The recovery index: A novel approach to measuring recovery and predicting remission in major depressive disorder

Waguih William IsHak[a],*, Wes Bonifay[b], Katherine Collison[c], Mark Reid[d], Haidy Youssef[d], Thomas Parisi[d], Robert M. Cohen[e], Li Cai[f]

[a] Department of Psychiatry at Cedars-Sinai Medical Center and David Geffen School of Medicine at UCLA, Los Angeles, California, United States
[b] College of Education, University of Missouri, United States
[c] Department of Psychology, Purdue University, United States
[d] Department of Psychiatry at Cedars-Sinai Medical Center, United States
[e] Department of Psychiatry at Emory University School of Medicine, United States
[f] UCLA Graduate School of Education and Information Studies, United States

ARTICLE INFO

ABSTRACT

Background: Clinicians view "recovery" as the reduction in severity of symptoms over time, whereas patients view it as the restoration of premorbid functioning level and quality of life (QOL). The main purpose of this study is to incorporate patient-reported measures of functioning and QOL into the assessment of patient outcomes in MDD and to use this data to define recovery.
Method: Using the STAR*D study of patients diagnosed with MDD, this present analysis grades patients' MDD severity, functioning level, and QOL at exit from each level of the study, as well as at follow-up. Using Item Response Theory, we combined patient data from functioning and QOL measures (WSAS, Q-LES-Q) in order to form a single latent dimension named the Recovery Index.
Results: Recovery Index – a latent measure assessing impact of illness on functioning and QOL – is able to predict remission of MDD in patients who participated in the STAR*D study.
Conclusions: By incorporating functioning and quality of life, the Recovery index creates a new dimension towards measuring restoration of health, in order to move beyond basic symptom measurement.

1. Introduction

The World Health Organization (WHO) defines health as "a state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity" (World Health Organization (WHO), 1948), suggesting that the ultimate goal of healthcare interventions ought to be the restoration of physical and psychosocial well-being, in addition to the reduction of symptoms. Nonetheless, most intervention studies measure outcomes purely in terms of changes in symptom presentation, a nosological tautology that may ignore functional impairment or other alterations in quality of life (QOL) if these are not featured in the disease definition. The psychiatric community, in particular, has struggled with how to define restoration of mental health during treatment of illnesses, such as major depressive disorder (MDD; Keller et al., 1982; Kivelä, et al. 2000; IsHak et al., 2002, 2011, 2012, 2013; Klein et al., 2004; Insel and Scolnick, 2006; Holma et al., 2008; Kessler et al., 2009; Gree et al., 2010; Galderisi et al., 2015).

Clinicians or researchers often define recovery from MDD as the reduction in symptom severity for a pre-determined period of time (Galderisi et al., 2015). Patients often define recovery as the return to baseline "normal" self, in which they regain the ability to function and to derive satisfaction and fulfillment from life activities (Zimmerman et al., 2006). Consistent with the WHO definition, the essence of recovery for patients is restoration of their QOL. Although improved QOL is correlated with symptom reduction, clinicians and researchers should include measures of psychosocial and physical functioning, in addition to symptom inventories, in order to assess patient recovery more completely.

This study aims at incorporating patient-reported measures of functioning and QOL into the 'Recovery Index' (RI) for the assessment of patient outcomes. We employed existing measures of QOL but significantly improved their utility by using item response theory (IRT) (Hambleton and Swaminathan, 1985). We tested our theory that the RI is associated with recovery by applying existing data from the

Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study, which is an examination of MDD treatment through four different treatment levels with patients moving to the next level of treatment if remission/response is not achieved (Rush et al., 2004). We premise that a measure incorporating functioning and QOL, such as the RI, can account for recovery outcomes as meaningfully as symptom measures, and that the RI may be useful for both psychiatric and physical illnesses.

2. Methods

2.1. Study population

The authors obtained a National Institute of Mental Health (NIMH) Data Use Certificate for the STAR*D Pub Ver3 dataset, containing data on acute treatment and follow-up of treatment-seeking 18–75 year old (mean age=42.6; SD=13) outpatients (80.9% Caucasians and 62.8% females) with a primary diagnosis of MDD (n=4041). The STAR*D study was funded by the NIMH and was conducted at 18 primary care and 23 psychiatric care settings in the United States, from 2001 to 2007, after obtaining institutional review board approval and informed consent from patients, and by following the principles outlined in the Declaration of Helsinki. Full details of the study methodology are described elsewhere (Fava et al., 2003; Rush et al., 2004). Patients who completed treatment at each level were enrolled in follow-up for 12 months with periodic outcome measurement at 3, 6, 9, and 12 months. In the present analysis, participants had complete data for each of the outcome measures detailed below, at exit from each level of the study, and at follow-up.

