The validity and reliability of screening measures for depression and anxiety disorders in multiple sclerosis

Ruth Ann Marrie\textsuperscript{ab}, Lixia Zhang\textsuperscript{b}, Lisa M. Lix\textsuperscript{b}, Lesley A. Graff\textsuperscript{c}, John R. Walker\textsuperscript{d}, John D. Fisk\textsuperscript{d}, Scott B. Patten\textsuperscript{e}, Carol A. Hitchon\textsuperscript{a}, James M. Bolton\textsuperscript{f}, Jitender Sareen\textsuperscript{f}, Renée El-Gabalawy\textsuperscript{c,g}, James J. Marriott\textsuperscript{h}, Charles N. Bernstein\textsuperscript{i}

\textsuperscript{a} Department of Internal Medicine, Max Rady College of Medicine, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Canada
\textsuperscript{b} Nova Scotia Health Authority, Departments of Psychiatry, Psychology & Neuroscience, and Medicine, Dalhousie University, Halifax, Canada
\textsuperscript{c} Department of Psychiatry, Max Rady College of Medicine, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Canada
\textsuperscript{d} Department of Community Health Sciences, Max Rady College of Medicine, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Canada
\textsuperscript{e} University of Calgary, Calgary, Canada
\textsuperscript{f} Departments of Anesthesia & Perioperative Medicine, Max Rady College of Medicine, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Canada
\textsuperscript{g} Departments of Community Health Sciences & Psychiatry, Cumming School of Medicine, University of Calgary, Calgary, Canada
\textsuperscript{h} Department of Community Health Sciences, Max Rady College of Medicine, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Canada

ARTICLE INFO

Keywords:
Multiple sclerosis
Depression
Anxiety
Validity
Reliability
Psychometrics

ABSTRACT

Objective: We aimed to evaluate the validity and reliability of multiple screening measures for depression and anxiety for use in the clinical care of people with multiple sclerosis (MS).

Methods: Participants with MS completed the Patient Health Questionnaire (PHQ-9), Hospital Anxiety and Depression Scale (HADS), Kessler-6 Distress Scale, PROMIS Emotional Distress Depression Short-Form 8a (PROMIS Depression) and Anxiety Short-Form 8a (PROMIS Anxiety), Generalized Anxiety Disorder 7-item Scale (GAD-7), and the Overall Anxiety and Severity Impairment Scale (OASIS). A subgroup repeated the screening measures two weeks later. All participants also completed a Structured Clinical Interview for DSM-IV-TR Axis I Disorders (SCID). For the screening measures we computed sensitivity, specificity, positive predictive and negative predictive value with SCID diagnoses as the reference standard and conducted receiver operating curve (ROC) analyses; we also assessed internal consistency and test-retest reliability.

Results: Of 253 participants, the SCID classified 10.3% with major depression and 14.6% with generalized anxiety disorder. Among the depression measures, the PHQ-9 had the highest sensitivity (84%). Specificity was generally higher than sensitivity, and was highest for the HADS-D with a cut-point of 11 (95%). In ROC analyses the area under the curve (AUC) did not differ between depression measures. Among the anxiety measures, sensitivity was highest for the HADS-A with a cut-point of 8 (95%). Specificity ranged from 83% to 86% for all measures except the HADS-A with a cut-point of 8 (86%). The AUC did not differ between anxiety measures.

Conclusion: Overall, performance of the depression and anxiety screening measures was very similar, with reasonable psychometric properties for the MS population, suggesting that other factors such as accessibility and ease of use could guide the choice of measure in clinical practice.

1. Introduction

Multiple sclerosis (MS) has a high prevalence of comorbid depression and anxiety disorders throughout the disease course (Marrie et al., 2015; Marrie et al., 2016). Comorbid depressive and anxiety disorders are associated with lower quality of life, and greater pain and health care utilization (Janssens et al., 2003; Fiest et al., 2015; Marrie et al., 2015). Therefore, emphasis has been placed on identifying these disorders promptly, and involving collaborative mental health services if needed.

Multiple potential case identification (aka screening) measures for assessing possible depression and anxiety disorders exist (Williams et al., 2002). However, somatic symptoms of depression such as fatigue, and difficulty sleeping captured in screening measures for depression are also common somatic symptoms of MS. Similar issues arise when screening for anxiety. For example, the Beck Anxiety Inventory (BAI) captures somatic symptoms of anxiety such as dizziness, numbness and tingling (Beck et al., 1988), which are common physical symptoms.
experienced in MS. This raises the question as to whether these measures adequately identify the depressive and anxiety disorders or whether they are confounded by physical symptoms of the MS, leading to overestimates and misclassification of these disorders. Problems with criterion contamination of depression scales have been reported in MS (Mohr et al., 1997).

A systematic review identified 21 studies which assessed the performance of nine depression screening measures in MS and found that further research was needed to assess the utility of most measures (Hind et al., 2016). A systematic review of screening measures for anxiety found relatively little support for the validity and reliability of three available instruments, the Hospital Anxiety and Depression Scale (HADS), Generalized Anxiety Disorder-7 (GAD-7) and BAI (Lüster et al., 2016). Therefore, we aimed to evaluate the validity and reliability of multiple screening measures for depression and anxiety for people with MS.

2. Materials and methods

As detailed elsewhere (Marrie et al., In press), from November 2014 through July 2016 we recruited individuals from the sole provincial MS Clinic with a definite diagnosis of MS (Poser et al., 1983; McDonald et al., 2001; Polman et al., 2005, 2011), who were aged ≥18 years, able to provide informed consent, and with an adequate knowledge of English to complete questionnaires and interviews. Ethics approval was provided by the University of Manitoba Health Research Ethics Board, Victoria General Hospital, the Health Sciences Centre, Seven Oaks General Hospital and St. Boniface Hospital.

