Journal of Psychiatric Research 60 (2015) 132–140

Contents lists available at ScienceDirect

Journal of Psychiatric Research

journal homepage: www.elsevier.com/locate/psychires

Review

Psychometric properties of the 16-item Quick Inventory of Depressive Symptomatology: A systematic review and meta-analysis

Thomas J. Reilly [a], Steve A. MacGillivray [b], Ian C. Reid [c], Isobel M. Cameron [c, *]

[a] School of Medicine and Dentistry, University of Aberdeen, Aberdeen, AB25 2ZD, UK
[b] Social Dimensions of Health Institute, University of Dundee, DD1 4HJ, UK
[c] Applied Medical Sciences (Psychiatry), University of Aberdeen, Royal Cornhill Hospital, Aberdeen, AB25 2ZH, UK

ARTICLE INFO

Article history:
Received 4 June 2014
Received in revised form
5 September 2014
Accepted 11 September 2014

Keywords:
Depression
Psychometric
Meta-analysis
Systematic review
Quick Inventory of Depressive Symptomatology

ABSTRACT

Effective management of depression is predicated upon reliable assessment. The Quick Inventory of Depressive Symptomatology (QIDS) is a depression severity scale with both self-rated ($QIDS\text{-}SR_{16}$) and clinician-rated ($QIDS\text{-}C_{16}$) versions. Although widely used in research, the psychometric properties of the $QIDS_{16}$ have not been systematically reviewed. We performed a systematic review of studies of the psychometric properties (factor structure, internal consistency, convergent validity, discriminant validity, test-retest reliability and responsiveness to change) of the $QIDS\text{-}SR_{16}$ or $QIDS\text{-}C_{16}$. Six databases were searched: MEDLINE, EMBASE, PsycINFO, CinAHL, Web of Science and the Cochrane Central Register of Controlled Trials. Findings were summarised, bias assessed and correlations with reference standards were pooled. 37 studies (17,118 participants) were included in the review. Both versions of the $QIDS_{16}$ were unidimensional. Cronbach's alpha ranged from 0.69 to 0.89 for the $QIDS\text{-}SR_{16}$ and 0.65 to 0.87 for the $QIDS\text{-}C_{16}$. The $QIDS\text{-}SR_{16}$ correlated moderately to highly with several depression severity scales. Seven studies were pooled where $QIDS\text{-}SR_{16}$ was correlated with the HRSD-17 ($r = 0.76$, CI 0.69, 0.81) in patients diagnosed with depression. Four studies examined convergent validity with the $QIDS\text{-}C_{16}$. Four studies examined discriminant validity, for the $QIDS\text{-}SR_{16}$ alone. Eighteen studies had at least one author who was a co-author of the original $QIDS_{16}$ study. Most studies were conducted in the USA ($n = 26$). The $QIDS\text{-}SR_{16}$ and the $QIDS\text{-}C_{16}$ are unidimensional rating scales with acceptable internal consistency. To justify the use of the $QIDS_{16}$ scale in clinical practice, more research is needed on convergent and discriminant validity, and in populations outside the USA.

1. Background

Depression is a common illness and a major burden of healthcare worldwide (Bromet et al., 2011). Guidelines for the treatment of depression highlight the importance of gauging symptom severity in relation to treatment recommendations (Bauer et al., 2007). Monitoring severity of symptoms using standardised, validated tools is advocated (Anderson et al., 2008; Mitchell et al., 2013; National Institute for Health and Clinical Excellence, 2009; New Zealand Guidelines Group, 2008). Such tools are thought to be more objective when prescribing antidepressants than clinical impression alone (Kendrick et al., 2005). Despite this, studies of the psychometric properties of commonly used depression severity scales (for example, Cameron et al., 2011, 2008; Hansson et al., 2009; Reddy et al., 2010; Zimmerman et al., 2012b) have identified shortfalls in their validity, thus raising concerns about their suitability for application in clinical practice.

The 16-item Quick Inventory of Depressive Symptomatology ($QIDS_{16}$) may have a role in monitoring depressive symptoms in clinical practice. It is derived from the 30-item Inventory of Depressive Symptomatology (IDS_{30}) and is available in both clinician ($QIDS\text{-}C_{16}$) and self-reported ($QIDS\text{-}SR_{16}$) formats (Rush et al., 2003). The scale contains all DSM-IV criterion symptoms for major depressive disorder (American Psychiatric Association, 1994) and rates the severity of these in the preceding seven days on a scale of 0–3. The nine symptom domains are: sad mood, concentration, self-criticism, suicidal ideation, general interest, energy/fatigue, sleep disturbance, decrease/increase in appetite/weight, psychomotor agitation/retardation. For six of these domains, one item covers each criterion. Additionally, four questions pertain to sleep symptoms, four to weight and appetite

* Corresponding author. Tel.: +44 1224 557966; fax: +44 1224 557400.
E-mail address: i.m.cameron@abdn.ac.uk (I.M. Cameron).

http://dx.doi.org/10.1016/j.jpsychires.2014.09.008
0022-3956/

symptoms and two to psychomotor symptoms. By calculating domain scores for these groups of closely related items, spurious influences on reliability statistics are minimised. The highest scored items within these domains are summed with the scores of the other items. Total scores range from 0 to 27. Summed scores indicate the following: 0–5 no depression, 6–10 mild, 11–15 moderate, 16–20 severe and 21–27 very severe depression. These severity cut offs were derived through item response theory analysis whereby the scores were calibrated against a reference standard in the form of the Hamilton Depression Rating Scale (Rush et al., 2003). The scale has been translated into 31 languages and is freely accessible from the QIDS/IDS website (www.ids.qids.org/).

