Journal of Clinical Epidemiology 155 (2023) 39−47

REVIEW ARTICLE

Mixed messages: most spinal pain and osteoarthritis observational research is unclear or misaligned

Connor Gleadhill[a,b,*], Hopin Lee[c], Steven J. Kamper[d,e], Aidan Cashin[f,g], Harrison Hansford[f,g], Adrian C. Traeger[h], Priscilla Viana Da Silva[a,b], Erin Nolan[a,i], Simon R.E. Davidson[a,b], Magdalena Wilczynska[a,b], Emma Robson[a,b], Christopher M. Williams[b,d,j]

[a]School of Medicine and Public Health, University of Newcastle, Callaghan, NSW, 2308, Australia
[b]Hunter New England Population Health, Hunter New England Local Health District, Wallsend, NSW, 2287, Australia
[c]Centre for Statistics in Medicine, Rehabilitation Research in Oxford, Nuffield Department of Orthopaedics Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK
[d]School of Health Sciences, University of Sydney, PO Box M179, Camperdown, NSW, 2050, Australia
[e]Allied Health Department, Nepean Blue Mountains Local Health District, Nepean Hospital, Penrith, NSW, 2750, Australia
[f]Centre for Pain IMPACT, Neuroscience Research Australia, Sydney, Australia
[g]School of Health Sciences, Faculty of Medicine & Health, University of New South Wales, Sydney, Australia
[h]Faculty of Medicine and Health, Institute for Musculoskeletal Health, Sydney School of Public Health, The University of Sydney, King George V Building, Camperdown, NSW, 2050, Australia
[i]Hunter Medical Research Institute, New Lambton Heights, NSW, 2305, Australia
[j]Research and Knowledge Translation Directorate, Mid North Coast Local Health District, Port Macquarie, Australia

Accepted 19 January 2023; Published online 1 February 2023

Abstract

Objectives: We assessed authors' language and methods to determine alignment between reported aims, methods, intent, and interpretations in observational studies in spinal pain or osteoarthritis.

Study Design and Setting: We searched five databases for observational studies that included people with spinal pain or osteoarthritis published in the last 5 years. We randomized 100 eligible studies, and classified study intent (aims and methods) and interpretations as causal, non-causal, unclear, or misaligned.

Results: Overall, 38% of studies were aligned regarding their intent and interpretation (either causally (22%) or non-causally (16%)). 29% of studies' aims and 29% of study methods were unclear. Intent was misaligned in 16% of studies (where aim differed to method) and 23% of studies had misaligned interpretations (where there were multiple conflicting claims). The most common kind of aim was non-causal (38%), and the most common type of method (39%), intent (38%), and interpretations (35%) was causal.

Conclusions: Misalignment and mixed messages are common in observational research of spinal pain and osteoarthritis. More than 6 in 10 observational studies may be uninterpretable, because study intent and interpretations do not align. While causal methods and intent are most common in observational research, authors commonly shroud causal intent in non-causal terminology.

Keywords: Methodology; Observational study; Causal inference; Language; Intent; Systematic review

Author contributions: Concept and design: HL, CG, CW, SK conceptualized the study, operationalized the construct of intent and developed the review protocol. Search and study randomization: CG performed the literature search with assistance from a librarian Jessica Birchall (University of Newcastle). CG deduplicated records, managed records, and performed the randomization. Title and abstract screening: CG, and AC screened titles and abstracts for eligibility. Full text screening: CG, AC, HH, and SD screened full text for eligibility. Data extraction: CG, AC, AT, CW, EN, ER, HH, MW, PVDS, SD, SK all performed data extraction. Data analysis: CG completed all data analysis. Manuscript: CG lead the manuscript writing and developed all tables and figures. AC, SK, CW assisted in early manuscript drafts. All authors contributed to final manuscript drafts. All authors approved the manuscript prior to submission.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of interests: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

* Corresponding author: School of Medicine and Public Health, University of Newcastle, University Drive, Callaghan, NSW, 2308, Australia, Tel.: +1-02-4924-6829; fax: +1-02-4924-6490.

E-mail address: connor.gleadhill@newcastle.edu.au (C. Gleadhill).

https://doi.org/10.1016/j.jclinepi.2023.01.005

What is new?

Key Findings

- Researchers' interpretations are misaligned with their intent in 62% of observational studies.
- In studies with causal methods, 56% use non-causal or unclear language when stating their aims.
- Thirty seven percent of studies with causal interpretations had either non-causal, unclear, or misaligned intent.
- In studies with causal intent, explanations for researchers' choice of method (like selection of confounders) are commonly incomplete or missing altogether.

What this adds to what was known?

- Researchers use methods that signal causal intent but shroud their intent in non-causal language.
- Researchers may overuse causal language when interpreting their results.
- Researchers commonly employ methods that signal causal intent but fail to explain the rationale behind their use.

