University of Groningen

No further research needed
Coyne, James C.; van Sonderen, Eric

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Link to publication in University of Groningen/UMCG research database

Correspondence

No further research needed: Abandoning the Hospital and Anxiety Depression Scale (HADS)

James C. Coyne [a,b,*], Eric van Sonderen [a]

[a] Health Psychology Section, Department of Health Sciences, University Medical Center, Groningen, University of Groningen, The Netherlands
[b] Department of Psychiatry, University of Pennsylvania School of Medicine, USA

Cosco and colleagues [this issue] provide a well done and transparently reported systematic review of the Hospital and Anxiety Depression Scale (HADS) literature of the past decade. They conclude that the underlying structure of the HADS is inconsistent across samples and highly dependent on the statistical methods used to establish that structure. The implication is that the HADS is not a dependable means of differentiating anxiety and depression for the purposes of assessing the absolute or relative levels of these variables. These results can also go far in explaining the confusing difficulties that have arisen in research concerning use of the HADS as the first stage of two-stage screening procedures for depression and anxiety disorders or case identification purposes.

Comparisons between results obtained with the HADS in a particular sample and the available literature can prove bewildering. It is quite uncommon to validate the presumed structure of cleanly separated anxiety and depression subscales, but there is little consistency whether the best fit is a unitary single factor solution, a two-factor solution without a clean separation of anxiety and depression, or a three factor solution. And then there is the anomalous finding in a large Danish study [1] that breast cancer patients had *lower* anxiety and depression scores on the HADS than women drawn from the general population. Do we accept that finding, dispute it on the basis of flaws we can identify in the study's methodology, or doubt the basic validity of the HADS?

Fundamental problems with the HADS are unintended consequences of deliberate decisions made in its construction. Cosco and colleagues provide only an easily missed clue to a basic problem by noting in passing, "Item 7 was found to anomalously load in 20 studies, indicating that it is a particularly poor item." The item, "I can sit at ease and feel relaxed," is one of six positively valence items, with greater endorsement representing less anxiety, but the response key is both reversed (1 equals definitely; 4 equals not at all) and has different anchors than the response key (1 equals not at all; 4 equals most of the time) for the item just prior to it. While such reversals of wording and varying response keys were intended to avoid effects of a response style, they are disorienting, and unless patients are particularly vigilant, they will miss the changes in direction of the items and scoring.

In constructing the HADS, there was also a deliberate reliance on colloquial British expressions (i.e., "butterflies in the stomach" for anxiety) that bedevils efforts to translate it into other languages. Curiously, there is usually no comment on how this problem was addressed in articles using translations of the HADS, casting suspicion on translated versions [2]. Other items intended to avoid the stigma of explicit reference to psychiatric symptoms are ambiguous as to whether they indicate psychopathology or normal individual differences in personality, such as "I feel restless as if I have to be on the move."

The deliberate avoidance of somatic items in construction of the HADS was based on unsubstantiated clinical lore and led to the exclusion of sleep and appetite disturbance items, the presence of which together might be a good indicator of depression. There is no good reason to believe that a combination of sleep and appetite disturbance is less revealing of depression among medical patients, than items like number 10, "I have lost interest in my appearance." which might reflect adaptation to a debilitating and disfiguring medical condition. Regardless, there is a lack of evidence that overall scores of medical patients on the conventional somatic items of other measures are higher than scores for psychiatric patients matched for age and cognitive/affective items [3,4]. The HADS's additional emphasis on anhedonia rather than sadness may be particularly problematic for making discriminations within a range of sadness, unhappiness, and demoralization in medical populations lacking full syndromal clinical depression.

The structural problems uncovered by Cosco and colleagues provide an explanation for recurring difficulties in research concerning the HADS as a screening or case finding instrument. Optimal cutpoints for clinically significant symptoms or decisions about the need for follow-up evaluation vary wildly and inexplicably across validation studies. As one review has documented [3], when cutpoints recommended in the literature are used as a basis for selecting patients for follow-up interviews, the anxiety subscale of the HADS may predict diagnoses of depression better than the depression subscale does or the total HADS score may better identify patients in need of further evaluation for anxiety or depressive disorders better than either of the anxiety or depression subscales.

