 23%, 3 predictors = 60%. Proportion recovered at 12 weeks with 0 predictors = 70%; 1 predictor = 85%; 2 predictors = 95%; 3 predictors = 95%.
Surgical intervention	23.	Predicting need for surgical intervention due to non-response to conservative treatment in patients with herniated nucleus pulposus	1. pain intensity; 2. duration of symptoms; 3. crossed straight leg raise test; 4. muscle power grade; 5. number of dermatome deficits	Score chart	Derivation Buranapanitkit et al 2003 Retrospective single cohort	n=251, patients admitted to hospital with a diagnosis with lumbar herniated nucleus pulposus, mean age 38 years (range 15-60), 60% female, 47% had symptoms less than 3 months, 67% prevalence of dependent outcome.	6 weeks of rest, analgesia, anti-inflammatory medication, and physical therapy.	Requiring surgical intervention as no improvement from 6 weeks of conservative treatment (dichotomous)		Scores < 45 = 15%; scores 45-64 = 53%; scores ≥ 65 = 96%.
Symptom persistence	24.	Identifying patients with acute LBP who are likely to develop persistent symptoms	1. characteristics of current episode (more points for exacerbation of chronic LBP and sciatica); 2. difficulty in walking a short distance or climbing stairs; 3. difficulty rising from bed or chair; 4. duration of certificate to remain off work > 8 days; 5. taking part in a sport	Score chart	Derivation Valat et al 2000 Single cohort	n=2487, employed patients with acute LBP (< 8 days) presenting to GPs or Rheumatologists, mean age 41 years (SD 9), 43% female, all had symptoms < 8 days, 6% prevalence of dependent outcome.	98% received medication, 52% strict bed rest for < 3 days.	Persistence of unchanged or worsened LBP at week 7 following the initial consultation (dichotomous)		Scores ≤3 = 2.9%; scores 4-6 = 7.9%; scores > 6 = 19.1%

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	25.	Identifying patients who are likely to develop long-term persistent symptoms	1. female; 2. dissatisfaction with employment situation; 3. history of LBP; 4. radiating leg pain; 5. widespread pain; 6. two or more restrictions in spinal movement	Count of predictors	Derivation Thomas et al 1999 Single cohort	n=180 (167 in multivariable analysis), patients presenting in general practice with new episode of LBP, 59% female, 66% aged between 30-59 years, 75% symptoms < 4 weeks, 34% prevalence of dependent outcome.	Not specified.	≥2 / 10 pain on VAS and Hanover score < 75% at 1 week and 3 and 12 months (dichotomous)	Not reported, but accuracy of 5 or more predictors present calculated to be sens = 0.41 (0.29 - 0.55); spec = 0.92 (0.86 - 0.96); +LR = 5.3 (2.6 - 10.8); -LR = 0.64 (0.50 - 0.81).	0 - 2 predictors present = 6%; 3 predictor present = 27%; 4 predictors present = 35%; 5-6 predictors present= 70%
Work	26.	Predicting probable work outcome following outpatient rehabilitation	1. MMPI†††††††† depression score; 2. age; 3. duration of problem; 4. duration of time off work; 5. gender	Classification functions	Derivation Cairns et al 1984 Single cohort	n=100, patients with LBP attending outpatient rehabilitation, mean age 43 years (dispersion data not reported), 50% female, mean 3.5 years duration of symptoms (no dispersion data reported), 52% prevalence of return to work.	Outpatient rehabilitation program 3 hour per day, 5 days a week x 4 weeks, consisting of conditioning exercises, stress management, nutrition advice, medication reduction and biofeedback.	Work status at 1 year post-discharge - working, ready for or in vocational rehabilitation, not working (trichotomous)	67% classification accuracy.	
	27.	Predicting probable work outcome following inpatient rehabilitation	1. MMPI hysteria score; 2. age; 3. duration of problem; 4. income source (no disability income, worker's compensation, social security disability, worker's compensation and social security disability); 5. gender	Classification functions	Derivation Cairns et al 1984 Single cohort	n=100, patients with LBP attending inpatient rehabilitation, mean age 46 years (dispersion data not reported), 67%%, female, mean 8 years duration of symptoms (no dispersion data reported), 15% prevalence of return to work.	Multidisciplinary inpatient treatment program.	Work status at 1 year post-discharge - working, ready for or in vocational rehabilitation, not working (trichotomous)	73% classification accuracy.	

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	28.	Identifying patients consulting in primary care who are at risk of an adverse occupational outcome	1. patient's recovery expectations; 2. radiating pain; 3. previous back surgery; 4. pain intensity; 5. frequent change in position because of back pain; 6. irritability and bad temper; 7. difficulty sleeping.	Algorithm	Derivation Dionne et al 2005 Single cohort with split sample	Training set n=354, validation set n=506, adults absent from work due to LBP consulting in primary care, full sample demographics - mean age 39 years (SD 11), 42% female, 78% with recurrent or persistent back pain, 17% failed to return to work in good health at 2 years.	Not specified	Return to work in good health at 2 years (success / partial success / failure) (categorical)	For failure vs partial success / success: Training set - sens = 0.79; spec = 0.64. LRs not reported, but calculated from study data to be +LR = 2.2 (1.8 – 2.7); -LR = 0.33 (0.20 – 0.55).	For algorithm predicted failure to return to work in good health = 31% (training set). 95%CrI calculated to be 23.8% - 38.7%.
	29.	Identifying patients off work with chronic LBP and participating in a functional restoration program who are likely to return to work by 6 months	1. duration of complaints (months); 2. functional disability (ODQ score); 3. presence of disc herniation with associated radiculopathy (MRI‡‡‡‡‡‡‡ confirmed herniation or extrusion, unilateral pain, unilateral paraesthesia or pain below the knee in 1 leg, and a SLR discrepancy of at least 15° between legs); 4. fear avoidance beliefs (FABQ)	Nomogram	Derivation Heymans et al 2007 Retrospective single cohort	n=194, patients attending a physiotherapy functional restoration program and on sick leave due to LBP, mean age 42 years (SD 10), 67% female, mean 21 months (SD 40) of symptoms, 70% prevalence of dependent outcome.	Physiotherapy functional restoration program 3 times per week for 4-8 weeks, consisting of progressive aerobic and resistance exercises delivered using a cognitive-behavioural approach.	Return to work (including modified duties) 6 months following completion of functional restoration program (dichotomous)	Calibration – Slope index = 0.91. Discrimination - Area under receiver operating characteristic curve = 0.76. For a threshold of the nomogram of ≥50% predicted probability of return to work, sens = 62%; spec = 78%. LRs not reported, but approximated from study data to be +LR = 2.8 (1.7 – 4.5); -LR = 0.49 (0.38 – 0.64).	For ≥50% threshold of predicted probability of return to work, post-test probability = 87%. 95%CrI approximated to be 78.5% - 92.0%.

Outcome construct	CPR	Function	Predictors	Format	Study / Design	Patients	Treatment	Dependent outcome	Accuracy / Results	Post-test probability / Notes
	30.	Identifying patients on sick leave due to LBP who are at risk of more than 6 months of sick leave.	1. job satisfaction (good, reasonable, moderate or poor); 2. fear avoidance beliefs (FABQ); 3. pain intensity (VAS 0-10); 4. duration of complaints (weeks); 5. gender	Score chart	Derivation Heymans et al 2009 Pooled results from 3 RCTs	n=628, patients on sick leave due to LBP < 8 weeks, mean age 41 years (SD 10), 29% female, median 6 months duration of complaint (IQR 13.3), 19% prevalence of dependent outcome.	RCT 1 (n=134) used behaviourally orientated graded activity program vs usual care; RCT 2 (n=195) used a workplace intervention and graded activity vs usual care; RCT 3 (n=299) used high and low intensity back schools vs usual care.	Prolonged sick leave > 6 months (dichotomous)	For scores ≥10, sens = 0.32; spec = 0.89. LRs not reported, but approximated from study data to be +LR = 2.8 (2.0 – 4.0); -LR = 0.77 (0.68 – 0.87).	For scores ≥10, post-test probability = 41%. 95%CrI approximated to be 31.2% - 50.5%.

---

* straight leg raise
† low back pain
‡ general practitioner
§ Roland Morris disability questionnaire
** sensitivity
†† specificity
‡‡ positive likelihood ratio
§§ negative likelihood ratio
*** 95% credible interval
††† standard deviation
‡‡‡ clinical prediction rule
§§§ fear avoidance beliefs questionnaire
**** range of motion
†††† Oswestry disability questionnaire
‡‡‡‡ randomized controlled trial
§§§§ interquartile range
***** numerical rating scale
††††† odds ratio
‡‡‡‡‡ 95% confidence interval
§§§§§ visual analogue scale
****** analysis of variance
†††††† Northwick Park neck pain questionnaire
‡‡‡‡‡‡ global rating of change
§§§§§§ body mass index

******* electromyography
††††††† Minnesota multiphasic personality inventory
‡‡‡‡‡‡‡ magnetic resonance imaging

REFERENCES

Abboud, P., & Cabana, M. D. (2001). Understanding barriers to the adoption of clinical decision rules. *Annals of Emergency Medicine, 38*(6), 703-704.

Abrams, D., Davidson, M., Harrick, J., Harcourt, P., Zylinski, M., & Clancy, J. (2006). Monitoring the change: current trends in outcome measure usage in physiotherapy. *Manual Therapy, 11*(1), 46-53.

Adams, S. T., & Leveson, S. H. (2012). Clinical prediction rules. *British Medical Journal, 344*, d8312.

Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., . . . Cohen, G. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist, 34*(3), 341-382.

Al-Sayegh, N. A., George, S. E., Boninger, M. L., Rogers, J. C., Whitney, S. L., & Delitto, A. (2010). Spinal mobilization of postpartum low back and pelvic girdle pain: an evidence-based clinical rule for predicting responders and nonresponders. *Pm & R, 2*(11), 995-1005.

Al Nezari, N. H., Schneiders, A. G., & Hendrick, P. A. (2013). Neurological examination of the peripheral nervous system to diagnose lumbar spinal disc herniation with suspected radiculopathy: a systematic review and meta-analysis. *The Spine Journal, 13*(6), 657-674.

Allan, D. B., & Waddell, G. (1989). An historical perspective on low back pain and disability. *Acta Orthopaedica, 60*(S234), 1-23.

Allison, S. C. (2009). On "A guide to interpretation of studies..." Hancock M, et al. Phys Ther. 2009;89:698-704. *Physical Therapy, 89*(10), 1098-1099.

Alonso-Blanco, C., Fernandez-de-las-Penas, C., & Cleland, J. A. (2009). Preliminary clinical prediction rule for identifying patients with ankylosing spondylitis who are likely to respond to an exercise program: a pilot study. *American Journal of Physical Medicine & Rehabilitation, 88*(6), 445-454.

American College of Radiology. (2011). ACR Appropriateness Criteria - Acute Trauma to the Knee. Retrieved December, 2014, from https://acsearch.acr.org/docs/69419/Narrative/

Andersson, G. B. J. (1998). Epidemiology of low back pain. *Acta Orthopaedica Scandinavica Supplementum, 281*, 28-31.

Andersson, G. B. J. (1999). Epidemiological features of chronic low-back pain. *The Lancet, 354*(9178), 581-585.

Anon. (1937). Mathematics and Medicine. *The Lancet, 229*(5914), 31.

Apeldoorn, A. T., Bosselaar, H., Ostelo, R. W., Blom-Luberti, T., van der Ploeg, T., Fritz, J. M., . . . van Tulder, M. W. (2012). Identification of patients with chronic low back pain who might benefit from additional psychological assessment. *The Clinical Journal of Pain, 28*(1), 23-31.

Arthritis & Osteoporosis New South Wales. (2014). Don't turn your back on it. Retrieved June, 2014, from http://www.dontturnyourbackonit.com.au/

Arthritis Care. (2011). Have you got the 'S' factor? Retrieved June, 2014, from http://www.arthritiscare.org.uk/AboutArthritis/HaveyougottheSfactor

Ash, L. M., Modic, M. T., Obuchowski, N. A., Ross, J. S., Brant-Zawadzki, M. N., & Grooff, P. N. (2008). Effects of diagnostic information, per se, on patient outcomes in acute radiculopathy and low back pain. *American Journal of Neuroradiology, 29*(6), 1098-1103.

Ashworth, J., Konstantinou, K., & Dunn, K. M. (2011). Prognostic factors in non-surgically treated sciatica: a systematic review. *BMC Musculoskeletal Disorders, 12*(1), 208.

Assendelft, W. J. J., Morton, S. C., Emily, I. Y., Suttorp, M. J., & Shekelle, P. G. (2003). Spinal manipulative therapy for low back painA meta-analysis of effectiveness relative to other therapies. *Annals of Internal Medicine, 138*(11), 871-881.

Assmann, S. F., Pocock, S. J., Enos, L. E., & Kasten, L. E. (2000). Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet, 355*(9209), 1064-1069.

Atlas, S. J., & Deyo, R. A. (2001). Evaluating and managing acute low back pain in the primary care setting. *Journal of General Internal Medicine, 16*(2), 120-131.

Attride-Stirling, J. (2001). Thematic networks: an analytic tool for qualitative research. *Qualitative Research, 1*(3), 385-405.

Auleley, G.-R., Ravaud, P., Giraudeau, B., Kerboull, L., Nizard, R., Massin, P., . . . Durieux, P. (1997). Implementation of the Ottawa ankle rules in

France: a multicenter randomized controlled trial. *Journal of the American Medical Association, 277*(24), 1935-1939.

Australian Acute Musculoskeletal Pain Guidelines Group. (2003). Evidence-based management of acute musculoskeletal pain. Bowen Hills: Australian Academic Press.

Australian Bureau of Statistics. (2009). 3201.0 Population by Age and Sex, Australian States and Territories, Jun 2009. Canberra.

Australian Bureau of Statistics. (2012). 4364.0.55.001 - Australian Health Survey: First Results, 2011-12. Retrieved October, 2014, from http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/4364.0.55.001Main+Features12011-12?OpenDocument

Australian Bureau of Statistics. (2013). Appendix - National Health Survey 2011-12 Questionnaire. In 4363.0.55.001 - Australian Health Survey: Users' Guide, 2011-13. Retrieved October, 2014, from http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/4363.0.55.012011-13?OpenDocument

Australian Institute of Health and Welfare. (1997). First report on National Health Priority Areas 1996. AIHW Cat. No. PHE 1. Canberra: Commonwealth of Australia.

Australian Institute of Health and Welfare. (2013a). Australian hospital statistics 2011 - 12. Health services series no. 50. Cat. no. HSE 134. Canberra.

Australian Institute of Health and Welfare. (2013b). How do back problems affect quality of life? Retrieved October, 2014, from http://www.aihw.gov.au/back-problems/quality-of-life/

Australian Institute of Health and Welfare. (2013c). National health priority areas. Retrieved September, 2014, from http://www.aihw.gov.au/national-health-priority-areas/

Australian Institute of Health and Welfare. (2013d). National hospital morbidity database. Retrieved October, 2014, from http://www.aihw.gov.au/hospitals-data/national-hospital-morbidity-database/

Australian Institute of Health and Welfare. (2013e). Principle diagnosis data cubes. Separation statistics by principal diagnosis in ICD-10-AM, Australia, 2011-12. Retrieved October, 2014, from https://reporting.aihw.gov.au/Reports/openRVUrl.do?rsRID=SBIP%3A%2F%2FMETASERVER%2FAIHW%2FReleasedPublic%2FHospitals%2FReports%2FHDU_PDX%201112.srx%28Report%29

Australian Institute of Health and Welfare. (2013f). What is the health burden of back problems? Retrieved October, 2014, from http://www.aihw.gov.au/back-problems/health-burden/

Australian Institute of Health and Welfare. (2013g). Who gets back problems? Retrieved October, 2014, from http://www.aihw.gov.au/back-problems/prevalence/

Australian Institute of Health and Welfare. (2014). Health-care expenditure on arthritis and other musculoskeletal conditions 2008-09. Arthritis series no. 20. Cat. no. PHE 177. Canberra.

Australian Physiotherapy Association. (2013). Retrieved January, 2013, from http://www.physiotherapy.asn.au/APAWCM/Careers/Career_Paths/Sp

ecialisation_Pathway/APAWCM/Careers/Career_Paths/Specialisation.aspx

Axen, I., Jones, J. J., Rosenbaum, A., Lovgren, P. W., Halasz, L., Larsen, K., & Leboeuf-Yde, C. (2005). The Nordic Back Pain Subpopulation Program: validation and improvement of a predictive model for treatment outcome in patients with low back pain receiving chiropractic treatment. *Journal of Manipulative & Physiological Therapeutics, 28*(6), 381-385.

Babyak, M. A. (2004). What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic medicine, 66*(3), 411-421.

Bachmann, L. M., Haberzeth, S., Steurer, J., & ter Riet, G. (2004). The accuracy of the Ottawa knee rule to rule out knee fractures: a systematic review. *Annals of Internal Medicine, 140*(2), 121-124.

Balague, F., Mannion, A. F., Pellise, F., & Cedraschi, C. (2012). Non-specific low back pain. *Lancet, 379*(9814), 482-491.

Baraliakos, X., & Braun, J. (2011). Spondyloarthritides. *Best Practice and Research Clinical Rheumatology, 25*(6), 825-842.

Barlow, G., Nathwani, D., Myers, E., Sullivan, F., Stevens, N., Duffy, R., & Davey, P. (2008). Identifying barriers to the rapid administration of appropriate antibiotics in community-acquired pneumonia. *Journal of Antimicrobial Chemotherapy, 61*(2), 442-451.

Barnard, S., & Wiles, R. (2001). Evidence-based physiotherapy: physiotherapists' attitudes and experiences in the Wessex area. *Physiotherapy, 87*(3), 115-124.

