



HHS Public Access

Author manuscript

Adm Policy Ment Health. Author manuscript; available in PMC 2016 December 13.

Published in final edited form as:

Adm Policy Ment Health. 2013 March ; 40(2): 76–86. doi:10.1007/s10488-011-0376-6.

Development and Validation of the Individual Burden of Illness Index for Major Depressive Disorder (IBI-D)

Waguih William IsHak,

Department of Psychiatry and Behavioral Neurosciences, Cedars-Sinai Medical Center and David Geffen School of Medicine at UCLA, 8730 Alden Drive, Thalians W-157, Los Angeles, CA 90048, USA

Jared Matt Greenberg,

David Geffen School of Medicine at UCLA and Keck School of Medicine at University of Southern California (USC), Los Angeles, CA, USA

Tammy Saah,

Emory University, Atlanta, GA, USA

Shakiba Mobaraki,

David Geffen School of Medicine at UCLA, Los Angeles, CA, USA

Hala Fakhry,

Department of Psychiatry, Faculty of Medicine, Cairo University, Cairo, Egypt

Qian (Vicky) Wu,

Biostatistics, UCLA School of Public Health, Los Angeles, CA, USA

Eunis Ngor,

Biostatistics, UCLA School of Public Health, Los Angeles, CA, USA

Fei Yu, and

Biostatistics, UCLA School of Public Health, Los Angeles, CA, USA

Robert M. Cohen

Department of Psychiatry and Biobehavioral Sciences, UCLA, Los Angeles, CA, USA

Psychiatry and Behavioral Neurosciences, S. Mark Taper Foundation Imaging Department, Cedars-Sinai Medical Center, Los Angeles, CA, USA

Abstract

This study aims at developing a single numerical measure that represents a depressed patient's individual burden of illness. An exploratory study examined depressed outpatients ($n = 317$) followed by a hypothesis confirmatory study using the NIMH STAR*D trial ($n = 2,967$). Eigenvalues/eigenvectors were obtained from the Principal Component Analyses of patient-reported measures of symptom severity, functioning, and quality of life. The study shows that a single principal component labeled as the Individual Burden of Illness Index for Depression (IBI-D) accounts for the vast majority of the variance contained in these three measures providing a

numerical z score for clinicians and investigators to determine an individual's burden of illness, relative to other depressed patients.

Keywords

Burden of illness; Major depressive disorder; Major depression; Outcome measurement; Quality of life; STAR*D

Introduction

The epidemiological concept of burden of disease is a familiar one that is meant to encompass the negative impact of a given disease on a societal scale, taking into account such factors as morbidity, mortality and direct and indirect costs including loss of productivity, among others. Several scales and statistics have been developed to assign a numerical score to this burden for purposes of health resource allocation and other public health and epidemiologic concerns. Among the most widely used are Quality Adjusted Life Years (QALY) (Zeckhauser and Shepard 1976; Sassi 2006) and Disability Adjusted Life Years (DALY) (Murray 1994; Anand and Hanson 1998). QALY is a health measure that incorporates both quantity and quality of life. Perfect health for a given year is given a score of 1 whereas death is scored as 0, multiplied by health-related quality of life (QOL) weight in this given year, e.g., $QALY = 1 \times QOL$. The DALY was created, based on QALY, by Murray and Lopez to assess population-based disease burden in the Global Burden of Disease project (Murray and Lopez 1997; Melse et al. 2000; Havelaar 2007; Gold et al. 2002). The DALY measures the difference between a current situation and an ideal situation where everyone lives up to the age of the standard life expectancy, and in perfect health (Prüss-Üstün et al. 2003). DALY is calculated by adding the number of years lost due to premature death, to the number of years lived with disability or illness. There are some significant differences between QALY and DALY (Sassi 2006). DALY includes a method of age weighting with different weights assigned to different age groups thus incorporating age of onset and duration of illness. Moreover, disability and quality of life weights are also different between the two measures.

