



NOVA

University of Newcastle Research Online

nova.newcastle.edu.au

Clover, Kerrie; Carter, Gregory Leigh; Mackinnon, Andrew; Adams, Catherine. "Is my patient suffering clinically significant emotional distress?: demonstration of a probabilities approach to evaluating algorithms for screening for distress" Supportive Care in Cancer Vol. 17, Issue 12, p. 1455-1462 (2009)

Available from: <http://dx.doi.org/10.1007/s00520-009-0606-6>

The final publication is available at www.springerlink.com

Accessed from: <http://hdl.handle.net/1959.13/806844>

Authors

Kerrie Clover, Research Manager¹, Conjoint Senior Lecturer²

Gregory Leigh Carter, Director¹, Conjoint Associate Professor²

Andrew Mackinnon, Professorial Fellow³

Catherine Adams, Senior Clinical Psychologist¹, Conjoint Clinical Teaching Fellow²,
Conjoint Senior Lecturer⁴

Title

Is my patient suffering clinically significant emotional distress? Demonstration of a probabilities approach to evaluating algorithms for screening for distress.

¹Psycho-Oncology Service, Calvary Mater Newcastle, #51 Locked Bag 10, Hunter Regional Mail Centre, 2310 NSW, Australia

²Centre for Brain and Mental Health, University of Newcastle, 2308 NSW, Australia

³Orygen Research Centre, University of Melbourne, Parkville, Vic, Australia

⁴School of Psychology. Faculty of Science and Information Technology, University of Newcastle, 2308 NSW, Australia

Address for Correspondence

Dr Kerrie Clover

Research Manager Psycho-Oncology Service

#51 locked Bag 10, Hunter Regional Mail Centre, 2310 NSW, Australia

Email: Kerrie.clover@newcastle.edu.au

Phone: 61 2 49 211 283

Fax: 61 2 49 211 870

ABSTRACT

GOALS OF WORK: Screening oncology patients for clinically significant emotional distress is a recommended standard of care in psycho-oncology. However, principles regarding the interpretation of screening and diagnostic tests developed in other areas of medicine have not been widely applied in psycho-oncology. This paper explores the application of the concepts of likelihood ratios and post-test probabilities to the interpretation of psychological screening instruments and demonstrates the development of an algorithm for screening for emotional distress and common psychopathology.

PATIENTS AND METHODS. 340 oncology/haematology outpatients at the Calvary Mater Newcastle, Australia completed the Distress Thermometer (DT), the PSYCH-6 subscale of the Somatic and Psychological Health Report (SPHERE-12) and the Kessler-10 scale. The Hospital Anxiety and Depression Scale (HADS) (cutoff 15+) was used as the gold standard.

MAIN RESULTS: Likelihood ratios showed that a score over threshold on the DT was 2.77 times more likely in patients who were cases on the HADS. These patients had a 53% post test probability of being cases on the HADS, compared with the pretest probability of 29%. Adding either the PSYCH-6 (3+) or the Kessler-10 (22+) to the DT (4+) significantly increased this post-test probability to 94% and 92%, respectively. The significance of these improvements was confirmed by logistic regression analysis.

CONCLUSIONS: This study demonstrated the application of probability statistics to develop an algorithm for screening for distress in oncology patients. In our sample a two-stage screening algorithm improved appreciably on the performance of the Distress Thermometer alone to identify distressed patients. Sequential administration

of a very brief instrument followed by selective use of a longer inventory may save time and increase acceptability.

KEYWORDS

Cancer, oncology, screening, emotional distress, depression, anxiety

INTRODUCTION

Distress in oncology patients has been defined in different ways. Generalised distress has been defined as “..an unpleasant emotional experience of a psychological ... social and/or spiritual nature that may interfere with the ability to cope effectively with cancer, its physical symptoms and its treatment” [1], whilst emotional distress refers more specifically to anxiety, depression and adjustment disorders related to experience of cancer [2].

Emotional distress is common in oncology patients and its negative consequences are well documented [3]. However, emotional distress frequently goes unrecognised and untreated [4]. A formalised program of screening has been recommended to improve detection and subsequent management of distress [4]. A variety of approaches to screening have been trialed [see for example 5, 6, 7] however the optimum method has not been established.

