Screening medical patients for distress and depression: does measurement in the clinic prior to the consultation overestimate distress measured at home?

C. H. Hansen[1], J. Walker[2], P. Thekkumpurath[1], A. Kleiboer[1], C. Beale[1], A. Sawhney[1], G. Murray[3] and M. Sharpe[2]*

[1] *Psychological Medicine Research, School of Molecular and Clinical Medicine, University of Edinburgh, UK*
[2] *Psychological Medicine Research, Department of Psychiatry, University of Oxford, UK*
[3] *Centre for Population Health Sciences, University of Edinburgh, UK*

Background. Medical patients are often screened for distress in the clinic using a questionnaire such as the Hospital Anxiety and Depression Scale (HADS) while awaiting their consultation. However, might the context of the clinic artificially inflate the distress score? To address this question we aimed to determine whether those who scored high on the HADS in the clinic remained high scorers when reassessed later at home.

Method. We analysed data collected by a distress and depression screening service for cancer out-patients. All patients had completed the HADS in the clinic (on computer or on paper) prior to their consultation. For a period, patients with a high score (total of $\geqslant 15$) also completed the HADS again at home (over the telephone) 1 week later. We used these data to determine what proportion remained high scorers and the mean change in their scores. We estimated the effect of 'regression to the mean' on the observed change.

Results. Of the 218 high scorers in the clinic, most [158 (72.5%), 95% confidence interval (CI) 66.6–78.4] scored high at reassessment. The mean fall in the HADS total score was 1.74 (95% CI 1.09–2.39), much of which could be attributed to the estimated change over time (regression to the mean) rather than the context.

Conclusions. Pre-consultation distress screening in clinic is widely used. Reassuringly, it only modestly over-estimates distress measured later at home and consequently would result in a small proportion of unnecessary further assessments. We conclude it is a reasonable and convenient strategy.

Introduction

There is increasing awareness of the importance of subjective measures including quality of life in medical care. Such measure are often referred to as patient-reported outcomes or 'PROs' (Greenhalgh, 2009). Emotional distress and depression are important PROs that have a major effect on quality of life (Moussavi *et al.* 2007). Consequently, it has been recommended that medical patients, such as those with cancer (Carlson *et al.* 2012), are screened for emotional distress and depression (Pignone *et al.* 2002 ; NICE, 2009), but only if there are facilities to provide

treatment for identified cases (USPSTF, 2009). Despite an extensive literature on such screening (Carlson *et al.* 2012), there is limited information on the practicalities of carrying it out, an important aspect of which is when and where to administer the screening measures.

The most convenient and widely used strategy is to administer a questionnaire, such as the Hospital Anxiety and Depression Scale (HADS; Zigmond & Snaith, 1983), in the medical clinic, taking advantage of the time patients spend waiting to go into their consultation. The patient's questionnaire score is then used to determine whether they have a significant level of distress that requires attention and whether they need a further assessment to determine whether they have a depressive disorder.

However, there is a potential problem with this strategy ; measuring distress in the clinic prior to the consultation might result in a transient inflation

---

 * Address for correspondence : Professor M. Sharpe, Psychological Medicine Research, Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford OX3 7JX, UK.

(Email : michael.sharpe@psych.ox.ac.uk)

of the score because of the clinical context and the anticipation of the consultation. This phenomenon would be similar to that referred to as the 'white-coat effect' in the measurement of blood pressure (Gerin *et al.* 2006). If such inflation were to occur it would result in false positives in the identification of patients suffering from significant distress and would lead to more patients than necessary being given assessment interviews for depression. Such an effect would therefore be important in increasing both inconvenience to patients and the costs to clinical services.

As far as we are aware, although there are studies of the test–retest reliability of measures of quality of life and distress (Hjermstad *et al.* 1995; Bakker *et al.* 2009), the course of distress over a series of cancer consultations (van Dooren *et al.* 2005) and of the influence of the content of the consultation on distress (van Dulmen *et al.* 1995), this particular question has not been specifically addressed in the published literature.