What should change now?

- Researchers should align their stated aims, methods, and interpretations.
- Researchers should precisely express their aim in observational research.
- Researchers should unambiguously interpret their results, in alignment with the intent of their study.
- Researchers should clearly explain their rationale behind study design elements like enrolling a control group, or methods like covariate selection and confounding adjustment.
- End-users should beware that non-causal or ambiguous language is used to shroud causal intent in observational research.

1. Background

1.1. Introduction

Observational research is a common and necessary part of health research. However, observational research is often seen as wasteful and ambiguous [1,2]. Historical shortcomings of observational research have prompted calls to improve wasteful research practices [2−4]. Recommendations on reducing wasteful observational research practices includes carefully planning research questions [2−4], anchoring methods to this research question [2−6], and unambiguous reporting [5]. Despite this, emerging evidence suggests that researchers continue to use inappropriate methods to answer their question or conflate terminology when interpreting their results in observational research [7]. The end-user has difficulty understanding what exactly observational health research can tell them when study methods, or interpretations do not match the research question [6]. Misalignment between study aims, methods, and interpretations creates confusion and adds to wasteful observational research practices [6].

Ambiguity, mixed messages, and misalignment may stem from authors choosing vague language. Researchers may avoid causal language because of a persistent and misguided idea that causal effects cannot be estimated in observational research [4,6−12]. Structural factors like journal editorial policies may compound researchers' selective use of causal language [13]. Evidence suggests that readers often comprehend 'non-causal' words, like *association*, as having causal implications when combined with certain methods, for example controlling for confounders in analysis [9]. Mixed messaging between authors' explicit description and the use of methods that implicitly signal a different intent creates confusion [6,9]. The usefulness of health research relies on clarity of language and the alignment between aims, methods, and interpretations.

1.2. Review objective

The objective of this review is to assess the degree of alignment between study intent and the interpretation of results in observational studies of spinal pain or osteoarthritis.

2. Methods

2.1. Study design

This review is reported in accordance with the preferred reporting items for systematic review and meta-analysis (PRISMA) guideline [14]. The review was prospectively registered on open science framework registries (registration link: https://doi.org/10.17605/OSF.IO/CAFRG) [15].

2.2. Search and random sample selection

We developed a search strategy to identify observational and nonrandomized trials in the field of spinal pain and osteoarthritis (Online Supplement 1). We chose to include only spinal pain and osteoarthritis to ensure our review team could appropriately understand study language and methods. Our research and clinical experience is in these health conditions and other conditions may include terms and methods that we could not accurately appraise. We searched Medline (Ovid), Embase (Ovid), PsychInfo

(Ovid), CINAHL (EbscoHost), Scopus, and Web of Science [16] to identify relevant studies published in the English language from January 2017 and 13th January 2022. We limited the inclusion criteria to the last 5 years to gain a contemporary sample of the literature reflecting current research practices and reporting. We removed duplicates through a recognized method [17], and then exported the references retrieved into an Excel workbook. We randomized studies by assigning a random number to each record using the Excel RAND function, then reordered in ascending value based on this random number.

After randomization, the titles and abstracts of studies were assessed for eligibility by two reviewers (CG and AC). Disagreements were resolved through discussion. Eligible full texts were screened independently in duplicate (CG and AC, HH, and SD) to obtain a sample of 100 studies. We chose to obtain a sample of 100 studies because it is within the range of similar reviews' samples [9—12] and to make data extraction feasible for our review team.

2.3. Eligibility criteria

We included observational cohort studies, cross-sectional studies, nonrandomized trials, quasi randomized studies, case control, and case series, which included human participants of any age with spinal pain (low back and neck pain, using the International Classification of Disease (ICD-10) code M54: Dorsalgia) or osteoarthritis (ICD-10 codes M15-19, M47) [18,19]. We excluded studies that used a truly random procedure to allocate exposure or treatments (as defined by PEDro criteria for randomization) [20], because these studies are least likely to present with misalignment [4]. We excluded systematic reviews and other kinds of reviews [21], opinion pieces, editorials, and qualitative studies. We excluded clinimetric studies and studies of diagnostic test accuracy because empirical evidence suggests these studies already present with a high degree of conflation (misalignment) between prediction (non-causal) and causation [7]. We excluded articles published in American Medical Association network journals because they prohibit the use of causal language for any study other than a randomized controlled trial [13]. We excluded studies including participants with rheumatoid arthritis, osteoporosis, systemic lupus erythematosus, polymyalgia rheumatica, gout, juvenile idiopathic arthritis, infectious arthritis, fracture, or cancer/metastases. We excluded studies not published in English.