Screening needs to be achieved using the minimum number of items to improve acceptability to patients and clinicians [8] while providing accurate results which are meaningful to both parties. A recent review of the psycho-oncology literature on ultra-short (fewer than five items) scales to detect psychiatric conditions recommended a two step approach to screening [9]. This review proposed that an ultra-short scale be used to rule out a psychiatric diagnosis and recommended a second step in patients who screened positive (over threshold) on the ultra-short scale.

A two step approach was used by Cull et al [6] to detect oncology patients with clinically significant levels of emotional distress. They recommended initial screening

with the five item Mental Health Inventory (MHI-5) followed by the 14 item Hospital Anxiety and Depression Scale for those patients who scored eleven or more on the MHI-5. Cull et al constructed a decision-tree to summarise their screening algorithm. Adding the HADS to the MHI increased the proportion of patients likely to be clinically significant cases on the Present State Exam from .35 to .47. Fann et al [10] have implemented a two-stage approach to screening for depression in cancer patients using the Patient Health Questionnaire-9 (PHQ-9), in which patients who respond positively to either of two 'cardinal' questions (anhedonia or depressed mood) go on to complete the full nine items while patients who answer no to both items receive no further questions. Fann et al [10] did not use statistical modeling to describe their algorithm.

Most approaches to validating instruments for distress screening in the psycho-oncology literature report sensitivity (the proportion of patients classified as ill by the tool that is truly ill according to a gold standard (the true positives)) and specificity (the proportion of patients that is classified as well by the test who truly are well according to a gold standard (the true negatives) [10]). However, sensitivity (SE) and specificity (SP) have some limitations in clinical application, since they are population measures, that is "they summarise the characteristics of the test over a population" [12, p111] rather than guide the interpretation of results for individual patients.

Two measures which are more clinically useful for the care of individual patients and which may be more readily interpreted by clinicians and understood by patients are the likelihood ratio and post-test probabilities. The likelihood ratio combines SE and

SP to estimate a patient's likelihood of having disease given a test result which is positive (positive likelihood ratio) or negative (negative likelihood ratio) [12]. For example, a positive result on a test with a positive likelihood ratio of five indicates the likelihood of disease is increased by a factor of five. A negative result on a test with a negative likelihood ratio of 0.1 indicates the likelihood of disease is decreased by a factor of ten. Tests with a likelihood ratio positive of greater than 10 or a likelihood ratio negative of less than 0.1 are considered highly useful in deciding whether a patient has a given condition [13].

The likelihood ratio can be converted to a probability of disease, which is easily interpreted by both patients and clinicians [14]. The probability of disease following a positive test is termed the post-test probability positive and the probability of disease following a negative test is termed the post-test probability negative. For example, a positive result on a test with a post-test probability positive of 77% would indicate a 77% probability of having emotional distress. A negative result on a test with a post-test probability negative of 2% would indicate a 2% probability having emotional distress according to a gold standard.

Calculating the post-test probability has been described by Jaeschke et al [13] as “not too difficult... but tedious and off-putting” (p354). It involves “converting pretest probability to odds, multiplying the result by the likelihood ratio, and converting the consequent posttest odds into a posttest probability” [13, p354]. An alternative method to estimate post-test probability, which is suitable for use in clinical practice, is to use a nomogram [15].

To use the nomogram requires estimation of the pre-test probability and calculation of the likelihood ratio. Pre-test probability is the probability that a patient has a disease, without reference to any tests. This will usually be the prevalence of the disorder in a given population. For example, if the prevalence of distress in a population of outpatients with mixed cancer types is 25%, then a clinician could reasonably estimate that an individual patient had a 25% probability of distress. That is, the pretest probability of distress is 25%. Pre-test probability usually must be informally estimated by the clinician [12].

Another advantage of post-test probabilities is that they can be used to integrate the results of multiple tests. To combine the results of two tests the post-test probability following the first test becomes the pre-test probability for the second test. It is important to note that combining the results of multiple tests in this way requires that the tests be independent [16]. Tests are independent if a patient is no more likely to get a positive result on one test given a positive result on the other [16]. In practice clinical tests are rarely independent [17]. Psychological tests in particular are likely to be highly correlated. Although the combination of tests can be powerful [17], violation of this assumption can result in a less accurate estimate of post-test probability than using the best single indicator [16]. Multiple logistic regression analysis should be used to examine whether integrating the results of more than one test leads to improved accuracy despite lack of independence between tests [18].