There is no widely accepted analog of the above population-based measures for clinicians and researchers concerned with an individual patient's burden of illness, a distinctly different concept. Population-based burden of disease measures such as QALY and DALY have been widely applied to help set health service and health research priorities, identify under-served groups, compare outcome of population interventions, and assess cost-effectiveness of health programs. However, these measures are not fully applicable to the individual patient's experience of the full burden of illness for a number of reasons. QALY and DALY are difficult to apply during the acute phase of illness, as they tend to measure the burden of the overall course of the illness. Moreover, QALY and DALY often focus on life expectancy, making them harder to apply in mental health as opposed to physical health, where morbidity prevails over mortality for the most part. The most significant limitation is that QALY and DALY calculate the burden of disease for a population and not an individual patient. Major depressive disorder (MDD), for instance, is an illness for which the enormous

societal burden is now well established, with current estimates placing depression as the second leading contributor to global burden of disease by the year 2020 [World Health Organization (WHO 2008)]. The individual burden of illness, however, remains relatively undefined. Accurate assessment of the burden of illness is one of the important needs in depression research as identified by the NIMH Affective Disorders Workgroup published in this journal in 2002 (McGuire et al. 2002). We conceptualize the individual burden of illness to be a reflection of the impact of the illness, i.e., suffering due to symptom severity (intensity, frequency, duration), functioning (occupational, social, and leisure activities), and quality of life (patient's satisfaction with health, occupational, social, and leisure activities) (IsHak et al. 2002, 2011), as shown in Fig. 1. Traditional initial assessment and outcome measurement in psychiatry has been focused on symptom severity. More recently, research findings have emphasized the importance of adding functioning and quality of life measures to adequately capture the full impact of depression and its treatment, including the most recently published American Psychiatric Association practice guideline for the treatment of patients with MDD (American Psychiatric Association (APA) 2010).

To establish empirical evidence in support of the above conceptualization and to develop a numerical score for the individual burden of illness we turned to Principal Component Analysis (PCA). The central concept of PCA is representation or summation, i.e., PCA is designed to evaluate whether the variance observed in measured variables can be captured with a smaller set of variables. In this instance we asked whether a substantial portion of the variance observed in the following patient-reported outcomes: Quick Inventory of Depressive Symptomatology-Self Report (QIDS-SR) (Rush et al. 2003) for depressive symptom severity, Work and Social Adjustment Scale (WSAS) (Mundt et al. 2002) for functioning, and the Quality of Life Enjoyment and Satisfaction Questionnaire—Short Form (Q-LES-Q) (Endicott et al. 1993) for quality of life, in depressed patients, could be significantly captured in one variable, i.e., a single principle component that we would label as the Individual's Burden of Illness in Depression (IBI-D). If that were the case, we expected to see that each of the scales shared substantial variance, i.e. variance explained by IBI-D, and smaller unique variances for each of the scales, i.e. the variances for each scale not explained by the concept of Burden of Illness.

We propose that IBI-D will prove to be a useful single measure for capturing the full multi-dimensional impact of depression on an individual in contrast to either using currently available single measures that traditionally focus primarily on one aspect, e.g., symptom severity, or using multiple independent measures with overlapping bases. IBI-D has the potential of providing more accurate assessments of the outcome of treatment interventions as well as informing clinical practice decisions.

Methods

Overview

An exploratory Principal Component Analysis (PCA) was initially conducted to evaluate the likelihood that a single principal component (IBI-D) could account for the majority of variance obtained from individual rating scales of symptom severity, functioning and quality of life in depressed patients who participated in the Cedars-Sinai Psychiatric Treatment

Outcome Registry (CS-PTR). To confirm the findings from the initial PCA and to determine the broader applicability of the findings a PCA was performed using the data from the Sequenced Treatment Alternatives to Relieve Depression trial (STAR*D) (Rush et al. 2004; Fava et al. 2003). The QIDS-SR, WSAS, and Q-LES-Q were selected for inclusion in the PCAs based on their demonstrated reliability and validity, wide acceptance and familiarity among clinicians, extensive use in the research literature, and ease of administration. Details of each measure are depicted in Table 1. As lower Q-LES-Q scores are associated with greater burden of illness whereas higher QIDS-SR and WSAS scores are associated with greater burden of illness, Q-LES-Q scores were first inverted by subtracting individual scores from 100. To reduce the likelihood that any one scale would have undue influence on the PCA analysis, individual scores from each scale were converted to z-scores prior to performance of PCA.

Populations Studied

Development Sample

All patients presenting for psychiatric evaluation and treatment at the Cedars-Sinai Medical Center were enrolled in the Cedars-Sinai Psychiatric Treatment Outcome Registry (CS-PTR), an IRB-approved ongoing research study to track the outcome of psychiatric interventions in a naturalistic clinical setting. Those patients who agreed to participate were initially evaluated using the Mini International Neuropsychiatric Interview (Sheehan et al. 1998). The evaluations were performed by psychiatric residents, psychology interns, and social work interns who had participated in a three-session diagnostic and structured interviewing. Each interview was monitored by a faculty member through a one-way mirror. A final diagnosis was made using consensus techniques by a team of mental health professionals led by a faculty member. Self-report measures were collected at the time of initial assessment and then on a quarterly basis for the following:

(1) Symptom Severity measures: QIDS-SR (Rush et al. 2003) for depressive symptoms, and Beck Anxiety Inventory (Beck et al. 1988) for anxiety symptoms; (2) Functioning measures: WSAS (Mundt et al. 2002), Global Assessment of Functioning (American Psychiatric Association (APA) 1994), Sheehan Disability Scale (Sheehan 1983), and Endicott Work Productivity Scale (Endicott and Nee 1997); and (3) Quality of Life measure: Quality of Life Enjoyment and Satisfaction—Short Form (Q-LES-Q) (Endicott 1993). All data were de-identified and entered into a secure database maintained by a data manager. Study staff monitored data completeness and integrity. For purposes of this study, 317 consecutive outpatients who had a primary diagnosis of major depressive disorder (MDD) and presented for initial outpatient evaluation between 2005 and 2008 were analyzed.

Validation Sample

The STAR*D trial is the largest study of MDD in modern times funded by the National Institute of Mental health (NIMH), utilizing sequential steps of antidepressants and measurement-based care. The details of the study methodology are described elsewhere (Rush et al. 2004; Fava et al. 2003). Briefly, STAR*D was conducted at 18 primary care settings and 23 psychiatric care settings in the United States, from July 2001 through April

2004, and managed to enroll 4,041 treatment-seeking outpatients from 18 to 75 years old who had a diagnosis of MDD. For the purpose of this study, the STAR*D patient sample was derived from the STAR*D data set for subjects who were at the entry point of Step 1 of the study. We analyzed the QIDS, WSAS, and Q-LES-Q data collected by the Interactive Voice Response system prior to starting citalopram (complete data $n = 2,967$). The authors obtained NIMH Data Use Certificate to use the STAR*D dataset (STAR*D Pub Ver1).

Principal Component Analysis vs. Factor Analysis

Because our goal was not to search for latent constructs in the data, i.e., that there are linear combinations of underlying factors (latent constructs) and unique factors responsible for the actual responses (scores) on the rating scales we did not believe that factor analysis was the appropriate multivariate technique for us to use. Rather our goal was to identify whether a single (component) score could capture the vast majority of the variance in those rating scales representing three global aspects of depression: symptom severity, functioning, and quality of life. PCA is meant for this type of analysis and its ease of use and absence of multiple flavors makes it easy to reproduce results across a range of statistical software packages and users. However, we recognize that PCA is not a true method of factor analysis and there is disagreement among statistical theorists about when it should be used, if at all. As delineated by Costello and Osborne (2005, and their cited references) “Some argue for severely restricted use of components analysis in favor of a true factor analysis method. Others disagree, and point out either that there is almost no difference between principal components and factor analysis, or that PCA is preferable” (Costello and Osborne 2005). Therefore in addition to the PCA we ran an exploratory factory analysis on the data in both Statistica and in R using AFDM (one of the flavors of factor analysis) and found a single factor with an Eigenvalue of 2.33 accounting for 77.76% of the variance with factor loadings (un-rotated) of 0.902 for Quality of Life, 0.896 for Symptom Severity and 0.847 for Functioning, numbers and results essentially identical to those we obtained with PCA and reported here.