Baron, J. A. (1994). Uncertainty in Bayes. *Medical Decision Making, 14*(1), 46-51.

Baxt, W. G. (1995). Application of artificial neural networks to clinical medicine. *The Lancet, 346*(8983), 1135-1138.

Bayes, T. R. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London, 53*, 370-418.

Beattie, P., & Nelson, R. (2006). Clinical prediction rules: what are they and what do they tell us? *Australian Journal of Physiotherapy, 52*(3), 157-163.

BenDebba, M., Torgerson, W. S., & Long, D. M. (2000). A validated, practical classification procedure for many persistent low back pain patients. *Pain, 87*(1), 89-97.

Beneciuk, J. M., Bishop, M. D., Fritz, J. M., Robinson, M. E., Asal, N. R., Nisenzon, A. N., & George, S. Z. (2013). The STarT Back Screening Tool and individual psychological measures: evaluation of prognostic capabilities for low back pain clinical outcomes in outpatient physical therapy settings. *Physical Therapy, 93*(3), 321-333.

Beneciuk, J. M., Bishop, M. D., & George, S. Z. (2009). Clinical prediction rules for physical therapy interventions: a systematic review. *Physical Therapy, 89*(2), 114-124.

Beneciuk, J. M., Fritz, J. M., & George, S. Z. (2014). The STarT Back Screening Tool for Prediction of 6-Month Clinical Outcomes: Relevance of Change Patterns in Outpatient Physical Therapy

Settings. *Journal of Orthopaedic & Sports Physical Therapy, 44*(9), 656-664.

Bener, A., Verjee, M., Dafeeah, E. E., Falah, O., Al-Juhaishi, T., Schlogl, J., . . . Khan, S. (2013). Psychological factors: anxiety, depression, and somatization symptoms in low back pain patients. *Journal of pain research, 6*, 95-101.

Bergstrom, G., Jensen, I. B., Bodin, L., Linton, S. J., & Nygren, A. L. (2001). The impact of psychologically different patient groups on outcome after a vocational rehabilitation program for long-term spinal pain patients. *Pain, 93*(3), 229-237.

Bero, L. A., Grilli, R., Grimshaw, J. M., Harvey, E., Oxman, A. D., & Thomson, M. A. (1998). Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. *British Medical Journal, 317*(7156), 465-468.

Beutel, B. G., Trehan, S. K., Shalvoy, R. M., & Mello, M. J. (2012). The Ottawa Knee Rule: Examining Use in an Academic Emergency Department. *Western Journal of Emergency Medicine, 13*(4), 366-372.

Bianchi, M. T., Alexander, B. M., & Cash, S. S. (2009). Incorporating uncertainty into medical decision making: an approach to unexpected test results. *Medical Decision Making, 29*(1), 116-124.

Billington, J., Fahey, T., & Galvin, R. (2012). Diagnostic accuracy of the STRATIFY clinical prediction rule for falls: A systematic review and meta-analysis. *BMC Family Practice, 13*(1), 76.

Billis, E. V., McCarthy, C. J., & Oldham, J. A. (2007). Subclassification of low back pain: a cross-country comparison. *European Spine Journal, 16*(7), 865-879.

Bishop, M. A., & Trout, J. D. (2002). 50 years of successful predictive modeling should be enough: Lessons for philosophy of science. *Philosophy of Science, 69*(S3), S197-S208.

Björk, J., Ekelund, U., & Ohlsson, M. (2012). Risk predictions for individual patients from logistic regression were visualized with bar–line charts. *Journal of Clinical Epidemiology, 65*, 335-342.

Boden, S. D., Davis, D. O., Dina, T. S., Patronas, N. J., & Wiesel, S. W. (1990). Abnormal magnetic-resonance scans of the lumbar spine in asymptomatic subjects. A prospective investigation. *Journal of Bone & Joint Surgery - American volume, 72*(3), 403-408.

Bogduk, N. (1999). Evidence-based clinical guidelines for the management of acute low back pain *The National Musculoskeletal Medicine Initiative*: NHMRC.

Bogduk, N. (2004a). Management of chronic low back pain. *Medical Journal of Australia, 180*(2), 79-83.

Bogduk, N. (2004b). *Practice guidelines for spinal diagnostic and treatment procedures*. San Francisco: International Spine Intervention Society.

Bogduk, N. (2005). *Clinical anatomy of the lumbar spine and sacrum* (Fourth ed.). Edinburgh: Elsevier Churchill Livingstone.

Bogduk, N., Dreyfuss, P., Baker, R., Yin, W., Landers, M., Hammer, M., & Aprill, C. N. (2008). Complications of spinal diagnostic and treatment procedures. *Pain Medicine, 9*(S1), S11-S34.

Bogduk, N., & Karasek, M. (2005). Precision diagnosis and treatment of back and neck pain. *CONTINUUM: Lifelong Learning in Neurology, 11*(6), 94-136.

Bombardier, C., van Tulder, M. W., Bronfort, G., Chou, R., Corbin, T., Deyo, R. A., . . . Tomlinson, G. (2014). Cochrane Back Group. *About the Cochrane Collaboration (Cochrane Review Groups (CRGs)), Issue 1. Art. No.: BACK*.

Boonen, A., van den Heuvel, R., van Tubergen, A., Goossens, M., Severens, J. L., van der Heijde, D., & van der Linden, S. (2005). Large differences in cost of illness and wellbeing between patients with fibromyalgia, chronic low back pain, or ankylosing spondylitis. *Annals of the rheumatic diseases, 64*(3), 396-402.

Boonen, A., & van der Linden, S. (2002). Rheumatism in the early 18th century: a doctor's diary. *Annals of the rheumatic diseases, 61*(5), 468-469.

Borkan, J. M., & Cherkin, D. C. (1996). An Agenda for Primary Care Research on Low Back Pain. *Spine, 21*(24), 2880-2884.

Borkan, J. M., Koes, B. W., Reis, S., & Cherkin, D. C. (1998). A Report From the Second International Forum for Primary Care Research on Low Back Pain: Reexamining Priorities. *Spine, 23*(18), 1992-1996.

Bossuyt, P. M., & Leeflang, M. (2008). Chapter 6. Developing criteria for including studies *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (Vol. Ver. 0.4): The Cochrane Collaboration.

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., . . . Standards for Reporting of Diagnostic Accuracy.

(2003). Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ, 326*(7379), 41-44.

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., . . . Lijmer, J. G. (2003). The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Annals of Internal Medicine, 138*(1), W1-12.

Bouter, L. M., Pennick, V., Bombardier, C., & Editorial Board of the Back Review Group. (2003). Cochrane Back Review Group. *Spine, 28*(12), 1215-1218.

Bouwmeester, W., van Enst, A., & van Tulder, M. (2009). Quality of low back pain guidelines improved. *Spine, 34*(23), 2562-2567.

Bouwmeester, W., Zuithoff, N. P. A., Mallett, S., Geerlings, M. I., Vergouwe, Y., Steyerberg, E. W., . . . Moons, K. G. M. (2012). Reporting and methods in clinical prediction research: a systematic review. *PLoS medicine, 9*(5), e1001221.

Brandt, H. C., Spiller, I., Song, I. H., Vahldiek, J. L., Rudwaleit, M., & Sieper, J. (2007). Performance of referral recommendations in patients with chronic back pain and suspected axial spondyloarthritis. *Annals of the rheumatic diseases, 66*(11), 1479-1484.

Braun, A., Saracbasi, E., Grifka, J., Schnitker, J., & Braun, J. (2011). Identifying patients with axial spondyloarthritis in primary care: how useful are items indicative of inflammatory back pain? *Annals of the rheumatic diseases, 70*(10), 1782-1787.

Braun, J., & Inman, R. (2010). Clinical significance of inflammatory back pain for diagnosis and screening of patients with axial spondyloarthritis. *Annals of the rheumatic diseases, 69*(7), 1264-1268.

Braun, J., & Sieper, J. (2007). Ankylosing spondylitis. *The Lancet, 369*(9570), 1379-1390.

Breasted, J. H. (1930). *The Edwin Smith Surgical Papyrus: Hieroglyphic transliteration, translation and commentary.* (Vol. 1): University of Chicago Press.

Brehaut, J. C., Graham, I. D., Wood, T. J., Taljaard, M., Eagles, D., Lott, A., . . . Kellerman, A. (2010). Measuring acceptability of clinical decision rules: validation of the Ottawa acceptability of decision rules instrument (OADRI) in four countries. *Medical Decision Making, 30*(3), 398-408.

Brehaut, J. C., Stiell, I. G., Graham, I. D., Brehaut, J. C., Stiell, I. G., & Graham, I. D. (2006). Will a new clinical decision rule be widely used? The case of the Canadian C-spine rule. *Academic Emergency Medicine, 13*(4), 413-420.

Brehaut, J. C., Stiell, I. G., Visentin, L., Graham, I. D., Brehaut, J. C., Stiell, I. G., . . . Graham, I. D. (2005). Clinical decision rules "in the real world": how a widely disseminated rule is used in everyday practice. *Academic Emergency Medicine, 12*(10), 948-956.

Brennan, G. P., Fritz, J. M., Hunter, S. J., Thackeray, A., Delitto, A., & Erhard, R. E. (2006). Identifying subgroups of patients with acute/subacute "nonspecific" low back pain: results of a randomized clinical trial. *Spine, 31*(6), 623-631.

Briggs, A. M., & Buchbinder, R. (2009). Back pain: a National Health Priority Area in Australia? *Med J Aust, 190*(9), 499-502.

Brock, A. J. (1916). Introduction. In C. Galenus & A. J. Brock (Eds.), *Galen. On the Natural Faculties* (pp. ix - xli). London: William Heinemann.

Brookes, S. T., Whitely, E., Egger, M., Smith, G. D., Mulheran, P. A., & Peters, T. J. (2004). Subgroup analyses in randomized trials: Risks of subgroup-specific analyses; power and sample size for the interaction test. *Journal of Clinical Epidemiology, 57*(3), 229-236.

Brown, T. (1828). On irritation of spinal nerves. *Glasgow Medical Journal, 1*, 131-160.

Buchbinder, R. (2008). Self-management education en masse: effectiveness of the Back Pain: Don't Take It Lying Down mass media campaign. *Medical Journal of Australia, 189*(10 Suppl), S29-32.

Buchbinder, R., & Jolley, D. (2004). Population based intervention to change back pain beliefs: three year follow up population survey. *British Medical Journal, 328*(7435), 321.

Buchbinder, R., Jolley, D., & Wyatt, M. (2001). Population based intervention to change back pain beliefs and disability: three part evaluation. *British Medical Journal, 322*(7301), 1516-1520.

Buderer, N. M. F. (1996). Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. *Academic Emergency Medicine, 3*(9), 895-900.

Buranapanitkit, B., Tautakul, K., Lim, A., Geater, A., & Chomchan, U. (2003). Clinical predictors for outcome of treatment of lumbar herniated nucleus palposus. *Journal of Musculoskeletal Research, 7*(1), 61-69.

Burgos-Vargas, R., & Braun, J. (2012). Inflammatory Back Pain. *Rheumatic Disease Clinics of North America, 38*(3), 487-499.

Bussieres, A. E., Taylor, J. A., & Peterson, C. (2007). Diagnostic imaging practice guidelines for musculoskeletal complaints in adults--an evidence-based approach. Part 1. Lower extremity disorders. *Journal of Manipulative and Physiological Therapeutics, 30*(9), 684-717.

Cabana, M. D., Rand, C. S., Powe, N. R., Wu, A. W., Wilson, M. H., Abboud, P. A., & Rubin, H. R. (1999). Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA, 282*(15), 1458-1465.

Cabana, M. D., Rushton, J. L., & Rush, A. J. (2002). Implementing practice guidelines for depression: applying a new framework to an old problem. *General Hospital Psychiatry, 24*(1), 35-42.

Cai, C., Pua, Y. H., & Lim, K. C. (2009). A clinical prediction rule for classifying patients with low back pain who demonstrate short-term improvement with mechanical lumbar traction. *European Spine Journal, 18*(4), 554-561.

Cairns, D., Mooney, V., & Crane, P. (1984). Spinal pain rehabilitation: inpatient and outpatient treatment results and development of predictors for outcome. *Spine, 9*(1), 91-95.

Carragee, E. J., Don, A. S., Hurwitz, E. L., Cuellar, J. M., Carrino, J. A., & Herzog, R. (2009). Does discography cause accelerated progression of degeneration changes in the lumbar disc: a ten-year matched cohort study. *Spine, 34*(21), 2338-2345.

Carragee, E. J., Haldeman, S., & Hurwitz, E. L. (2007). The pyrite standard: the Midas touch in the diagnosis of axial pain syndromes. *The Spine Journal, 7*(1), 27-31.

Carragee, E. J., Tanner, C. M., Khurana, S., Hayward, C., Welsh, J., Date, E., . . . Hagle, C. (2000). The rates of false-positive lumbar discography in select patients without low back symptoms. *Spine, 25*(11), 1373-1381.

Carter, G. M., & Polger, P. D. (1986). *A 20-year summary of National Weather Service verification results for temperature and precipitation*: US Department of Commerce, National Oceanic and Atmospheric Administration.

Cassidy, J. D., Côté, P., Carroll, L. J., & Kristman, V. (2005). Incidence and course of low back pain episodes in the general population. *Spine, 30*(24), 2817-2823.

Chan, C. C., Inrig, T., Molloy, C. B., Stone, M. A., & Derzko-Dzulynsky, L. (2012). Prevalence of inflammatory back pain in a cohort of patients with anterior uveitis. *American Journal of Ophthalmology, 153*(6), 1025-1030.e1021.

Cheung, K. M. C., Karppinen, J., Chan, D., Ho, D. W. H., Song, Y.-Q., Sham, P., . . . Luk, K. D. K. (2009). Prevalence and pattern of lumbar magnetic resonance imaging changes in a population study of one thousand forty-three individuals. *Spine, 34*(9), 934-940.

Childs, J. D., & Cleland, J. A. (2006). Development and application of clinical prediction rules to improve decision making in physical therapist practice. *Physical Therapy, 86*(1), 122-131.

Childs, J. D., Flynn, T. W., & Fritz, J. M. (2006). A perspective for considering the risks and benefits of spinal manipulation in patients with low back pain. *Manual Therapy, 11*(4), 316-320.

Childs, J. D., Flynn, T. W., & Wainner, R. S. (2012). Low back pain: do the right thing and do it now. *Journal of Orthopaedic & Sports Physical Therapy, 42*(4), 296-299.

Childs, J. D., Fritz, J. M., Flynn, T. W., Irrgang, J. J., Johnson, K. K., Majkowski, G. R., & Delitto, A. (2004). A clinical prediction rule to identify patients with low back pain most likely to benefit from spinal manipulation: A validation study. *Annals of Internal Medicine, 141*(12), 920-928.

Childs, J. D., Fritz, J. M., Piva, S. R., & Erhard, R. E. (2003). Clinical decision making in the identification of patients likely to benefit from spinal manipulation: a traditional versus an evidence-based approach. *Journal of Orthopaedic and Sports Physical Therapy, 33*(5), 259-272.

Chou, R., Deyo, R. A., & Jarvik, J. G. (2012). Appropriate use of lumbar imaging for evaluation of low back pain. *Radiol Clin North Am, 50*(4), 569-585.

Chou, R., Fu, R., Carrino, J. A., & Deyo, R. A. (2009). Imaging strategies for low-back pain: systematic review and meta-analysis. *The Lancet, 373*(9662), 463-472.

Chou, R., Qaseem, A., Owens, D. K., & Shekelle, P. (2011). Diagnostic imaging for low back pain: advice for high-value health care from the American College of Physicians. *Annals of Internal Medicine, 154*(3), 181-189.

Chou, R., Qaseem, A., Snow, V., Casey, D., Cross, J. J. T., Shekelle, P., & Owens, D. K. (2007). Diagnosis and Treatment of Low Back Pain: A Joint Clinical Practice Guideline from the American College of Physicians and the American Pain Society. *Annals of Internal Medicine, 147*(7), 478-491.

Cieza, A., Stucki, G., Weigl, M., Disler, P., Jackel, W., van der Linden, S., . . . de Bie, R. (2004). ICF Core Sets for low back pain. *Journal of Rehabilitation Medicine, Suppl. 44*, 69-74.

Cleland, J. A., Fritz, J. M., Kulig, K., Davenport, T. E., Eberhart, S., Magel, J., & Childs, J. D. (2009). Comparison of the effectiveness of three manual physical therapy techniques in a subgroup of patients with low back pain who atisfy a clinical prediction rule: A randomized clinical trial. *Spine, 34*(25), 2720-2729.

Cleland, J. A., Fritz, J. M., Whitman, J. M., Childs, J. D., & Palmer, J. A. (2006). The use of a lumbar spine manipulation technique by physical therapists in patients who satisfy a clinical prediction rule: a case series. *Journal of Orthopaedic & Sports Physical Therapy, 36*(4), 209-214.

Clopper, C. J., & Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika, 26*(4), 404-413.

Cochrane Back Group. (2009). Search strategies for Back Group Specialized Registry. Retrieved January, 2009, from http://www.mrw.interscience.wiley.com/cochrane/clabout/articles/BACK/frame.html

Cohen, S. P., Larkin, T. M., Barna, S. A., Palmer, W. E., Hecht, A. C., & Stojanovic, M. P. (2005). Lumbar discography: a comprehensive review of outcome studies, diagnostic accuracy, and principles. *Regional anesthesia and pain medicine, 30*(2), 163-183.

Concato, J., Feinstein, A. R., & Holford, T. R. (1993). The risk of determining risk with multivariable models. *Annals of Internal Medicine, 118*(3), 201-210.

