Journal of Affective Disorders 148 (2013) 77–83

Contents lists available at SciVerse ScienceDirect

Journal of Affective Disorders

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

Research report

A comparison of the PRIME-MD PHQ-9 and PHQ-8 in a large military prospective study, the Millennium Cohort Study

Timothy S. Wells [a], Jaime L. Horton [a,*], Cynthia A. LeardMann [a], Isabel G. Jacobson [a], Edward J. Boyko [b]

[a] Deployment Health Research Department, Naval Health Research Center, San Diego, CA, USA
[b] Seattle Epidemiologic Research and Information Center, Department of Veterans Affairs Puget Sound Health Care System, Seattle, WA, USA

ARTICLE INFO

Article history:
Received 24 July 2012
Accepted 20 November 2012
Available online 14 December 2012

Keywords:
Depression
Epidemiology
Assessment/diagnosis
Measurement/psychometrics
Mood disorders

ABSTRACT

Background: In light of increased concerns about suicide in the military, institutional review boards have mandated increased scrutiny of the final item on the depression screening tool, the PHQ-9, which asks about suicidal thoughts. Since real-time monitoring of all individual responses in most observational studies is not feasible, many investigators have adopted the PHQ-8, choosing to remove the ninth item. This study compares the performance of the PHQ-8 with the PHQ-9 in a population-based sample of military or nonmilitary subjects.
Methods: The Millennium Cohort Study administers a self-reported questionnaire that includes the PHQ-9 at 3-year intervals to current and former U.S. military personnel. PHQ-9 responses of 143,705 Millennium Cohort members were investigated. Cross-sectional comparisons of the PHQ-9 and PHQ-8 and prospective analyses to detect a 5-unit change in these measures were performed.
Results: Greater than substantial agreement was found between the PHQ-8 and 9 instruments (kappas, 0.966–0.974 depending on survey cycle). There was similarly high agreement between the PHQ-8 and 9 in detecting a 5-point increase (κ=0.987) or decrease (κ=0.984) in score.
Limitations: One potential limitation of this study is that participants completed the PHQ-9, and PHQ-8 scores were extrapolated from the PHQ-9. In addition, the Millennium Cohort may not fully represent the U.S. military; though previous evaluations have shown the cohort to be a well-representative sample.
Conclusions: Since excellent agreement was detected between the PHQ-8 and PHQ-9 instruments, the PHQ-8 would capture nearly all the same cases of depression as the PHQ-9 in populations similar to the one in this study.

1. Introduction

The Primary Care Evaluation of Mental Disorders Patient Health Questionnaire (PHQ) is a standardized instrument that provides an assessment of mental health status based on scores of several health concepts (Spitzer et al., 1999, 2000, 1994). A 9-item scale from this instrument (PHQ-9) used to screen for major depressive disorder (Spitzer et al., 1999), has been shown to have high sensitivity (0.93) and specificity (0.89) (Fann et al., 2005), and correlates well with a diagnosis of depression, as outlined in the Diagnostic and Statistical Manual of Mental Disorders (2000). The ninth and final item on the scale asks about thoughts of being better off dead or hurting oneself, which is known to be related to suicidal thoughts and ideation (Corson et al., 2004). Exclusion of the final item results in the PHQ-8, which has also been shown to be a valid instrument for evaluating depression symptoms in specific populations (Kroenke et al., 2009).

The Millennium Cohort Study (Ryan et al., 2007; Smith, 2009), the largest population-based longitudinal cohort study in military history, has included the PHQ-9 as part of the standard questionnaire since the study launched in 2001. Symptoms of depression have previously been examined in this population using the PHQ-9 (Ryan et al., 2007; Wells et al., 2010), with positive screens for new-onset depression occurring in approximately 4% of men, and 8% of women at the time of the first follow-up questionnaire (Wells et al., 2010). However, in light of increased concerns about suicide in the military (Kuehn, 2009; Oquendo et al., 2005), institutional review boards have mandated increased scrutiny of this item with the intent of initiating provider referral when respondents positively endorse it. In nonclinical research settings, many investigators have adopted the PHQ-8, choosing to remove the final item that may indicate suicidal thoughts. Since real-time monitoring of more than 150,000 responses from Millennium Cohort members is not feasible, starting in 2011 the final item

* Corresponding author. Tel.: +1 619 767 4905; fax: +1 619 553 7601.
E-mail address: jaime.horton@med.navy.mil (J.L. Horton).

0165-0327/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jad.2012.11.052

was removed from the questionnaire and the study currently uses the PHQ-8 to screen for depression. Although the PHQ-8 has been shown to have similar operating characteristics as the PHQ-9 (Kroenke and Spitzer, 2002), this was done using a sample of 6000 individuals seeking treatment in primary care or obstetrics-gynecology clinics (Kroenke et al., 2001). To our knowledge, no study has validated the PHQ-8 in a large, population-based cohort, military or otherwise. The objective of this study was to compare the PHQ-8 with the PHQ-9 to understand differences in depression screening capability between these two instruments in the Millennium Cohort Study, which may be generalizable to other similar population-based studies. Unique features of this study include investigating differences in a large military population that is probably healthier than a general population sample. Also, with the large sample size, subgroups such as differences by sex can be explored to determine performance of the PHQ-8 compared with the PHQ-9.

2. Methods

2.1. Study population

The Millennium Cohort Study began collection of self-reported health outcome and exposure data in 2001, prior to the start of the operations in Iraq and Afghanistan. The Millennium Cohort currently includes over 150,000 U.S. service members who enrolled during three separate cycles (panels) between 2001 and 2008. With the goal of evaluating long-term health outcomes related to military service, participants are surveyed every 3 years throughout a 21-year planned follow-up period. Detailed descriptions of methodology for the Millennium Cohort Study have been published elsewhere (Ryan et al., 2007; Smith, 2009; Gray et al., 2002).

The first invited panel of the Millennium Cohort Study consisted of a weighted random sample of U.S. military personnel serving in October 2000, with oversampling of women, service members previously deployed to Bosnia, Kosovo, or Southwest Asia, and Reserve and National Guard members. Of the 77,047 participants who consented and enrolled, 55,021 (71%) completed the first follow-up questionnaire from 2004–2006, and 54,790 (71%) completed the second follow-up from 2007–2008. A second panel was randomly selected from military personnel with 1–2 years of service as of October 2003, with oversampling again performed for women and Marine Corps members. The second panel enrolled 31,110 consenting members from 2004–2006, 17,152 (55%) of whom completed a follow-up questionnaire from 2007–2008. Also, a third panel of 43,440 participants with 1–3 years of service as of October 2006 and oversampling of women and Marine Corps members was enrolled in 2007–2008. Demographic and military-specific data obtained from electronic personnel files include sex, birth date, highest education achieved, marital status, race/ethnicity, deployment experience in support of the operations in Iraq and Afghanistan, pay grade, service component, service branch, and duty occupations. Data for education level, marital status, deployment experience, and occupations were supplemented by self-reported data when missing from electronic files. Participants were excluded if they were missing information from the PHQ or other covariate information.

2.2. Depression

Depression was investigated using the PHQ (Spitzer et al., 1999), which is embedded in the Millennium Cohort questionnaire and provides a psychosocial assessment based on scores of several health concepts (Spitzer et al., 2000, 1994; Kroenke and Spitzer, 2002). Using a 4-point Likert scale, participants rated the severity of each depressive symptom from "not at all" to "nearly every day" during the 2 weeks prior to questionnaire completion, where a higher score indicates greater severity (Spitzer et al., 1999). Using the PHQ 9-item scale, participants screened positive for depression if they met the following two criteria: (1) responded "more than half the days" or "nearly every day" to at least five of the nine depressive items, with "thoughts that you would be better off dead or of hurting yourself in some way" being counted if present at all, and (2) one of the items endorsed is having depressed mood or anhedonia (Kroenke and Spitzer, 2002). Similarly, using the PHQ-8, which does not include the item "thoughts that you would be better off dead or of hurting yourself in some way," individuals screen positive for depression if five or more of the eight depressive symptom criteria have been present "more than half the days" or "nearly every day" in the past 2 weeks, and one of the symptoms is depressed mood or anhedonia (Kroenke et al., 2009). Using both the PHQ-8 and the PHQ-9 instruments, participants were categorized as screening positive or negative for depression. Those who screened positive on the PHQ-9 and PHQ-8 and those who screened negative on the PHQ-9 and PHQ-8 were termed "concordant positive" and "concordant negative," respectively, while the discordant group was termed the PHQ-9 positive/PHQ-8 negative group for the purposes of this study.

2.3. Covariates

Variables considered for this study mirrored those considered for the models in the first depression study conducted using Millennium Cohort data (Wells et al., 2010). Covariates included sex, birth year, education, marital status, race/ethnicity, military pay grade, branch of service, service component, occupational category, smoking status, alcohol-related problems (Yes if one or more items endorsed on the CAGE questionnaire) (Dhalla and Kopec, 2007), a positive screen for posttraumatic stress disorder (PTSD) assessed using the 17-item PTSD Checklist-Civilian Version (PCL-C) (Weathers et al., 1993), deployment experience with or without combat, and cumulative days deployed. All independent variables were evaluated at baseline. Defense Manpower Data Center (DMDC) provided deployment data to classify individuals as nondeployed and deployed with or without combat experience. Participants were considered deployed in support of the operations in Iraq and Afghanistan if they completed at least one deployment prior to their baseline questionnaire. Combat experience was determined by an affirmative response to ever witnessing at least one of the following items on the baseline questionnaire: a person's death due to war, instances of physical abuse, dead and/or decomposing bodies, maimed soldiers or civilians, or prisoners of war or refugees. Cumulative days deployed were assessed by calculating the number of days each participant was deployed prior to baseline using in and out of theater dates from data provided by DMDC.

2.4. Statistical analyses

Univariate analyses were used to describe frequencies and proportions of those who screened positive on both instruments, those who screened negative on both instruments, and those who screened PHQ-9 positive but PHQ-8 negative. To determine the degree of nonrandom agreement between the PHQ-9 and PHQ-8, the kappa statistic (Cohen, 1960) was calculated between the two instruments for each survey period (2001–2003, 2004–2006, and 2007–2008), providing a cross-sectional comparison of performance of the PHQ-9 versus the PHQ-8 at each survey cycle.