Principal Component Analysis Procedure

Before conducting the PCA analysis on the 317 patients from CS-PTR, we evaluated the Pearson product moment correlations among the three rating scales. If the correlations among the three rating scales were all very high this would suggest that all three rating scales were measuring the same clinical phenomenon with no unique variance among the scales. Therefore, performing a PCA would be of little value as using data from any one rating scale would be the equivalent of a principle component, and data from the other two scales would be superfluous. Alternatively if there were very little shared variance among the scales, that is, the correlations among the three rating scales were very small, it would suggest that each rating scale was measuring uniquely different clinical phenomena, suggesting that a PCA would not be of value. We created a correlation matrix for QIDS-SR, WSAS, and Q-LES-Q to examine the appropriateness of a PCA and the likelihood that a single principle component is likely to account for most of the variance shared among the variables. The quantitative change in going from the correlation matrix to the partial correlation matrix is used as the basis of the Kaiser–Meyer–Olkin test, which has a range of 0 to 1. Measures above 0.6 for an individual item, in this instance rating scale and for all

items together, support the use of a PCA or factor analysis (Hair et al. 2006; Gorsuch 1983). The data reported is from a PCA performed using the open source R programming language version 2.10.1 (The R foundation for Statistical Computing); however, essentially identical results were obtained when checked using both SAS version 9.2 (Cary, NC) and Statistica® (StatSoft, Tulsa, OK) packages.

Calculation of IBI-D Index

Values for the weighted means and weighted standard deviations of each rating scale were used to determine an individual patient's IBI-D index. These values were obtained from the two sample populations weighted by size. To obtain an individual's IBI-D index the individual's score on each scale must be converted to a *z*-score by first subtracting the scale's mean and dividing by the SD of the scale:

$$\begin{aligned} \text{For the QIDS: } zQIDS - SR &= (QIDS - SR - 15.6) / 5.1. \\ \text{For the WSAS: } zWSAS &= (WSAS - 23.9) / 9.3 \end{aligned}$$

For the QLES-Q we correct for the inversion of the scale by subtracting score from a 100 and inverting the sign to obtain the following formula for the *z*-score: $zInvQ-LES-Q = (41.4 - QLES-Q) / 15.3$.

The *z* scores are then substituted into the following formula that uses the weightings (factor loadings) obtained from the PCA to obtain the IBI-D index and then dividing by the overall SD of the IBI-D index obtained from our two sample populations.

$$IBI-D \text{ index} = [0.57 * (zQIDS - SR) + 0.58 * (zWSAS) + 0.59 * (zInvQ - LES - Q)] / 1.51.$$

Results

The demographic and clinical characteristics of the development of the depressed patient sample obtained from the Cedars-Sinai Medical Center clinic (CS-PTR) and the validation sample obtained from the STAR*D trial are presented in Table 2.

The comparison of individual rating scales between the CS-PTR and STAR*D samples show no significant differences in ratings of depressive symptoms severity (QIDS-SR), functioning (as measured by WSAS), and quality of life (as measured by Q-LES-Q), between the two samples as shown in Table 3. The correlation matrix, as shown in Table 4, for QIDS-SR, WSAS, and Q-LES-Q, demonstrated modest correlations of nearly equal strength among the three scales with substantial reductions in the partial correlation matrix. This strongly supported the likelihood that a single principle component accounted for most of the variance shared among the variables.

When performed, the PCA on the data from CS-PTR sample confirmed the importance of a single principle component: PC1 that we label as IBI-D. The analysis summarized in Table 5 shows that the first principal component extracted (PC1) had an eigenvalue of 2.33, and accounted for 77.8% of the variance in the data set, with subsequent extracted components

demonstrating a dramatic fall off in eigenvalues to 0.409 and 0.258. Two criteria are generally used to determine the likely validity and reliability of principal components, i.e., if they account for less variance than any of the original variables and their location with respect to a bend in the slope of the scree plot, a plot of the variance accounted for by a PC on the y -axis vs. the order of PCs from highest variance accounted for to lowest (Fig. 2a, b). Since the variance of a standardized variable is 1 principal components with eigenvalues less than one would be omitted by the first criterion. In this instance both approaches to determining the validity and reliability of PCs suggest that only PC1 is likely to be a reliable and valid PC. The eigenvector for PC1 shows that the direction of the vector is a linear summation of near equal contributions from each of the three rating scales. Further correlations (loadings) between rating scale scores and factor scores support that a very large percentage of the variance in each rating scale is accounted for by PC1 (Fig. 3).

