

# Validation of an item bank for detecting and assessing psychological distress in cancer patients

Adam B. Smith<sup>1\*</sup>, Robert Rush<sup>2</sup>, Penny Wright<sup>3</sup>, Dan Stark<sup>3</sup>, Galina Velikova<sup>3</sup> and Michael Sharpe<sup>4</sup>

<sup>1</sup>Centre for Health & Social Care, University of Leeds, UK

<sup>2</sup>Centre for Integrated Health Research, Queen Margaret University, Edinburgh, UK

<sup>3</sup>Cancer Research UK—Clinical Centre, St. James's University Hospital, Leeds, UK

<sup>4</sup>School of Molecular & Clinical Medicine, University of Edinburgh, Edinburgh, UK

\*Correspondence to: Centre for Health & Social Care, University of Leeds, Charles Thackrah Building, 101 Clarendon Road, Leeds LS2 9JL, UK. E-mail: a.b.smith@leeds.ac.uk

## Abstract

**Objective:** To validate an item bank for assessing and detecting psychological distress in cancer patients by (1) identifying whether additional items are required in the full item bank; (2) identifying any item bias in the existing item bank; (3) linking levels of distress against thresholds derived from gold-standard psychiatric interviews (PSE/SCAN/SCID).

**Method:** A Rasch analysis was conducted on a heterogeneous sample of cancer patients ( $n = 4919$ ) who had completed a combination of eight psychological distress screening instruments. A subset of patients had completed a psychiatric interview along with the HADS ( $n = 381$ ) or PHQ-9 ( $n = 440$ ). Item thresholds were plotted along the latent trait. Furthermore, items were assessed for differential item functioning (DIF) by age and gender. Finally, optimum thresholds were derived for the HADS and PHQ-9 and plotted along the latent trait distribution for the entire item bank.

**Result:** Item thresholds exceeded the range of person measures, although a gap was still present along the latent trait. No DIF was observed for either age or gender. Putative cut-offs were derived for the item bank detecting moderate to severe levels of psychological distress.

**Conclusion:** The item bank covers the majority of levels of emotional distress reported by cancer patients. Additionally, initial thresholds have been derived on the item bank, which correspond to a formal psychiatric assessment. Further work is required to ascertain the stability of the item bank over time and by diagnosis and stage of disease, as well as to determine additional thresholds for levels of distress.

Copyright © 2008 John Wiley & Sons, Ltd.

**Keywords:** cancer; oncology; item bank; psychiatric interview; emotional distress

Received: 14 February 2008  
Revised: 12 June 2008  
Accepted: 17 June 2008

## Background

Despite the prevalence of psychological or emotional distress in cancer patients, which ranges in estimates from 20 to 35% [1,2], and the potential impact on patients' quality of life if left untreated or not managed, figures suggest that only 10% of cancer patients are being referred for specialist help [3].

Both the National Comprehensive Cancer Network (NCCN) [4] and the National Academy of Sciences [5] in the USA and the National Institute for Health and Clinical Excellence in the UK (NICE) [6] have highlighted the importance of identifying emotional distress by screening cancer patients and tailoring management and intervention to individual patient needs and levels of distress. In the absence of a formal psychiatric assessment, which remains impractical for high throughput screening, the most commonly used method for assessing and screening levels of emotional distress remains by self-completed questionnaire. However, the diagnostic accuracy of these instruments remains modest [7–9].