Aim

The aim of this paper was to explore the use of likelihood ratios and post-test probabilities in the detection of emotional distress among oncology outpatients. The

statistical methods described in this paper are well established (see for example 19, 20, 21) however, we are not aware of any other paper which has applied this method to detecting cases of emotional distress among oncology outpatients.

The specific aims of the study were to

- i. Apply the concepts of likelihood ratios and post-test probabilities to the evaluation of individual tests for emotional distress
- ii. Demonstrate the use of post-test probabilities to evaluate the performance of a two-stage (sequential) scoring algorithm to screen for emotional distress
- iii. Explore the effect of varying the thresholds for defining cases on the performance of a sequential screening algorithm

PATIENTS AND METHODS

Setting

The study was conducted at the Calvary Mater Newcastle hospital, NSW, Australia during an eight week period (April - May 2005). This hospital provides comprehensive secondary and tertiary treatment services for adults with malignant disease in Medical Oncology, Surgical Oncology, Radiation Oncology and Haematology. The project was approved by the Hunter Area Health Service Human Research Ethics Committee (Protocol Number 02/12/11/3.25).

Sample

All patients attending non-surgical Oncology and Haematology outpatient clinics were eligible for the study with the exclusion of patients attending for the first time.

Measures

Hospital Anxiety and Depression Scale (HADS): The HADS was chosen as the comparator since it is widely used in oncology [6, 9] and considerable validation work has been undertaken [22, 23]. The HADS is a 14 item scale composed of two subscales containing seven questions on non-somatic anxiety symptoms and seven questions on non-somatic symptoms of depression over the past week [24]. Each item is scored between zero and three with a maximum possible total score of 42 and a minimum of zero. A number of cut off points have been proposed for the HADS [22]. For this study a person with a total score of 15 or more (15+) on the HADS was considered a case of clinically significant emotional distress [25]. This cut off point was found to have sensitivity of 80% and specificity of 76% for a diagnosis of either generalised anxiety disorder or major depressive illness using a structured clinical interview (the Psychiatric Assessment Schedule) in a sample of over 500 English cancer patients [25].

Distress Thermometer: The Distress Thermometer (DT) has been widely evaluated for use in oncology populations [3]. It is a single item visual analogue scale on which patients indicate the level of distress they have been feeling over the past week from a minimum of zero to a maximum of ten [1]. The recommended cut off point on the DT for clinically significant emotional distress is a score of four or more [1]. A multi-centre evaluation with 380 patients [5] found that at this cut off point the DT had SE of 0.77 and SP of 0.68 in detecting cases as defined by the HADS. The Distress Thermometer is available free of charge.

Somatic and Psychological HEalth REport (SPHERE-12): The SPHERE-12 has been validated in a large scale (n=46,515 patients) Australian survey of general practice patients [26]. It is available free of charge. The “PSYCH” is a six item sub-scale of the SPHERE-12 and measures aspects of anxiety and depression [26]. It is based on the General Health Questionnaire and was initially developed for use in general practice. Each item is scored on a three point scale between zero and two, with a maximum possible score of 12 and a minimum of zero. The cut off point for PSYCH-6 caseness recommended by the SPHERE-12’s developers is a score of two or more (2+). However, we have previously established that a score of three or more (3+) on the PSYCH-6 sub scale had greater concordance with HADS caseness among oncology outpatients at our facility (unpublished data). Thus we used score of three or more (3+) to designate caseness on the PSYCH. The timeframe was “over the past few weeks”.

Kessler-10: The Kessler-10 (K-10) has been widely used in Australia and is the standard measure of psychological distress used in population surveys by the Australian Bureau of Statistics [27]. The K-10 is a ten-item scale which measures nonspecific psychological distress, first developed by Kessler & Mroczek in 1992-1994 [28]. The items encompass symptoms of anxiety and depression and are scored on a five point scale from “None of the time” to “All of the time”. In this study the minimum score of 10 represented better functioning and the maximum score of 50 represented poorer functioning. We used the Australian version of the K-10 [27] and the timeframe was the past four weeks. A cutoff score of 22 and above (22+) was used to define caseness [28]. The K-10 is available free of charge.