As the correlation and partial correlation matrices for the 2,967 patient STAR*D sample had a similar pattern as those based on the CS-PTR sample, a second PCA was justified and performed with almost identical results to those found in the initial PCA, i.e., an eigenvector for PC1 with almost identical directionality to the initial PCA with an eigenvalue of 2.29 that accounted for 76.2% of the total variance of the three rating scales followed by eigenvalues of 0.403 and 0.309 for PC2 and PC3 (Table 5).

Using a weighted average from the two samples to determine the eigenvector, means and standard deviations leads to the following formula for calculation of the z score for IBI-D, i.e., the IBI-D index:

$$IBI-D \text{ index} = [0.57 * (zQIDS - SR) + 0.58 * (zWSAS) + 0.59 * (zInvQ - LESQ)] / 1.51$$

where

$$\begin{aligned} zQIDS - SR &= (QIDS - SR - 15.6) / 5.1, \\ zWSAS &= (WSAS - 23.9) / 9.3, \quad \text{and} \\ zInvQ - LES - Q &= (41.4 - Q - LES - Q) / 15.3, \end{aligned}$$

As the IBI-D index is based on a z score, one can readily calculate an individual's burden of illness relative to other depressed patients. For example an IBI-D index of 0 indicates that an individual's burden of illness is the average burden of illness for an individual seeking treatment, an IBI-D index of -2 indicates that 98% of depressed patients seeking treatment have a higher burden of illness whereas an index of 2 indicates that the burden of illness in this patient is only exceeded by 2% of depressed patients.

Discussion

Data from the WHO, APA Practice Guidelines, ICD-10, and DSM-IV, define mental illness or disorder as experiencing symptoms, signs, or behaviors resulting in significant distress and/or impairment of functioning. 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”, implying that quality of life is an essential aspect of health. In this

report we suggest the need for a burden of illness index for depression (IBI-D) at the level of the individual patient that parallels population based burden of illness scales incorporating symptom severity, impairment of functioning, and quality of life. We propose the IBI-D index (numerical score) based on a PCA analysis of three previously validated and reliable instruments, the QIDS-SR (symptom severity), WSAS (functional impairment), and Q-LES-Q (quality of life), all of which we conceptualize as contributing to burden of illness, to meet this need. Some may question whether it would be better to omit symptom severity from the index as the effect of symptoms might be fully accounted for by measures of functionality and quality of life; however, we have presented correlation and partial correlation data across two depressed population samples demonstrating that is not the case.

That these scales (variables) are related to each other is not surprising as symptoms of depression surely contribute to functional impairment (Kessler et al. 2003) and quality of life (Fleck et al. 2005), but it is important to note that the strength of the correlations among the scales may vary with socioeconomic circumstances, genetic background, ethnicity, and with time. For example, when MDD is treated to response (50% reduction in symptom severity) or remission (reduction in severity to a minimum), functioning and quality of life may lag behind symptom improvement (Angermeyer et al. 2002). More importantly quality of life and functioning measures by themselves may not sufficiently weight state vs. trait issues, i.e., individuals may differ in quality of life measures independent of depression. For example using the IBI-D in patients with identical ratings on quality of life (Q-LES-Q) and “disability” (WSAS), but with higher symptom severity scores (QIDS-SR) will be determined to have a higher burden of illness, that is the index will attribute their present ratings with respect to disability and quality of life as more likely the result of a greater burden of illness, i.e., depression rather than the result of other factors than disease burden. These other factors could include distinguishing traits among the patients as well as the impact of each patient's circumstances, e.g., work situation, relationships, etc. Conversely patients with identical disease severity measures who differ on quality of life and disability measures should and would be viewed as having varying burdens of illness. Disproportionate burdens might, for example, well arise from circumstances patients find themselves in, such as social support or the ability to return to a previous job.