The popularity of the $QIDS_{16}$ in research settings is shown by a large number of clinical trials which have used the tool. Notably, it was used in the STAR*D trial, (Rush et al., 2008, 2006b; Warden et al., 2007). Despite the widespread use, the psychometric properties of the $QIDS_{16}$ scale have not been systematically reviewed. The purpose of this review is to address this need by comprehensively searching the literature for psychometric data using the $QIDS_{16}$ and evaluating the evidence for using the scale to measure depression severity. Specifically, we address the following psychometric properties of the QIDS-SR_{16} and the QIDS-C_{16}: factor structure, internal consistency, test-retest reliability, convergent validity, discriminant validity and responsiveness to change. *A priori* we aimed to conduct a meta-analysis regarding convergent validity where data allowed.

2. Method

The aims and method of this review were pre-specified and the registered protocol can be accessed at: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42013004011.

2.1. Inclusion criteria

We included original research which used either the QIDS-SR_{16} or QIDS-C_{16} and provided psychometric data including factor analysis, measures of internal consistency, test-retest reliability, convergent validity, discriminant validity, and responsiveness to change. No restrictions were applied on the basis of publication status or language. Studies where data pertained to adult participants (over 15 years), in clinical (for example, primary-care, secondary-care or inpatients) or non-clinical (for example, general population) settings were all considered relevant, as were intervention studies which provided psychometric data.

2.2. Exclusion criteria

We excluded studies which assessed modified versions of the QIDS-SR_{16} or QIDS-C_{16} with the exception of direct translations into other languages.

2.3. Literature search

The following databases were searched from inception to August 2014: MEDLINE, EMBASE, PsycINFO, CinAHL, Web of Science and the Cochrane Central Register of Controlled Trials. Search terms were: “QIDS” OR “quick ADJ1 inventory” as text words. In addition we reviewed the references of included papers and references listed on the $QIDS_{16}$ website (www.ids-qids.org). We contacted authors of relevant conference abstracts and grey literature to request unpublished data. Additionally, we searched the publication lists of the websites of key investigators in this field.

2.4. Study selection

Working independently, two authors examined the titles and abstracts of the initial search results. At this stage, studies which clearly did not use the $QIDS_{16}$ were excluded, disagreement was resolved through discussion. Full text was obtained for papers which were considered potentially relevant. The same two investigators inspected the full text articles and excluded those which did not provide psychometric data. Once again, disagreement was resolved through discussion.

2.5. Data extraction

Using standardised bespoke pro-formas, data were independently extracted from all relevant articles by two authors and checked by a third author. Setting, year of study, sample size and psychometric data were extracted and tabulated.

Psychometric data included:

- Method and results of factor analysis (dimensionality);
- Internal consistency (Cronbach's alpha and item–total correlations);
- Convergent validity (correlation of the $QIDS_{16}$ scores with comparator measure and discriminant validity from comparator. Where convergent validity was assessed against the IDS_{30}, data were not included due to the high risk of incorporation bias as the $QIDS_{16}$ is a subset of this scale. We also disregarded convergence data where the QIDS-SR_{16} or QIDS-C_{16} were the “reference standards” against which another scale was assessed);
- Responsiveness to change (relative effect sizes) in relation to validated comparator measures.

2.6. Data analysis

Cronbach's alpha and item-total correlations were considered acceptable if between 0.7 and 0.9 and item-total correlations if > 0.3 (Nunnally and Bernstein, 1994). For convergent validity, correlation of the $QIDS_{16}$ scores with comparator measures were considered moderate if > 0.6 and high if > 0.8. Discriminant validity was considered acceptable where correlations were <0.6. Regarding convergent validity, correlations with reference standards were pooled provided data were sufficiently similar with respect to: comparator measure, correlation method, diagnosis, treatment setting. Where individual studies provided correlations for baseline and exit time points, the data from exit time point was used as this was deemed to have a greater range. For example, data at baseline may be restricted to meet entry criteria. Using MedCalc Version 12.7.7 (2013), the weighted summary correlation coefficient (with a Fisher Z transformation) was calculated following the Hedges and Olkin (1985) method. We assessed statistical heterogeneity with the I^2 test with 95% Confidence Intervals [CI] (Higgins and Thompson, 2002) where 25%, 50% and 75% indicate low, medium, and high statistical heterogeneity respectively. In the presence of medium or high statistical heterogeneity we employed a random-effects model (DerSimonian and Laird, 1986).

2.7. Assessment of risk of bias

Risk of bias within individual studies was assessed according to the following criteria: investigator bias (researchers not being independent of the scale's development team), sampling method (evidence of random, consecutive or complete method), and sufficiency of sample size (in the case of Cronbach α $n > 100$ based on the assumption that the construct will be unidimensional and as