We therefore aimed to find out whether oncology patients who were high scorers on the HADS questionnaire, completed while waiting for their cancer consultation in clinic, remained high scorers when completing a repeat HADS questionnaire a week later at home. Specifically, we aimed to determine: (*a*) what proportion of the patients who scored high (total score of $\geqslant 15$) on the HADS prior to their consultation still had a high score when reassessed at home 1 week later; and (*b*) how much the mean HADS score had changed between these two occasions and how much any fall could be accounted for by regression to the mean.

Method

To address the research question we analysed data that had been routinely collected by an established distress and depression screening service operating in multiple cancer out-patient clinics in Scotland, UK.

Routine screening procedure

The screening service was in operation in numerous clinics, each specializing in one of a variety of cancer types including breast, colorectal, gynaecological, lung and genito-urinary. All patients attending the clinics were asked to complete the HADS on touch-screen computers (or, where computers were not available, on paper) prior to their medical consultation. The results of screening were given to their cancer clinician at the time of the consultation. In addition, all patients who had scored high on the HADS in clinic were telephoned at home, approximately 1 week later, and assessed for depression using the major depression

component of the Structured Clinical Interview for DSM-IV (SCID; First *et al.* 1999).

Collection of repeat HADS scores

As part of routine clinical service data collection during March and April 2009, patients who had scored $\geqslant 15$ on the HADS in the clinic were asked to complete the HADS again at home over the telephone, immediately before they were given the routine interview to assess them for depression. We analysed these clinical data to address the research question.

Ethical approval

We obtained ethical approval from the local Research Ethics Committee to use the data in this way and also obtained each patient's permission to use their anonymized clinical data for research.

Measure

The HADS is the most extensively studied distress scale in cancer patients and is very widely used as a first stage in screening medical patients for depression (Vodermaier *et al.* 2009). The HADS asks patients how they have been feeling over the past 2 weeks. It has 14 items: seven on each of the anxiety and depression subscales. Each item is rated from 0 to 3, resulting in a total HADS score between 0 and 42, with higher scores indicating more severe symptoms (Zigmond & Snaith, 1983) A recent review concluded that the HADS was an effective measure of emotional distress but that the subscales were unable to differentiate consistently between anxiety and depression (Cosco *et al.* 2012). A total HADS score of $\geqslant 15$ has been reported to be optimal to identify cancer patients likely to have major depression on further assessment (Walker *et al.* 2007).

Analysis

We analysed these data to determine whether patients with high HADS scores measured in the clinic prior to their consultation still had high scores when measured later at home. We therefore calculated the proportion of patients who still had a high score ($\geqslant 15$) when the HADS was repeated at home. We also determined the mean change in the total HADS score between clinic and home.

Individual patient distress scores vary over time. Patients scoring high or low are likely to score closer to the mean score of all assessed patients on later reassessment, a phenomenon known as 'regression to the mean'. If all patients who completed a first HADS also completed a second HADS, we would expect the effect of these variations on the mean score of the
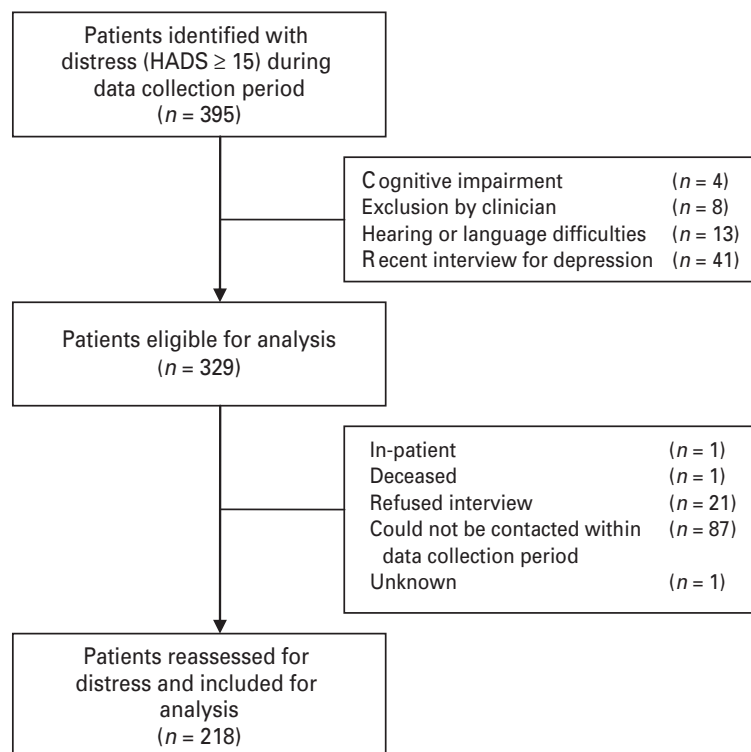
Fig. 1. Derivation of patient sample. The patients initially identified with distress (HADS ⩾15) were screened during the period from 25 February 2009 to 31 March 2009.