Patients' prospective studies confirm the importance of considering all three aspects of depression: when considering positive mental health: a return to one's perceived self, and a return to baseline level of function, in addition to absence of depressive symptoms (Zimmerman et al. 2006). Depression remission is often defined in terms of symptom resolution, which does not usually (while it should), include consideration of functioning level and quality of life (Rush et al. 2006; Zimmerman et al. 2008). Thus, although the above three components, as captured in the above three scales are correlated, each scale still captures different aspects of the illness as shown in both research (Rapaport et al. 2005) and clinical settings (Trivedi et al. 2006) as well as our own analyses.

The data that we report in this manuscript adds empirical support for this conceptualization: (1) The reduction in magnitude of associations in the partial correlation matrix compared to that in the Pearson product moment correlation matrix of the rating scales. (2) A single common principal component (factor or PC1) accounts for the vast majority of shared

variance within each scale and (3) that there is additional non-shared variance among the three rating scales of disease severity, functioning, and quality of life. A new finding, and perhaps one that might not have been predicted from earlier studies, was the finding that the three scales contributed nearly equally to PC1 which is reflected in the eigenvector for PC1 (IBI-D). Perhaps equally surprising was the finding that the means and standard deviations for the each of the rating scales across the two sample populations obtained under such differing circumstances, would be nearly identical and have PCAs yielding nearly identical eigenvectors for PC1 (IBI-D) with each of the three rating scales accounting for nearly the same amount of variance in PC1.

The idea of combining quality of life measures with functioning and symptom severity has been explored before. In the Netherlands, a study by Kruijshaar et al. examined 3 degrees of severity of depression (mild, moderate, and severe) to 3 stages of esophageal cancer, 2 stages of OCD, 5 for prostate cancer, and 5 for vision disorders (Kruijshaar et al. 2005). They utilized the adapted EuroQol 5D + C5L, expanding on the original EuroQoL 5D-3L international classification by adding the assessment of cognition, mobility, self-care, usual activities, pain/discomfort, and anxiety/depression to determine the patient's quality of life and disability depression causes. In Prague Psychiatric Centre, Goppoldova et al. (2008) investigated the subjective quality of life of patients with psychotic, mood, anxiety disorders using the 10-item Schwartz Outcome Scale, and severity of illness/global improvement using the Clinical Global Impression scale. Although the investigators did not combine the measures in one scale, they noted that there was a discrepancy between physicians' ratings of improvement (as measured by symptom severity) and patients' ratings of improvement (as measured by quality of life). This discrepancy was influenced by diagnostic categories and illness manifestations suggesting the need to measure the above dimensions and not rely solely on clinicians' rating of symptom improvement. Waern et al. (2002) investigated the disease burden of geriatric depression and suicide by conducting semi-structured interviews with geriatric patients or relatives of seniors who had committed suicide. They used organ-specific guidelines of the Cumulative Illness Rating Scale for geriatrics to rate burden of illness. The interview examined social situations, mental and physical health including psychiatric symptoms in past month and dementia symptoms, life events, alcohol and drug use. They found a strong correlation between mental illness and geriatric suicide, with 89% of those who had committed suicide having a level 3 or 4 (out of 4) in the mental illness category (Waern et al. 2002). Molenaar et al. (2007) evaluated pharmacotherapy efficacy using the Hamilton Rating Scale for Depression, 17-item version, the CGI, the EuroQoL, the SF-36 and its modified SF-12. They also used the Quality of Life Depression Scale, the Depression subscale of the 90-item Symptom Checklist, and the Groningen Social Disability Schedule. When used both before treatment and after treatment onset, efficacy of treatment could be inferred indirectly (Molenaar et al. 2007).

While there are precedents as outlined above, for studying symptom severity, level of functioning, and quality of life (QOL) together, the vast majority of studies have not systematically evaluated them. For example, the NIMH Collaborative Depression Study measured depressive symptom severity in correlation with level of function. The study assessed weekly symptoms by the 6-point Psychiatric Status Rating Scale for major depressive episodes, and a 3-point scale for minor depressive or dysthymic episodes.