2.4. Data extraction

We developed data extraction forms for reviewers to use (Online Supplement 2 and 3). To obtain a balance between interpretations, all study data were extracted in duplicate (CG and AC, AT, CW, EN, ER, MW, PVS, SD, or SK). All disagreements between reviewers were resolved through discussion. We extracted study characteristics (e.g., publication year, study design), methods (e.g., method of covariate selection, confounding mitigation), and analysis (e.g., regression, t-test), for full details, see Online Supplement 2. Reviewers extracted aims and claims verbatim and provided a rationale behind each classification.

We classified study sections into the following categories according to authors' use of language, study design elements and methods.

- Causal (cause and effect).
- Non-causal (predictive or descriptive aims and methods).
- Unclear (used when authors used ambiguous terminology, the meaning was not clear, or contained elements that may be either causal or non-causal).

Study intent and interpretations could be classified as one extra category.

- Misaligned (used when studies had different methods to their aim, or vice-versa, and when studies had multiple conflicting interpretations).

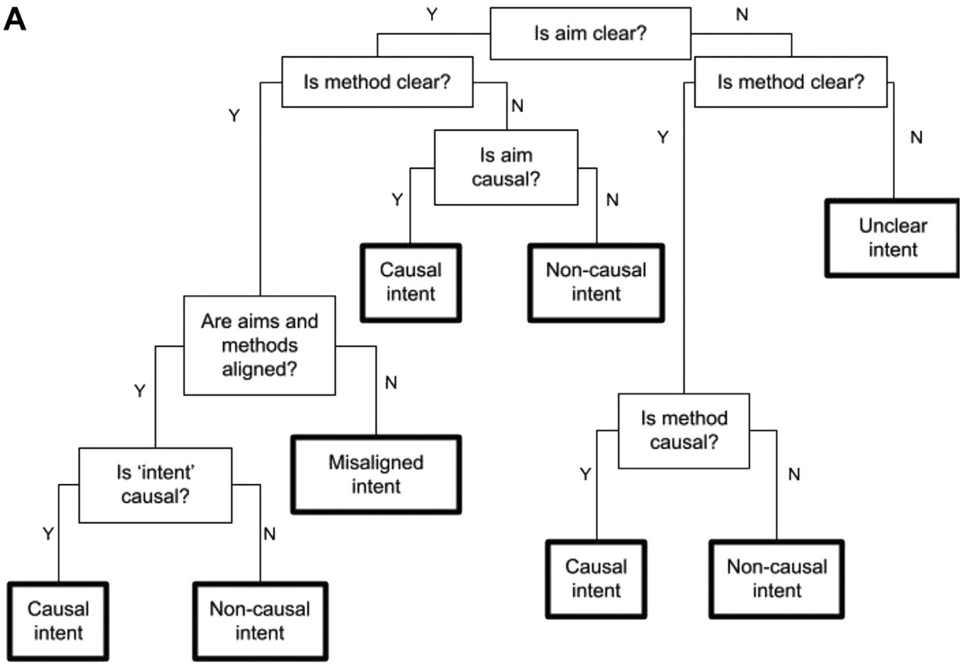
2.4.1. Step 1: classifying aims

We classified primary aims only. A priori, we established a list of causal, non-causal, and ambiguous linking words [15] (Table 1). We first placed terms in this list according to existing literature [9,22,23]. Through multiple rounds of discussion among the authors, we reallocated certain terms based on author judgment and finalized this list for protocol registration [15]. During data extraction, reviewers used this list as guidance only. Decisions on study section categorization were made through a holistic approach of assessing linking and modifying language as well as considering the sentence or clause more generally.

Table 1. Causal, non-causal, and ambiguous linking terms

Initial list of causal, non-causal, and ambiguous linking terms		
Causal the effect of the impact of the causal effect/relationship	Non-causal predict describe	Ambiguous the influence of the benefit of leads to links to results in contributes to is a determinant of is a risk factor for improves
Terms added following data extraction		
Causal mediate Affect	Non-causal relate	Ambiguous odds ratio (lower/higher odds) likelihood difference

A



B

Rule	*Condition (If)*	*Decision (then)*
1	Aim is unclear	Rely on methods to make a decision about intent
2	Method is unclear	Rely on aims to make a decision about intent
3	Aim and methods are both unclear	Unclear intent
4	Aim and methods are clear, but misaligned	Misaligned intent
5	Aim and methods are clear, and aligned	Intent is either causal or non-causal

Fig. 1. A) Flowchart outlining rules and application for making decisions about study intent. B) Decision making table (using if, then format) to operationalize study intent.

For example, a causal linking word may have been used but the modifying language implies prediction, and thus the aim was categorized as unclear. We added to this list if reviewers frequently came across terms used in consistently causal, non-causal, or ambiguous ways.