There have been recent attempts to improve efficacy of screening instruments using modern psychometrics, such as Rasch models [10]. These have found that instruments may be shortened in length without significantly decreasing diagnostic accuracy. However, the ability of these adapted instruments to identify levels of distress warranting intervention remains at best modest [10]. Combining items drawn from a number of emotional distress instruments into an item bank may improve diagnostic accuracy, while at the same time minimising the number of questions patients are required to answer and consequently reducing patient burden. Item banks such as these and computer-adaptive tests (CAT), which tailor the questions presented to patients' responses, have already been successfully developed for assessing emotional distress in a psychiatric population [11,12]. In cancer patients, an initial bank has been developed to assist in both screening and assessing emotional distress [13]. However, the results demonstrated that there was limited overlap between the severity of distress targeted by the

items (item locations ranged from  $-3.5$  to  $+2$ ) and distress reported by patients (person locations ranged from  $-6$  to  $+3$ ). This may reflect perhaps the fact that most instruments had been designed for screening, rather than assessment, in predominantly psychiatric populations and suggesting that additional items may be required to assess lower levels of emotional distress.

There have also been recent attempts to apply Rasch models to the standardised psychiatric interview schedule for major depression (MD) [14]. A modified SCID interview [15] was used on a large sample twins from the Virginia Twin Registry ( $n = 2163$ ). Participants were asked to report whether they had experienced any of the 14 disaggregated DSM-III-R criteria for MD [14]. The Rasch model was used to derive liability thresholds (the point at which there is a 50% probability of a given diagnostic category being endorsed) for the 10 symptom criteria for MD. The results demonstrated an uneven spacing between liability thresholds where 'depressed mood' was easiest to endorse ( $-1.8$  logits), and 'suicidal ideation' at the other end of the latent trait ( $2.5$  logits) was hardest to endorse, suggesting a tentative link between the latent trait, as measured by the Rasch model and that derived from a formal psychiatric interview.

The overall aim of this study therefore was to further develop an existing item bank designed for assessing emotional distress in cancer patients. The specific aims were (1) identifying whether additional items are required in the full item bank by investigating the distribution of thresholds between response categories across levels of distress; (2) identifying any item bias in the existing item bank by determining whether estimates of item location are constant across sub-groups of patients (age and gender); and (3) linking levels of distress (person measures or scores along the latent trait) against thresholds derived from gold-standard psychiatric interviews (PSE/SCAN/SCID).

## Methods

### Patients and item bank

A total of 4919 cancer patients had completed combinations of eight questionnaires (Hospital Anxiety & Depression Scale; the Emotional Functioning Scale from the EORTC QLQ-C30; Mental Health Inventory-5; Beck Depression Inventory; Spielberger State-Trait Anxiety Inventory; General Health Questionnaire-12; Patient Health Questionnaire-9; the Emotional Well-being Scale from the FACT-G).

All patients had completed at least the HADS and one other instrument. Age and gender was available for 4097 of these patients. Patient details, as well as the methodology and development of the

initial item bank have been previously described in detail [13]. However, in summary, both item fit and dimensionality of the initial 83 items in the item bank have been assessed. This process led to 20 items being removed due to misfit, leaving a final bank of 63 items (infit mean squares  $< 1.3$ ). Furthermore, no additional factors were evident in the set of items.

Two subsets of these patients had received a formal psychiatric interview consisting of either the Present State Examination (PSE)/Schedule for Clinical Neuropsychiatry (SCAN) or the depression subsection of the Structural Clinical Interview for DSM IV (SCID) [15–18]. One subset of patients ( $n = 381$ ) completed the HADS followed by a SCAN interview approximately within (but no later than) 2 weeks. Another group of patients ( $n = 440$ ) completed the HADS and PHQ-9 and patients scoring 15 or more on the questionnaire received the SCID interview by telephone within 2 weeks of their clinic visit. These data have been described previously [10,19].

### Rasch model

The Rating Scale Model (RSM) [20] was applied to the items in the item bank grouped by screening instrument. The RSM is part of the family of Rasch models, which map the probabilistic relationship between responses to items from instruments, such as patients' levels of emotional distress, or person measures, to the location of the items along a single latent trait. In Rasch models, the distribution of person measures (or scores) may be plotted alongside the mean item location for each item (both parameters are measured in logits or log-odds). However, a more detailed plot may also be derived by plotting person measures against location for each item threshold. These 50% cumulative probability or Rasch–Thurstone thresholds describe the intersections between adjacent response categories for each item, and represent the point or level of latent trait at which the choice between adjacent categories is equiprobable for a given item. For instance, this analysis may be used to determine the level of emotional distress at which point the probability of choosing either 'Quite a bit' and 'Very much', e.g. on the Emotional Functioning Scale of the EORTC QLQ-C30 is equivalent. Given the level of detail in these threshold plots they may therefore provide a more comprehensive overview of the distribution of items across the latent trait, i.e. levels of emotional distress and may be used to determine more accurately whether (and where) additional items may be required to supplement the item bank.

In addition to the criteria whether items fit the Rasch model and whether the instrument or set of items describe a unidimensional scale it also important to ascertain whether item location

estimates hold for different groups. Item invariance or differential item functioning (DIF) may be ascertained by estimating item parameters separately by groups of patients (e.g. age group, gender, diagnostic category) and statistically testing differences in estimated item location (paired *t*-tests).

### Analysis

*Winsteps* [21] software was used for the Rasch analysis. The item locations were anchored using previous estimates generated for the item bank [13].

### Distribution of item category thresholds across levels of psychological distress

Item thresholds were plotted against levels of psychological distress for each item from the item bank.

### Differential item functioning (DIF)

Patients were categorised into two age groups based on the median age of 56 years. In addition to this DIF was also explored by gender. The criteria used for the DIF analysis were DIF contrast  $>0.50$  [22] and  $p < 0.001$  to correct for multiple comparisons.

### Determining cut-off points for the HADS and PHQ-9 on the item bank

Optimum thresholds were derived for both the HADS and PHQ-9 by plotting ROC curves and determining sensitivity and specificity for cases of psychological distress, as identified by the PSE/SCAN (cases were identified on the basis of a CATEGO score  $\geq 5$ ) [10], and major depressive disorder (MDD) based on the SCID.

A Rasch analysis was applied to both the HADS and PHQ-9 for each full instrument and for each instrument with misfitting items removed (misfit as previously identified from the entire item bank) [13]. Person measures were generated for each version of the questionnaires, and raw score thresholds converted to logits (or log-odds) scores.

## Results

### Distribution of item category thresholds across levels of psychological distress

The item category thresholds ranged from  $-10$  (GHQ7, 'Able to enjoy') to  $+4$  (BDI6, 'Self-hatred'). The majority fell between  $-4$  and  $+4$ , representing 85% of the person measures. However, a gap was observed between  $-4$  and  $-7.5$  logits (Figure 1).

### Analysis of differential item functioning

No DIF was observed for any items by Gender. Similarly, there was no DIF for Age with the exception of a single item HADS-Anxiety1 ('I feel tense' or 'wound-up') which was easier to endorse by older patients (DIF contrast = 0.58,  $t(4096) = 11.46$ ,  $p < 0.0001$ ).

### Cut-off points for the HADS and PHQ-9 on the item bank

In the original item bank four items from the HADS ('I feel cheerful'; 'I have my interest in my appearance'; 'I feel restless as if I have to be on the move'; 'I can enjoy a good book or radio of TV programme') and a single item from the PHQ-9 misfitted ('Feeling bad about yourself, or that you are a failure, or have let yourself or your family down') [13]. The results from the ROC analysis are shown in Table 1 along with raw score and logit score threshold for each questionnaire. The thresholds identified from the ROC analysis plotted against the full item bank are included in Figure 1.

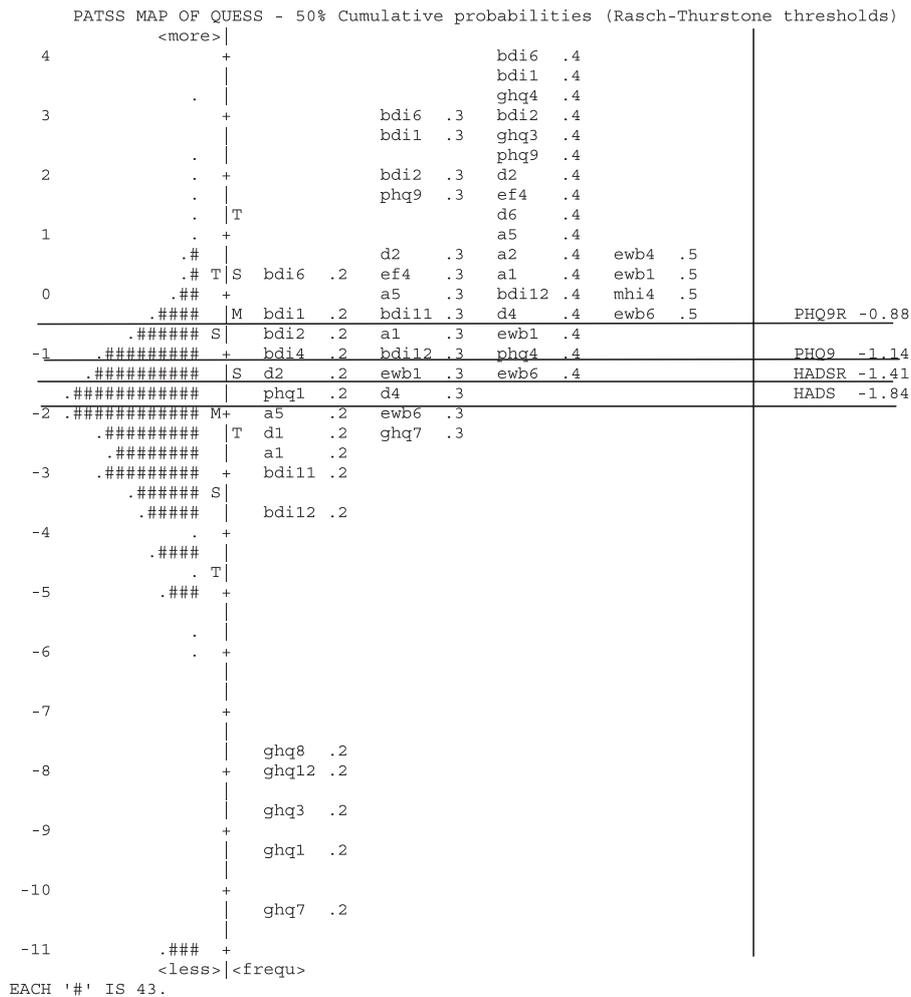
For the PSE/SCAN defined distress the logit cut-off was  $-1.38$  (raw score 12) for the 14-item HADS and  $-1.84$  (raw score of 8) for the 10-item HADS. This corresponds to 29 and 43%, respectively, of person measures from the entire item bank, i.e. 29% of patients score  $-1.38$  or higher on the item bank.

Similarly, for the SCID defined MD the figures were  $-0.88$  (raw score 11) for the PHQ-9 and  $-1.14$  (raw score 9) for the PHQ-8, corresponding to 16 and 23% of person measures, respectively. The SCID identified fewer scores as the sample used was based on patients scoring 15 or more on HADS, and therefore presumably these logit scores corresponded to high levels of distress.

## Discussion

This paper describes the further development of an item bank to assess psychological distress in cancer patients. An analysis of the distribution of the item thresholds demonstrated that levels of emotional distress between  $-10$  and  $+4$  logits were represented by the items. This range exceeds the range of person measures in the sample ( $-6$  to  $+3$  logits) and although there was a gap in the thresholds between approximately  $-4$  and  $-7$  logits, given that this represented less than 15% of person measures (since person measures beyond this point were also represented) it may therefore be concluded that no additional items are required for the item bank. In addition to this no clinically relevant differential item functioning was observed by items by either age group or gender.