Cook, C. (2008). Potential pitfalls of clinical prediction rules. *Journal of Manual and Manipulative Therapy, 16*(2), 69-71.

Cook, C., Brismee, J. M., Pietrobon, R., Sizer Jr, P., Hegedus, E., & Riddle, D. L. (2010). Development of a Quality Checklist Using Delphi Methods for Prescriptive Clinical Prediction Rules: The QUADCPR. *Journal of Manipulative and Physiological Therapeutics, 33*(1), 29-41.

Cook, C., Brown, C., Michael, K., Isaacs, R., Howes, C., Richardson, W. S., . . . Hegedus, E. (2011). The clinical value of a cluster of patient history and observational findings as a diagnostic support tool for lumbar spine stenosis. *Physiotherapy Research International, 16*(3), 170-178.

Cook, C., Learman, K., O'Halloran, B., Showalter, C., Kabbaz, V., Goode, A., & Wright, A. (2013). Which prognostic factors for low back pain are generic predictors of outcome across a range of recovery domains? *Physical Therapy, 93*(1), 32-40.

Cook, E. F., & Goldman, L. (1984). Empiric comparison of multivariate analytic techniques: Advantages and disadvantages of recursive partitioning analysis. *Journal of Chronic Diseases, 37*(9–10), 721-731.

Cooke, G., Valenti, L., Glasziou, P., & Britt, H. (2013). Common general practice presentations and publication frequency. *Aust Fam Physician, 42*(1-2), 65-68.

Costa, L. d. C. M., Koes, B. W., Pransky, G., Borkan, J. M., Maher, C. G., & Smeets, R. J. (2013). Primary Care Research Priorities in Low Back Pain: An Update. *Spine, 38*(2), 148-156.

Costa, L. d. C. M., Maher, C. G., Hancock, M. J., McAuley, J. H., Herbert, R. D., & Costa, L. O. P. (2012). The prognosis of acute and persistent low-back pain: a meta-analysis. *Canadian Medical Association Journal*.

Coste, J., Paolaggi, J. B., & Spira, A. (1992). Classification of nonspecific low back pain. II. Clinical diversity of organic forms. *Spine, 17*(9), 1038-1042.

Côté, A. M., Durand, M. J., Tousignant, M., & Poitras, S. (2009). Physiotherapists and use of low back pain guidelines: a qualitative study of the barriers and facilitators. *Journal of Occupational Rehabilitation, 19*(1), 94-105.

Coxe, J. R. (1846). *The Writings of Hippocrates and Galen. Epitomised from the original latin translations.* Philadelphia: Linday and Blakiston.

Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory Into Practice, 39*(3), 124-130.

Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic Medicine, 78*(8), 775-780.

Croskerry, P. (2009). Clinical cognition and diagnostic error: applications of a dual process model of reasoning. *Advances in health sciences education, 14*(1), 27-35.

Dagenais, S., Caro, J., & Haldeman, S. (2008). A systematic review of low back pain cost of illness studies in the United States and internationally. *Spine Journal, 8*(1), 8-20.

Dahri, K., & Loewen, P. (2007). The risk of bleeding with warfarin: a systematic review and performance analysis of clinical prediction rules. *Thrombosis & Haemostasis, 98*(5), 980-987.

Dana, J., & Thomas, R. (2006). In defense of clinical judgment … and mechanical prediction. *Journal of Behavioral Decision Making, 19*(5), 413-428.

Dannapfel, P., Peolsson, A., & Nilsen, P. (2013). What supports physiotherapists' use of research in clinical practice? A qualitative study in Sweden. *Implementation Science, 8*(1), 31.

Davidson, M. (2002). The interpretation of diagnostic tests: A primer for physiotherapists. *Australian Journal of Physiotherapy, 48*(3), 227-232.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgement. *Science, 243*, 1668-1674.

de Graaf, I., Prak, A., Bierma-Zeinstra, S., Thomas, S., Peul, W. C., & Koes, B. W. (2006). Diagnosis of lumbar spinal stenosis: a systematic review of the accuracy of diagnostic tests. *Spine, 31*(10), 1168-1176.

de Morton, N. A. (2009). The PEDro scale is a valid measure of the methodological quality of clinical trials: a demographic study. *Australian Journal of Physiotherapy, 55*(2), 129-133.

de Schepper, E. I. T., Overdevest, G. M., Suri, P., Peul, W. C., Oei, E. H. G., Koes, B. W., . . . Luijsterburg, P. A. J. (2013). Diagnosis of lumbar spinal stenosis: an updated systematic review of the accuracy of diagnostic tests. *Spine, 38*(8), E469-E481.

de Vet, H. C. W., Heymans, M. W., Dunn, K. M., Pope, D. P., van der Beek, A. J., Macfarlane, G. J., . . . Croft, P. R. (2002). Episodes of low back pain: a proposal for uniform definitions to be used in research. *Spine, 27*(21), 2409-2416.

Delitto, A., Erhard, R. E., & Bowling, R. W. (1995). A treatment-based classification approach to low back syndrome: identifying and staging patients for conservative treatment. *Physical Therapy, 75*(6), 470-485.

Delitto, A., George, S. Z., Van Dillen, L. R., Whitman, J. M., Sowa, G., Shekelle, P., . . . Godges, J. J. (2012). Low back pain. Clinical Practice Guidelines Linked to the International Classification of Functioning, Disability, and Health from the Orthopaedic Section of the American Physical Therapy Association. *Journal of Orthopaedic & Sports Physical Therapy, 42*(4), A1-57.

Dembe, A. E. (1996). *Occupation and Disease: How Social Factors Affect the Conception of Work-Related Disorders*. USA: Yale University Press.

Deming, D. (2010). *Science and Technology in World History, Vol. 1: The Ancient World and Classical Civilization.* North Carolina: McFarland and Company.

Dennison, C. R., Mendez-Tellez, P. A., Wang, W., Pronovost, P. J., & Needham, D. M. (2007). Barriers to low tidal volume ventilation in

acute respiratory distress syndrome: Survey development, validation, and results. *Critical Care Medicine, 35*(12), 2747-2754.

DePalma, M. J., Ketchum, J. M., & Saullo, T. (2011). What is the source of chronic low back pain and does age play a role? *Pain Medicine, 12*(2), 224-233.

Derby, R., Lee, S. H., Kim, B. J., Chen, Y., Aprill, C. N., & Bogduk, N. (2005). Pressure-Controlled Lumbar Discography in Volunteers Without Low Back Symptoms. *Pain Medicine, 6*(3), 213-221.

Derogatis, L. R. (1977). *Symptoms Checklist-90. Administration, scoring and procedures manual for the revised version.* Baltimore: Clinical Psychometric Research.

DeRosa, C. P., & Porterfield, J. A. (1992). A physical therapy model for the treatment of low back pain. *Physical Therapy, 72*(4), 261-269.

Deville, W. L., van der Windt, D. A., Dzaferagic, A., Bezemer, P. D., & Bouter, L. M. (2000). The test of Lasegue: systematic review of the accuracy in diagnosing herniated discs. *Spine, 25*(9), 1140-1147.

Deyo, R. A., Mirza, S. K., & Martin, B. I. (2006). Back pain prevalence and visit rates: estimates from US national surveys, 2002. *Spine, 31*(23), 2724-2727.

Deyo, R. A., Mirza, S. K., Turner, J. A., & Martin, B. I. (2009). Overtreating chronic back pain: time to back off? *The Journal of the American Board of Family Medicine, 22*(1), 62-68.

Deyo, R. A., Rainville, J., & Kent, D. L. (1992). What can the history and physical examination tell us about low back pain? *Journal of the American Medical Association, 268*(6), 760-765.

Deyo, R. A., & Weinstein, J. N. (2001). Low back pain. *New England Journal of Medicine, 344*(5), 363-370.

Dionne, C. E. (2005). Psychological distress confirmed as predictor of long-term back-related functional limitations in primary care settings. *Journal of Clinical Epidemiology, 58*(7), 714-718.

Dionne, C. E., Bourbonnais, R., Frémont, P., Rossignol, M., Stock, S. R., & Larocque, I. (2005). A clinical return-to-work rule for patients with back pain. *Canadian Medical Association Journal, 172*(12), 1559-1567.

Dionne, C. E., Dunn, K. M., Croft, P. R., Nachemson, A. L., Buchbinder, R., Walker, B. F., . . . Leboeuf-Yde, C. (2008). A consensus approach toward the standardization of back pain definitions for use in prevalence studies. *Spine, 33*(1), 95-103.

Dionne, C. E., Koepsell, T. D., Von Korff, M., Deyo, R. A., Barlow, W. E., & Checkoway, H. (1997). Predicting long-term functional limitations among back pain patients in primary care settings. *Journal of Clinical Epidemiology, 50*(1), 31-43.

Dionne, C. E., Le Sage, N., Franche, R.-L., Dorval, M., Bombardier, C., & Deyo, R. A. (2011). Five questions predicted long-term, severe, back-related functional limitations: evidence from three large prospective studies. *Journal of Clinical Epidemiology, 64*(1), 54-66.

Doust, J. (2009). Diagnosis in general practice: using probabilistic reasoning. *BMJ, 339*(7729), 1080-1082.

Downie, A., Williams, C. M., Henschke, N., Hancock, M. J., Ostelo, R. W., de Vet, H. C., . . . Maher, C. G. (2013). Red flags to screen for

malignancy and fracture in patients with low back pain: systematic review. *British Medical Journal, 347*, f7095.

Du Bois, M., & Donceel, P. (2008). A screening questionnaire to predict no return to work within 3 months for low back pain claimants. *European Spine Journal, 17*(3), 380-385.

Dunn, K. M., Jordan, K. P., & Croft, P. R. (2010). Recall of medication use, self-care activities and pain intensity: a comparison of daily diaries and self-report questionnaires among low back pain patients. *Primary Health Care Research & Development, 11*(1), 93-102.

Eagles, D., Stiell, I. G., Clement, C. M., Brehaut, J., Kelly, A. M., Mason, S., . . . Perry, J. J. (2008). International survey of emergency physicians' priorities for clinical decision rules. *Academic Emergency Medicine, 15*(2), 177-182.

Eagles, D., Stiell, I. G., Clement, C. M., Brehaut, J., Taljaard, M., Kelly, A. M., . . . Perry, J. J. (2008). International survey of emergency physicians' awareness and use of the Canadian Cervical-Spine rule and the Canadian Computed Tomography Head Rule. *Academic Emergency Medicine, 15*(12), 1256-1261.

Eddy, D. M. (2011). The origins of evidence-based medicine—a personal perspective. *Virtual Mentor, 13*(1), 55-60.

Edwards, I., Jones, M. A., Carr, J., Braunack-Mayer, A., & Jensen, G. M. (2004). Clinical reasoning strategies in physical therapy. *Physical Therapy, 84*(4), 312-330.

Edwards, P., Clarke, M., DiGuiseppi, C., Pratap, S., Roberts, I., & Wentz, R. (2002). Identification of randomized controlled trials in systematic

reviews: accuracy and reliability of screening records. *Statistics in Medicine, 21*(11), 1635-1640.

Ehrlich, G. E., & Khaltaev, N. G. (1999). *Low back pain initiative*. Geneva: World Health Organization.

Einhorn, H. J. (1986). Accepting error to make less error. *Journal of Personality Assessment, 50*(3), 387-395.

Elstein, A. S., & Schwarz, A. (2002). Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *British Medical Journal, 324*(7339), 729-732.

Ely, J. W., Graber, M. L., & Croskerry, P. (2011). Checklists to reduce diagnostic errors. *Academic Medicine, 86*(3), 307-313.

Endean, A., Palmer, K. T., & Coggon, D. (2011). Potential of magnetic resonance imaging findings to refine case definition for mechanical low back pain in epidemiological studies: a systematic review. *Spine, 36*(2), 160-169.

Engel, A., MacVicar, J., & Bogduk, N. (2014). A philosophical foundation for diagnostic blocks, with criteria for their validation. *Pain Medicine, 15*(6), 998-1006.

Erichsen, J. E. (1867). *On railway and other injuries of the nervous system*. Philadelphia: Henry C. Lea.

Eva, K. W. (2005). What every teacher needs to know about clinical reasoning. *Medical education, 39*(1), 98-106.

Fade, S., & Swift, J. (2011). Qualitative research in nutrition and dietetics: data analysis issues. *Journal of Human Nutrition and Dietetics, 24*(2), 106-114.

Fairbank, J., Gwilym, S. E., France, J. C., Daffner, S. D., Dettori, J., Hermsmeyer, J., & Andersson, G. B. (2011). The role of classification of chronic low back pain. *Spine, 36*, S19-S42.

Feinstein, A. R. (2002). Misguided efforts and future challenges for research on "diagnostic tests". *Journal of Epidemiology and Community Health, 56*(5), 330-332.

Field, J., & Newell, D. (2012). Relationship between STarT Back Screening Tool and prognosis for low back pain patients receiving spinal manipulative therapy. *Chiropractic & Manual Therapies, 20*, 17.

Flynn, T., Fritz, J., Whitman, J., Wainner, R., Magel, J., Rendeiro, D., . . . Allison, S. (2002). A clinical prediction rule for classifying patients with low back pain who demonstrate short-term improvement with spinal manipulation. *Spine, 27*(24), 2835-2843.

Flynn, T., Smith, B., & Chou, R. (2011). Appropriate use of diagnostic imaging in low back pain: a reminder that unnecessary imaging may do as much harm as good. *Journal of Orthopaedic & Sports Physical Therapy, 41*(11), 838-846.

Fortin, J. D., Dwyer, A. P., West, S., & Pier, J. (1994). Sacroiliac joint: pain referral maps upon applying a new injection/arthrography technique: Part I: asymptomatic volunteers. *Spine, 19*(13), 1475-1482.

Foster, N. E., Dziedzic, K. S., Windt, D. A. W. M., Fritz, J. M., & Hay, E. M. (2009). Research priorities for non-pharmacological therapies for common musculoskeletal problems: nationally and internationally agreed recommendations. *BMC Musculoskeletal Disorders, 10*(1), 3.

Foster, N. E., Hill, J. C., & Hay, E. M. (2011). Subgrouping patients with low back pain in primary care: are we getting any better at it? *Manual Therapy, 16*(1), 3-8.

Foster, N. E., Hill, J. C., O'Sullivan, P., & Hancock, M. (2013). Stratified models of care. *Best Practice & Research Clinical Rheumatology, 27*(5), 649-661.

Foster, N. E., Mullis, R., Hill, J. C., Lewis, M., Whitehurst, D. G. T., Doyle, C., . . . Sowden, G. (2014). Effect of Stratified Care for Low Back Pain in Family Practice (IMPaCT Back): A Prospective Population-Based Sequential Comparison. *The Annals of Family Medicine, 12*(2), 102-111.

Freburger, J. K., Holmes, G. M., Agans, R. P., Jackman, A. M., Darter, J. D., Wallace, A. S., . . . Carey, T. S. (2009). The rising prevalence of chronic low back pain. *Archives of Internal Medicine, 169*(3), 251-258.

French, S. D., Charity, M. J., Forsdike, K., Gunn, J. M., Polus, B. I., Walker, B. F., . . . Britt, H. C. (2013). Chiropractic Observation and Analysis Study (COAST): providing an understanding of current chiropractic practice. *Medical Journal of Australia, 199*(10), 687-691.

Fritz, J. M. (2009). Clinical prediction rules in physical therapy: coming of age? *Journal of Orthopaedic & Sports Physical Therapy, 39*(3), 159-161.

Fritz, J. M., Beneciuk, J. M., & George, S. Z. (2011). Relationship between categorization with the STarT Back Screening Tool and prognosis for people receiving physical therapy for low back pain. *Physical Therapy, 91*(5), 722-732.

Fritz, J. M., Brennan, G. P., Clifford, S. N., Hunter, S. J., & Thackeray, A. (2006). An examination of the reliability of a classification algorithm for subgrouping patients with low back pain. *Spine, 31*(1), 77-82.

Fritz, J. M., Brennan, G. P., & Leaman, H. (2006). Does the evidence for spinal manipulation translate into better outcomes in routine clinical care for patients with occupational low back pain? A case-control study. *Spine Journal: Official Journal of the North American Spine Society, 6*(3), 289-295.

Fritz, J. M., Childs, J. D., & Flynn, T. W. (2005). Pragmatic application of a clinical prediction rule in primary care to identify patients with low back pain with a good prognosis following a brief spinal manipulation intervention. *BMC Family Practice, 6*(29).

Fritz, J. M., Delitto, A., & Erhard, R. E. (2003). Comparison of classification-based physical therapy with therapy based on clinical practice guidelines for patients with acute low back pain: a randomized clinical trial. *Spine, 28*(13), 1363-1371.

Fritz, J. M., Hancock, M., Herbert, R., & Riddle, D. (2009a). July 2009 Debate: Clinical Prediction Rules - Part 1. Retrieved 27th Feb 2010, from http://www.scienceaudio.net/ptj/200907/ptj_200907_debate_1.mp3

Fritz, J. M., Hancock, M., Herbert, R., & Riddle, D. (2009b). July 2009 Debate: Clinical Prediction Rules - Part 2. Retrieved 27th Feb 2010, from http://www.scienceaudio.net/ptj/200907/ptj_200907_debate_2.mp3

Fritz, J. M., & Irrgang, J. J. (2001). A comparison of a modified Oswestry low back pain disability questionnaire and the Quebec back pain disability scale. *Physical Therapy, 81*(2), 776-788.

Fritz, J. M., Lindsay, W., Matheson, J. W., Brennan, G. P., Hunter, S. J., Moffit, S. D., . . . Rodriquez, B. (2007). Is there a subgroup of patients with low back pain likely to benefit from mechanical traction? Results of a randomized clinical trial and subgrouping analysis. *Spine, 32*(26), E793-E800.