2.4.2. *Step 2: classifying methods*

We assessed study methods with the following priming questions about study design, method of confounding mitigation, and method of statistical estimation (Online Supplement 3). Reviewers were encouraged to form a general impression about whether methods were causal, non-causal, or unclear based on a combination of these factors.

2.4.3. *Step 3: classifying intent*

We first classified aims and methods as causal, non-causal, or unclear. We then operationalized study intent through the rules and flowchart in Figure 1.

2.4.4. *Step 4: classifying interpretation*

We classified interpretations by categorizing claims related to the primary aim. Claims were sentences or passages, which included authors' interpretation of a result. We classified interpretations as causal, non-causal, or unclear by assessing linking and modifying language in the same way as study aims (Table 1). Reviewers classified all interpretations authors made about the primary result, so some studies had multiple claims. We recorded when authors made statements about whether a study design could not allow causal interpretation of results, for example, "*as our study is observational in nature, causation cannot be inferred*" [23]. We also recorded whether authors provide appropriate discussion of causal assumptions for causal, unclear, or misaligned interpretations (non-causal interpretations do not need to discuss causal assumptions).

2.4.5. *Step 5: assessing alignment*

We assessed the alignment between a study's intent and interpretation. Studies that had the same intent and interpretation classification were aligned (casually or noncasually). A study with a different interpretation to its intent classification was misaligned.

2.5. Primary and secondary outcomes

Our primary outcome was the proportion of studies aligned with causal intent and causal interpretation. Our secondary outcome was the proportion of studies aligned with non-causal intent and non-causal interpretation.

2.6. Data synthesis and analysis

We summarized each categorical item with frequencies and proportions and continuous variables with median and interquartile range. We analysed data using R (*version 4.1.0*). We thematically analyzed the free text of reviewers' rationales behind classifications to understand common themes behind conflicts and help develop signaling questions (see discussion) for study method classification.

2.7. Sensitivity analyses

We undertook two post hoc sensitivity analyses. To determine the effect of changing the categorization of certain ambiguous words on the proportion of alignment, we selected all terms we listed as ambiguous (Table 1) and reclassified them as causal. To determine whether operationalizing intent caused the proportion of alignment to change, we assessed alignment of studies without including intent.

3. Results

After removing duplicates, we identified 18,066 records. We screened 600 titles and abstracts and 196 full texts for

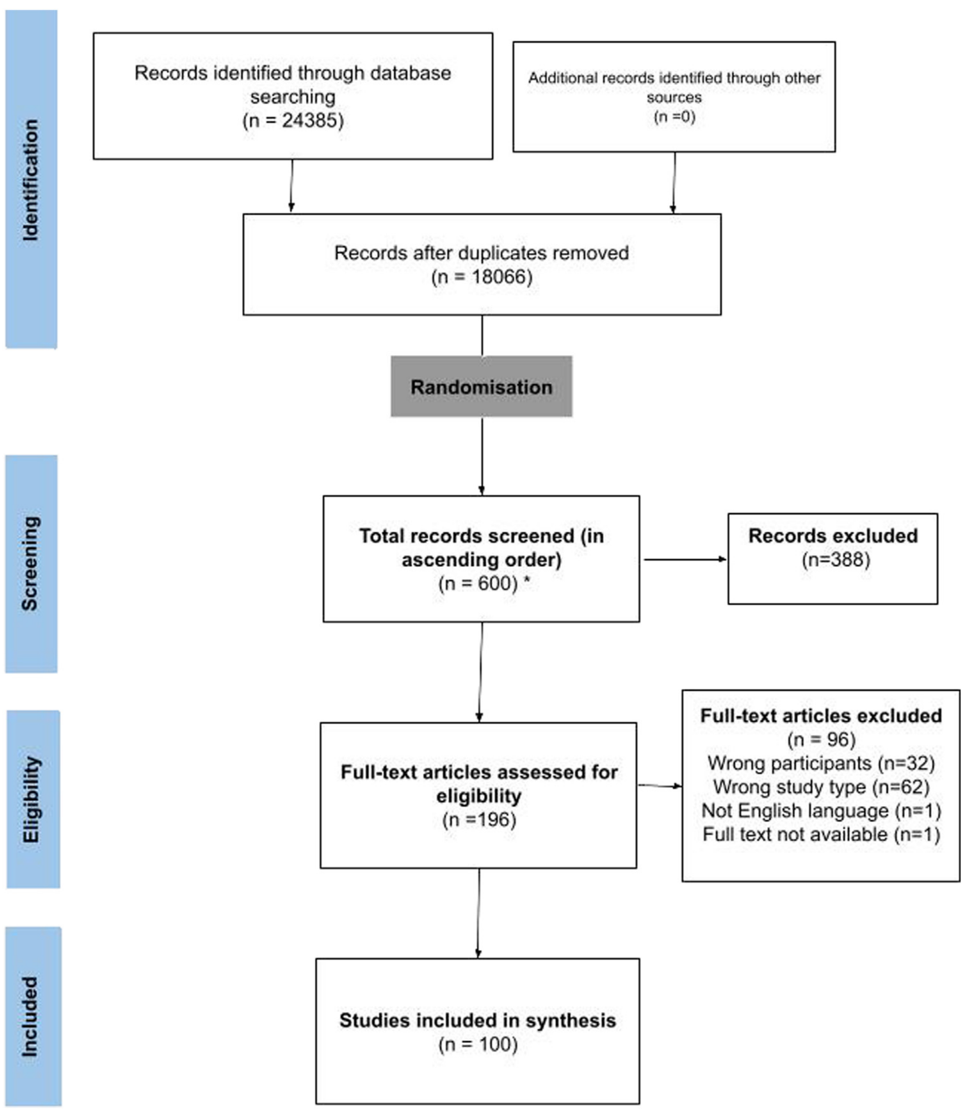


Fig. 2. PRISMA flow diagram showing record screening and inclusion.

Table 2. Summary of included studies

	n = 100
Year of publication	
2017	22
2018	19
2019	22
2020	17
2021	20
Health condition	
Spinal pain	41
Knee osteoarthritis	38
Hip osteoarthritis	11
Mixed osteoarthritis	10
Sample number	
Min	15
Max	358,121
Median (Interquartile range)	175.5 (78.0 - 461.2)
Study design	1
Nonrandomized controlled trial	4
'Quasi' randomized controlled trial	28
Observational cohort: Prospective	22
Observational cohort: Retrospective	23
Case-control	20
Cross-sectional	2
Statistical estimation method	
Assessment of correlation	69
Regression	48
Other	27
No assessment of correlation	31
Assessment of difference (t-test, one way analysis of variance ANOVA, chi-square etc.)	65
T-test	35
Chi-square	24
ANOVA	17
Only assessment of correlation (regression, Spearman, Pearson etc.)	33
Only assessment of difference	29
Both assessment of difference and correlation	36
No statistical method used (descriptive statistics)	2
Summary of methods used in studies with causal intent	*n* = 38
Selection of confounding	Causal intent
Directed acyclic graph	1 (2.6%)
Cite previous research	3 (7.9%)
Cite previous research AND stepwise selection	1 (2.6%)
Stepwise selection	4 (10.5%)
Change-in-estimate between adjusted and unadjusted model	1 (2.6%)
Not explained	14 (36.8%)
Not selected	15 (39.5%)
Baseline confounding mitigation	
By design	14 (36.8%)
By analysis	9 (23.7%)
Both	9 (23.7%)
No baseline confounding mitigation	6 (15.8%)
Time-varying confounding mitigation	
By design	13 (34.2%)
By analysis	7 (18.4%)
Both	9 (23.7%)
No time-varying confounding mitigation	9 (23.7%)
Causal inference methods	
Propensity score adjustment/ matching/weighting	2 (5.2%)
Instrumental variable methods	1 (2.6%)

inclusion. Online Supplement 4 lists all full texts screened and reasons for their exclusions. Figure 2 shows the flow of records through screening and inclusion.

3.1. Summary of included studies

Table 2 provides a summary of the included studies (further detail can be found in Online Supplement 2). Spinal pain accounted for 41% (41/100) of conditions examined in studies and knee pain 38%. The most common study design was prospective cohort (28%), followed by case-control (23%), cross-sectional (22%), and retrospective cohort (22%). The median sample size included was 175 participants (min = 15, max = 358,121; interquartile range = 78.0–461.2).

3.2. Statistical analysis, covariate selection, and confounding mitigation

The most common statistical analysis method was regression (48%), and the second most common was t-test (35%). Covariate selection was not explained in 36 studies. Covariates were not selected in 33 of studies. 11% studies cited either previous research, cited a theory or model, or used a directed acyclic graph (DAG). 68% studies attempted to mitigate baseline confounding (26% by design, 24% by analysis, and 18% by both), 60% studies attempted to mitigate time-varying confounding (25% by design, 19% by analysis, and 16% by both). Five percent studies used statistical methods that are recommended for causal inference (4% used weighting, matching, propensity score adjustment; 1% used instrumental variable methods) [24,25].

3.3. Categorization of study sections

The classification of study sections is provided in Figure 3. Non-causal aims were the most common (38%),

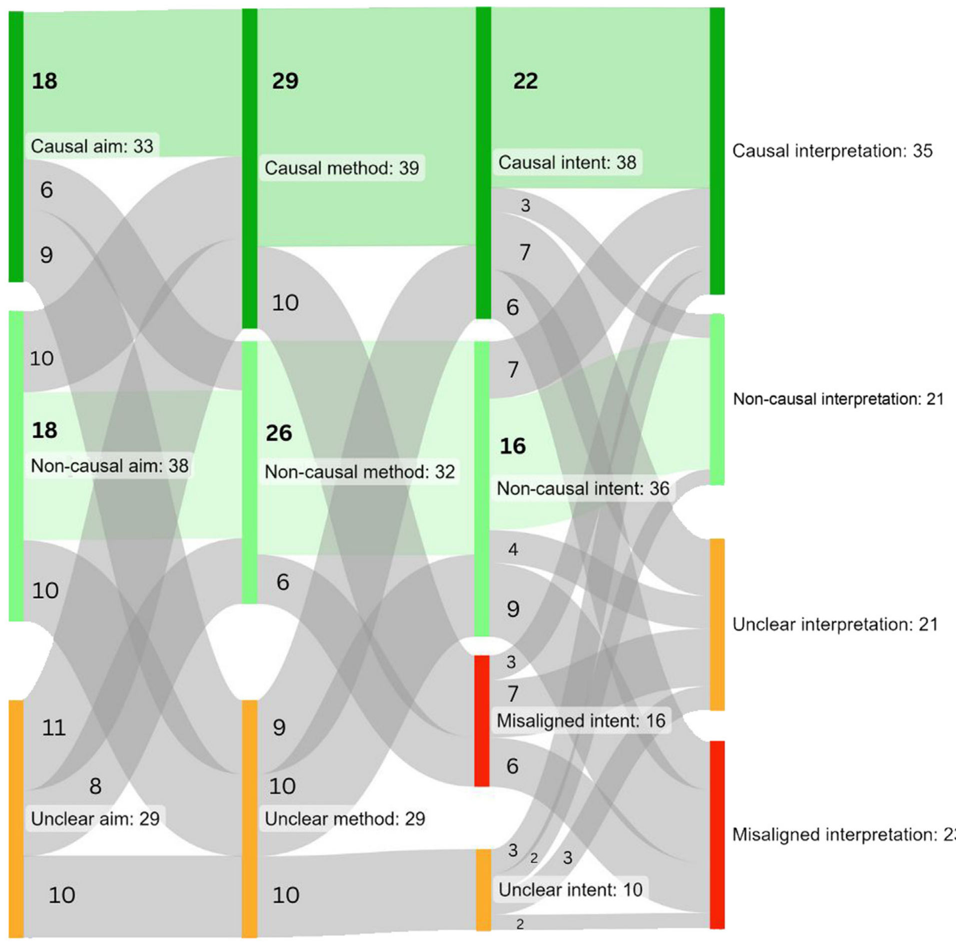


Fig. 3. Diagram demonstrating the classification of study aims, methods, intent, and interpretation (nodes). The links track the flow of studies that are aligned across aims, methods, intent, and interpretation (dark green indicates causal alignment and light green indicates non-causal alignment). All other gray links indicate that either studies are misaligned across sections or are unclear.

followed by causal aims (33%) and unclear aims ((29%). Causal methods were the most described (39/100 (39%)), followed by non-causal methods (32%), then unclear methods ((29%). Causal intent was most common ((38%), followed by non-causal ((36%), misaligned (16%), and unclear (10%). The most common kind of study interpretation was causal (35%), 23% studies had multiple claims of different kinds (causal, non-causal, or unclear) and so were misaligned. Twenty one percent of studies had non-causal interpretations, and 21% studies had unclear interpretations.

3.4. Alignment

There were 22% of studies with causal alignment (causal intent and interpretations) and 16% of studies with non-causal alignment (non-causal intent and interpretations) (Figure 3).

Seventy-nine percent of studies had causal, unclear, or misaligned interpretations, and therefore we assessed whether authors reported satisfying causal assumptions. Only 7/79 (8.9%) provided a rationale for their claim that mentioned necessary assumptions for causal inference.

Table 3. Signaling questions to assist in classifying observational study methods

Study section	Signaling causal intent
Study design	1. Do authors enroll a control group? *If yes, this signals causal intent.*
	1a. Authors enroll a control group with an attempt to match on exposure (i.e., considering exchangeability)? *If yes, causal intent may be more likely.*
Statistical analysis	2. Do Authors adjust for confounding? *If yes, this signals causal intent.*
	2 a. Do authors systematically adjust for confounding (by citing previous research, theory, or using a DAG)? *If yes, causal intent may be more likely.*

However, all 7/79 (8.9%) only briefly mentioned causal assumptions and were ambiguous (not clear in their causal reasoning). 7/79 (8.9%) studies included a statement that refutes the ability of the study design to make causal inference.

