Journal of Psychosomatic Research xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Journal of Psychosomatic Research

A comparison of multiple patient reported outcome measures in identifying major depressive disorder in people with multiple sclerosis

Dagmar Amtmann *, Alyssa M. Bamer, Kurt L. Johnson, Dawn M. Ehde, Meghan L. Beier, Jamie L. Elzea, Charles H. Bombardier

Department of Rehabilitation Medicine, University of Washington, Seattle, WA, USA

ARTICLE INFO

Article history:
Received 30 January 2015
Received in revised form 20 August 2015
Accepted 26 August 2015
Available online xxxx

Keywords:
Depression
PHQ
CESD
PROMIS
Neuro-QOL
Beck depression
SCID
Diagnostic accuracy

ABSTRACT

Background: Depression is one of the most prominent and debilitating symptoms in individuals with multiple sclerosis (MS), yet there is currently no consensus on the best instruments for depression screening in MS. More head to head comparisons of available screening instruments are needed to advise MS researchers and clinicians.
Methods: A cross-sectional comparison of the effectiveness of screening for MDD using multiple patient reported outcome (PRO) screeners against a modified SCID telephone interview was completed in 164 individuals with MS. Stratum goals were set for depression levels to ensure participation by people with borderline and higher levels of depression. Criterion standard was a modified SCID MDD module. PRO measures included the PHQ-9, BDI-FS, PROMIS depression, Neuro-QOL depression, M-PHQ-2, PHQ-2, and CESD.
Results: 48 (29%) individuals met the modified SCID criteria for MDD. The sensitivity of the PRO measures ranged from 60% to 100% while specificity ranged from 46% to 86%. The ROC area for the PRO measures ranged from 0.79 to 0.83. Revised (higher) cutoff scores were suggested by the ROC analyses for most self-reported screeners.
Limitations: Enrollment was stopped early because of difficulties with recruitment. Several SCID recording could not be reviewed and diagnosis confirmed.
Conclusions: CESD-10 and PHQ9 had the best diagnostic performance using optimal cutoffs, but no one PRO measure stood out as significantly better than any other. Even when revised cutoff scores were used, none of the self-reported screeners identified people with MDD with adequate accuracy. More accurate self-reported screeners would facilitate diagnosing of MDD for both research and clinical purposes.

1. Introduction

Multiple sclerosis (MS) is a chronic neurologic disease characterized by inflammation, demyelination, and neurodegeneration in the central nervous system. People with MS often report physical, cognitive and psychological symptoms, of which depression is one of the most prominent and debilitating [49]. The life-time prevalence of major depressive disorder (MDD) in individuals with MS has been estimated at over 50% [38]. In addition, the 12-month prevalence of MDD is approximately twice that of the general population at 15.7% [38]. Depression is associated with a host of poor outcomes in people with MS, including poorer overall health, non-adherence to disease modifying medications [54], loss of employment or reduction in work hours [39], an increased risk of suicidal ideation and completed suicide [17,18,42], greater cognitive dysfunction [4], and an overall reduction in quality of life [15,25]. Despite these poor outcomes in MS, MDD remains under-recognized and undertreated [16,20,25].

Numerous instruments are available to assess depressive symptoms in people with MS, and they can also be used to identify depression cases in need of treatment. In the MS literature these measures are commonly referred to as depression "screening" instruments or measures [19,34]. However, only a few published studies have compared the agreement between depression measures and structured diagnostic interviews for MDD in people with MS. In a series of newly diagnosed individuals with MS, the original Beck Depression Inventory (BDI) [6] (cutoff 13) produced 71% sensitivity, 79% specificity when compared to the Diagnostic Interview Schedule for DSM-III disorders [53]. A more recent similar study of the BDI-II in a clinical MS population produced similar results at 85% sensitivity and 76% specificity compared to the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) [56]. A two-item measure adapted from the Primary Care Evaluation of Mental Disorders (PRIME-MD) [51,58] was compared to MDD diagnoses derived from the Structured Clinical Interview for DSM-IV Disorders (SCID) [23] and reported a 99% sensitivity and 87% specificity [35] in one study and 80% sensitivity and 93% specificity in a second

* Corresponding author at: University of Washington, Box 354237, Seattle, WA 98195, USA.
E-mail address: dagmara@uw.edu (D. Amtmann).

http://dx.doi.org/10.1016/j.jpsychores.2015.08.007
0022-3999/

[40]. The depression subscale of the Hospital Anxiety and Depression Scale (HADS) has also been validated against the SCID or the Schedules for Clinical Assessment in Neuropsychiatry. A cutoff of eight in two studies found 90% sensitivity and 87.3% specificity [28] and 85% sensitivity and 82% specificity [40] while a third study found 77% sensitivity and 81% specificity at a cutoff of 11 [56]. Recently a study of the Center for Epidemiologic Studies Depression scale (CESD) found that it provided 95% sensitivity and 73% specificity compared to the SCID [40]. These findings provide important information regarding the utility of each measure, however many instruments have only been examined once and in small clinical samples and still more depression instruments are available that have not been examined.