Procedure

Patient recruitment followed a two-step procedure approved by the local ethics committee. Reception staff offered patients a slip of paper containing study information. Patients who indicated an interest in participating met with research staff and were given a written information sheet. Patients were then asked for written consent and completed the measures on a computer. Clinical information was extracted from participants' medical records with their written consent by a qualified medical practitioner: cancer type, stage, and current (at the time of assessment) treatment with chemotherapy, radiotherapy or other therapy (other medications, haematological support and other treatments). Reasons for non-participation could not be formally collected as the Ethics Committee required that we did not approach patients directly until they had indicated a willingness to talk to research staff.

Statistical Analysis

SPSS for Windows Version 14 was used to calculate frequencies and the cross-tabulations (two by two tables) used in the analyses. Pretest probability was defined as the prevalence of cases of emotional distress as determined by a score of 15 or more (15+) on the HADS and the HADS was used as the gold standard. The SE and SP of individual tests (DT, PSYCH, K-10) were calculated using the DAG-Stat program [11] in order to provide comparison with other papers. The likelihood ratio positive, likelihood ratio negative, post-test probability positive and post-test probability negative for the individual tests were calculated using Microsoft Excel spreadsheets devised by KC, AM and Cynthia Millar.

The likelihood ratio of a positive test was calculated by dividing the SE of a test by one minus the SP [29]. The likelihood ratio of a negative test was calculated by dividing one minus the SE by the SP [29]. Formulae given in relevant texts [for example 30] were used for the calculation of post-test probabilities.

Combining test results to develop a scoring algorithm

To explore the benefits of adding each of the tests to the DT, the post-test probability positive and post-test probability negative for each of two combinations of tests were calculated. To model sequential administration of tests, cases were classified as those with positive scores on Test 1 (the DT) *and* Test 2. Thus cases were defined as those who were cases on (i) both the DT and the PSYCH and (ii) on both the DT and the K-10. Non cases were those who scored negative on the DT and those who scored positive on the DT but negative on the second test.

To calculate post-test probability after two tests the pre to post-test formula was used iteratively. For the post-test probability positive the post-test probability after a positive score on the DT was used as the pre-test probability and the likelihood ratio positive of the second test (as a single test) was used. To calculate post-test probability negative with the sequential model (that is after a negative DT or a positive DT and negative second test), the post-test probability after a negative score on the DT was used as the pre-test probability and the LR negative of the second test (as a single test) were used.

To determine whether the models of sequential test administration significantly enhanced performance compared with the DT alone two logistic regressions were

performed, each containing the DT alone as well as the sequential variable. The two regressions were (i) the DT alone and the sequential DT and PSYCH-6 variable and (ii) the DT alone and the sequential DT and K-10 variable.

Post test probability positive and post-test probability negative were calculated at a range of cut off points in order to explore the optimum combination. Since a screening algorithm ideally has a sensitive first step followed by a specific second step, the effect of lowering the threshold on the DT and raising the threshold of the second step measures (PSYCH, K-10) was explored.

RESULTS

Sample characteristics

A total of 1707 potentially eligible patients attended clinics during the recruitment period of the study. There were 393 eligible participants who consented to commence the interview (23% response rate) and complete data were obtained for 377. Of the 377 with complete data, 340 participants had malignant disease (Table 1) and this paper is restricted to data from these participants.

The sample had a mean age of 60 years (standard deviation 12 years); range 18 to 88 years. About half (52%, n=177) of the participants were male, most (74%, n=251) were married or living as married and were not working (73%, n=250). Sixty two percent (n=214) of participants were currently undergoing some kind of treatment: chemotherapy for 36% (n=122); radiotherapy for 24% (n=83) and 24% (n=83) 'other' therapy. Patients could be receiving more than one type of therapy. The two main

diagnosis groups were breast cancer 24% (n=80) and haematological malignancy 24% (n=82) (Table 1).

Caseness at published cut off points

Twenty nine percent of the sample scored as cases on the HADS (15+). Thus the pre-test probability used in the calculations for individual tests was 29%. Caseness on the other measures was 29% on the DT, 27% on the PSYCH and 26% on the Kessler-10.

Sensitivity and specificity

The SE of the tests to HADS caseness ranged from .72 for the K-10 to .86 for the DT. Conversely, SP ranged from .69 for the DT to .94 with the PSYCH-6 (Table 2).