The results demonstrated that items may be removed from both the HADS and PHQ-9 without



**Figure 1.** Distribution of person measures and item thresholds. The figure shows a representative sample of items from the item bank and highlights the distribution of categories across the latent trait. Note: Not all items are shown; for clarity items have been removed at each level of psychological distress

**Table 1.** Results of the ROC analysis

Questionnaire	ROC AUC <sup>a</sup>	Threshold Raw score (≥)	Threshold Logit score	Sensitivity	Specificity
HADS (14) <sup>b</sup>	0.72	12	-1.38	0.74	0.65
HADSR (10)	0.68	8	-1.84	0.78	0.61
PHQ (9)	0.78	11	-0.88	0.71	0.69
PHQR (8)	0.78	9	-1.14	0.75	0.63

<sup>a</sup>Area under the curve.  
<sup>b</sup>Number of items in each questionnaire.

affecting diagnostic accuracy adversely (see also [10]), suggesting that it is therefore possible to reduce patient burden without losing diagnostic accuracy of the instrument. In addition, the results also provide putative cut-offs for cases of psychological distress and major depression along the latent trait mapped by the item bank, suggesting that moderate levels of psychological distress may be identified between -1.8 and -1.4. It is interesting to note that the lower threshold corresponds to the liability threshold for depressed mood deter-

mined by Aggen *et al.* [14]. Similarly, cases of MD scores were suggested at scores above -0.8 on the item bank.

A number of limitations of the study should be addressed. The sample used to derive the diagnostic accuracy was not independent from the sample employed to estimate item and particularly, person parameters. Furthermore, for the PSE/SCAN interviews took place within 2 weeks of, rather than contemporaneously with the HADS assessment. Additionally, the DIF analysis was limited to

age and gender. Given that the prevalence of emotional distress may vary by cancer site [2] additional work is required to determine whether item invariance holds for different diagnoses and stages of disease. Additionally, since levels of distress may vary over time further work is required to determine whether item locations remain constant longitudinally.

The study describes the continuing validation of an item bank for assessing emotional distress in cancer patients. The use of Rasch models has the potential to improve the clinical utility of self-report questionnaires. For instance, the Rasch analysis to date has demonstrated that this item bank forms a unidimensional instrument [13]. The Rasch model predicts that within a unidimensional scale person estimates should be independent of the subset of item used. Given that tentative cut-offs were derived for emotional distress, ranging between  $-1.80$  and  $-0.80$  logits, there is also the potential to link subsets of items (either entire scales or items drawn from different scales) to this. Therefore, items drawn from the item bank could be used to create shorter, parallel forms reducing the number of items presented to patients without reducing the accuracy of the assessment.

In addition, item banks form the basis of computer adaptive tests (CAT). CAT systems have already been developed for use with psychiatric populations to identify emotional distress [11,12] and may be tailored for screening or assessment. If, for example, clinical screening is required greater numbers of items closer to a clinically relevant threshold may be presented to the patient. On the other hand the broad spectrum of the trait can be the aim if measurement of change in the whole population is required, such as before and after an intervention. The next step in the development of the item bank will be to develop CAT systems. An important corollary to this will be to continue to map the item bank, in particular levels of emotional distress to both psychiatric diagnoses of clinical anxiety and MD, as well as to the levels of intervention recommended in the NICE guidelines for assessing and managing psychological distress [6]. This will not only provide a potentially more sensitive instrument for assessing and screening for distress, but will also assist in tailoring the management of distress and associated interventions to individual patients.

## References

1. Strong V, Waters R, Hibberd C *et al.* Emotional distress in cancer patients: the Edinburgh Cancer Centre symptom study. *Br J Cancer* 2007; **96**:868–874.
2. Zabora J, BrintzenhofeSzoc K, Curbow B, Hooker C, Piantadosi S. The prevalence of psychological distress by cancer site. *Psycho-Oncology* 2001; **10**:19–28.