Fritz, J. M., Piva, S. R., & Childs, J. D. (2005). Accuracy of the clinical examination to predict radiographic instability of the lumbar spine. *European Spine Journal, 14*(8), 743-750.

Fritz, J. M., & Wainner, R. S. (2001). Examining diagnostic tests: an evidence-based perspective. *Physical Therapy, 81*(9), 1546-1564.

Fritz, J. M., Whitman, J. M., Flynn, T. W., Wainner, R. S., & Childs, J. D. (2004). Factors Related to the Inability of Individuals with Low Back Pain to Improve with a Spinal Manipulation. *Physical Therapy, 84*(2), 173-190.

Fullerton, J. B., & Silverman, M. E. (2009). Claudius Galen of Pergamum: authority of medieval medicine. *Clinical cardiology, 32*(11), E82-E84.

Galvin, R., Geraghty, C., Motterlini, N., Dimitrov, B. D., & Fahey, T. (2011). Prognostic value of the ABCD2 clinical prediction rule: a systematic review and meta-analysis. *Family practice, 28*(4), 366-376.

Gatchel, R. J., Peng, Y. B., Peters, M. L., Fuchs, P. N., & Turk, D. C. (2007). The biopsychosocial approach to chronic pain: scientific advances and future directions. *Psychological bulletin, 133*(4), 581-624.

Geersing, G.-J., Bouwmeester, W., Zuithoff, P., Spijker, R., Leeflang, M. M. G., & Moons, K. G. M. (2012). Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. *PloS one, 7*(2), e32844.

George, S. Z., Bialosky, J. E., & Donald, D. A. (2005). The centralization phenomenon and fear-avoidance beliefs as prognostic factors for acute low back pain: a preliminary investigation involving patients classified for specific exercise. *Journal of Orthopaedic & Sports Physical Therapy, 35*(9), 580-588.

Gibson, J. N., & Waddell, G. (2007). Surgical interventions for lumbar disc prolapse: updated Cochrane Review. *Spine, 32*(16), 1735-1747.

Glynn, P. E., & Weisbach, P. D. (2010). Chapter 9. Case Studies. *Clinical Prediction Rules. A Physical Therapy Reference Manual* (pp. 226-232). Sudbury, Massachusetts: Jones and Barlett Publishers.

Golder, V., & Schachna, L. (2013). Ankylosing spondylitis. An update. *Aust Fam Physician, 42*(11), 780-784.

Gore, M., Sadosky, A., Stacey, B. R., Tai, K., & Leslie, D. (2012). The burden of chronic low back pain: clinical comorbidities, treatment patterns, and health care costs in usual care settings. *Spine, 37*(11), E668-E677.

Graber, M. L., Franklin, N., & Gordon, R. (2005). Diagnostic error in internal medicine. *Archives of Internal Medicine, 165*(13), 1493-1499.

Graber, M. L., Gordon, R., & Franklin, N. (2002). Reducing Diagnostic Errors in Medicine: What's the Goal? *Academic Medicine, 77*(10), 981-992.

Graham, I. D., Logan, J., Harrison, M. B., Straus, S. E., Tetroe, J., Caswell, W., & Robinson, N. (2006). Lost in knowledge translation: time for a map? *Journal of Continuing Education in the Health Professions, 26*(1), 13-24.

Graham, I. D., Stiell, I. G., Laupacis, A., McAuley, L., Howell, M., Clancy, M., . . . Wells, G. (2001). Awareness and use of the Ottawa ankle and knee rules in 5 countries: can publication alone be enough to change practice? *Annals of Emergency Medicine, 37*(3), 259-266.

Graham, I. D., Stiell, I. G., Laupacis, A., O'Connor, A. M., & Wells, G. A. (1998). Emergency physicians' attitudes toward and use of clinical decision rules for radiography. *Academic Emergency Medicine, 5*(2), 134-140.

Greenwood, J., & Parsons, M. (2000). A guide to the use of focus groups in health care research: Part 2. *Contemporary Nurse, 9*(2), 181-191.

Gregg, C. D., Dean, S., & Schneiders, A. G. (2009). Variables associated with active spondylolysis. *Physical Therapy in Sport, 10*(4), 121-124.

Grobman, W. A., & Stamilio, D. M. (2006). Methods of clinical prediction. *American journal of obstetrics and gynecology, 194*(3), 888-894.

Grol, R., & Wensing, M. (2004). What drives change? Barriers to and incentives for achieving evidence-based practice. *Medical Journal of Australia, 180*(6), S57-S60.

Grotle, M., Brox, J. I., Veierød, M. B., Glomsrød, B., Lønn, J. H., & Vøllestad, N. K. (2005). Clinical course and prognostic factors in acute low back pain: patients consulting primary care for the first time. *Spine, 30*(8), 976-982.

Grove, W. M., & Lloyd, M. (2006). Meehl's contribution to clinical versus statistical prediction. *Journal of Abnormal Psychology, 115*(2), 192-194.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, Public Policy, and Law, 2*(2), 293-323.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: a meta-analysis. *Psychological assessment, 12*(1), 19-30.

Gruber, P., & Boeni, T. (2008). History of Spinal Disorders. In N. Boos & M. Aebi (Eds.), *Spinal Disorders: Fundamentals of Diagnosis and Treatment* (pp. 1-37). Berlin: Springer-Verlag.

Guyatt, G. H. (2008). Preface. In G. H. Guyatt, D. Rennie, M. O. Meade & D. J. Cook (Eds.), *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. USA: McGraw-Hill Medical.

Guyatt, G. H., Cairns, J., Churchill, D., Cook, D. J., Haynes, B., Hirsh, J., . . . Nishikawa, J. (1992). Evidence-based medicine: a new approach to teaching the practice of medicine. *Journal of the American Medical Association, 268*(17), 2420-2425.

Guyatt, G. H., Sackett, D., Cook, D. J., & Evidence Based Medicine Working Group. (1994). Users' guides to the medical literature. II. How to use an article about therapy or prevention. B. What were the results and how will they help me in caring for my patients? *Journal of the American Medical Association, 271*, 59-63.

Hall, C. C., Ariss, L., & Todorov, A. (2007). The illusion of knowledge: When more information reduces accuracy and increases confidence. *Organizational Behavior and Human Decision Processes, 103*(2), 277-290.

Hall, H., McIntosh, G., & Boyle, C. (2009). Effectiveness of a low back pain classification system. *Spine Journal: Official Journal of the North American Spine Society, 9*(8), 648-657.

Hallegraeff, J. M., de Greef, M., Winters, J. C., & Lucas, C. (2009). Manipulative therapy and clinical prediction criteria in treatment of acute nonspecific low back pain. *Perceptual & Motor Skills, 108*(1), 196-208.

Hamilton, L., Macgregor, A., Warmington, V., Pinch, E., & Gaffney, K. (2014). The prevalence of inflammatory back pain in a UK primary care population. *Rheumatology, 53*(1), 161-164.

Hamm, L., Mikkelsen, B., Kuhr, J., Støvring, H., Munck, A., & Kragstrup, J. (2003). Danish physiotherapists' management of low back pain. *Advances in Physiotherapy, 5*(3), 109-113.

Hancock, M. J., Herbert, R. D., & Maher, C. G. (2009). A guide to interpretation of studies investigating subgroups of responders to physical therapy interventions. *Physical Therapy, 89*(7), 698-704.

Hancock, M. J., Koes, B. W., Ostelo, R., & Peul, W. C. (2011). Diagnostic accuracy of the clinical examination in identifying the level of herniation in patients with sciatica. *Spine, 36*(11), E712-E719.

Hancock, M. J., Maher, C. G., & Herbert, R. D. (2008). Answer to the letter to the editor of J. Hebert et al. concerning "Hancock MJ, Maher CG,

Latimer J, Herbert RD, McAuley JH (2008) Independent evaluation of a clinical prediction rule for spinal manipulative therapy: A randomised controlled trial. *European Spine Journal, 17*(10), 1403-1404.

Hancock, M. J., Maher, C. G., & Latimer, J. (2008). Spinal manipulative therapy for acute low back pain: a clinical perspective. *Journal of Manual & Manipulative Therapy, 16*(4), 198-203.

Hancock, M. J., Maher, C. G., Latimer, J., Herbert, R. D., & McAuley, J. H. (2008). Independent evaluation of a clinical prediction rule for spinal manipulative therapy: a randomised controlled trial. *European Spine Journal, 17*(7), 936-943.

Hancock, M. J., Maher, C. G., Latimer, J., Herbert, R. D., & McAuley, J. H. (2009). Can rate of recovery be predicted in patients with acute low back pain? Development of a clinical prediction rule. *European Journal of Pain, 13*(1), 51-55.

Hancock, M. J., Maher, C. G., Latimer, J., Spindler, M. F., McAuley, J. H., Laslett, M., & Bogduk, N. (2007). Systematic review of tests to identify the disc, SIJ or facet joint as the source of low back pain. *European Spine Journal, 16*(10), 1539-1550.

Hancock, M. J., Maher, C. G., Macaskill, P., Latimer, J., Kos, W., & Pik, J. (2012). MRI findings are more common in selected patients with acute low back pain than controls? *European Spine Journal, 21*(2), 240-246.

Hannes, K., Staes, F., Goedhuys, J., & Aertgeerts, B. (2009). Obstacles to the implementation of evidence-based physiotherapy in practice: A focus group-based study in Belgium (Flanders). *Physiotherapy Theory and Practice, 25*(7), 476-488.

Harkness, E. F., Macfarlane, G., Silman, A., & McBeth, J. (2005). Is musculoskeletal pain more common now than 40 years ago?: Two population-based cross-sectional studies. *Rheumatology, 44*(7), 890-895.

Harper, B. E., & Reveille, J. D. (2009). Spondyloarthritis: Clinical Suspicion, Diagnosis, and Sports. *Current Sports Medicine Reports, 8*(1), 29-34.

Harting, J., Rutten, G. M. J., Rutten, S. T. J., & Kremers, S. P. (2009). A qualitative application of the diffusion of innovations theory to examine determinants of guideline adherence among physical therapists. *Physical Therapy, 89*(3), 221-232.

Haskins, R., Osmotherly, P. G., & Rivett, D. A. (2015a). Diagnostic clinical prediction rules for specific subtypes of low back pain: A systematic review. *Journal of Orthopaedic & Sports Physical Therapy, 45*(2), 61-76.

Haskins, R., Osmotherly, P. G., & Rivett, D. A. (2015b). Validation and impact analysis of prognostic clinical prediction rules for low back pain is needed: a systematic review. *Journal of Clinical Epidemiology, 68*(7), 821-832.

Haskins, R., Osmotherly, P. G., Southgate, E., & Rivett, D. A. (2014). Physiotherapists' knowledge, attitudes and practices regarding clinical prediction rules for low back pain. *Manual Therapy, 19*(2), 142-151.

Haskins, R., Osmotherly, P. G., Southgate, E., & Rivett, D. A. (2015). Australian physiotherapists' priorities for the development of clinical prediction rules for low back pain: a qualitative study. *Physiotherapy, 101*(1), 44-49.

Haskins, R., Osmotherly, P. G., Tuyl, F., & Rivett, D. A. (2014). Uncertainty in Clinical Prediction Rules: The Value of Credible Intervals. *Journal of Orthopaedic & Sports Physical Therapy, 44*(2), 85-91.

Haskins, R., Rivett, D. A., & Osmotherly, P. G. (2012). Clinical prediction rules in the physiotherapy management of low back pain: a systematic review. *Manual Therapy, 17*(1), 9-21.

Hayden, J. A., Côté, P., & Bombardier, C. (2006). Evaluation of the quality of prognosis studies in systematic reviews. *Annals of Internal Medicine, 144*(6), 427-437.

Hayden, J. A., van Tulder, M. W., Malmivaara, A. V., & Koes, B. W. (2005). Meta-analysis: exercise therapy for nonspecific low back pain. *Annals of Internal Medicine, 142*(9), 765-775.

Healy, P. J., & Helliwell, P. S. (2005). Classification of the spondyloarthropathies. *Current Opinion in Rheumatology, 17*(4), 395-399.

Hebert, J. J., & Perle, S. M. (2008). Letter to the editor concerning "Independent evaluation of a clinical prediction rule for spinal manipulative therapy: a randomised controlled trial" (M. Hancock et al.). *European Spine Journal, 17*(10), 1401-1402.

Heinrich, R. L., Cohen, M. J., Naliboff, B. D., Collins, G. A., & Bonebakker, A. D. (1985). Comparing physical and behavior therapy for chronic low back pain on physical abilities, psychological distress, and patients' perceptions. *Journal of behavioral medicine, 8*(1), 61-78.

Henry, S. M., Fritz, J. M., Trombley, A. R., & Bunn, J. Y. (2012). Reliability of a treatment-based classification system for subgrouping people with

low back pain. *Journal of Orthopaedic & Sports Physical Therapy, 42*(9), 797-805.

Henschke, N., Maher, C. G., Ostelo, R. W., de Vet, H. C., Macaskill, P., & Irwig, L. (2013). Red flags to screen for malignancy in patients with low-back pain. *Cochrane Database of Systematic Reviews, 2*, CD008686.

Henschke, N., Maher, C. G., & Refshauge, K. M. (2007). Screening for malignancy in low back pain patients: a systematic review. *European Spine Journal, 16*(10), 1673-1679.

Henschke, N., Maher, C. G., & Refshauge, K. M. (2008). A systematic review identifies five "red flags" to screen for vertebral fracture in patients with low back pain. *Journal of Clinical Epidemiology, 61*(2), 110-118.

Henschke, N., Maher, C. G., Refshauge, K. M., Das, A., & McAuley, J. H. (2007). Low back pain research priorities: a survey of primary care practitioners. *BMC Family Practice, 8*, 40.

Henschke, N., Maher, C. G., Refshauge, K. M., Herbert, R. D., Cumming, R. G., Bleasel, J., . . . McAuley, J. H. (2009). Prevalence of and screening for serious spinal pathology in patients presenting to primary care settings with acute low back pain. *Arthritis & Rheumatism, 60*(10), 3072-3080.

Hermann, J., Giessauf, H., Schaffler, G., Ofner, P., & Graninger, W. (2009). Early spondyloarthritis: usefulness of clinical screening. *Rheumatology, 48*(7), 812-816.

Hess, E. P., Thiruganasambandamoorthy, V., Wells, G. A., Erwin, P., Jaffe, A. S., Hollander, J. E., . . . Stiell, I. G. (2008). Diagnostic accuracy of

clinical prediction rules to exclude acute coronary syndrome in the emergency department setting: a systematic review. *CJEM Canadian Journal of Emergency Medical Care, 10*(4), 373-382.

Hestbaek, L., Leboeuf-Yde, C., & Manniche, C. (2003). Low back pain: what is the long-term course? A review of studies of general patient populations. *European Spine Journal, 12*(2), 149-165.

Hewitt, J. A., Hush, J. M., Martin, M. H., Herbert, R. D., & Latimer, J. (2007). Clinical prediction rules can be derived and validated for injured Australian workers with persistent musculoskeletal pain: an observational study. *Australian Journal of Physiotherapy, 53*(4), 269-276.

Heymans, M. W., Anema, J. R., van Buuren, S., Knol, D. L., van Mechelen, W., & De Vet, H. C. W. (2009). Return to work in a cohort of low back pain patients: development and validation of a clinical prediction rule. *Journal of Occupational Rehabilitation, 19*(2), 155-165.

Heymans, M. W., Ford, J. J., McMeeken, J. M., Chan, A., de Vet, H. C. W., & van Mechelen, W. (2007). Exploring the contribution of patient-reported and clinician based variables for the prediction of low back work status. *Journal of Occupational Rehabilitation, 17*(3), 383-397.

Hicks, E. P., & Kluemper, G. T. (2011). Heuristic reasoning and cognitive biases: Are they hindrances to judgments and decision making in orthodontics? *American Journal of Orthodontics & Dentofacial Orthopedics, 139*(3), 297-304.

Hicks, G. E., Fritz, J. M., Delitto, A., & McGill, S. M. (2005). Preliminary development of a clinical prediction rule for determining which patients

with low back pain will respond to a stabilization exercise program. *Archives of Physical Medicine and Rehabilitation, 86*(9), 1753-1762.

Higgins, J. P. T., & Green, S. (2011). *Cochrane Handbook for Systematic Reviews of Interventions (updated March 2011)*. Available from www.cochrane-handbook.org: The Cochrane Collaboration.

Higgs, J., Jones, M. A., Loftus, S., & Christensen, N. (2008). *Clinical Reasoning in the Health Professions*: Elsevier.

Hill, A. B. (1965). The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine, 58*, 295-300.

Hill, J. C., Dunn, K. M., Lewis, M., Mullis, R., Main, C. J., Foster, N. E., & Hay, E. M. (2008). A primary care back pain screening tool: identifying patient subgroups for initial treatment. *Arthritis & Rheumatism, 59*(5), 632-641.

Hill, J. C., & Fritz, J. M. (2011). Psychosocial Influences on Low Back Pain, Disability, and Response to Treatment. *Physical Therapy, 91*(5), 712-721.

Hill, J. C., Konstantinou, K., Egbewale, B. E., Dunn, K. M., Lewis, M., & van der Windt, D. A. (2011). Clinical outcomes among low back pain consulters with referred leg pain in primary care. *Spine, 36*(25), 2168-2175.

Hill, J. C., Whitehurst, D. G. T., Lewis, M., Bryan, S., Dunn, K. M., Foster, N. E., . . . Somerville, S. (2011). Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. *The Lancet, 378*(9802), 1560-1571.

Holland, J. L., Wilczynski, N. L., & Haynes, R. B. (2005). Optimal search strategies for identifying sound clinical prediction studies in EMBASE. *BMC medical informatics and decision making, 5*(1), 11.

Holt, R. R. (1970). Yet another look at clinical and statistical prediction: or, is clinical psychology worthwhile? *American Psychologist, 25*(4), 337-349.

Horvitz, E. (2010). From Data to Predictions and Decisions: Enabling Evidence-Based Healthcare. *Computing Community Consortium, 6*.

Houghton, P. E., Nussbaum, E. L., & Hoens, A. M. (2010). Electrophysical agents. Contraindications and precautions: an evidence-based approach to clinical decision making in physical therapy. *Physiotherapy Canada, 62*(5), 1-80.

Hoy, D., Bain, C., Williams, G., March, L., Brooks, P., Blyth, F., . . . Buchbinder, R. (2012). A systematic review of the global prevalence of low back pain. *Arthritis & Rheumatism, 64*(6), 2028-2037.

Hoy, D., Brooks, P., Blyth, F., & Buchbinder, R. (2010). The epidemiology of low back pain. *Best Practice & Research Clinical Rheumatology, 24*(6), 769-781.

Hoy, D., March, L., Brooks, P., Blyth, F., Woolf, A., Bain, C., . . . Barendregt, J. (2014). The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. *Annals of the rheumatic diseases, 73*(6), 968-974.

Hoy, D., March, L., Brooks, P., Woolf, A., Blyth, F., Vos, T., & Buchbinder, R. (2010). Measuring the global burden of low back pain. *Best Practice & Research Clinical Rheumatology, 24*(2), 155-165.

Ingui, B. J., & Rogers, M. A. (2001). Searching for clinical prediction rules in MEDLINE. *Journal of the American Medical Informatics Association, 8*(4), 391-397.

Institute for Health Metrics and Evaluation. (2013). GBD 2010 Heat Map. Retrieved October, 2014, from http://vizhub.healthdata.org/irank/heat.php

Irwig, L., Tosteson, A. N., Gatsonis, C., Lau, J., Colditz, G., Chalmers, T. C., & Mosteller, F. (1994). Guidelines for meta-analyses evaluating diagnostic tests. *Annals of Internal Medicine, 120*(8), 667-676.

Irwin, R. W., Watson, T., Minick, R. P., & Ambrosius, W. T. (2007). Age, body mass index, and gender differences in sacroiliac joint pathology. *American Journal of Physical Medicine & Rehabilitation, 86*(1), 37-44.

Itz, C. J., Geurts, J. W., Kleef, M., & Nelemans, P. (2013). Clinical course of non-specific low back pain: A systematic review of prospective cohort studies set in primary care. *European Journal of Pain, 17*(1), 5-15.

Jaeschke, R., Guyatt, G. H., Sackett, D., & Evidence-Based Medicine Working Group. (1994). Users' Guides to the Medical Literature: III. How to Use an Article About a Diagnostic Test B. What Are the Results and Will They Help Me in Caring for My Patients? *Journal of the American Medical Association, 271*(9), 703-707.

Jellema, P., van der Horst, H. E., Vlaeyen, J. W. S., Stalman, W. A. B., Bouter, L. M., & van der Windt, D. A. W. M. (2006). Predictors of outcome in patients with (sub)acute low back pain differ across treatment groups. *Spine, 31*(15), 1699-1705.

Jensen, M. C., Brant-Zawadzki, M. N., Obuchowski, N. A., Modic, M. T., Malkasian, D., & Ross, J. S. (1994). Magnetic resonance imaging of the lumbar spine in people without back pain. *New England Journal of Medicine, 331*(2), 69-73.

Jette, D. U., Bacon, K., Batty, C., Carlson, M., Ferland, A., Hemingway, R. D., . . . Volk, D. (2003). Evidence-based practice: Beliefs, attitudes, knowledge, and behaviors of physical therapists. *Physical Therapy, 83*(9), 786-805.

Johnson, D. D. P., Blumstein, D. T., Fowler, J. H., & Haselton, M. G. (2013). The evolution of error: error management, cognitive constraints, and adaptive decision-making biases. *Trends in ecology & evolution, 28*(8), 474-481.

Johnston, S. C., Rothwell, P. M., Nguyen-Huynh, M. N., Giles, M. F., Elkins, J. S., Bernstein, A. L., & Sidney, S. (2007). Validation and refinement of scores to predict very early stroke risk after transient ischaemic attack. *The Lancet, 369*(9558), 283-292.

Jones, C. M., & Athanasiou, T. (2009). Diagnostic accuracy meta-analysis: review of an important tool in radiological research and decision making. *British Journal of Radiology, 82*(978), 441-446.

Jones, M. A., & Rivett, D. A. (2004). *Clinical Reasoning for Manual Therapists*: Butterworth Heinemann.

Jones, S., Carley, S., & Harrison, M. (2003). An introduction to power and sample size estimation. *Emergency Medicine Journal, 20*(5), 453-458.

Junila, J., Niinimäki, T., & Tervonen, O. (1997). Epidural abscess after lumbar discography: a case report. *Spine, 22*(18), 2191-2193.

Justice, A. C., Covinsky, K. E., & Berlin, J. A. (1999). Assessing the generalizability of prognostic information. *Annals of Internal Medicine, 130*(6), 515-524.

Kain, T., Zochling, J., Taylor, A., Manolios, N., Smith, M. D., Reed, M. D., . . . Schachna, L. (2008). Evidence-based recommendations for the diagnosis of ankylosing spondylitis: results from the Australian 3E initiative in rheumatology. *Medical Journal of Australia, 188*(4), 235-237.

Kamper, S. J., Maher, C. G., Hancock, M. J., Koes, B. W., Croft, P. R., & Hay, E. (2010). Treatment-based subgroups of low back pain: a guide to appraisal of research studies and a summary of current evidence. *Best Practice & Research Clinical Rheumatology, 24*(2), 181-191.

Karayannis, N. V., Jull, G. A., & Hodges, P. W. (2012). Physiotherapy movement based classification approaches to low back pain: comparison of subgroups through review and developer/expert survey. *BMC Musculoskeletal Disorders, 13*(1), 24.

Kato, Y., Kawakami, T., Kifune, M., Kishimoto, T., Nibu, K., Oda, H., . . . Tsue, K. (2009). Validation study of a clinical diagnosis support tool for lumbar spinal stenosis. *Journal of Orthopaedic Science, 14*(6), 711-718.

Kattan, M. (2002). Statistical prediction models, artificial neural networks, and the sophism “I am a patient, not a statistic”. *Journal of clinical oncology, 20*(4), 885-887.

Katz, M. H. (2003). Multivariable analysis: a primer for readers of medical research. *Annals of Internal Medicine, 138*(8), 644-650.

Katz, M. H. (2006). *Multivariable Analysis. A Practical Guide for Clinicians.* (2nd ed.). New York: Cambridge University Press.

Keefe, F. J., Bradley, L. A., & Crisson, J. E. (1990). Behavioral assessment of low back pain: identification of pain behavior subgroups. *Pain, 40*(2), 153-160.

Keith, M. P. (2012). Inflammatory back pain and the diagnosis of axial spondyloarthritis. *American Journal of Orthopedics, 41*(12), E166-168.

Keller, A., Hayden, J. A., Bombardier, C., & van Tulder, M. (2007). Effect sizes of non-surgical treatments of non-specific low-back pain. *European Spine Journal, 16*(11), 1776-1788.

Keller, T., & Chappell, T. (1996). The Rise and Fall of Erichsen's Disease (Railroad Spine). *Spine, 21*(13), 1597-1601.

Kent, P. M., Hancock, M., Petersen, D. H., & Mjøsund, H. L. (2010). Clinimetrics corner: choosing appropriate study designs for particular questions about treatment subgroups. *The Journal of Manual & Manipulative Therapy, 18*(3), 147-152.

Kent, P. M., & Keating, J. L. (2004). Do Primary-Care Clinicians Think That Nonspecific Low Back Pain Is One Condition? *Spine, 29*(9), 1022-1031.

Kent, P. M., & Keating, J. L. (2005). Classification in nonspecific low back pain: What methods do primary care clinicians currently use? *Spine, 30*(12), 1433-1440.

Kent, P. M., Keating, J. L., & Leboeuf-Yde, C. (2010). Research methods for subgrouping low back pain. *BMC medical research methodology, 10*(1), 62.

Kent, P. M., Keating, J. L., & Taylor, N. F. (2009). Primary care clinicians use variable methods to assess acute nonspecific low back pain and usually focus on impairments. *Manual Therapy, 14*(1), 88-100.

Keogh, C., Wallace, E., O'Brien, K. K., Galvin, R., Smith, S. M., Lewis, C., . . . Fahey, T. (2014). Developing an international register of clinical prediction rules for use in primary care: a descriptive analysis. *Ann Fam Med, 12*(4), 359-366.

Keogh, C., Wallace, E., O'Brien, K. K., Murphy, P. J., Teljeur, C., McGrath, B., . . . Fahey, T. (2011). Optimized retrieval of primary care clinical prediction rules from MEDLINE to establish a web-based register. *Journal of Clinical Epidemiology, 64*(8), 848-860.

Ketelslegers, E., Collard, X., Vande Berg, B., Danse, E., El-Gariani, A., Poilvache, P., & Maldague, B. (2002). Validation of the Ottawa knee rules in an emergency teaching centre. *Eur Radiol, 12*(5), 1218-1220.

Kirkaldy-Willis, W. H., & Hill, R. J. (1979). A more precise diagnosis for low-back pain. *Spine, 4*(2), 102-109.

Klapow, J. C., Slater, M. A., Patterson, T. L., Doctor, J. N., Atkinson, J. H., & Garfin, S. R. (1993). An empirical evaluation of multidimensional clinical outcome in chronic low back pain patients. *Pain, 55*(1), 107-118.

Kocher, M. S., Mandiga, R., Zurakowski, D., Barnewolt, C., & Kasser, J. R. (2004). Validation of a clinical prediction rule for the differentiation between septic arthritis and transient synovitis of the hip in children. *The Journal of Bone & Joint Surgery, 86*(8), 1629-1635.

Koenig, H. G. (2000). Religion and medicine I: historical background and reasons for separation. *International journal of psychiatry in medicine, 30*(4), 385-398.

Koes, B. W., van Tulder, M. W., Lin, C. W., Macedo, L. G., McAuley, J., & Maher, C. (2010). An updated overview of clinical guidelines for the management of non-specific low back pain in primary care. *European Spine Journal, 19*(12), 2075-2094.

Koes, B. W., van Tulder, M. W., Ostelo, R., Burton, A. K., & Waddell, G. (2001). Clinical guidelines for the management of low back pain in primary care: an international comparison. *Spine, 26*(22), 2504-2513.

Koes, B. W., van Tulder, M. W., & Peul, W. C. (2007). Diagnosis and treatment of sciatica. *British Medical Journal, 334*(7607), 1313-1317.

Koes, B. W., van Tulder, M. W., & Thomas, S. (2006). Diagnosis and treatment of low back pain. *British Medical Journal, 332*(7555), 1430-1434.

Konno, S., Hayashino, Y., Fukuhara, S., Kikuchi, S., Kaneda, K., Seichi, A., . . . Kawai, S. (2007). Development of a clinical diagnosis support tool to identify patients with lumbar spinal stenosis. *European Spine Journal, 16*(11), 1951-1957.

Kovacs, F. M., Seco, J., Royuela, A., Corcoll Reixach, J., Abraira, V., & Spanish Back Pain Research Network. (2012). Predicting the evolution of low back pain patients in routine clinical practice: results from a registry within the Spanish National Health Service. *Spine Journal: Official Journal of the North American Spine Society, 12*(11), 1008-1020.

Kovacs, F. M., Urrútia, G., & Alarcón, J. D. (2011). Surgery versus conservative treatment for symptomatic lumbar spinal stenosis: a systematic review of randomized controlled trials. *Spine, 36*(20), E1335-E1351.

Kraemer, H. C., Frank, E., & Kupfer, D. J. (2006). Moderators of treatment outcomes: Clinical, research, and policy importance. *Journal of the American Medical Association, 296*(10), 1286-1289.

Kraemer, H. C., & Gibbons, R. D. (2009). Why does the randomized clinical trial methodology so often mislead clinical decision making? Focus on moderators and mediators of treatment. *Psychiatric Annals, 39*(7), 736-745.

Kraemer, H. C., Wilson, G. T., Fairburn, C. G., & Agras, W. S. (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry, 59*(10), 877-883.

Krause, N., & Ragland, D. R. (1994). Occupational disability due to low back pain: a new interdisciplinary classification based on a phase model of disability. *Spine, 19*(9), 1011-1020.

Krefting, L. (1991). Rigor in qualitative research: The assessment of trustworthiness. *The American journal of occupational therapy, 45*(3), 214-222.

Kreiner, D. S., Shaffer, W. O., Baisden, J. L., Gilbert, T. J., Summers, J. T., Toton, J. F., . . . Reitman, C. A. (2013). An evidence-based clinical guideline for the diagnosis and treatment of degenerative lumbar spinal stenosis (update). *The Spine Journal, 13*(7), 734-743.

Krismer, M., & van Tulder, M. (2007). Low back pain (non-specific). *Best Practice & Research Clinical Rheumatology, 21*(1), 77-91.

Krueger, R. A., & Casey, M. A. (2009). *Focus Groups. A Practical Guide for Applied Research.* (4th ed.). Thousand Oaks, California: Sage Publications, Inc.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174.

Lang, E. S., Wyer, P. C., & Haynes, R. B. (2007). Knowledge translation: closing the evidence-to-practice gap. *Annals of Emergency Medicine, 49*(3), 355-363.

Laplace, P. S. (1902). *A Philosophical Essay on Probabilities. Translated from the sixth French edition.* London: Chapman & Hall, Ltd.

Larson, E. (2004). A tool to assess barriers to adherence to hand hygiene guideline. *American Journal of Infection Control, 32*(1), 48-51.

Laslett, M. (2008). Evidence-based diagnosis and treatment of the painful sacroiliac joint. *Journal of Manual & Manipulative Therapy, 16*(3), 142-152.

Laslett, M., Aprill, C. N., McDonald, B., & Oberg, B. (2006). Clinical predictors of lumbar provocation discography: a study of clinical predictors of lumbar provocation discography. *European Spine Journal, 15*(10), 1473-1484.

Laslett, M., Aprill, C. N., McDonald, B., & Young, S. B. (2005). Diagnosis of Sacroiliac Joint Pain: Validity of individual provocation tests and composites of tests. *Manual Therapy, 10*(3), 207-218.

Laslett, M., McDonald, B., Aprill, C. N., Tropp, H., & Oberg, B. (2006). Clinical predictors of screening lumbar zygapophyseal joint blocks: development of clinical prediction rules. *Spine Journal, 6*(4), 370-379.

Laslett, M., McDonald, B., Tropp, H., Aprill, C. N., & Oberg, B. (2005). Agreement between diagnoses reached by clinical examination and available reference standards: a prospective study of 216 patients with lumbopelvic pain. *BMC Musculoskeletal Disorders, 6*, 28.

Laslett, M., & van Wijmen, P. (1999). Low back and referred pain: diagnosis and a proposed new system of classification. *New Zealand Journal of Physiotherapy, 27*, 5-14.

Laslett, M., Young, S. B., Aprill, C. N., & McDonald, B. (2003). Diagnosing painful sacroiliac joints: A validity study of a McKenzie evaluation and sacroiliac provocation tests. *Australian Journal of Physiotherapy, 49*(2), 89-97.

Laupacis, A., Sekar, N., & Stiell, I. G. (1997). Clinical prediction rules: A review and suggested modifications of methodological standards. *Journal of the American Medical Association, 277*(6), 488-494.

Laupacis, A., Wells, G., Richardson, W. S., Tugwell, P., Guyatt, G. H., Browman, G., . . . Hayward, R. (1994). Users' guides to the medical literature: V. How to use an article about prognosis. *Journal of the American Medical Association, 272*(3), 234-237.

Learman, K., Showalter, C., & Cook, C. (2012). Does the use of a prescriptive clinical prediction rule increase the likelihood of applying inappropriate treatments? A survey using clinical vignettes. *Manual Therapy, 17*(6), 538-543.

Leavitt, F., Garron, D. C., Whisler, W. W., & Sheinkop, M. B. (1978). Affective and sensory dimensions of back pain. *Pain, 4*, 273-281.

Leboeuf-Yde, C., Gronstvedt, A., Borge, J. A., Lothe, J., Magnesen, E., Nilsson, O., . . . Larsen, K. (2004). The nordic back pain subpopulation program: demographic and clinical predictors for outcome in patients receiving chiropractic treatment for persistent low back pain. *Journal of Manipulative & Physiological Therapeutics, 27*(8), 493-502.

Ledley, R. S., & Lusted, L. B. (1959). Reasoning foundations of medical diagnosis. *Science, 130*(3366), 9-21.

Leeflang, M. M. G., Bossuyt, P. M. M., & Irwig, L. (2009). Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. *Journal of Clinical Epidemiology, 62*(1), 5-12.

Legare, F., Ratte, S., Gravel, K., & Graham, I. D. (2008). Barriers and facilitators to implementing shared decision-making in clinical practice: update of a systematic review of health professionals' perceptions. *Patient education and counseling, 73*(3), 526.

Lemon, S. C., Roy, J., Clark, M. A., Friedmann, P. D., & Rakowski, W. (2003). Classification and regression tree analysis in public health: methodological review and comparison with logistic regression. *Annals of behavioral medicine, 26*(3), 172-181.

Lewis, S. (2007). Regression analysis. *Practical Neurology, 7*(4), 259-264.

Liao, L., & Mark, D. B. (2003). Clinical prediction models: are we building better mousetraps? *Journal of the American College of Cardiology, 42*(5), 851-853.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., . . . Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Annals of Internal Medicine, 151*(4), W65-W94.

Loeser, J. D., & Treede, R. (2008). The Kyoto protocol of IASP basic pain terminology. *Pain, 137*(3), 473-477.

Long, A., Donelson, R., & Fung, T. (2004). Does it matter which exercise? A randomized control trial of exercise for low back pain. *Spine, 29*(23), 2593-2602.

Lopez, A. D., & Murray, C. C. J. L. (1998). The global burden of disease, 1990-2020. *Nature Medicine, 4*(11), 1241-1243.

Lubetzky-Vilnai, A., Ciol, M., & McCoy, S. W. (2014). Statistical analysis of clinical prediction rules for rehabilitation interventions: current state of the literature. *Archives of Physical Medicine and Rehabilitation, 95*(1), 188-196.

Luo, X., Pietrobon, R., X Sun, S., Liu, G. G., & Hey, L. (2004). Estimates and Patterns of Direct Health Care Expenditures Among Individuals With Back Pain in the United States. *Spine, 29*(1), 79-86.

Lurie, J. D., Tosteson, T. D., Tosteson, A. N. A., Zhao, W., Morgan, T. S., Abdu, W. A., . . . Weinstein, J. N. (2014). Surgical Versus Nonoperative Treatment for Lumbar Disc Herniation: Eight-Year Results for the Spine Patient Outcomes Research Trial. *Spine, 39*(1), 3-16.

Lusted, L. B. (1975). Editorial: In the process of solution. *New England Journal of Medicine, 293*(5), 255-256.

Lutz, G. K., Butzlaff, M., & Schultz-Venrath, U. (2003). Looking back on back pain: trial and error of diagnoses in the 20th century. *Spine, 28*(16), 1899-1905.

Macaskill, P., Gatsonis, C., Deeks, J., Harbord, R., & Takwoingi, Y. (2010). Chapter 10: Analysing and presenting results *Cochrane handbook for systematic reviews of diagnostic test accuracy version* (Vol. 1.0): The Cochrane Collaboration.

Machado, L. A. C., De Souza, M., Ferreira, P. H., & Ferreira, M. L. (2006). The McKenzie method for low back pain: a systematic review of the literature with a meta-analysis approach. *Spine, 31*(9), E254-E262.

Machado, L. A. C., Kamper, S. J., Herbert, R. D., Maher, C. G., & McAuley, J. H. (2009). Analgesic effects of treatments for non-specific low back pain: a meta-analysis of placebo-controlled randomized trials. *Rheumatology, 48*(5), 520-527.

MacKinnon, D. P., & Luecken, L. J. (2008). How and for Whom? Mediation and Moderation in Health Psychology. *Health Psychology, 27*(2 SUPPL. 2), S99-S100.

Magilvy, J. K., & Thomas, E. (2009). A first qualitative project: Qualitative descriptive design for novice researchers. *Journal for Specialists in Pediatric Nursing, 14*(4), 298-300.

Maher, C. G., Sherrington, C., Herbert, R. D., Moseley, A. M., & Elkins, M. (2003). Reliability of the PEDro scale for rating quality of randomized controlled trials. *Physical Therapy, 83*(8), 713-721.

Maher, C. G., Williams, C., Lin, C., & Latimer, J. (2011). Managing low back pain in primary care. *Australian Prescriber, 34*, 128-132.

Maigne, J., Aivaliklis, A., & Pfefer, F. (1996). Results of sacroiliac joint double block and value of sacroiliac pain provocation tests in 54 patients with low back pain. *Spine, 21*(16), 1889-1892.

Main, C. J., Wood, P. L., Hollis, S., Spanswick, C. C., & Waddell, G. (1992). The Distress and Risk Assessment Method. A simple patient classification to identify distress and evaluate the risk of poor outcome. *Spine, 17*(1), 42-52.

Malmqvist, S., Leboeuf-Yde, C., Ahola, T., Andersson, O., Ekström, K., Pekkarinen, H., . . . Wedderkopp, N. (2008). The Nordic back pain subpopulation program: predicting outcome among chiropractic patients in Finland. *Chiropr & Osteopat, 16*(13), Online access only 12 p.

Manchikanti, L., Abdi, S., Atluri, S., Benyamin, R. M., Boswell, M. V., Buenaventura, R. M., . . . Hirsch, J. A. (2013). An update of comprehensive evidence-based guidelines for interventional techniques in chronic spinal pain. Part II: guidance and recommendations. *Pain Physician, 16*(2 Suppl), S49-283.

Manchikanti, L., Datta, S., Derby, R., Wolfer, L. R., Benyamin, R. M., & Hirsch, J. A. (2010). A critical review of the American Pain Society clinical practice guidelines for interventional techniques: Part 1. Diagnostic interventions. *Pain Physician, 13*, E141-E174.

Manchikanti, L., Pampati, S., & Cash, K. A. (2010). Making sense of the accuracy of diagnostic lumbar facet joint nerve blocks: an assessment

of the implications of 50% relief, 80% relief, single block, or controlled diagnostic blocks. *Pain Physician, 13*(2), 133-143.

Manchikanti, L., Singh, V., Pampati, V., Damron, K. S., Barnhill, R. C., Beyer, C., & Cash, K. A. (2001). Evaluation of the relative contributions of various structures in chronic low back pain. *Pain Physician, 4*(4), 308-316.

Marras, W. S., Ferguson, S. A., Burr, D., Schabo, P., & Maronitis, A. (2007). Low Back Pain Recurrence in Occupational Environments. *Spine, 32*(21), 2387-2397.

Martin, B. I., Deyo, R. A., Mirza, S. K., Turner, J. A., Comstock, B. A., Hollingworth, W., & Sullivan, S. D. (2008). Expenditures and health status among adults with back and neck problems. *JAMA, 299*(6), 656-664.

Mathers, C., Fat, D. M., & Boerma, J. T. (2008). *The global burden of disease: 2004 update*. Geneva: World Health Organization.

Mathew, J., Singh, S. B., Garis, S., & Diwan, A. D. (2013). Backing up the stories: The psychological and social costs of chronic low-back pain. *The International Journal of Spine Surgery, 7*(1), e29-e38.

May, S., Gardiner, E., Young, S., & Klaber-Moffett, J. (2008). Predictor variables for a positive long-term functional outcome in patients with acute and chronic neck and back pain treated with a McKenzie approach: A secondary analysis. *Journal of Manual and Manipulative Therapy, 16*(3), 155-160.

May, S., & Rosedale, R. (2009). Prescriptive Clinical Prediction Rules in Back Pain Research: A Systematic Review. *Journal of Manual and Manipulative Therapy, 17*(1), 36-45.

Mays, N., & Pope, C. (2000). Qualitative research in health care: assessing quality in qualitative research. *British Medical Journal, 320*(7226), 50-52.

McCall, I. W., Park, W. M., & O'Brien, J. P. (1979). Induced pain referral from posterior lumbar elements in normal subjects. *Spine, 4*(5), 441-446.

McCarthy, C., Arnall, F., Strimpakos, N., Freemont, A., & Oldham, J. (2004). The biopsychosocial classification of non-specific low back pain: a systematic review. *Physical Therapy Reviews, 9*(1), 17-30.

McCarthy, P. J., Ball, D., & Purcell, W. (2007). *Project Phoenix—Optimizing the machine-person mix in high-impact weather forecasting.* Paper presented at the 22nd Conference on Weather Analysis and Forecasting/18th Conference on Numerical Weather Prediction.

McCaslin, P. T., Nakazawa, T., Swinbank, R., & Toth, Z. (2010). Improving cyclone warning Case study: Philippines. *WMO Bulletin, 59*(2).

McGinn, T. G., Guyatt, G. H., Wyer, P. C., Naylor, C. D., Stiell, I. G., & Richardson, W. S. (2000). Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-Based Medicine Working Group. *JAMA, 284*(1), 79-84.

McGinn, T. G., Wyer, P., Wisnivesky, J., Devereauz, P. J., Stiell, I., Richardson, S., & Guyatt, G. (2008). Advanced Topics in Diagnosis: Clinical Prediction Rules. In G. Guyatt, D. Rennie, M. O. Meade & D. J. Cook (Eds.), *Users' Guides to the Medical Literature. A manual for*

evidence-based clinical practice. Second Edition. (pp. 491-505). USA: McGraw Hill Medical.

McGrayne, S. B. (2011). *The theory that would not die: how Bayes' rule cracked the enigma code, hunted down Russian submarines, & emerged triumphant from two centuries of controversy*. London: Yale University Press.

McKenzie, R. A. (1981). *The Lumbar Spine: Mechanical Diagnosis and Therapy*. Waikanae, New Zealand: Spinal Publications.

McKenzie, R. A., & May, S. (2003). *The Lumbar Spine: Mechanical Diagnosis and Therapy*. Waikanae, New Zealand.

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*: University of Minnesota Press.

Meehl, P. E. (1986). Causes and effects of my disturbing little book. *Journal of Personality Assessment, 50*(3), 370-375.

Mehta, N. B. (2004). The doctors' challenge: how can we follow guidelines better. *Cleveland Clinic Journal of Medicine, 71*(2), 81-82.

Merskey, H., & Bogduk, N. (1994). *Classification of Chronic Pain. Descriptions of Chronic Pain Syndromes and Definition of Pain Terms* (Second ed.). Seattle: IASP Press.

Met Office. (2014). How accurate are our public forecasts? Retrieved December, 2014, from http://www.metoffice.gov.uk/about-us/who/accuracy/forecasts

Mixter, W. J., & Ayre, J. B. (1935). Herniation or rupture of the intervertebral disc into the spinal canal: Report of thirty-four cases. *New England Journal of Medicine, 213*(9), 385-393.

Mixter, W. J., & Barr, J. S. (1934). Rupture of the intervertebral disc with involvement of the spinal canal. *New England Journal of Medicine, 211*(5), 210-215.

Moffroid, M. T., Haugh, L. D., Henry, S. M., & Short, B. (1994). Distinguishable groups of musculoskeletal low back pain patients and asymptomatic control subjects based on physical measures of the NIOSH Low Back Atlas. *Spine, 19*(12), 1350-1358.

Mooney, V. (1989). The classification of low back pain. *Annals of Medicine, 21*(5), 321-325.

Moons, K. G. M., Royston, P., Vergouwe, Y., Grobbee, D. E., & Altman, D. G. (2009). Prognosis and prognostic research: what, why, and how? *British Medical Journal, 338*, b375.

Morse, J. M., Barrett, M., Mayan, M., Olson, K., & Spiers, J. (2008). Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Methods, 1*(2), 13-22.

Moseley, G. L., & Vlaeyen, J. W. S. (2015). Beyond nociception: the imprecision hypothesis of chronic pain. *Pain, 156*(1), 35-38.

Mossman, D., & Berger, J. O. (2001). Intervals for posttest probabilities: a comparison of 5 methods. *Medical Decision Making, 21*(6), 498-507.

Murray, C. J. L. (1994). Quantifying the burden of disease: the technical basis for disability-adjusted life years. *Bulletin of the World Health Organization, 72*(3), 429-445.

Murray, C. J. L., Ezzati, M., Flaxman, A. D., Lim, S., Lozano, R., Michaud, C., . . . Vos, T. (2012). GBD 2010: a multi-investigator collaboration for

global comparative descriptive epidemiology. *The Lancet, 380*(9859), 2055-2058.

Murray, C. J. L., Vos, T., Lozano, R., Naghavi, M., Flaxman, A. D., Michaud, C., . . . Lopez, A. D. (2012). Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet, 380*(9859), 2197-2223.

Nachemson, A. L., & Andersson, G. B. J. (1982). Classification of low-back pain. *Scandinavian journal of work, environment & health, 8*, 134-136.

Naghavi, M., Wang, H., Lozano, R., Davis, A., Liang, X., Zhou, M., . . . Abd-Allah, F. (2015). Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet, 385*(9963), 117-171.

National Ankylosing Spondylitis Society. (2014). Inflammatory back pain. Retrieved June, 2014, from http://nass.co.uk/about-as/getting-my-diagnosis/inflammatory-arthritis/

National Institute for Health and Care Excellence. (2008). Stroke: Diagnosis and initial management of acute stroke and transient ischaemic attack (TIA). Retrieved Dec, 2014, from http://www.nice.org.uk/guidance/CG68

National Institute of Clinical Studies. (2006). Identifying barriers to evidence uptake. Melbourne, Australia: NICS.

National Stoke Foundation. (2010). Clinical Guidelines for Stroke Management 2010. Retrieved Dec, 2014, from

http://strokefoundation.com.au/site/media/clinical_guidelines_stroke_managment_2010_interactive.pdf

Nee, R. J., & Coppieters, M. W. (2011). Interpreting research on clinical prediction rules for physiotherapy treatments. *Manual Therapy, 16*(2), 105-108.

Neergaard, M. A., Olesen, F., Andersen, R. S., & Sondergaard, J. (2009). Qualitative description–the poor cousin of health research? *BMC medical research methodology, 9*, 52.

Newcombe, R. G. (1998). Two-sided confidence intervals for the single proportion: comparison of seven methods. *Statistics in Medicine, 17*(8), 857-872.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology, 2*(2), 175-220.

Novak, D. R., Bailey, C., Brill, K., Eckert, M., Petersen, D., Rausch, R., & Schichtel, M. (2011). *Human improvement to numerical weather prediction at the Hydrometeorological Prediction Center.* Paper presented at the 91th Annual AMS Meeting, AMS 24th Conference on Weather and Forecasting/20th Conference on Numerical Weather Prediction.

O'Connor, G. T., & Sox, H. C. (1991). Bayesian Reasoning in Medicine The Contributions of Lee B. Lusted, MD. *Medical Decision Making, 11*(2), 107-111.

O'Hearn, M., Lowry, C., Emerson-Kavchak, A., & Courtney, C. (2009). Re:“Interrater reliability of a new classification scheme for patients with neural low back-related leg pain” Schafer A et al. J Man Manip Ther

2009; 17: 109–116. *The Journal of Manual & Manipulative Therapy, 17*(4), e115.

O'Sullivan, M. J., & O'Sullivan, I. (2006). Application of the Ottawa Knee Rules in assessing acute knee injuries. *Ir Med J, 99*(7), 199-200.

O'Sullivan, P. (2005). Diagnosis and classification of chronic low back pain disorders: maladaptive movement and motor control impairments as underlying mechanism. *Manual Therapy, 10*(4), 242-255.

O'Neill, C. W., Kurgansky, M. E., Derby, R., & Ryan, D. P. (2002). Disc stimulation and patterns of referred pain. *Spine, 27*(24), 2776-2781.

Osmond, M. H., Klassen, T. P., Wells, G. A., Correll, R., Jarvis, A., Joubert, G., . . . Stiell, I. G. (2010). CATCH: a clinical decision rule for the use of computed tomography in children with minor head injury. *Canadian Medical Association Journal, 182*(4), 341-348.

Ozgocmen, S., & Khan, M. A. (2012). Current concept of spondyloarthritis: special emphasis on early referral and diagnosis. *Curr Rheumatol Rep, 14*(5), 409-414.

Ozguler, A., Gueguen, A., Leclerc, A., Landre, M. F., Piciotti, M., Le Gall, S., . . . Boureau, F. (2002). Using the dallas pain questionnaire to classify individuals with low back pain in a working population. *Spine, 27*(16), 1783-1789.

Parisien, R. C., & Ball, P. A. (1998). William Jason Mixter (1880-1958). Ushering in the "dynasty of the disc". *Spine, 23*(21), 2363-2366.

Parker, M. (2005). False dichotomies: EBM, clinical freedom, and the art of medicine. *Medical Humanities, 31*(1), 23-30.

Patel, S., Friede, T., Froud, R., Evans, D. W., & Underwood, M. (2013). Systematic review of randomized controlled trials of clinical prediction rules for physical therapy in low back pain. *Spine, 38*(9), 762-769.

Pengel, L. H., Herbert, R. D., Maher, C. G., & Refshauge, K. M. (2003). Acute low back pain: systematic review of its prognosis. *British Medical Journal, 327*(7410), 323.

Perry, J. J., Eagles, D., Clement, C. M., Brehaut, J., Kelly, A. M., & Stiell, I. G. (2009). An international study of emergency physicians' practice for acute headache management and the need for a clinical decision rule. *Canadian Journal of Emergency Medicine, 11*(6), 516-522.

Petersen, T., Laslett, M., Thorsen, H., Manniche, C., Ekdahl, C., & Jacobsen, S. (2003). Diagnostic classification of non-specific low back pain. A new system integrating patho-anatomic and clinical categories. *Physiotherapy Theory and Practice, 19*(4), 213-237.

Petersen, T., Thorsen, H., Manniche, C., & Ekdahl, C. (1999). Classification of non-specific low back pain: a review of the literature on classifications systems relevant to physiotherapy. *Physical Therapy Reviews, 4*(4), 265-281.

Petty, N. J., Thomson, O. P., & Stew, G. (2012a). Ready for a paradigm shift? Part 1: Introducing the philosophy of qualitative research. *Manual Therapy, 17*, 267-274.

Petty, N. J., Thomson, O. P., & Stew, G. (2012b). Ready for a paradigm shift? Part 2: Introducing qualitative research methodologies and methods. *Manual Therapy, 17*, 378-384.

Peul, W. C., van Houwelingen, H. C., van den Hout, W. B., Brand, R., Eekhof, J. A. H., Tans, J. T. J., . . . Koes, B. W. (2007). Surgery versus prolonged conservative treatment for sciatica. *New England Journal of Medicine, 356*(22), 2245-2256.

Player, R. P. (1821). On irritation of the spinal nerves. *Quarterly Journal of Science, 12*, 428-430.

Plüddemann, A., Wallace, E., Bankhead, C., Keogh, C., Van der Windt, D. A., Lasserson, D., . . . O'Brien, K. K. (2014). Clinical prediction rules in practice: review of clinical guidelines and survey of GPs. *British Journal of General Practice, 64*(621), e233-e242.

Poddubnyy, D., Vahldiek, J., Spiller, I., Buss, B., Listing, J., Rudwaleit, M., & Sieper, J. (2011). Evaluation of 2 screening strategies for early identification of patients with axial spondyloarthritis in primary care. *Journal of Rheumatology, 38*(11), 2452-2460.

Pogorzelska, M., & Larson, E. L. (2008). Assessment of attitudes of intensive care unit staff toward clinical practice guidelines. *DCCN - Dimensions of Critical Care Nursing, 27*(1), 30-38.

Poitras, S., Blais, R., Swaine, B., & Rossignol, M. (2005). Management of work-related low back pain: a population-based survey of physical therapists. *Physical Therapy, 85*(11), 1168-1181.

Poitras, S., Blais, R., Swaine, B., & Rossignol, M. (2007). Practice patterns of physiotherapists in the treatment of work-related back pain. *J Eval Clin Pract, 13*(3), 412-421.

Poynton, A. R., Hinman, A., Lutz, G., & Farmer, J. C. (2005). Discography-induced acute lumbar disc herniation: a report of five cases. *Journal of spinal disorders & techniques, 18*(2), 188-192.

Prüss-Ustün, A., Mathers, C., Corvalán, C., & Woodward, A. (2003). *Introduction and methods: assessing the environmental burden of disease at national and local levels*. Geneva: World Health Organization.

Randolph, A. G., Guyatt, G. H., Calvin, J. E., Doig, G., Richardson, W. S., & Cook, D. J. (1998). Understanding articles describing clinical prediction tools. *Critical Care Medicine, 26*(9), 1603-1612.

Raney, N. H., Petersen, E. J., Smith, T. A., Cowan, J. E., Rendeiro, D. G., Deyle, G. D., & Childs, J. D. (2009). Development of a clinical prediction rule to identify patients with neck pain likely to benefit from cervical traction and exercise. *European Spine Journal, 18*(3), 382-391.

Rathmell, J. P., Saal, J. S., & Saal, J. (2008). Discography, IDET, Percutaneous Discectomy, and Nucleoplasty: Complications and Their Prevention. *Pain Medicine, 9*(S1), S73-S81.

Reilly, B. M., & Evans, A. T. (2006). Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Annals of Internal Medicine, 144*(3), 201-209.

Reitsma, J., Rutjes, A., Whiting, P., Vlassov, V., Leeflang, M., & Deeks, J. (2009). Chapter 9: Assessing methodological quality *Cochrane handbook for systematic reviews of diagnostic test accuracy version 1.0.0* (Vol. 1.0.0): The Cochrane Collaboration.

Reserve Bank of Australia. (2014). Inflation Calculator. Retrieved October, 2014, from http://www.rba.gov.au/calculator/annualDecimal.html

Revel, M. E., Listrat, V. M., Chevalier, X. J., Dougados, M., N'guyen, M. P., Vallee, C., . . . Amor, B. (1992). Facet joint block for low back pain: identifying predictors of a good response. *Archives of Physical Medicine and Rehabilitation, 73*(9), 824-828.

Revel, M. E., Poiraudeau, S., Auleley, G. R., Payan, C., Denke, A., Nguyen, M., . . . Fermanian, J. (1998). Capacity of the clinical picture to characterize low back pain relieved by facet joint anesthesia. Proposed criteria to identify patients with painful facet joints. *Spine, 23*(18), 1972-1976; discussion 1977.

Rezaian, S. M., Spector, I. M., & Collins, A. (1993). A new clinical classification of low back pain: a practical guide for diagnosing etiology by patient history. *JOURNAL OF NEUROLOGICAL AND ORTHOPAEDIC MEDICINE AND SURGERY, 14*, 205-205.

Richardson, W. S., & Wilson, M. C. (2008). Chapter 14: The Process of Diagnosis. In G. Guyatt, D. Rennie, M. O. Meade & D. G. Cook (Eds.), *Users' Guides to the Medical Literature* (2nd ed., pp. 399-406). USA: McGraw Hill Medical.

Richardson, W. S., Wilson, M. C., Guyatt, G. H., Cook, D. J., Nishikawa, J., & Evidence-Based Medicine Working Group. (1999). Users' guides to the medical literature: XV. How to use an article about disease probability for differential diagnosis. *Journal of the American Medical Association, 281*(13), 1214-1219.

Riddle, D. L. (1998). Classification and low back pain: a review of the literature and critical analysis of selected systems. *Physical Therapy, 78*(7), 708-737.

Riihimaki, H. (1996). Epidemiology and pathogenesis of non-specific low back pain: what does the epidemiology tell us? *Bulletin of the Hospital for Joint Diseases, 55*(4), 197-198.

Rivett, D. A., & Haskins, R. (2009). Weighing up CPRs *In Touch* (pp. 4). Victoria, Australia: Musculoskeletal Physiotherapy Australia, Australian Physiotherapy Association.

Rodin, A. E. (1962). A historical survey of disease concepts. *Canadian Medical Association Journal, 87*(3), 124-128.

Roebber, P. J., & Bosart, L. F. (1996). The complex relationship between forecast skill and forecast value: A real-world analysis. *Weather and Forecasting, 11*(4), 544-559.

Roland, M., Morrell, D. C., & Morris, R. (1983). Can general practitioners predict the outcome of episodes of back pain? *British Medical Journal, 286*(6364), 523-526.

Roland, M., & Morris, R. (1983). A study of the natural history of back pain: part I: development of a reliable and sensitive measure of disability in low-back pain. *Spine, 8*(2), 141-144.

Roman, M., Brown, C., Richardson, W. S., Isaacs, R., Howes, C., & Cook, C. (2010). The development of a clinical decision making algorithm for detection of osteoporotic vertebral compression fracture or wedge deformity. *Journal of Manual & Manipulative Therapy, 18*(1), 44-49.

Rossignol, M., Arsenault, B., Dionne, C. E., Poitras, S., Tousignant, M., Truchon, M., . . . Neveu, A. (2007). Clinic on low-back pain in interdisciplinary practice (CLIP) guidelines. Montréal: Direction de santé publique, Agence de la santé et des services sociaux de Montréal.

Roux, C., Priol, G., Fechtenbaum, J., Cortet, B., Liu-Leage, S., & Audran, M. (2007). A clinical tool to determine the necessity of spine radiography in postmenopausal women with osteoporosis presenting with back pain. *Annals of the rheumatic diseases, 66*(1), 81-85.

Rubinson, L., Wu, A. W., Haponik, E. F., & Diette, G. B. (2005). Why is it that internists do not follow guidelines for preventing intravascular catheter infections? *Infection Control and Hospital Epidemiology, 26*(6), 525-533.

Rudwaleit, M., Listing, J., Brandt, J., Braun, J., & Sieper, J. (2004). Prediction of a major clinical response (BASDAI 50) to tumour necrosis factor α blockers in ankylosing spondylitis. *Annals of the rheumatic diseases, 63*(6), 665-670.

Rudwaleit, M., Metter, A., Listing, J., Sieper, J., & Braun, J. (2006). Inflammatory back pain in ankylosing spondylitis: a reassessment of the clinical history for application as classification and diagnostic criteria. *Arthritis & Rheumatism, 54*(2), 569-578.

Rudwaleit, M., van der Heijde, D., Khan, M. A., Braun, J., & Sieper, J. (2004). How to diagnose axial spondyloarthritis early. *Annals of the rheumatic diseases, 63*(5), 535-543.

Rudwaleit, M., Van der Heijde, D., Landewé, R., Listing, J., Akkoc, N., Brandt, J., . . . Dougados, M. (2009). The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. *Annals of the rheumatic diseases, 68*(6), 777-783.

Runyon, M. S., Richman, P. B., & Kline, J. A. (2007). Emergency medicine practitioner knowledge and use of decision rules for the evaluation of patients with suspected pulmonary embolism: variations by practice setting and training level. *Academic Emergency Medicine, 14*(1), 53-57.

Sachs, B. L., Vanharanta, H., Spivey, M. A., Guyer, R. D., Videman, T., Rashbaum, R. F., . . . Mooney, V. (1987). Dallas discogram description a new classification of CT/discography in low-back disorders. *Spine, 12*(3), 287-294.

Sackett, D., Hayes, R., Guyatt, G., & Tugwell, P. (1991). *Clinical Epidemiology: A Basic Science for Clinical Medicine. 2nd Ed.* Boston, USA: Brown & Co.

Safe Work Australia. (2013). Compendium of Workers' Compensation Statistics Australia 2010-11. Canberra.

Salomon, J. A., Vos, T., Hogan, D. R., Gagnon, M., Naghavi, M., Mokdad, A., . . . Kosen, S. (2013). Common values in assessing health outcomes from disease and injury: disability weights measurement study for the Global Burden of Disease Study 2010. *The Lancet, 380*(9859), 2129-2143.

Sandelowski, M. (2000). Whatever happened to qualitative description? *Research in Nursing & Health, 23*(4), 334-340.

Sandelowski, M. (2010). What's in a name? Qualitative description revisited. *Research in Nursing & Health, 33*(1), 77-84.

Schäfer, A., Hall, T., & Briffa, K. (2009). Classification of low back-related leg pain—a proposed patho-mechanism-based approach. *Manual Therapy, 14*(2), 222-230.

Schellingerhout, J. M., Heymans, M. W., de Vet, H. C. W., Koes, B. W., & Verhagen, A. P. (2009). Categorizing continuous variables resulted in different predictors in a prognostic model for nonspecific neck pain. *Journal of Clinical Epidemiology, 62*(8), 868-874.

Schenk, R., Dionne, C., Simon, C., & Johnson, R. (2012). Effectiveness of mechanical diagnosis and therapy in patients with back pain who meet a clinical prediction rule for spinal manipulation. *Journal of Manual and Manipulative Therapy, 20*(1), 43-49.

Schneider, G. M., Jull, G. A., Thomas, K., Smith, A., Emery, C., Faris, P., . . . Salo, P. (2014). Derivation of a Clinical Decision Guide in the Diagnosis of Cervical Facet Joint Pain. *Archives of Physical Medicine and Rehabilitation, 95*(9), 1695-1701.

Schofield, D. J., Callander, E. J., Shrestha, R. N., Percival, R., Kelly, S. J., & Passey, M. E. (2012). Labor Force Participation and the Influence of Having Back Problems on Income Poverty in Australia. *Spine, 37*(13), 1156-1163.

Schofield, D. J., Kelly, S. J., Shrestha, R. N., Callander, E. J., Passey, M. E., & Percival, R. (2012). The impact of back problems on retirement wealth. *Pain, 153*(1), 203-210.

Schofield, D. J., Shrestha, R. N., Passey, M. E., Earnest, A., & Fletcher, S. L. (2008). Chronic disease and labour force participation among older Australians. *Medical Journal of Australia, 189*(8), 447-450.

Schofield, D. J., Shrestha, R. N., Percival, R., Callander, E. J., Kelly, S. J., & Passey, M. E. (2011). Early retirement and the financial assets of individuals with back problems. *European Spine Journal, 20*(5), 731-736.

Schofield, D. J., Shrestha, R. N., Percival, R., Passey, M. E., Callander, E. J., & Kelly, S. J. (2012). The personal and national costs of early retirement because of spinal disorders: impacts on income, taxes, and government support payments. *The Spine Journal, 12*(12), 1111-1118.

Scholz, J., Mannion, R. J., Hord, D. E., Griffin, R. S., Rawal, B., Zheng, H., . . . Woolf, C. J. (2009). A novel tool for the assessment of pain: validation in low back pain. *PLoS Medicine / Public Library of Science, 6*(4), e1000047.

Schouten, J. A., Hulscher, M. E. J. L., Natsch, S., Kullberg, B. J., Van Der Meer, J. W. M., & Grol, R. P. T. M. (2007). Barriers to optimal antibiotic use for community-acquired pneumonia at hospitals: A qualitative study. *Quality and Safety in Health Care, 16*(2), 143-149.

Schulz, K. F., Altman, D. G., Moher, D., & Consort Group. (2010). CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Annals of Internal Medicine, 152*(11), 726-732.

Schuman, H., & Presser, S. (1979). The open and closed question. *American sociological review, 44*(October), 692-712.

Schuman, H., & Scott, J. (1987). Problems in the use of survey questions to measure public opinion. *Science, 236*(4804), 957-959.

Schwarzer, A. C., Aprill, C. N., Derby, R., Fortin, J. D., Kine, G., & Bogduk, N. (1995). The prevalence and clinical features of internal disc disruption in patients with chronic low back pain. *Spine, 20*(17), 1878-1883.

Schwind, J., Learman, K., O'Halloran, B., Showalter, C., & Cook, C. (2013). Different minimally important clinical difference (MCID) scores lead to different clinical prediction rules for the Oswestry disability index for the same sample of patients. *Journal of Manual and Manipulative Therapy, 21*(2), 71-78.

Seel, R. T., Steyerberg, E. W., Malec, J. F., Sherer, M., & Macciocchi, S. N. (2012). Developing and evaluating prediction models in rehabilitation populations. *Archives of Physical Medicine and Rehabilitation, 93*(8), S138-S153.

Shekelle, P. G., Adams, A. H., Chassin, M. R., Hurwitz, E. L., & Brook, R. H. (1992). Spinal manipulation for low-back pain. *Annals of Internal Medicine, 117*(7), 590-598.

Sieper, J. (2012). How to screen for axial spondyloarthritis in primary care? *Current Opinion in Rheumatology, 24*(4), 359-362.

Sieper, J., Rudwaleit, M., Baraliakos, X., Brandt, J., Braun, J., Burgos-Vargas, R., . . . Maksymowych, W. (2009). The Assessment of SpondyloArthritis international Society (ASAS) handbook: a guide to

assess spondyloarthritis. *Annals of the rheumatic diseases, 68*(Suppl 2), ii1-ii44.

Sieper, J., Srinivasan, S., Zamani, O., Mielants, H., Choquette, D., Pavelka, K., . . . Vastesaeger, N. (2013). Comparison of two referral strategies for diagnosis of axial spondyloarthritis: the Recognising and Diagnosing Ankylosing Spondylitis Reliably (RADAR) study. *Annals of the rheumatic diseases, 72*(10), 1621-1627.

Sieper, J., van der Heijde, D., Landewe, R., Brandt, J., Burgos-Vagas, R., Collantes-Estevez, E., . . . Rudwaleit, M. (2009). New criteria for inflammatory back pain in patients with chronic back pain: a real patient exercise by experts from the Assessment of SpondyloArthritis international Society (ASAS). *Annals of the rheumatic diseases, 68*(6), 784-788.

Sikorski, J. M. (1985). A rationalized approach to physiotherapy for low-back pain. *Spine, 10*(6), 571-579.

Simon, H. A. (1990). Invariants of human behavior. *Annual review of psychology, 41*(1), 1-20.

Skargren, E. I., Carlsson, P. G., & Oberg, B. E. (1998). One-Year Follow-up Comparison of the Cost and Effectiveness of Chiropractic and Physiotherapy as Primary Management for Back Pain: Subgroup Analysis, Recurrence, and Additional Health Care Utilization. *Spine, 23*(17), 1875-1883.

Slaboda, J., Boston, J. R., & Rudy, T. E. (2006). *Evaluating reliability of hidden Markov models that describe the lifting patterns of chronic lower back pain patients and controls.* Paper presented at the

Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE.

Slade, S. C., Molloy, E., & Keating, J. L. (2009). People with non-specific chronic low back pain who have participated in exercise programs have preferences about exercise: a qualitative study. *Australian Journal of Physiotherapy, 55*(2), 115-121.

Sowden, G., Hill, J. C., Konstantinou, K., Khanna, M., Main, C. J., Salmon, P., . . . Foster, N. E. (2011). Subgrouping for targeted treatment in primary care for low back pain: the treatment system and clinical training programmes used in the IMPaCT Back study. *Family practice*, cmr037.

Sparks, C. L., McGehee, W. F., Buettner, M. A., & Scott, K. M. (2010). Use of a clinical prediction rule for low back pain by outpatient physical therapists. *Journal of Manual & Manipulative Therapy, 18*(4), 237-237.

Spitzer, W. (1987). Scientific approach to the assessment and management of activity-related spinal disorders. A monograph for clinicians. Report of the Quebec Task Force on Spinal Disorders. *Spine (Phila Pa 1976), 12*(7 Suppl), S1-59.

Spoto, M. M., & Collins, J. (2008). Physiotherapy diagnosis in clinical practice: a survey of orthopaedic certified specialists in the USA. *Physiotherapy Research International, 13*(1), 31-41.

Stanton, T. R., Fritz, J. M., Hancock, M. J., Latimer, J., Maher, C. G., Wand, B. M., & Parent, E. C. (2011). Evaluation of a treatment-based classification algorithm for low back pain: a cross-sectional study. *Physical Therapy, 91*(4), 496-509.

Stanton, T. R., Hancock, M. J., Apeldoorn, A. T., Wand, B. M., & Fritz, J. M. (2013). What characterizes people who have an unclear classification using a treatment-based classification algorithm for low back pain? A cross-sectional study. *Physical Therapy, 93*(3), 345-355.

Stanton, T. R., Hancock, M. J., Maher, C. G., & Koes, B. W. (2010). Critical appraisal of clinical prediction rules that aim to optimize treatment selection for musculoskeletal conditions. *Physical Therapy, 90*(6), 843-854.

Stanton, T. R., Henschke, N., Maher, C. G., Refshauge, K. M., Latimer, J., & McAuley, J. H. (2008). After an episode of acute low back pain, recurrence is unpredictable and not as common as previously thought. *Spine, 33*(26), 2923-2928.