Likelihood ratios

The likelihood ratio of a positive score on the DT was 2.77 indicating that a positive score on the DT occurred 2.77 times more frequently among HADS cases than among non-cases. The likelihood ratio of a negative score on the DT was 0.20 indicating that a negative score on the DT occurred 0.20 times as often among HADS cases compared with non-cases. The PSYCH-6 and K-10 had high likelihood ratios for positive scores (13.28 and 10.79, respectively). The likelihood ratios for a negative score on the PSYCH-6 and K-10 were slightly higher than for the DT (0.24 and 0.30).

Post-test probability

The post-test probability of HADS caseness following a positive score on the DT was .53, indicating that a person who scored positive on the DT had a 53% chance of being a case on the HADS (Table 2). The post-test probability of HADS caseness

after a negative score was .08, indicating that a person who scored negative on the DT had an 8% chance of being a case on the HADS. The post-test probability of HADS caseness for a positive score was over 80% for the PSYCH-6 and the K-10 (Table 2). The post-test probability of HADS caseness for a negative score was similar for the PSYCH-6 and K-10 (9% and 11% respectively) (Table 2).

Combining test results to develop a screening algorithm

The model of sequential administration, using the DT followed by a second test resulted in higher post-test probability positive and lower post-test probability negative than obtained with the single item DT.

Using the formula iteratively, at the cut off points shown in Table 2, following the DT (4+) by the PSYCH-6 (3+) increased the post-test probability of a positive test to 94% from 53% and decreased the post-test probability of a negative test to 2% from 8%. Similarly, when the DT (4+) was followed by the K-10 (22+), post-test probability positive increased to 92% (from 53%) and post-test probability negative decreased to 2% (from 8%).

Logistic regression analyses confirmed that the model of sequential administration of the DT followed by the PSYCH-6 had significantly improved performance compared with the DT as a single test (change in -2 log likelihood = 81.00, df=1, p<.001).

Similarly the model of the DT followed by the K-10 had significantly improved performance compared with the DT as a single test (change in -2 log likelihood = 53.03, df=1, p<.001).

Effect of varying the thresholds for defining cases

The effect of lowering the DT and increasing the threshold on the PSYCH-6 and K-10 instruments is shown in Table 3. At the thresholds examined all combinations of the DT with either the PSYCH-6 or the K-10 performed within 6% of each other. Post-test probability increased from 90% to 95% for the PSYCH-6 at the highest threshold and increased from 88% to 94% for the K-10. Post-test probability negative was never more than 6%.

DISCUSSION

This paper demonstrated the application of likelihood ratios and post-test probabilities to screening for emotional distress in oncology patients and modeled the effect of sequential administration of tests. As single tests, the PSYCH-6 and the K-10 outperformed the DT in terms of likelihood ratios and the post-test probability positive of being a case on the HADS. The three scales had similar post-test probability negative.

Logistic regression analysis confirmed that adding a second measure to the DT improved the detection of HADS caseness, despite the lack of independence between tests. In our sample, patients with positive scores on the DT followed by the PSYCH-6 or the DT followed by the K-10 had greater than 90% probability of HADS caseness. Non-cases had around 2% probability of being HADS cases. The effect of lowering the threshold on the DT and raising the threshold on the PSYCH-6 or K-10 varied performance by up to 6%.

Clinical Application

In clinical practice, the sequential administration of tests would be the ideal to minimise the number of items asked and therefore improve acceptability to patients and staff. In devising a screening strategy a trade off must be made between maximizing SE and maximizing SP. The optimal choice depends on the context of screening. In terms of screening for distress it could be argued that maximum SE is required in order to offer patients the best possible care and in this context the DT alone might be chosen. However, centres implementing this strategy are likely to require considerable resources. Based on our sample 29% of patients would require some kind of follow-up and for close to half of them this would be misdirected.

Restricting screening to a single item such as the DT would save time for patients and staff, especially if screening is done repeatedly, at successive clinic visits, thereby improving acceptability. For example if a single item takes 30 seconds to complete, compared with 3 minutes for a 6 item scale, then for every 1,000 occasions of screening a one item scale saves around 40 hours. Compared with a 14-item scale like the HADS, a one item scale would save staff and patients over 100 hours per 1,000 occasions of screening. In our regional hospital where we screen approximately one third of outpatients using a sequential model (DT and the PSYCH), described in more detail below, we completed over 3000 occasions of screening in the first six months of operation. Based on 71% scoring under threshold on the DT, screening has taken three minutes less time on 2,130 occasions, a saving of over 100 hours for patients and staff.

In centres where resources for dealing with emotional distress are limited, a strategy to maximize SP might better suit the circumstances. In this case the sequential method of screening could be employed. Resources would be closely targeted to clinical cases of depression and anxiety, however a small percentage of cases would be missed.

Using the iterative formula as applied to our sample, patients identified by screening would have a 92-94% probability of being cases and those not identified by screening would have a 2-3% probability of being cases.

We have implemented the sequential screening algorithm in our clinical service.

Patients complete a computerized touchscreen survey while waiting for their oncology or haematology outpatient appointment [31]. Patients who score under threshold on the DT complete the associated problem list but do not have written, scored feedback generated for their treating oncologist or haematologist. Patients who score over threshold on the DT are then given the six items from the PSYCH. A written, scored report is generated for all patients scoring over threshold on the DT and is placed with the notes for the treating oncologist/haematologist to use during the consultation. A recommendation to consider referral to the Psycho-Oncology service is made for patients scoring three or above on the PSYCH-6 and a recommendation to consider referral to social work is made for patients scoring two or less on the PSYCH. A recommendation to explore all items checked on the problem list from the DT is also made. Following screening clinicians are able to further explore patient concerns and apply a tailored approach to intervention; from information provision which is universally required, to therapy with specialist mental health professionals for the small but important minority of patients with significant mental health problems [32].

Patient acceptance of screening is high, with 87% of approached patients agreeing to undertake screening [Unpublished data].

Strengths and limitations of the study

In interpreting the results of this paper the limitations associated with the use of the HADS as a gold standard for emotional distress must be considered. Although widely used the HADS, like any relatively brief self-report emotional distress measure, has limitations. Therefore the values resulting from our analysis should not be considered definitive measures of the performance of the other tools. A more accurate estimate of the performance of the other tools would be obtained by comparison with a structured interview or clinician's diagnosis as the gold standard. However, the purpose of the paper was to demonstrate the application of these statistical methods to detection of emotional distress in oncology patients and this demonstration is not affected by the quality of the gold standard.

A second point to note is that while SE, SP and likelihood ratios are unaffected by the prevalence of disease (that is, pre-test probability), post-test probabilities are affected by prevalence. Thus the performance of these indices will need to be reestablished for new patient populations since prevalence may vary across populations [33]. Further to this, the 23% response rate to the survey would affect how well our results might generalize to other centres in other countries and cultures. Consequently our data should not be considered to provide the definitive validation of the tests that were

studied. However, the statistical methods described are not affected by the external validity of our results.

A final point to reiterate is that the simple combination of tests requires that the tests be independent [16]. Since psychological tests are likely to be highly correlated, multiple logistic regression analysis should be used, as in this study, to examine whether combining tests leads to improved accuracy despite lack of independence between tests. Additionally, despite the confirmation derived from the logistic regression, the lack of independence between tests means that the post-test probabilities reported for the combined tests represent an over-estimation of the accuracy of the algorithm.

CONCLUSION

Our clinical aim has been to minimize the burden on patients and staff while maximizing screening effectiveness. The algorithm we have developed reduces the number of items completed by each patient to the minimum necessary. Since the majority (71% in this study) of patients score below threshold on the DT the burden of screening for patients and staff alike is reduced.

The application of the statistical methods described in this paper to the detection of emotional distress among oncology outpatients provides a useful basis for evaluating screening algorithms for referral to clinical services such as Psycho-Oncology.

Important future steps in this process are the use of an improved “gold standard” such as a structured clinical interview and calculation of the statistics for different patient populations in other centres.

ACKNOWLEDGEMENTS

The authors wish to thank Mr Stuart Allen for devising the computerized versions of the questionnaires and data support, Dr Heidi Reichler for extracting clinical information, Ms Cynthia Millar for assistance with the statistical analysis, staff and volunteers in the outpatients' department for their support of the project and the research assistants Ms Danielle Anthony, Ms Neta Moses and Mr Jason O'Connor.

REFERENCES

- [1] National Comprehensive Cancer Network, 2004. Distress management guidelines. Jenkintown, Pennsylvania: National Comprehensive Cancer Network.
- [2] Carlson LE, Bultz BD. Cancer distress screening. Needs, models, and methods. *J Psychosom Res.* 2003;55:403-9.
- [3] Institute of Medicine (IOM) (2008). Cancer care for the whole patient: Meeting psychosocial health needs. The National Academies Press: Washington, DC:
- [4] Holland JC. An algorithm for rapid assessment and referral of distressed patients. American Society of Clinical Oncology 2000 Educational Book. Thirty-sixth Annual Meeting May 19-23, 2000. New Orleans, LA.
- [5] Jacobsen PB, Donovan KA, Trask PC, Fleishman SB, Zabora J, Baker F, Holland JC Screening for psychologic distress in ambulatory cancer patients: A multicenter evaluation of the Distress Thermometer. *Cancer* 2005;103:1494-502.
- [6] Cull A, Gould A, House A, Smith A, Strong V, Velikova G, Wright P, Selby P. Validating automated screening for psychological distress by means of computer touchscreens for use in routine oncology practice. *British Journal of Cancer* 2001;85(12):1842-1849.
- [7] Gil F, Grassi L, Travado L, Tomamichel M, Gonzalez JR. Southern European Psycho-Oncology Study Group. Use of distress and depression thermometers to measure psychosocial morbidity among southern European cancer patients. *Supportive Care in Cancer.* 2005;13:600-6.

- [8] Mitchell AJ, Kaar S, Coggan C, Herdman J. Acceptability of common screening methods used to detect distress and related mood disorders – preferences of cancer specialists and non-specialists. *Psycho-Oncology* 2008;17:226-36.
- [9] Mitchell, AJ. Pooled results from 38 analyses of the accuracy of Distress Thermometer and other ultra-short methods of detecting cancer-related mood disorders. *J Clin Oncol* 2007;25: 4670-4681.
- [10] Fann JR, Berry DL, Wolpin S, Austin-Seymour M, Bush N, Halpenny B, Lober WB, McCorkle R. Depression screening using the Patient Health Questionnaire-9 administered on a touch screen computer. *Psycho-Oncology* 2009;18:14-22.
- [11] Mackinnon, A. A spreadsheet for the calculation of comprehensive statistics for the assessment of diagnostic tests and inter-rater agreement. *Computers in Biology and Medicine* 2000;30:127-134.
- [12] Attia, J. Moving beyond sensitivity and specificity: using likelihood ratios to help interpret diagnostic tests. *Australian Prescriber* 2003;26:111-113.
- [13] Jaeschke R, Guyatt GH, Barratt A, Walter S, Cook D, McAlister F, Attia J. Measures of association. *In* Guyatt G, Rennie D, editors. *Users' guide to the medical literature: Evidence-based clinical practice*. Chicago: American Medical Association Press, 2002: 351-368.
- [14] Panzer RJ, Black ER, Griner PF. Interpretation of diagnostic tests and strategies for their use in quantitative decision making. *In* *Diagnostic Strategies for Common Medical Problems*. Black ER, Bordley DR, Tape TG, Panzer RJ (Eds). 1999 p31-36. American College of Physicians – American Society of Internal Medicine, Philadelphia.

- [15] Fagan, TJ. Nomogram for Bayes's theorem. *N Engl J Med* 1975;293:257.
- [16] Holleman DR, Simel DL. Quantitative assessments from the clinical examination. How should clinicians integrate the numerous results? *J Gen Intern Med* 1997;12:165-171.
- [17] Gill CJ, Sabin L, Schmid CH. Why clinicians are natural Bayesians. *BMJ* 2005;330:1080-1083.
- [18] Mackinnon A, Mulligan R. Combined cognitive testing and informant report to increase accuracy in screening for dementia. *Am J Psychiatry* 1998;155:1529-1535.
- [19] 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. *JAMA* 1994;271:703-707.
- [20] Greenhalgh, T. How to read a paper: Papers that report diagnostic or screening tests. *BMJ* 1997;315:540-543.
- [21] Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. *BMJ* 2004;329:168-169.
- [22] Bjelland I, Dahl AA, Haug TT, Neckelman D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. *J Psychosom Res* 2002;52:69-77
- [23] Herrmann C International experiences with the Hospital Anxiety and Depression Scale – A review of validations data and clinical results. *J Psychosom Res* 1997;42:17 - 41
- [24] Zigmond A, Snaith R. The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavia* 1983;67:361-370.

- [25] Ibbotson T, Maguire P, Selby P, Priestman T, Wallace L. Screening for anxiety and depression in cancer patients: the effects of disease and treatment. *Eur J Cancer*. 1994;30:37-40.
- [26] Hickie IB, Davenport T, Hadzi-Pavlovic D, Koschera A, Naismith SL, Scott EM, Wilhelm KA. Development of a simple screening tool for common mental disorders in general practice. *MJA* 2001;175:S10-S17.
- [27] Andrews G, Slade T. Interpreting scores on the Kessler Psychological Distress Scale (K10). *Australian & New Zealand Journal of Public Health*. 2001;25(6):494-7.
- [28] Australian Bureau of Statistics. Information paper – Use of the Kessler Psychological Distress Scale in ABS Health Surveys. Cat No 4817.0.55.001 Australian Bureau of Statistics, Commonwealth Publishing, 2003.
- [29] Suchman AL, Dolan JG. Odds and likelihood ratios. In *Diagnostic Strategies for Common Medical Problems*. Black ER, Bordley DR, Tape TG, Panzer RJ (Eds). 1999 p31-36. American College of Physicians – American Society of Internal Medicine, Philadelphia.
- [30] Guyatt G, Rennie D, editors. *Users' guide to the medical literature: Evidence-based clinical practice*. Chicago: American Medical Association Press, 2002.
- [31] Clover K, Rogers K, Carter C, Adams C. QUICA-TOUCH: The first 12 months of screening for distress, pain and psychopathology. *Asia-Pac J Clin Oncol* 2008;4 (Suppl 2): A60
- [32] Hutchison SD, Steginga SK, Dunn J. The tiered model of psychosocial intervention in cancer: A community based approach. *Psycho-Oncology* 2006;15:541-546

- [33] Irwig L, Boussuyt P, Glasziou P, Gatsonis C, Lijmer J. Designing studies to ensure that estimates of test accuracy are transferable. *BMJ* 2002;324:669-671.

Table 1. Demographic and disease characteristics of sample (n=340)

	N	%
Marital status		
Married or defacto	251	74
Separated or divorced	39	12
Widowed	29	9
Never married	21	5
Employment status		
Pension	168	49
Full-time	58	18
Part-time	32	9
Sickness Benefit	11	3
Unemployed	5	2
Other	66	19
Diagnosis		
Haematological malignancy	82	24
Breast	80	24
Urological	51	15
Colo-rectal	38	11
Lung	25	7
Other malignancy	64	19
Stage		
Haematological		
Advancing Disease (Palliative)	5	2
Active Treatment Investigations	38	11
Stable Disease / Partial Response	19	6
Complete Remission	18	5
Non- Haematological		
Metastatic	47	14
Locally advanced (palliative)	28	7
Limited/local	104	31
None	81	24

Table 2: Indices of test performance for individual tests compared with the HADS total score using a cutoff of 15+

Performance index	Test		
	DT Cutoff 4+	PSYCH-6 Cutoff 3+	K-10 Cutoff 22+
Sensitivity (95% CI)	0.86 (.77-.92)	0.77 (.68-.84)	0.72 (.62-.80)
Specificity (95% CI)	0.69 (.62-.75)	0.94 (.90-.97)	0.93 (.89-.96)
Likelihood Ratio Positive	2.77	13.28	10.79
Likelihood Ratio Negative	0.20	0.24	0.30
*Post-test probability positive	0.53	0.84	0.82
*Post-test probability negative	0.08	0.09	0.11

DT = Distress thermometer, PSYCH-6 = PSYCH-6 subscale of the SPHERE-12,
K-10 = Kessler-10; (95% CI) = 95% Confidence Interval

† Using HADS total as gold standard. 29% prevalence of distress equals a pre-test probability of 0.29 and pre-test odds of 2.9:7.1

Table 3. Post test probabilities for sequential screening obtained by lowering the threshold on the Distress Thermometer and/or increasing the threshold on the other scales

Scale	Cut-off	Distress Thermometer Cut-off Point					
		2		3		4	
		+ve	-ve	+ve	-ve	+ve	-ve
Psych-6	3	90	1	92	1	94	2
	4	95	1	95	1	97	3
	5	95	1	96	2	97	4
K-10	22	88	1	90	1	92	2
	23	89	1	91	1	93	3
	24	91	1	93	2	94	4

+ve = Post test probability positive, -ve = Post test probability negative

PSYCH-6 = PSYCH-6 subscale of the SPHERE-12, K-10 = Kessler-10;