There is currently no consensus on best practices for what instruments to use to assess depression in people with MS. As noted by the American Academy of Neurology in their recent review of the evidence for depression screening measures in MS [34], a number of depression measures used in the MS field lack strong evidence for their utility in identifying cases of depression, particularly relative to other commonly used measures. More head to head comparisons of measures are needed to advise researchers and clinicians. Among the many common measures whose utility for case identification in MS have not been well studied are the Beck Depression Inventory-Fast Screen (BDI-FS) [5], Patient Health Questionnaire-9 (PHQ-9) [30], and the CESD [57]. The BDI-FS is made up of a subset of items from the Beck scale proposed for use in MS [7], while the CESD and the PHQ-9 are two of the most commonly used instruments in the literature. Additionally, new instruments were recently developed with modern psychometric methods, and their clinical utility as MDD screeners in MS has yet to be examined. These include the Patient Reported Outcomes Measurement Information Systems (PROMIS®) and Neurological Quality of Life (Neuro-QOL) depression item banks. Both instruments provide population norms and have an added benefit of the availability of administration through computerized adaptive testing (CAT) [24,41]. The Neuro-QOL measure was specifically developed for use in neurological populations, including MS, and – the psychometric properties of the PROMIS Depression short form have been examined in people with MS and found to be acceptable [2,46].

A recently published evidence-based guideline on the assessment and management of psychiatric disorders in individuals with MS emphasized the need for research comparing different self-report and diagnostic instruments for identifying psychiatric disorders, including MDD [34]. Therefore, the purpose of this study was to: (1) examine the correspondence between the standard diagnostic interview (SCID) and multiple self-report depression measures which are commonly used as tools for identifying MDD in MS; and (2) examine the published cutoffs for each measure and potentially identify optimal cutoffs for identifying people with MS for MDD.

2. Methods

2.1. Participants

Between September 2011 and March 2012 individuals with MS were recruited through invitation letters, print advertisements, and referrals from active research studies at the University of Washington (UW) in Seattle, WA. Individuals were sent invitation letters if they had participated in past research studies at the UW and indicated interest in future studies or were members of the UW disability registry. Individuals were required to be 18 years or older, self-report a definitive MS diagnosis by a physician, be able to read and understand English, and have access to a telephone from which they could answer sensitive questions. In order to ensure the participant pool represented all levels of depressive symptoms and included more participants with borderline depression, enrollment targets were stratified by PHQ-9 scores at the time of screening. This was done in order to ensure that the performance of the instruments was examined in participants across the whole depression continuum and individuals with no or low depressive symptoms, who are the most likely to volunteer for research studies, were not overrepresented. Initial recruitment targets were set at 200 total with 10% having no or minimal depressive symptoms (PHQ-9 score <5), 20% mild (PHQ-9 score 5–9), 30% moderate (PHQ-9 score 10–14), 30% moderate-severe (PHQ-9 score 15–19), and 10% severe (PHQ-9 score ≥20). Initial mailings to past research participants were done through random selection. However, individuals reporting a moderate to high level of depressive symptoms in past survey studies were specifically targeted in later recruitment mailings in order to try to meet recruitment goals for those strata. If a participant did not respond to the invitation recruitment mailing an attempt was made to call them approximately two weeks after the mailing. Once recruitment goals were met for a stratum, individuals scoring within that range upon screening were determined ineligible for study enrollment. All procedures, including written informed consent, were approved by the UW Human Subjects Division. Participants were paid $25 upon study completion.

2.2. Procedures

At screening, potential participants completed the PHQ-9 on the telephone in order to determine study eligibility. Those who were eligible and interested were mailed a packet containing the six self-report depression measures described below. They opened and completed the self-report measures at the scheduled time of their telephone interview. The average time between initial screening into the study and the interview (that included both the SCID and responding to self-report measures) was 10 days (range: three to 29 days). The research interviewer was available for questions but did not otherwise participate or interact with the participants while they responded to the self-reported measures. Upon completion of the self-report measures, the researcher conducted the SCID MDD module with the participant [22]. Thus, all self-report measures were completed within minutes of the SCID as they were done while the interviewer waited on the phone prior to administering the interview. Participants were then instructed to mail the self-report packet back to research staff.

2.3. Measures

2.3.1. Disease and demographic characteristics

Participants were asked to provide their age, gender, race and ethnicity, education level, and employment status. In addition, participants reported the year of their MS diagnosis and completed a self-report version of the mobility section of the Expanded Disability Status Scale (EDSS) [9] in order to estimate MS severity level.

2.3.2. Self-report depression measures

Participants completed six different paper and pencil self-report measures of depressive symptoms. The following measures were selected for inclusion: PROMIS Depression short form (SF), Neuro-QOL depression SF, PHQ-9, modified PHQ-2, BDI-FS, and the CESD. Order of administration was counterbalanced to address potential order effects [14].

2.3.2.1. PROMIS depression (PROMIS-D) SF. The Patient Reported Outcomes Measurement Information System (PROMIS®, www.nihpromis.org) depression SF version 1.0 includes eight items that were selected from the PROMIS-D item bank using CAT simulation results, item information, and content [41]. The PROMIS-D item bank was developed using item response theory (IRT), and scores on the SF are directly comparable to CAT scores or other SF scores. Unlike most depression measures, the PROMIS-D item bank does not measure behavioral and somatic indicators [41]. All eight questions are rated on a five point Likert scale (1 = never; 5 = always), and respondents are asked to recall how they felt over the past seven days. Scores are reported on a T-score metric [mean = 50; standard deviation (SD) = 10] that is

Please cite this article as: D. Amtmann, et al., A comparison of multiple patient reported outcome measures in identifying major depressive disorder in people with multiple sclerosis, J Psychosom Res (2015), http://dx.doi.org/10.1016/j.jpsychores.2015.08.007