2.2. Measures

2.2.1. MDD severity
We assessed severity of depression symptoms at time of entry, and at the end of each level of STAR-D using the Quick Inventory of Depressive Symptomatology-Self Report (QIDS-SR) (Rush et al., 2003). QIDS-SR scores range from 0 (not depressed) to 27 (most severe depression). We defined remission as a score of 5 or less.

2.2.2. Functioning
We assessed functional impairment at time of entry, and at the end of each level of STAR-D using the Work and Social Adjustment Scale (WSAS) (Mundt et al., 2002). WSAS scores range from 0 (no impairment) to 40 (severe impairment). WSAS has fairly strong psychometric properties, with a Cronbach's alpha ranging from 0.70 to 0.94 and a test-retest reliability of r =0.73 (Mundt et al., 2002). Scores above 20 indicate moderate to severe impairment, scores between 10 and 20 indicate significant impairment, and scores below 10 are considered subclinical (Mundt et al., 2002).

2.2.3. QOL
We assessed QOL, satisfaction and fulfillment with activities at time of entry, and at the end of each level of STAR-D using the Quality of Life Enjoyment and Satisfaction Questionnaire (Q-LES-Q) (Endicott et al., 1993). Q-LES-Q scores range from 0 to 100, where lower scores indicate lower quality of life. The measure demonstrates strong psychometric properties (Cronbach's alpha=0.90; test-retest reliability=0.74). Norming data obtained from community samples show an average Q-LES-Q score of 78.3 (SD=11.3). Scores within 10% of this average are considered "within-normal" (Q-LES-Q > =70.47) (Endicott et al., 1993; Rapaport et al., 2005; Schechter et al., 2007). Severe impairment is defined as Q-LES-Q scores of at least 2 SDs below the community norm average (i.e., 55.7 or lower signifies severe impairment (Endicott et al., 1993; Rapaport et al., 2005; Schechter et al., 2007).

2.3. IRT methods

Item response theory comprises an assortment of nonlinear statistical models that are designed to investigate the psychometric structure of a set of categorically scaled items (Embretson and Reise, 2000; Thissen and Wainer, 2001 for overviews of IRT; Chang and Reeve, 2005 for the use of IRT in patient-reported health outcomes measurement). For the current study we are estimating a graded response model (GRM) (Samejima, 1969), a polytomous IRT model used on ordinal response data that measure a latent trait (Thissen and Wainer, 2001). We are combining patient data from functioning and QOL measures (WSAS, Q-LES-Q) in order to form a single latent RI dimension.

Prior to conducting IRT analyses, we recoded WSAS item responses in the same direction as the Q-LES-Q (higher values are associated with greater social adjustment) and we removed nine cases in the STAR*D Pub Ver3 data that were missing data. Both instruments demonstrated strong reliability (WSAS α=0.85; Q-LES-Q α=0.87) and internal consistency, so no items were dropped.

We explored three possible structures of the RI: a unidimensional model, wherein all items represented a single latent construct; a two-dimensional structure, in which the WSAS and Q-LES-Q items were treated as separate correlated traits; an item bifactor model, consisting of a global dimension (primary factor) and several smaller dimensions (or specific factors; see Gibbons and Hedeker, 1992). These nested structures were evaluated according to relative model fit, as indexed by the −2loglikelihood (−2LL) value (where lower −2LL indicates better relative fit to the data).

2.4. Statistical Methods

After specifying the RI structure, we converted raw summed scores on the WSAS and Q-LES-Q to IRT-scaled scores. Using a recursive algorithm developed by Lord and Wingersky (1984), generalized by Thissen and Wainer (2001), and updated more recently for bifactor models (Cai et al., 2011), the likelihood of a particular summed score is found by summing the likelihoods of each of the response patterns that produces that specific score.

Next, we assessed the degree to which these latent recovery trait scores were associated with patient remission or relapse. We used structural equation modeling (SEM)[1] to assess the direct and indirect effects of RI and mean-centered covariates (QIDS score at intake, sex, and Hispanic ethnicity) on these outcomes. To address missing data, we based parameter estimation on full information maximum likelihood using an EM algorithm with Monte Carlo integration.

We expressed summary values as means (SD) for continuous variables and frequencies (%) for categorical variables. We considered tests with *p*-values less than or equal to 0.05 to be statistically significant. We performed analyses using IRTPRO software, version 2.1 (Cai et al., 2013) and Mplus software, version 7.0 (Muthén and Muthén, 2012).

3. Results

3.1. IRT Analysis

The item-level descriptive and summed-score statistics of the WSAS and Q-LES-Q for this patient sample (listwise complete n=3,695) are presented in Table 1.

In GRM IRT analysis, the bifactor model provided significantly better fit to the data (−2LL=209,188.06) than either the two-dimen-

---

[1] The latent IRT-scaled scores were included as observed values in the SEM; this method (while technically path analysis) is referred to in the literature as the two-step approach to embedding IRT scores in a structural model (e.g., Lu et al., 2004; Kim and Nicewander, 1993).