Symptom severity was then ascertained by placing each week's responses on a 4-point scale. To add to the data collection, trained interviewers assessed patients every 6 months for 5 years and yearly thereafter using an adapted version of the Longitudinal Interval Follow-up Evaluation. This version explored 9 domains of function and analyzed where, during each month, the patient was most functionally debilitated by depressive symptoms. They showed that a significant increase in psychosocial disability was seen in proportion to increasing severity of depressive symptoms, and it was noted that this correlation was seen increasing in stepwise fashion along the entire gradient of depressive symptom severity from sub-threshold depression to MDD (Judd et al. 2000). Due to the fact that it is a chronic and recurrent illness with relapses throughout life, MDD persists as a burden on health and living, especially when left untreated (Parikh and Lam 2001; Skärsäter et al. 2006).

Thus, while a number of disease-specific and non-disease-specific scales have been developed and used to assess symptom severity, functional impairment, and quality of life for the individual patient, none has addressed all of these domains in a single measure or index. Because symptoms are specific to a given disease, and because of the need to quantify individual burden of illness in depression to guide both clinical practice and treatment outcome studies, we chose to assess burden of illness in samples of depressed patients and assessed severity of depressive symptoms using the QIDS-SR. Future applications of this approach to create burden of illness indices for other psychiatric and medical disorders would, of course, require disease-specific scales of symptom severity.

Limitations and Utility

The authors struggled with the optimum number of rating scales to include in the Burden of Illness Index and whether these scales should be patient-rated or provider-rated or some combination of patient and provider rated scales. For example, we considered whether to include specific measures for the side effects associated with treatment. In the end we decided against having a separate scale to measure side effects concluding that such effects would be adequately brought into the IBI-D through the functioning and quality of life rating scales. We worried that the use of a separate measure would create comparability problems across patients who were either not receiving treatment or who were receiving such diverse treatments as cognitive therapy, antidepressants, atypical antipsychotics, and electroconvulsive treatment. As an alternative we would suggest that when users of the IBI-D observe a disconnection between the IBI-D and severity of illness in a patient receiving medications they look at psychotropic medication side effects as one possible explanation for the discrepancy and that this could be done by using scales directed at those side effects most closely associated with the specific medications that their patient is receiving. For instance, cognitive functioning measures may be indicated in patients reporting significant memory loss after ECT, whereas regular weight/BMI/waist measurements might be important to use in patients at risk for developing metabolic syndrome with atypical antipsychotics often used to augment antidepressants.

Analogously, we chose not to include independent rating scales of other psychiatric and medical disorders that may burden the depressed patient, believing that these effects will be carried into the IBI-D through affects primarily on the functioning and quality of life rating

scales. In making these decisions we had to keep in mind the goal of having a Burden of Illness Index that would not present an additional burden on already burdened patients and care providers. This was an important factor in choosing patient-rated scales and limiting the number of scales to be included in the Index.

Having simplicity and ease of use is part of our goals; we recognize that the calculation of the IBI-D might prove to be a barrier to its use. As a result the authors will make available an Excel spreadsheet that would automate the calculation upon entry of the scale scores for interested clinicians or investigators with the future possibility of creating a web page calculator and/or a smart phone app if demand warrants such approaches. The value of implementing such technological tools is to help treating clinicians and researchers determine how does their own patients' /subjects' burden compare with patients with MDD in the above described real world depressive populations and subsequently monitor their progress.

With simplicity and utility for the provider foremost in our minds, we foresee the possibility of the IBI-D index being used in an analogous way to a battery of tests administered by a neuropsychologist. While we expect most depressed patients to have similar z scores on the three individual scales and on the IBI-D considerable insight is to be gained from discordant z scores. For example, if subjects have a relatively high z score on quality of life compared to the other z-scores, it may be that antidepressant treatment alone will have only a modest effect on quality of life suggesting that adding adjunctive psychotherapies earlier than later may be called for. However, additional studies are required to confirm the utility of the IBI-D index.

Conclusions

This is the first study introducing and validating a composite calculation of individual burden of illness in major depressive disorder. It is based on the recognition of the inadequacy of symptom-oriented care in producing patient-centered outcomes and the need to include a focus on functioning and quality of life as the ultimate goals of healthcare interventions. We demonstrate via statistical means, the contribution of each of the three domains of symptom severity, functional disability, and quality of life, to the variability in the overall burden of illness and establish a single weighted composite numerical score that adequately represents burden of illness. While we feel that the IBI-D index will prove useful in both research and clinical settings studies need to be performed with the IBID to look at its usefulness as a numerical score of burden of illness in treated depressed patients, for predicting treatment response, for sensitivity and specificity in clinical trials, and for deciding whether augmentation or shift in antidepressants are warranted. The success of the approach with respect to depression may well provide the impetus for the creation and implementation of IBI indices for other psychiatric and medical illnesses. These additional disease-specific versions would, of course, require the same type of analysis and validation as performed and outlined for the study of depression.