When ambiguous linking terms were reclassified as causal, the proportion of causally aligned studies did not meaningfully change (1% increase in causal alignment) (Online Supplement 1). The proportion of aligned studies when considering only aims, methods, and interpretation was 23% (14% causal alignment, and 9% non-causal alignment).

4. Discussion

The majority (62%) of observational studies involving spinal pain and osteoarthritis published in the last 5 years were either unclear or did not interpret their results in a way that matched their aims and methods (the intent of their research). Our study suggests that researchers are evasive with causal terminology in their stated aims but often use methods that signal causal intent. Authors then tend to interpret their results in a causal way.

4.1. Relation to previous research

There is an emerging evidence base indicating that authors may selectively use causal language [9–12,26]. Haber et al. (2018) [9], and Cofield et al. (2010) [26] indicate authors may overstate the causal implications of their research in interpretation sections [3,26]. In our study, causal interpretations were most common (35%). Authors may be eager to use causal language when interpreting their results. Our study adds a new finding. We found non-causal aims most common (38%), indicating that authors may underuse causal language in stating their aims.

4.2. Strengths and limitations

We included a representative sample of observational studies published in the last 5 years to assess contemporary research and reporting practice. We used a comprehensive search strategy and broad inclusion criteria to ensure all observational study types were included. In data extraction, we allowed for, and achieved a balance between, different interpretations of linking and modifying language [11]. Two authors independently extracted data and resolved disagreements through discussion. Our team included experienced researchers and statisticians, which provides some validity to our study categorization.

Including study intent may have led us to underestimate study misalignment and may not capture the authors' original intent. However, we transparently operationalized intent a priori [15], and consider it an important construct because it provides an objective way to navigate ambiguous wording and the potential variation when interpreting study aims [11]. Within our classification system, we may not have captured important differences between lack of clarity and misalignment. We classified study aims and interpretations as unclear when sentences used ambiguous language. But we also classified sentences that used both causal and non-causal language *within the same sentence*.

4.3. Implications

4.3.1. Research implications

To improve clarity, researchers should choose unambiguous language when writing research reports. Our study highlights the importance of appropriate use of language, both causal and non-causal. There may always be heterogeneity in how readers interpret language in health research [11]; however, researchers should work toward consensus on what encompasses causal language. Frameworks outlining the causal implication of terms have already been developed [4,17,18]. Journal's support of these terms may lead researchers to use less ambiguous language. When researchers are not explicit about their aims, the reader cannot make sense of why certain methods are chosen or gauge the accuracy of study results. More direct language in research questions may lead to methodological improvements [2]. For example, a precise causal question forces researchers to rigorously consider selection and adjustment for confounding. Conversely, if the aim is non-causal, consistent language in aims and interpretations may free the researcher from attempting (erroneous) adjustment for confounding and lead the reader to better understand the implications of results.

Researchers often apply causal methods, but it is difficult to interpret how or why they are used. In our study, researchers often applied methods that signaled causal intent but failed to provide sufficient rationale behind their choices. Initially, one key improvement is the consistent use of observational reporting guidelines, which recommend authors clearly define confounders [5]. Evidence suggests adherence has remained suboptimal among researchers and endorsement low among journals [27,28]. Additionally, researchers should provide clear rationale behind variable selection [5]. DAGs force researchers to make causal assumptions explicit and should be based on prior evidence and knowledge [29,30]. DAGs can assist researchers to carefully reason through their choice of confounders [29,30]. When seeking to estimate causal effects, researchers should consider and transparently report DAGs as early as possible to provide clarity in study design and statistical analyses [31].

4.3.2. End-user implications

End-users must be able to rely on observational research. Although we have highlighted interpretability issues with a contemporary sample of spinal pain and osteoarthritis studies, end-users can take two key messages from our study to ensure observational research remains useful to them. First, end-users should beware that non-causal or vague language can be used when authors intend to understand cause and effect relationships. Second, it may be unclear when methods are used to understand causal effects, because researchers often fail to provide sufficient rationale for their choice of methods. Although this is a clear area for improvement for researchers, we provide questions to assist readers to assess the causal implications of common observational methods (Table 3). In this way, end-users can navigate whether or not causal effect estimation is being signaled by the study in question.

5. Conclusions

In contemporary observational research, most of the time authors' interpretation of results does not match the intent of their study. Causal methods and interpretations are most common in observational research, but authors may avoid being explicit and shroud causal intent in non-causal terminology. Misalignment and mixed messaging is common, which confounds the reader and undermines the usefulness of observational research.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2023.01.003.