Defenders of the HADS as a screening instrument can nonetheless point to an exceptionally large literature concerning its performance. Yet, much of the literature provides inflated estimates of the performance of the HADS because cutpoints were allowed to freely vary and therefore capitalized on idiosyncratic features of the particular

* Corresponding author at: Department of Psychiatry, University of Pennsylvania School of Medicine, 3535 Market St. Rm 676, Philadelphia, PA 19104, USA. Tel.: +1 215 662 7035; fax: +1 215 349 5067.
E-mail address: jcoyne@mail.med.upenn.edu (J.C. Coyne).

sample or sampling in ways that cannot be expected to generalize [5]. A further problem shared with the evaluation of other screening instruments is that almost no studies exclude cases that are already identified into account in estimating the performance of the HADS [6]. But ambiguity in the language of items, particularly the use of British colloquialisms aggravates this problem in the HADS literature and introduces differences in performance of the HADS across languages and cultures [2] that should discourage uncritical integration of cross-cultural studies into systematic reviews and meta-analyses. Strangely, reviewers keep stumbling upon evidence pointing to these issues, sometimes noting them and usually not. They inevitably continue to recommend the HADS as a screening instrument or major depression and anxiety, although sometimes with the suggestion that the HADS needs to be recalibrated in new samples [7], which is no small task because such re-calibrations need replication that is rarely done.

Finally, we should not make too much of the brevity of the HADS. Fourteen items may be too many to win sustained use in busy clinical settings where even the PHQ-9 is considered too long [8], particularly when the subscales of the HADS do not function as anticipated.

It is customary for systematic reviews to make suggestions for future research based on the substantive conclusions they have reached. However, Cosco and colleagues may have provided the final authoritative evidence that no further research is needed concerning the HADS because of its lack of ability to uncover a reliable, generalizable underlying structure, and certainly not matching the anxiety and depression subscales. There are abundant reasons why the field should move on, leave the HADS literature behind, and select any of a number of alternative instruments in its place. Tradition and the HADS still being the most widely used screening and assessment instrument with medically ill patients are insufficient reasons to continue to recommend it.

References

[1] Groenvold M, Fayers PM, Sprangers MAG, Bjorner JB, Klee MC, Aaronson NK, et al. Anxiety and depression in breast cancer patients at low risk of recurrence compared with the general population: a valid comparison? Journal of Clinical Epidemiology 1999;52(6):523–30.

[2] Maters, G.A., Sanderman, R., Kim, A.Y. & Coyne, JC. Problems in the Cross-cultural Adaptation of the Hospital Anxiety and Depression Scale: "No Butterflies in the Desert". Manuscript Under Review, 2011.

[3] Mitchell AJ, Meader N, Symonds P. Diagnostic validity of the Hospital Anxiety and Depression Scale (HADS) in cancer and palliative settings: a meta-analysis. Journal of Affective Disorders 2010;126:335–48.

[4] Thombs BD, Ziegelstein RC, et al. Somatic symptom overlap in Beck Depression Inventory-II scores following myocardial infarction. British Journal of Psychiatry 2011;197(1):61–6.

[5] Thombs BD, de Jonge P, Coyne JC, Whooley MA, Frasure-Smith N, Mitchell AJ. Depression screening and patient outcomes in cardiovascular care: a systematic review. Journal of the American Medical Association 2008;300(18):2161–71.

[6] Thombs BD, Arthurs E, El-Baalbaki G, Meijer A, Ziegelstein RC, Steele RJ, et al. Risk of bias from inclusion of patients who already have diagnosis of or are undergoing treatment for depression in diagnostic accuracy studies of screening tools for depression: systematic review. BMJ 2011:343.

[7] Vodermaier A, Linden W, Siu C. Screening for emotional distress in cancer patients: a systematic review of assessment instruments. Journal of the National Cancer Institute 2009;101(21):1464–88.

[8] Bermejo I, Frey C, Kriston L, Schneider F, Gaebel W, Hegerl U, et al. Stability of the effects of guideline training in primary care on the identification of depressive disorders. Primary Care & Community Psychiatry 2007;12(3–4):99–107.