whole group to even out. However, as we only had follow-up data the subsample of initial high scorers we would expect the average of the reassessed scores in this subsample to be lower because of this 'regression to the mean' effect (Barnett *et al.* 2005). Therefore, to isolate the effect of the clinic from this phenomenon we estimated the size of the anticipated regression to the mean. This involved using more than 5000 HADS scores that had been collected by the screening service in similar clinics from 2007 to 2010 to obtain details of the overall distribution of HADS scores in this population. These details included the variance and covariance of repeated scores. The technical details of this approach are provided in the Appendix and described elsewhere (Das & Mulder, 1983). Finally, we conducted an exploratory analysis describing and comparing the changes in the HADS anxiety and depression subscales to determine whether these differed in the amount they changed.

Results

The service had offered screening to all patients attending the cancer clinics except for a small number (<5%) who were unable to complete questionnaires because they were too unwell or had severe cognitive or communication problems. A further 10% of patients were missed by the service, mainly because they were taken straight to their consultation before being screened, and an additional 7% refused to participate in screening.

A total of 1691 patients were screened in clinic during the period from which the data analysed were derived. Of these, 395 scored high on the HADS in clinic and 329 were listed for further assessment at home (the remainder were not listed for a variety of reasons including a recent depression assessment, cognitive or communication problems or exclusion by their clinician, usually because they were considered to be too ill). Repeat HADS were not available on 111 of these patients for several reasons, but mainly because they were not contacted by the screening service within the 1-month time window used for the analysis. The final patient sample is shown in Fig. 1. A total of 218 patients were given a repeat HADS at home by the screening service during the data collection period. This is the sample analysed.

In the analysed sample, 159 (73%) patients were female and the median age was 61 years [interquartile range (IQR) 53–70 years]. Almost all of the patients were attending follow-up appointments. The median interval between the clinic and repeat HADS assessments was 6 days (IQR 5–8 days). The 111 patients who did not have a repeat HADS at home had similar

Table 1. *Characteristics of the analysed sample compared with those in the group of eligible patients not included*

	Eligible patients included for analysis (n=218)	Eligible patients not included for analysis (n=111)	p value[a]
Age (years)			0.954
Mean (S.D.)	61.4 (11.5)	61.3 (12.2)	
Median (range)	61.4 (25.3–87.7)	62.5 (28.7 to 89.8)	
Age categories, n (%)			0.736
⩽50 years	38 (17)	23 (21)	
51–60 years	67 (31)	28 (25)	
61–70 years	66 (30)	35 (32)	
⩾71 years	47 (22)	25 (23)	
Gender, n (%)			0.396
Male	59 (27)	35 (32)	
Female	159 (73)	76 (68)	
Cancer clinic type, n (%)			0.383
Breast	83 (38)	39 (35)	
Gynaecology	33 (15)	15 (14)	
Lung	47 (22)	24 (22)	
Colorectal	18 (8)	8 (7)	
Urology	14 (6)	12 (11)	
Gastrointestinal	15 (7)	4 (4)	
Other	8 (4)	9 (8)	
Appointment type[b], n (%)			0.025
First appointment	30 (14)	6 (6)	
Return appointment	183 (86)	100 (94)	
Poor prognosis[c], n (%)			<0.001
Yes	19 (9)	25 (23)	
No	195 (91)	84 (77)	
HADS scores			0.356
Mean (S.D.)	20.1 (4.7)	20.6 (4.8)	
Median (range)	19 (15–37)	19 (15–34)	
HADS score categories, n (%)			0.604
15–19	115 (53)	59 (53)	
20–24	66 (30)	29 (26)	
⩾ 25	37 (17)	23 (21)	

HADS, Hospital Anxiety and Depression Scale; S.D., standard deviation.

[a] Age in years and HADS scores were compared using the Wilcoxon rank sum test. All other *p* values were from χ^2 tests.

[b] Appointment type was unknown for 10 patients.

[c] Poor prognosis was defined for lung (non-lung) cancer patients as a life expectancy of <3 (12) months. Prognosis was unknown for six patients.

distributions of sex, age and clinic HADS scores and attended similar types of cancer clinics. However, there were more new and good prognosis patients included in the sample reassessed. The patients' characteristics and the comparison of those with and without a HADS rated at home are shown in Table 1.

Figure 2 shows the distributions of HADS scores when patients were (*a*) assessed in clinic and (*b*) reassessed at home. Fig. 3 shows the change in HADS scores for each individual patient. As a result of the large variance in the HADS scores, there was also considerable variability in the change scores between the two assessments despite a high intra-class correlation between repeated measurements (ICC=0.83).

Almost three-quarters (72.5%; 158/218) of the initial high-scoring patients were still high scorers at reassessment [95% confidence interval (CI) 66.6–78.4]. The mean change in total HADS score was a reduction of 1.74 points (95% CI 1.09–2.39).
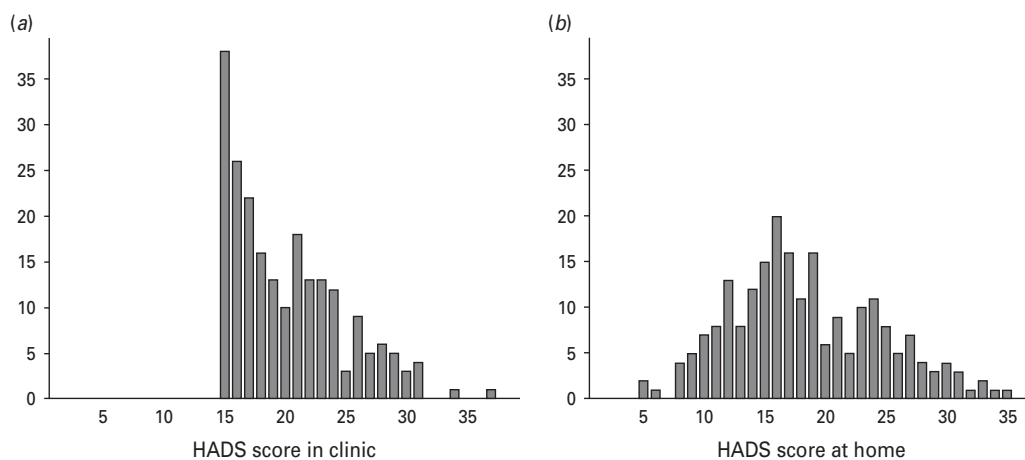
(a)

(b)

Fig. 2. Hospital Anxiety and Depression Scale (HADS) scores of patients (*n* = 218) in the study sample (*a*) when assessed in clinic and (*b*) when reassessed at home.
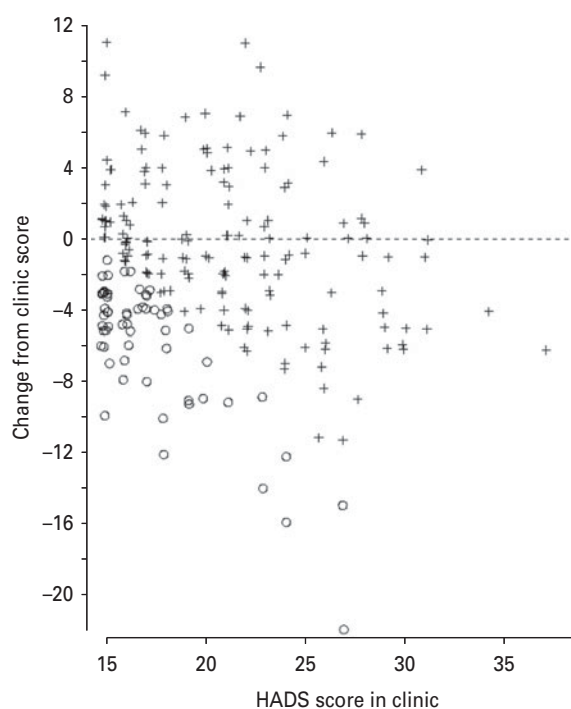

Fig. 3. Change in Hospital Anxiety and Depression Scale (HADS) total score from clinic to follow-up at home plotted against initial HADS score in clinic. Circles indicate patients whose reassessment score fell below 15. Patients plotted above the dashed line had a higher HADS score on reassessment whereas those below the line had a lower score. A degree of 'jitter' was applied to separate out overlapping data points.