Stanton, T. R., Latimer, J., Maher, C. G., & Hancock, M. (2009). Definitions of recurrence of an episode of low back pain: a systematic review. *Spine, 34*(9), E316-E322.

Stanton, T. R., Maher, C. G., & Hancock, M. (2009). On Clinical prediction rules for physical therapy interventions. *Physical Therapy, 89*(4), 394.

Stiell, I. G., Brehaut, J., Clement, C., Grimshaw, J., Graham, I., Brison, R., . . . Wells, G. A. (2006). Perceived barriers to the implementation of the Canadian C-spine rule and the Canadian CT head rule. *Canadian Journal of Emergency Medicine, 8*(3), 206-207.

Stiell, I. G., Greenberg, G. H., McKnight, R. D., Nair, R. C., McDowell, I., & Worthington, J. R. (1992). A study to develop clinical decision rules for the use of radiography in acute ankle injuries. *Annals of Emergency Medicine, 21*(4), 384-390.

Stiell, I. G., Greenberg, G. H., Wells, G. A., McKnight, R. D., Cwinn, A. A., Cacciotti, T., . . . Smith, N. A. (1995). Derivation of a decision rule for the use of radiography in acute knee injuries. *Annals of Emergency Medicine, 26*(4), 405-413.

Stiell, I. G., McKnight, R. D., Greenberg, G. H., McDowell, I., Nair, R. C., Wells, G. A., . . . Worthington, J. R. (1994). Implementation of the Ottawa ankle rules. *Journal of the American Medical Association, 271*(11), 827-832.

Stiell, I. G., Wells, G., Laupacis, A., Brison, R., Verbeek, R., Vandemheen, K., & Naylor, C. D. (1995). Multicentre trial to introduce the Ottawa ankle rules for use of radiography in acute ankle injuries. *British Medical Journal, 311*(7005), 594-597.

Stiell, I. G., & Wells, G. A. (1999). Methodologic standards for the development of clinical decision rules in emergency medicine. *Annals of Emergency Medicine, 33*(4), 437-447.

Stiell, I. G., Wells, G. A., Hoag, R. H., Sivilotti, M. L. A., Cacciotti, T. F., Verbeek, P. R., . . . Greenberg, G. H. (1997). Implementation of the Ottawa Knee Rule for the use of radiography in acute knee injuries. *Journal of the American Medical Association, 278*(23), 2075-2079.

Stiell, I. G., Wells, G. A., McDowell, I., Greenberg, G. H., McKnight, R. D., Cwinn, A. A., . . . Yeats, A. (1995). Use of radiography in acute knee injuries: need for clinical decision rules. *Academic Emergency Medicine, 2*(11), 966-973.

Stolze, L. R., Allison, S. C., & Childs, J. D. (2012). Derivation of a preliminary clinical prediction rule for identifying a subgroup of patients with low

back pain likely to benefit from Pilates-based exercise. *Journal of Orthopaedic & Sports Physical Therapy, 42*(5), 425-436.

Stratford, P. W. (2010). The Added Value of Confidence Intervals. *Physical Therapy, 90*(3), 333-335.

Stroke Foundation of New Zealand. (2008). New Zealand Guideline for the Assessment and Management of People with Recent Transient Ischaemic Attack (TIA). Retrieved Dec, 2014, from http://www.stroke.org.nz/resources/TIA_GuidelineForweb.pdf?bcsi_scan_9CF786ACA806128D=0&bcsi_scan_filename=TIA_GuidelineForweb.pdf

Strong, J., Ashton, R., & Stewart, A. (1994). Chronic low back pain: toward an integrated psychosocial assessment model. *Journal of Consulting & Clinical Psychology, 62*(5), 1058-1063.

Sugioka, T., Hayashino, Y., Konno, S., Kikuchi, S., & Fukuhara, S. (2008). Predictive value of self-reported patient information for the identification of lumbar spinal stenosis. *Family practice, 25*(4), 237-244.

Sullivan, R. (1995). A brief journey into medical care and disease in ancient Egypt. *Journal of the Royal Society of Medicine, 88*(3), 141-145.

Sun, X., Briel, M., Walter, S. D., & Guyatt, G. H. (2010). Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *British Medical Journal, 340*, c117.

Sutlive, T. G., Lopez, H. P., Schnitker, D. E., Yawn, S. E., Halle, R. J., Mansfield, L. T., . . . Childs, J. D. (2008). Development of a clinical prediction rule for diagnosing hip osteoarthritis in individuals with

unilateral hip pain. *Journal of Orthopaedic & Sports Physical Therapy, 38*(9), 542-550.

Sutlive, T. G., Mabry, L. M., Easterling, E. J., Durbin, J. D., Hanson, S. L., Wainner, R. S., & Childs, J. D. (2009). Comparison of short-term response to two spinal manipulation techniques for patients with low back pain in a military beneficiary population. *Military Medicine, 174*(7), 750-756.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000a). Better decisions through science. *Scientific American, 283*(4), 82-87.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000b). Psychological science can improve diagnostic decisions. *Psychological science in the public interest, 1*(1), 1-26.

Sydenham, T. (1848). *The Works of Thomas Sydenham. Vol 1 of 2. Translated by R. G. Latham.* London: The Sydenham Society.

Szadek, K. M., van der Wurff, P., van Tulder, M. W., Zuurmond, W. W., & Perez, R. S. (2009). Diagnostic validity of criteria for sacroiliac joint pain: a systematic review. *Journal of Pain, 10*(4), 354-368.

Tamariz, L. J., Eng, J., Segal, J. B., Krishnan, J. A., Bolger, D. T., Streiff, M. B., . . . Bass, E. B. (2004). Usefulness of clinical prediction rules for the diagnosis of venous thromboembolism: a systematic review. *American Journal of Medicine, 117*(9), 676-684.

Teyhen, D. S., Flynn, T. W., Childs, J. D., & Abraham, L. D. (2007). Arthrokinematics in a subgroup of patients likely to benefit from a lumbar stabilization exercise program. *Physical Therapy, 87*(3), 313-325.

The Cochrane Collaboration. (2012). Cochrane Primary Health Care Field: More about us. Retrieved December, 2014, from http://www.cochraneprimarycare.org/more-about-us

Thiruganasambandamoorthy, V., Stiell, I. G., Sivilotti, M. L. A., Murray, H., Rowe, B. H., Lang, E., . . . Wells, G. A. (2014). Risk stratification of adult emergency department syncope patients to predict short-term serious outcomes after discharge (RiSEDS) study. *BMC Emergency Medicine, 14*, 8.

Thomas, E., Silman, A. J., Croft, P. R., Papageorgiou, A. C., Jayson IV, M., & Macfarlane, G. J. (1999). Predicting who develops chronic low back pain in primary care: a prospective study. *British Medical Journal, 318*(7199), 1662-1667.

Thorpe, A. J., & Petersen, G. N. (2006). Predictability and targeted observations. In T. Palmer & R. Hagedorn (Eds.), *Predictability of Weather and Climate*: Cambridge University Press.

Toll, D. B., Janssen, K. J. M., Vergouwe, Y., & Moons, K. G. M. (2008). Validation, updating and impact of clinical prediction rules: A review. *Journal of Clinical Epidemiology, 61*(11), 1085-1094.

Tu, J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology, 49*(11), 1225-1231.

Turner, J. A., Holtzman, S., & Mancl, L. (2007). Mediators, moderators, and predictors of therapeutic change in cognitive-behavioral therapy for chronic pain. *Pain, 127*(3), 276-286.

Tuyl, F. (2009). *Diagnostic testing: a simple approximate interval for the positive predictive value*. ASEARC: Proceedings of the Third Annual ASEARC Research Conference, Newcastle, NSW, Australia.

Tuyl, F., Gerlach, R., & Mengersen, K. (2008). A Comparison of Bayes–Laplace, Jeffreys, and Other Priors. *The American Statistician, 62*(1), 40-44.

Underwood, M. R., & Dawes, P. (1995). Inflammatory back pain in primary care. *Rheumatology, 34*(11), 1074-1077.

US National Library of Medicine. (2009). PubMed Clinical Queries. Retrieved December, 2014, from http://www.ncbi.nlm.nih.gov/pubmed/clinical/

Vaismoradi, M., Turunen, H., & Bondas, T. (2013). Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing & health sciences, 15*(3), 398-405.

Valat, J. P., Goupille, P., Rozenberg, S., Urbinelli, R., Allaert, F., & Spine group of the Societe Francaise. (2000). Acute low back pain: predictive index of chronicity from a cohort of 2487 subjects. *Joint, Bone, Spine: Revue du Rhumatisme, 67*(5), 456-461.

van der Hulst, M., Vollenbroek-Hutten, M. M. R., Groothuis-Oudshoorn, K. G. M., & Hermens, H. J. (2008). Multidisciplinary rehabilitation treatment of patients with chronic low back pain: a prognostic model for its outcome. *Clinical Journal of Pain, 24*(5), 421-430.

van der Windt, D. A., Simons, E., Riphagen, II, Ammendolia, C., Verhagen, A. P., Laslett, M., . . . Aertgeerts, B. (2010). Physical examination for

lumbar radiculopathy due to disc herniation in patients with low-back pain. *Cochrane Database of Systematic Reviews*(2), CD007431.

van der Wurff, P., Buijs, E. J., & Groen, G. J. (2006). A multitest regimen of pain provocation tests as an aid to reduce unnecessary minimally invasive sacroiliac joint procedures. *Archives of Physical Medicine & Rehabilitation, 87*(1), 10-14.

van Dillen, L. R., Sahrmann, S. A., Norton, B. J., Caldwell, C. A., Fleming, D. A., McDonnell, M. K., & Woolsey, N. B. (1998). Reliability of physical examination items used for classification of patients with low back pain. *Physical Therapy, 78*(9), 979-988.

van Middendorp, J. J., Sanchez, G. M., & Burridge, A. L. (2010). The Edwin Smith papyrus: a clinical reappraisal of the oldest known document on spinal injuries. *European Spine Journal, 19*(11), 1815-1823.

van Oort, L., van den Berg, T., Koes, B. W., de Vet, R. H. C. W., Anema, H. J. R., Heymans, M. W., & Verhagen, A. P. (2012). Preliminary state of development of prediction models for primary care physical therapy: a systematic review. *Journal of Clinical Epidemiology, 65*(12), 1257-1266.

Verwoerd, A. J. H., Luijsterburg, P. A. J., Lin, C. W. C., Jacobs, W. C. H., Koes, B. W., & Verhagen, A. P. (2013). Systematic review of prognostic factors predicting outcome in non-surgically treated patients with sciatica. *European Journal of Pain, 17*(8), 1126-1137.

Vos, T., Flaxman, A. D., Naghavi, M., Lozano, R., Michaud, C., Ezzati, M., . . . Murray, C. J. L. (2012). Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic

analysis for the Global Burden of Disease Study 2010. *The Lancet, 380*(9859), 2163-2196.

Vroomen, P. C. A. J., de Krom, M. C. T. F. M., & Knottnerus, J. A. (1999). Diagnostic value of history and physical examination in patients suspected of sciatica due to disc herniation: a systematic review. *Journal of neurology, 246*(10), 899-906.

Vroomen, P. C. A. J., De Krom, M. C. T. F. M., & Knottnerus, J. A. (2002). Predicting the outcome of sciatica at short-term follow-up. *British Journal of General Practice, 52*(475), 119-123.

Waddell, G. (1996). Low back pain: a twentieth century health care enigma. *Spine, 21*(24), 2820-2825.

Waddell, G., & Allan, D. B. (2004). Chapter 4. Back pain through history. In G. Waddell (Ed.), *The Back Pain Revolution* (2nd ed., pp. 47-69). Edinburgh: Churchill Livingstone.

Waddell, G., Newton, M., Henderson, I., Somerville, D., & Main, C. J. (1993). A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. *Pain, 52*(2), 157-168.

Waldrop, M. A. (2006). Diagnosis and treatment of cervical radiculopathy using a clinical prediction rule and a multimodal intervention approach: a case series. *Journal of Orthopaedic & Sports Physical Therapy, 36*(3), 152-159.

Walker, B. F. (1999). The prevalence of low back pain in Australian adults. A systematic review of the literature from 1966-1998. *Asia-Pacific Journal of Public Health, 11*(1), 45-51.

Walker, B. F., Muller, R., & Grant, W. D. (2003). Low back pain in Australian adults: the economic burden. *Asia-Pacific Journal of Public Health, 15*(2), 79-87.

Walker, B. F., Muller, R., & Grant, W. D. (2004a). Low back pain in Australian adults. health provider utilization and care seeking. *Journal of Manipulative and Physiological Therapeutics, 27*(5), 327-335.

Walker, B. F., Muller, R., & Grant, W. D. (2004b). Low back pain in Australian adults: prevalence and associated disability. *Journal of Manipulative and Physiological Therapeutics, 27*(4), 238-244.

Walker III, J., El Abd, O., Isaac, Z., & Muzin, S. (2008). Discography in practice: a clinical and historical review. *Current reviews in musculoskeletal medicine, 1*(2), 69-83.

Wallace, E., Smith, S. M., Perera-Salazar, R., Vaucher, P., McCowan, C., Collins, G. A., . . . Fahey, T. (2011). Framework for the impact analysis and implementation of Clinical Prediction Rules (CPRs). *BMC medical informatics and decision making, 11*(1), 62.

Walsh, T. R., Weinstein, J. N., Spratt, K. F., Lehmann, T. R., Aprill, C. N., & Sayre, H. (1990). Lumbar discography in normal subjects. A controlled, prospective study. *The Journal of Bone & Joint Surgery, 72*(7), 1081-1088.

Wasiak, R., Pransky, G. S., & Webster, B. S. (2003). Methodological challenges in studying recurrence of low back pain. *Journal of Occupational Rehabilitation, 13*(1), 21-31.

Wassenaar, M., van Rijn, R. M., van Tulder, M. W., Verhagen, A. P., van der Windt, D. A. W. M., Koes, B. W., . . . Ostelo, R. W. J. G. (2012).

Magnetic resonance imaging for diagnosing lumbar spinal pathology in adult patients with low back pain or sciatica: a diagnostic systematic review. *European Spine Journal, 21*(2), 220-227.

Wasson, J. H., Sox, H. C., Neff, R. K., & Goldman, L. (1985). Clinical prediction rules. Applications and methodological standards. *New England Journal of Medicine, 313*(13), 793-799.

Webster, B. S., & Snook, S. H. (1994). The cost of 1989 workers' compensation low back pain claims. *Spine, 19*(10), 1111-1115.

Weinstein, J. N., Tosteson, T. D., Lurie, J. D., Tosteson, A. N. A., Blood, E., Herkowitz, H., . . . Hilibrand, A. (2010). Surgical versus non-operative treatment for lumbar spinal stenosis four-year results of the Spine Patient Outcomes Research Trial (SPORT). *Spine, 35*(14), 1329 - 1338.

Weisman, M. H. (2012). Inflammatory back pain: the United States perspective. *Rheumatic Disease Clinics of North America, 38*(3), 501-512.

Wells, P. S., Hirsh, J., Anderson, D. R., Lensing, A. W. A., Foster, G., Kearon, C., . . . Prandoni, P. (1995). Accuracy of clinical assessment of deep-vein thrombosis. *The Lancet, 345*(8961), 1326-1330.

Werner, E. L., & Indahl, A. (2005). Knowledge, practice and attitudes to back pain among doctors, physiotherapists and chiropractors. *Tidsskrift for Den Norske Laegeforening, 125*(13), 1794-1797.

Whitehurst, D. G. T., Bryan, S., Lewis, M., Hill, J., & Hay, E. M. (2012). Exploring the cost–utility of stratified primary care management for low

back pain compared with current best practice within risk-defined subgroups. *Annals of the rheumatic diseases, 71*(11), 1796-1802.

Whiting, P. F., Rutjes, A. W. S., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., . . . Bossuyt, P. M. M. (2011). QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine, 155*(8), 529-536.

Willett, G., Johnson, G., & Jones, K. (2011). The effect of a hybrid continuing education course on outpatient physical therapy for individuals with low back pain. *The Internet Journal of Allied Health Sciences and Practice, 9*(1), 1-11.

Williams, C. M., Henschke, N., Maher, C. G., van Tulder, M. W., Koes, B. W., Macaskill, P., & Irwig, L. (2013). Red flags to screen for vertebral fracture in patients presenting with low-back pain. *Cochrane Database of Systematic Reviews, 1*, CD008643.

Wisnivesky, J. P., Serebrisky, D., Moore, C., Sacks, H. S., Iannuzzi, M. C., & McGinn, T. (2005). Validity of clinical prediction rules for isolating inpatients with suspected tuberculosis. A systematic review. *Journal of General Internal Medicine, 20*(10), 947-952.

Wolmar, C. (2012). *The Great Railroad Revolution: The History of Trains in America*. USA: PublicAffairs.

Wong, S. S.-L., Wilczynski, N. L., Haynes, R. B., Ramkissoonsingh, R., & Hedges Team. (2003). *Developing optimal search strategies for detecting sound clinical prediction studies in MEDLINE.* Paper presented at the AMIA Annual Symposium Proceedings.

WorkCover New South Wales. (2010). New South Wales Workers Compensation Statistical Bulletin 2008/09.

World Health Organization. (2014a). Disability weights, discounting and age weighting of DALYs. Retrieved October, 2014, from http://www.who.int/healthinfo/global_burden_disease/daly_disability_weight/en/

World Health Organization. (2014b). Metrics: Disability-Adjusted Life Year (DALY). Retrieved October, 2014, from http://www.who.int/healthinfo/global_burden_disease/metrics_daly/en/

Zhang, H., & Pu, Z. (2010). Beating the uncertainties: ensemble forecasting and ensemble-based data assimilation in modern numerical weather prediction. *Advances in Meteorology, 2010*, Article ID 432160.