References

[1] Harper S. A Future for observational epidemiology: clarity, credibility, transparency. Am J Epidemiol 2019;188:840–5.
[2] Smith GD, Ebrahim S. Epidemiology—is it time to call it a day? Int J Epidemiol 2001;30:1.
[3] Lawlor DA, George DS, Bruckdorfer KR, Kundu D, Shah E. Those confounded vitamins: what can we learn from the differences between observational versus randomised trial evidence? The Lancet 2004;363(9422):1724–7.
[4] Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol 2016;183:758–64.
[5] Elm EV, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ 2007;335:806–8.
[6] Hernan MA. The c-word: scientific euphemisms do not improve causal inference from observational data. Am J Public Health 2018;108:616–9.
[7] Ramspek CL, Steyerberg EW, Riley RD, et al. Prediction or causality? A scoping review of their conflation within current observational research. Eur J Epidemiol 2021;36:889–98.
[8] Hernán M, Hsu J, Healy B. A second chance to get causal inference right: a classification of data science tasks. CHANCE 2019;32(1):42–9.
[9] Haber N, Smith E, Moscoe E, et al. Causal language and strength of inference in academic and media articles shared in social media (CLAIMS): a systematic review. PLoS One 2018;13:e0196346.
[10] Olarte Parra C, Bertizzolo L, Schroter S, et al. Consistency of causal claims in observational studies: a review of papers published in a general medical journal. BMJ Open 2021;11:e043339.
[11] Haber N, Wieten S, Rohrer J, et al. Causal and associational language in observational health research: a systematic evaluation. Am J Epidemiol 2022;191:2084–97.
[12] Han MA, Leung G, Storman D, et al. Causal language use in systematic reviews of observational studies is often inconsistent with intent: a systematic survey. J Clin Epi 2022;148:65–73.
[13] Journal of American medicine network. Instructions for authors. Webpage. Available at: https://jamanetwork.com/journals/jama/pages/instructions-for-authors. Accessed February 25, 2021.
[14] Liberati A, Altman DG, Tetzlaf J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epi 2009;62(10):e1–34.
[15] Gleadhill C, Lee H, Kamper S, Williams CM. Mixed messages: An audit of alignment between study intent and interpretations. Open Science Framework. Available at https://doi.org/10.17605/OSF.IO/CAFRG. Accessed December 8, 2021.
[16] Bramer WM, Rethlefsen ML, Kleijnen J, Franco O. Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. Syst Rev 2017;6:245.
[17] Bramer WM, Giustini D, de Jonge GB, Holland L, Bekhuis T. Deduplication of database search results for systematic reviews in EndNote. J Med Libr Assoc 2016;104:240–3.
[18] World Health Organinization. ICD-10-CM Codes M54*: dorsalgia. 2021. Available at: https://www.icd10data.com/ICD10CM/Codes/M00-M99/M50-M54/M54. Accessed August 1, 2021.
[19] World Health Organinization. ICD-10-CM Codes M15-19*: osteoarthritis. 2021. Available at: https://www.icd10data.com/ICD10CM/Codes/M00-M99/M15-M19. Accessed August 1, 2021.
[20] Physiotherapy evidence database (PEDro). PEDro scale. Available at: https://pedro.org.au/english/resources/pedro-scale/. Accessed June 8, 2021.
[21] Dahabreh I, Steingrimsson JA, Robertson SE, Petio LC, Hernan M. Towards causally interpretable meta-analysis: transporting inferences from multiple randomized trials to target population. Epidemiology 2020;31:334–44.
[22] Mackenzie D, Pearl J. The book of why: the new science of cause and effect. New York: Basic Books; 2017.
[23] Thapa DK, Visentin DC, Hunt GE, Watson R, Cleary M. Being honest with causal language in writing for publication. J Adv Nurs 2020;76(6):1285–8.
[24] Charney M, Paxton E, Stradiotto R, et al. A comparison of risk of dislocation and cause-specific revision between direct anterior and posterior approach following elective cementless total hip arthroplasty. J Arthroplasty 2020;35:1651–7.
[25] Pearl J. Causal inference in statistics: an overview. Stat Surv *Statist. Surv* 2009;3:96–146.
[26] Cofield SS, Corona RV, Allison DB. Use of causal language in observational studies of obesity and nutrition. Obes Facts 2010;3:353–6.
[27] Pouwels KB, Widyakusuma NN, Groenwold RHH, Hak E. Quality of reporting of confounding remained suboptimal after the STROBE guideline. J Clin Epi 2016;69:217–24.
[28] Sharp MK, Tokalic R, Gomez G, et al. A cross-sectional bibliometric study showed suboptimal journal endorsement rates of STROBE and its extensions. J Clin Epi 2019;107:42–50.
[29] Pearl J, Verma T. The logic of representing dependencies by directed graphs. University of California. Los Angeles: Computer Science Department; 1987.
[30] Lipsky AM, Greenland S. Causal directed acyclic graphs. JAMA 2022;327:1083–4.
[31] Tennant PWG, Murray EJ, Arnold KF, et al. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. Int J Epidemiol 2021;50:620–32.