Our estimate of the regression-to-mean effect was an average reduction of 1.21 points (95% CI 1.02–1.43). Hence regression to the mean potentially accounts for the majority of this observed fall in mean score, meaning that the effect of measuring in clinic was very

small. The exploratory analysis of changes in HADS subscales found a mean reduction in the anxiety subscale of 1.26 points (95% CI 0.84–1.67) and in the depression subscale of 0.48 points (95% CI 0.12–0.85). The difference between the scales in the reduction in scores was statistically significant ($p < 0.001$).

Discussion

We had hypothesized that patients' HADS scores might be transiently inflated when measured in the clinic prior to the consultation because of the potentially stressful clinical surroundings and anticipation of the upcoming appointment. If that were the case it would question the utility of this widely used strategy for screening for distress and depression in medical clinics. We found that the majority of the patients who scored high on the HADS in clinic prior to their cancer consultation (72.5%) were still high scorers when reassessed at home a week later. That also means that 27.5% of patients who had scored high in the clinic were no longer high scorers when reassessed later at home. However, further analysis indicates that despite large variability at the individual patient level, the mean HADS total score in the sample fell by only 1.74 points between the two assessments, most of which could reasonably be attributed to the natural tendency for individuals who score high on an initial measurement to score lower on later reassessment (regression to the mean), independent of the setting in which the measurement was made. Our hypothesis was therefore not supported and measuring distress in the clinic prior to the consultation is a reasonable strategy to adopt.

There was considerable individual variability in the size of change scores between the two assessments

despite a high intra-class correlation between repeated measurements from the same patient. This was due to large overall variance in the scores, a property common to measures of psychological distress. It is unclear whether this variation is due to a large random error in the measurements or a reflection of actual fluctuations in the severity of distress over time. Nonetheless, our sample of 218 patients was sufficiently large to estimate the mean change for the sample with reasonable accuracy.

It is notable that, whereas the screening service used the total score in the HADS to define significant distress, the fall in score was slightly larger on the anxiety subscale. This may be because the consultation has a greater transient effect on anxiety than on depression. It may also imply that scales that measure only depressive symptoms are even less subject to a clinic effect.

We are not aware of any studies that have directly addressed the question we have posed. We identified a test–retest reliability study of the European Organization for Research and Treatment of Cancer (EORTC) quality of life measures, which include emotional functioning, that compared questionnaire scores administered to 270 patients attending routine post-treatment follow-up visits to cancer clinics with their score at home 4 days later and found generally good agreement (Hjermstad *et al.* 1995). Other studies that have administered repeated psychological assessment have examined distress trajectories over longer periods of time (Hinnen *et al.* 2008) or before and after consultations (van Dooren *et al.* 2005) but we found none that directly addressed the possible effect of the clinical context on the measurement score.

There were limitations to this study. First, we analysed data collected by a routine screening service operating in cancer clinics; the findings may not therefore generalize to other clinical settings. Second, the service administered a second HADS only to patients who had scored high in clinic. This meant that our observed HADS scores obtained at home underestimated the true proportion of patients who would have scored high had all patients been reassessed, as it would be likely that some of the patients who scored low in clinic would have scored high on the second occasion. This limitation was addressed by estimating the regression to the mean. Third, there were missing data from patients who could not be contacted during the limited time window in which repeat HADS were administered. However, the characteristics of patients on whom we had analysable data and those on whom we did not were mostly similar; systematic bias is therefore unlikely. Fourth, there may be limits to the intrinsic test–retest reliability of the HADS (as opposed to real changes in

symptoms) but this is unlikely to be large over this time period, or to represent a systematic bias. Fifth, patients completed the HADS on a touch-screen computer or on paper in the clinic, but the follow-up assessment was carried out by reading out the scale over the telephone. It is possible that administering the HADS over the telephone causes patients to score differently. Previous studies have found good agreement between self-completed and verbally completed distress screening questionnaires, with a tendency for the latter to record a lower score (Pinto-Meza *et al.* 2005; Cheung *et al.* 2006). Such a bias, if present, would reduce further the observed fall in HADS score attributable to the effect of measurement in the clinic. Future studies could use the same mode of administration to avoid this issue. Sixth, the content of the consultation and its meaning for the patient, whether positive or negative, might have accounted for some of the changes in scores and we were not able to assess this. However, most of the consultations were for follow-up and not for the communication of new diagnoses. The effect of consultation type could be addressed in future studies. Seventh, because the results of the screening were given to the clinician before the consultation, it is possible that they might have taken action to address the distress, for example by referring the patient for psychological treatment. This is, however, very unlikely to have occurred within 1 week. Finally, although the average change in scores was small, the intra-patient variability was high, with some patients scoring very differently on reassessment. It is possible, therefore, that a minority of patients are affected considerably by the clinic setting. Consequently, we cannot rule out the possibility of an important 'clinic effect' for some individuals.

Conclusions

In conclusion, most patients who scored high on the HADS administered in clinic prior to their medical consultation remained high scorers when reassessed at home a week later. There was only a small reduction in mean score, most of which could be attributed to regression to the mean. Therefore, the widely used strategy of asking patients to complete a screening questionnaire for distress while they wait for their clinic appointment is a reasonable method of identifying those who have significant distress and also a useful first step in identifying those who require an interview for the assessment of possible depressive disorder. The increasing use of telephones and the internet provides opportunities to screen patients away from the clinic, thereby potentially avoiding the issue of clinic context. However, the pre-consultation waiting time has long provided an opportunity to

undertake such clinic-based screening, and is likely to continue to do so in the future.

Appendix

Estimating the regression-to-the-mean effect

As only patients with an initial high score were followed up, the scores on reassessment were subject to regression to the mean. We estimated the average drop in scores caused by this effect as follows.

Suppose that a patient's HADS score, H, is the sum of their (constant) true underlying score, S, and an independent error term, e, where S is distributed according to some arbitrary density function with variance σ_s^2 and the errors are normally distributed with mean 0 and variance σ_e^2. The total variance is then $\mathrm{Var}(H) = \sigma_t^2 = \sigma_s^2 + \sigma_e^2$ and $\rho = \sigma_s^2/\sigma_t^2$ is the intra-patient correlation between repeated scores from the same individual.

We wanted to estimate the expected difference between a pair of repeated HADS scores, H_1 and H_2, conditional on H_1 being $\geqslant 15$. That is, we wanted to estimate $E[H_1 - H_2 | H_1 \geqslant 15]$.

For a continuous H it can be shown that

$$E[H_1 - H_2 | H_1 > h_c] = (1-\rho)\sigma_t^2 \frac{g(h_c)}{1 - G(h_c)},$$

where $g(h_c)$ is the probability density function for H evaluated at h_c, and $G(h_c)$ is the corresponding cumulative distribution function (Das & Mulder, 1983). From the large sample of scores collected by the screening service in similar clinics from 2007 to 2010, we obtained empirical estimates of $g(h_c)$ and $G(h_c)$. Using data from the 5215 patients who had HADS scores measured in subsequent clinic visits during this period, we estimated σ_t^2 and ρ as the correlation of scores obtained 1 week apart. We did this by modelling the covariance matrix of repeated scores in a linear regression with random intercept and exponential covariance structure to account for a decreasing correlation over time. A 95% quantile-based CI for the regression-to-mean estimate was derived through bootstrapping.

Although technically a discrete scale, we applied the HADS (range 0–42) with the above result, introducing a continuity correction by evaluating $g(.)$ and $G(.)$ at $h_c = 14.5$ by approximating a theoretical continuous curve. The approach was verified through simulation studies and sensitivity analysis.

Acknowledgements

Declaration of Interest

None.

References

Bakker IM, Terluin B, van Marwijk HW, van Mechelen W, Stalman WA (2009). Test-retest reliability of the PRIME-MD: limitations in diagnosing mental disorders in primary care. *European Journal of Public Health* **19**, 303–307.