After providing informed consent, participants completed questionnaires, and underwent physical assessments as described below. If possible, they participated in the Structured Clinical Interview for DSM-IV-TR Axis I Disorders – Research version (SCID) the same day (First et al., 2002). If not, the SCID was completed within two to four weeks of enrollment. A subgroup of participants completed the screening measures again within two weeks of initial administration.

2.1. Sociodemographic and clinical characteristics

Participants reported their sex, date of birth, ethnicity, and highest level of education attained. Ethnicity was categorized as white or non-white. Education was categorized as less than high school, high school/ GED, college, technical/trade, and Bachelor’s degree or higher. Participants also reported their age at MS symptom onset. We determined clinical course by medical records review. Participants underwent a neurologic examination for determination of disability status as measured by the Expanded Disability Status Scale (EDSS) (Kurtzke, 1983).

2.2. Screening measures

Each participant completed the Patient Health Questionnaire (PHQ-9) from which we also derived a score for the PHQ-2, the HADS, Kessler-6 Distress Scale, Patient-Reported Outcomes Measurement Information System Emotional Distress Depression Short-Form 8a (PROMIS Depression) and Anxiety Short-Form 8a (PROMIS Anxiety), GAD-7 and Overall Anxiety and Severity Impairment Scale (OASIS) (Zigmond and Snaith, 1983; Spitzer et al., 1999; Norman et al., 2006; Spitzer et al., 2006; Cairney et al., 2007). When selecting these measures we considered properties including face validity, ease of use, availability for self-administration, and copyright restrictions.

The PHQ-9 includes nine items with response options of 0 (not at all) to 3 (nearly every day), and assesses depressive symptoms over the last two weeks (Spitzer et al., 1999). Total scores range from 0 to 27. The PHQ-2 includes the first two items from the PHQ-9 and has been promoted as a brief screen for depression (Kroenke et al., 2003). Scores range from 0 to 6. The HADS includes 14 items, 7 for depression and 7 for anxiety, which assess symptoms over the past week (Zigmond and Snaith, 1983). Two cut-points are commonly used for the HADS (8, 11) therefore we tested both. Total scores for each of the two subscales range from 0 to 21. The Kessler-6 includes 6 items which measure non-specific distress over the past 30 days; we classified it with depression measures since five of its six items are common depressive symptoms (hopelessness, agitation, depressed mood, low energy, worthlessness). Using the alternative scoring method (https://www.hcp.med.harvard.edu/ncs/k6_scales.php), scores range from 6 to 30. The PROMIS Depression and Anxiety measures include 8 items with response options ranging from 1 (never) to 5 (always) (Pilkonis et al., 2011). These items assess symptoms over the past 7 days. Total scores for the PROMIS Depression measure are transformed into T scores with values ranging from 38.2 to 81.3, while they are transformed into T scores with values ranging from 37.1 to 83.1 for the PROMIS Anxiety measure. A score of 50 is average for the United States general population. The GAD-7 includes 7 items which assess symptoms of anxiety over the last two weeks. Response options range from 0 (not at all) to 3 (nearly every day); total scores range from 0 to 21. The OASIS includes 5 items which assess anxiety and fear over the past week (Norman et al., 2006). Response options range from 0 to 4 and total scores from range 0–20. For all measures, higher scores indicate more severe symptoms.

2.3. Questionnaires assessing related constructs

We assessed fatigue using the Fatigue Impact Scale for Daily Use (DFIS), a validated instrument which includes 8 items scored on an ordinal scale from 0 (no) to 4 (extreme problem) (Fisk and Doble, 2002). We assessed pain using from MOS-Modified Pain Effects Scale, a valid and reliable instrument with scores ranging from 6 to 30 (Ritvo et al., 1997a, 1997b) higher scores indicate greater pain.

2.4. Interview

The SCID is a semi-structured interview to identify DSM-IV diagnoses including anxiety, and major depression. Trained interviewers, blinded to the results of the screening measures, administered the SCID to determine the current histories of depressive and anxiety disorders. For this study, SCID-based diagnoses of current major depression and generalized anxiety disorder served as the reference standard in analyses of criterion validity. In a complementary analysis, we used a SCID diagnosis of any anxiety disorder in the last month (generalized anxiety disorder, panic disorder, social phobia, specific phobia, anxiety disorder due to general medical condition, anxiety disorder due to substance use, stress disorder) instead of generalized anxiety disorder.

2.5. Analysis

We summarized the characteristics of study participants using frequency (percent (%)) for categorical variables, and mean (standard deviation (SD)) or median (interquartile range [IQR]) for continuous variables. Missing data were not imputed; individuals with missing values for a measure were excluded from analyses of that measure.

Based on the taxonomy proposed by the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) we assessed criterion validity, construct validity (through hypothesis testing), content validity, internal consistency reliability, and test-retest reliability of the selected measures (Mokkink et al., 2010).

Criterion validity indicates how well the scores of the screening tool reflect the reference (criterion) standard. First, we compared depression and anxiety status based on the (i) SCID (criterion standard) and (ii) self-reported screening measures. Based on published cut-points for depression/anxiety for these measures, we computed sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) for the screening measures versus the criterion standard. Second, we used receiver operating curve (ROC) analysis to identify the best
دریافت فوری
متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات