Journal of Psychiatric Research 68 (2015) 114–119

Contents lists available at ScienceDirect

Journal of Psychiatric Research

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

Early identification of treatment non-response utilizing the Patient Health Questionnaire (PHQ-9)

J. Christopher Fowler [a, b, *], Michelle Patriquin [a, b], Alok Madan [a, b], Jon G. Allen [a, b], B. Christopher Frueh [a, c], John M. Oldham [a, b]

[a] The Menninger Clinic, 12301 Main Street Houston, TX 77035, USA
[b] Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
[c] University of Hawaii, Department of Psychology, 200 West Kawili St., Hilo, HI 96720, USA

ARTICLE INFO

Article history:
Received 30 March 2015
Received in revised form
19 June 2015
Accepted 20 June 2015

Keywords:
PHQ-9
Prediction
Treatment non-response
Diagnostic efficiency

ABSTRACT

Background: Treatment non-response among high-risk, psychiatric patients exposes those suffering to suicidal risk as well as persistent social and occupational difficulties. Strategies for identification of treatment non-response are limited.
Aims: Diagnostic efficiency of a self-report, cross-cutting symptom measure was assessed as a marker of treatment non-response.
Method: 835 inpatients at a specialist psychiatric hospital completed the Patient Health Questionnaire – Depression (PHQ-9) at admission and every two weeks during hospitalization.
Results: For patients admitted with severe depression (PHQ-9 $\geq$ 20), results indicated good accuracy of 2-week PHQ-9 change score in identifying treatment non-response (AUC = 0.80, SE = 0.04, $p < .0001$; sensitivity = 85%; specificity = 73%; OR = 14.91).
Conclusions: The search for predictors of non-response to psychiatric treatment has a long and generally unfulfilled history. The PHQ-9 change score holds promise as a cost-effective test with comparable diagnostic characteristics to other medical tests.

1. Introduction

Despite decades of treatment innovation and research, rates of treatment non-response for individuals with disorders such as depression or schizophrenia approaches 50% of all patients (McEvoy et al., 2006; Newnham et al., 2007, 2010; Rush and Trivedi, 1995; Rush et al., 2006; Stroup et al., 2006; Thase et al., 2007). While major depressive disorder (MDD) continues to be among the most prevalent psychiatric disorders worldwide (Bromet et al., 2011) symptoms of depression are so pervasive among individuals with psychiatric disorders in general that depression severity has been identified as a cross-cutting domain of psychopathology (Löwe et al., 2008) and has been included as a DSM-5 cross-cutting measure of psychopathology (Narrow et al., 2013). Persistent depressive symptoms (regardless of a formal MDD diagnosis) expose those affected to continued suffering as well as social and occupational difficulties – factors that increase the likelihood of treatment dropout and suicidal (National Commitee for Quality Assurance, 2007). As such, depression severity is an excellent marker of treatment non-response among heterogeneous patient populations encountered in outpatient and inpatient settings. Accurate identification of individuals at risk of persistent depressive symptoms could lead to more rapid implementation of augmentation strategies to optimize clinical outcomes (Dunner et al., 2014; Holtzheimer and Mayberg, 2012). Biologically-based strategies for early identification of treatment resistant depression (TRD) have been explored including: neuroimaging (Konarski et al., 2009; Mayberg et al., 1997; Siegle et al., 2006, 2012; Stroup et al., 2006), genetic (Ising et al., 2009, Uher et al., 2009), electroencephalographic (Leuchter et al., 2009), and immune-related measures (Raison et al., 2013). These approaches have noteworthy appeal due to their objectivity; however, the significant cost, time, and diagnostic skill needed to correctly implement these techniques limit their application at present. A clinically-based test with the potential to identify treatment non-response across a spectrum of disorders is an attractive option because self-report

* Corresponding author. The Menninger Clinic, 12301 Main Street Houston, TX 77035, USA.
E-mail address: cfowler@menninger.edu (J.C. Fowler).

http://dx.doi.org/10.1016/j.jpsychires.2015.06.018
0022-3956/

measures require minimal financial investment, are quickly completed, and can be administered via computer surveys.

The current study leveraged multi-wave data from the Menninger Outcomes Project (Allen et al., 2009; Fowler et al., 2013), a patient-centered quality improvement and research endeavor, to explore the possibility of an early identifier of treatment non-response. Based on similar patient-centered approaches that demonstrate improved outcomes for psychiatric patients (Howard et al., 1996; Whipple et al., 2003; Newnham et al., 2010; Crits-Cristoph et al., 2012; Shimokawa et al., 2010) patient-profiling allows Menninger practitioners to compare individual patient trajectories with the average treatment trajectory, prompting clinical consultation when flat or deteriorating courses are identified. The next step in the quality and outcomes progression was to build accurate, low-cost systems for early identification of non-response.

The current study assessed the viability of a screening method for predicting treatment non-response by assessing the diagnostic efficiency of 2-week change score quantified by a single self-report depression measure (Patient Health Questionnaire, PHQ-9; Löwe et al., 2004). The PHQ-9 was selected because it is a simple, easily scored measure that exhibits excellent psychometric properties (including sensitivity to change), and has established severity cut-points such that practicing clinicians can identify treatment non-response. Using a large ($N = 835$) inpatient sample with histories of treatment non-response, we assessed the potential utility of the PHQ-9 to predict treatment non-response at point-of-discharge.

2. Methods

2.1. Participants

Participants were 835 adult patients consecutively admitted (July 2012–August 2014) to the Menninger Clinic, a psychiatric hospital specializing in complex treatment resistant neuropsychiatric disorders. Typical lengths of stay in the hospital range from four to eight weeks. Treatment included psychiatric and medical care, medication management, individual and group psychotherapy integrating third-wave cognitive-behavioral, mentalization-based, and dialectical behavioral therapies, psycho-educational groups, family work, continuous nursing care, health promotion, physical exercise, and leisure-time social/recreational activities.

Gender distribution was relatively even with 54.6% female. Average age was 35.3 years ($SD = 14.9$). Participants were Caucasian (90.5%), multiracial (5.3%), African American (1.6%), Asian (1.8%), American Indian (0.5%) and Pacific Islander (0.4%). Sixty patients identified as being of Hispanic or Latino ethnicity (7.2%). Education level was above the national average with 83% indicating some college experience. The majority (59.2%) of participants were not working in the 30 days prior to admission. The single exclusion criterion was length of hospitalization greater than 56 days (the upper limit of the treatment program).

2.2. Procedures

Data were collected as part of the hospital's ongoing Adult Outcomes Project to assess treatment response (Allen et al., 2009, Clapp et al., 2013, Fowler et al., 2013). All measures used in the current study were collected within 72 h of admission and the PHQ-9 was re-administered every 2 weeks and at point of discharge. Two week interval was chosen based on prior findings (Clapp et al., 2013) that 2-week assessments of depression yielded significant rates of change with inpatients with histories of treatment resistant conditions. Assessments were conducted via hospital-wide web survey on laptop computers. This project was a hybrid clinical quality and research outcomes project, conducted with all patients; accordingly, all assessments were designed and implemented as an element of routine clinical care and integrated into treatment planning and monitoring of progress such that less than 1% of patients declined participation. Patients and their treatment teams were provided with profile scores and feedback within 24 h of each assessment point with the expressed intention that individual patient profiles would be used to inform treatment decisions. Patients and teams were informed that the findings would be used to evaluate the overall effectiveness of treatment and for research purposes. Use of the project's data was approved by Baylor College of Medicine's Institutional Review Board.

2.3. Measures

Demographic variables and history of psychiatric hospitalization and psychiatric service usage were assessed using a standardized patient information survey (Fowler et al., 2013). Patient Health Questionnaire-Depression (PHQ-9) is a 9-item screen for depression severity with excellent internal consistency, construct validity, and test-retest reliability (Kroenke and Spitzer, 2002) and has demonstrated robust sensitivity to change (Löwe et al., 2006). PHQ-9 depression severity scores of 5, 10, 15, and 20 are indicative of mild, moderate, moderately severe, and severe depression, respectively (Kroenke et al., 2001). In the current sample internal consistency of the PHQ-9 was high (Cronbach's $\alpha = .90$). Psychiatric disorders including personality disorder diagnoses were assessed using the research versions of the Structured Clinical Interview for DSM-IV Disorders (SCID-I/II). The SCID-I (First et al., 1997b) and SCID II (First et al., 1997a) were performed by master's level researcher assessors after reviewing pertinent psychiatric and psychosocial evaluations as well as consulting with the attending psychiatrist.

2.4. Data analysis

All analyses were conducted in IBM SPSS version 22.0. Receiver Operating Characteristics (ROC) analyses were carried out to obtain the area under the curve (AUC) and standard error (SE) using the non-parametric method to assess the accuracy of 2-week PHQ-9 change score in predicting treatment non-response (defined as a discharge PHQ-9 score $\geq$ 15; moderately severe depression or greater). While Youden's index score (Youden, 1950) is commonly computed to determine the optimal diagnostic cutoff score for identifying the presence of a binary outcome, we employed diagnostic efficiency statistics because this method provides more granular information in determining the optimal cut points for screening and diagnostic tools (Streiner, 2003).

Diagnostic efficiency statistics (Kessel and Zimmerman, 1993) were calculated for four groups: all patients (N = 835), patients with at least moderate (PHQ-9 score $\geq$ 10: n = 506), moderately severe (PHQ-9 score $\geq$ 15: n = 374), or severe (PHQ-9 score $\geq$ 20: n = 207) depression at admission. Within these groups, diagnostic efficiency statistics were compared between those with PHQ-9 change scores (from admission to week 2 of inpatient psychiatric treatment). Given the relatively poor diagnostic efficiency statistics of all possible change scores, only change scores of $\geq$6 and $\geq$ 7 are presented. Again, treatment non-response was defined as having a discharge score of at least moderately severe depression (i.e.,PHQ-9 score $\geq$ 15). These comparisons determine the PHQ-9's ability to differentiate treatment non-response patients from those demonstrating an end point response of moderate, mild, or no depression. Five statistics were calculated: 1. Sensitivity (SN: the ability of a "positive" test result to correctly identify treatment non-response); 2. Specificity (SP: the ability of a "negative" test result to correctly identify those individuals without treatment non-response); 3.