Barnett AG, van der Pols JC, Dobson AJ (2005). Regression to the mean: what it is and how to deal with it. *International Journal of Epidemiology* **34**, 215–220.

Carlson LE, Waller A, Mitchell AJ (2012). Screening for distress and unmet needs in patients with cancer: review and recommendations. *Journal of Clinical Oncology* **30**, 1160–1177.

Cheung YB, Goh C, Thumboo J, Khoo KS, Wee J (2006). Quality of life scores differed according to mode of administration in a review of three major oncology questionnaires. *Journal of Clinical Epidemiology* **59**, 185–191.

Cosco TD, Doyle F, Ward M, McGee H (2012). Latent structure of the Hospital Anxiety and Depression Scale: a 10-year systematic review. *Journal of Psychosomatic Research* **72**, 180–184.

Das P, Mulder PGH (1983). Regression to the mode. *Statistica Neerlandica* **37**, 15–20.

First MB, Spitzer RL, Gibbon M, Williams JBW (1999). *Structured Clinical Interview for DSM-IV Axis I Disorders*. Biometrics Research Department, New York State Psychiatric Institute: New York.

Gerin W, Ogedegbe G, Schwartz JE, Chaplin WF, Goyal T, Clemow L, Davidson KW, Burg M, Lipsky S, Kentor R, Jhalani J, Shimbo D, Pickering TG (2006). Assessment of the white-coat effect. *Journal of Hypertension* **24**, 67–74.

Greenhalgh J (2009). The applications of PROs in clinical practice: what are they, do they work, and why? *Quality of Life Research* **18**, 115–123.

Hinnen C, Ranchor AV, Sanderman R, Snijders TA, Hagedoorn M, Coyne JC (2008). Course of distress in breast cancer patients, their partners, and matched control couples. *Annals of Behavioral Medicine* **36**, 141–148.

Hjermstad MJ, Fossa SD, Bjordal K, Kaasa S (1995). Test/retest study of the European Organization for Research and Treatment of Cancer Core Quality-of-Life Questionnaire. *Journal of Clinical Oncology* **13**, 1249–1254.

Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, Ustun B (2007). Depression, chronic diseases, and decrements in health: results from the World Health Surveys. *Lancet* **370**, 851–858.

NICE (2009). *Depression in Adults with a Chronic Physical Health Problem: Treatment and Management*. National Institute for Health and Clinical Excellence: London.

Pignone MP, Gaynes BN, Rushton JL, Burchell CM, Orleans CT, Mulrow CD, Lohr KN (2002). Screening for depression in adults : a summary of the evidence for the U.S. Preventive Services Task Force. *Annals of Internal Medicine* **136**, 765–776.

Pinto-Meza A, Serrano-Blanco A, Penarrubia MT, Blanco E, Haro JM (2005). Assessing depression in primary care with the PHQ-9 : can it be carried out over the telephone ? *Journal of General Internal Medicine* **20**, 738–742.

USPSTF (2009). Screening for depression in adults : U.S. Preventive Services Task Force recommendation statement. *Annals of Internal Medicine* **151**, 784–792.

van Dooren S, Seynaeve C, Rijnsburger AJ, Duivenvoorden HJ, Essink-Bot ML, Tilanus-Linthorst MM, Klijn JG, de Koning HJ, Tibben A (2005). Exploring the course of psychological distress around two successive control visits in women at hereditary risk of breast cancer. *European Journal of Cancer* **41**, 1416–1425.

van Dulmen AM, Fennis JF, Mokkink HG, van der Velden HG, Bleijenberg G (1995). Doctor-dependent changes in complaint-related cognitions and anxiety during medical consultations in functional abdominal complaints. *Psychological Medicine* **25**, 1011–1018.

Vodermaier A, Linden W, Siu C (2009). Screening for emotional distress in cancer patients : a systematic review of assessment instruments. *Journal of the National Cancer Institute* **101**, 1464–1488.

Walker J, Postma K, McHugh GS, Rush R, Coyle B, Strong V, Sharpe M (2007). Performance of the Hospital Anxiety and Depression Scale as a screening tool for major depressive disorder in cancer patients. *Journal of Psychosomatic Research* **63**, 83–91.

Zigmond AS, Snaith RP (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica* **67**, 361–370.