3. Kadan-Lottick NS, Vanderwerker LC, Block SD, Zhang B, Prigerson HG. Psychiatric disorders and mental health service use in patients with advanced cancer. *Cancer* 2005; **104**:2872–2881.
4. NCCN. *Clinical Practice Guidelines in Oncology: Distress Management*. National Comprehensive Cancer Network: Rockledge, PA, 2007.
5. Adler NE, Page AEK. *Cancer Care for the whole patient: Meeting psychosocial health needs*. The National Academies Press: Washington, DC: 2008.
6. NICE. *Guidance on Cancer Services. Improving Supportive and Palliative Care for Adults with Cancer*. National Institute for Clinical Excellence: London, 2004.
7. Hall A, A'Hern R, Fallowfield L. Are we using appropriate self-report questionnaires for detecting anxiety and depression in women with early breast cancer? *Eur J Cancer* 1999; **35**:79–85.
8. Cull A, Gould A, House A *et al.* Validating automated screening for psychological distress by means of computer touchscreens for use in routine oncology practice. *Br J Cancer* 2001; **85**:1842–1849.
9. Love AW, Kissane DW, Bloch S, Clarke D. Diagnostic efficiency of the hospital anxiety and depression scale in women with early stage breast cancer. *Aust N Z J Psychiatry* 2002; **36**:246–250.
10. Smith AB, Wright EP, Rush R, Stark DP, Velikova G, Selby PJ. Rasch analysis of the dimensional structure of the hospital anxiety and depression scale. *Psycho-Oncology* 2006; **15**:817–827.
11. Fliege H, Becker J, Walter OB, Bjorner JB, Klapp BF, Rose M. Development of a computer-adaptive test for depression (D-CAT). *Qual Life Res* 2005; **14**:2277–2291.
12. Walter OB, Becker J, Bjorner JB, Fliege H, Klapp BF, Rose M. Development and evaluation of a computer adaptive test for 'Anxiety' (Anxiety-CAT). *Qual Life Res* 2007; **16**:S143–S155.
13. Smith AB, Rush R, Velikova G *et al.* The initial development of an item bank to assess and screen for psychological distress in cancer patients. *Psycho-Oncology* 2007; **16**:724–732.
14. Aggen SH, Neale MC, Kendler KS. DSM criteria for major depression: evaluating symptom patterns using latent-trait item response models. *Psychol Med* 2005; **35**:475–487.
15. First MB, Spitzer RL, Williams JBW, Gibbon M. *Structured Clinical Interview for DSM-IV (SCID)*. American Psychiatric Association: Washington, DC, 1995.
16. World Health Organization. *Schedules for Clinical Assessment in Neuropsychiatry*. WHO—Assessment, Classification and Epidemiology, WHO. Geneva, Switzerland, 1999.
17. Wing JK, Cooper JE, Sartorius N. *The Measurement and Classification of Psychiatric Symptoms: An Instructional Manual for the Pse and Catego Program*. Cambridge University Press: London, 1974.
18. APA. *Diagnostic and Statistical Manual of Mental Disorders (3rd revised edn.)*. American Psychological Association: Washington, DC, 1987.
19. Sharpe M, Strong V, Allen K *et al.* Major depression in outpatients attending a regional cancer centre: screening and unmet treatment needs. *Br J Cancer* 2004; **90**:314–320.
20. Andrich D. A rating formulation for ordered response categories. *Psychometrika* 1978; **43**:561–573.
21. Linacre JM. *A User's Guide to Winsteps/Ministeps Rasch-Model Programs*. MESA Press: Chicago, IL, 2005.
22. Lai JS, Cella D, Chang CH, Bode RK, Heinemann AW. Item banking to improve, shorten and computerize self-reported fatigue: an illustration of steps to create a core item bank from the FACIT-Fatigue Scale. *Qual Life Res* 2003; **12**:485–501.