References

- American Psychiatric Association (APA). Diagnostic and statistical manual of mental disorders. 4th ed.. American Psychiatric Publishing Inc; Washington, DC: 1994. Global assessment of functioning..
- American Psychiatric Association (APA). [Dec 23, 2010] Practice guideline for the treatment of patients with major depressive disorder. 3rd ed.Oct. 2010 p. 26from: <http://www.psych.org/guidelines/mdd2010>
- Anand S, Hanson K. DALYs: Efficiency versus equity. *World Development*. 1998; 26(2):307–310.
- Angermeyer MC, Holzinger A, Matschinger H, Stenger-Wenzke K. Depression and quality of life: Results of a follow-up study. *International Journal of Social Psychiatry*. 2002; 48(3):189–199. [PubMed: 12413247]
- Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology*. 1988; 56(6):893–897. [PubMed: 3204199]
- Costello AB, Osborne JW. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*. 2005; 10(7) [15 Sep 2011] Available online <http://pareonline.net/getvn.asp?v=10&n=7>.
- Endicott J, Nee J. Endicott work productivity scale (EWPS): A new measure to assess treatment effects. *Psycho-pharmacology Bulletin*. 1997; 33(1):13–16.
- Endicott J, Nee J, Harrison W, Blumenthal R. Quality of life enjoyment and satisfaction questionnaire: A new measure. *Psychopharmacology Bulletin*. 1993; 29(2):321–326. [PubMed: 8290681]
- Fava M, Rush AJ, Trivedi MH, Nierenberg AA, Thase ME, Sackeim HA, et al. Background and rationale for the sequenced treatment alternatives to relieve depression (STAR*D) study. *Psychiatric Clinics of North America*. 2003; 26(2):457–494. [PubMed: 12778843]
- Fleck M, Simon G, Herrman H. Major depression and its correlates in primary care settings in six countries. *British Journal of Psychiatry*. 2005; 186(1):41–47. [PubMed: 15630122]
- Gold MR, Stevenson D, Fryback DG. HALYs and QALYs and DALYs, Oh My: Similarities and differences in summary measures of population health. *Annual Review of Public Health*. 2002; 23:115–134.
- Goppoldova E, Dragomirecka E, Motlova L, Hajek T. Subjective quality of life in psychiatric patients: Diagnosis and illness-specific profiles. *Canadian Journal of Psychiatry*. 2008; 53(9):587–593. [PubMed: 18801221]
- Gorsuch, RL. Factor analysis. Lawrence Erlbaum; Hillsdale: 1983.
- Hair, JF.; Anderson, RE.; Tatham, RL.; Black, WC. Multivariate data analysis. 6th ed.. Prentice-Hall; Upper Saddle River: 2006.
- Havelaar, A. [Dec 21, 2010] Methodological choices for calculating the disease burden and cost-of-illness of foodborne zoonoses in European countries. Aug. 2007 from Netherlands: Med-Vet-Net Workpackage 23: http://www.medvetnet.org/pdf/Reports/Reports_07_002.pdf
- IsHak, WW.; Burt, T.; Sederer, LI. Outcome measurement in psychiatry: A critical review. American Psychiatric Publishing, Inc.; Washington: 2002.
- IsHak WW, Greeberg JM, Balayan K, Kapitanski N, Jeffrey J, Fathy H, et al. Quality of life: The ultimate outcome measure of interventions in major depressive disorder. *Harvard Review of Psychiatry*. 2011; 19:229–239. [PubMed: 21916825]
- Judd LL, Akiskal HS, Zeller PJ, Paulus M, Leon AC, Maser JD. Psychosocial disability during the long-term course of unipolar major depressive disorder. *Archives of General Psychiatry*. 2000; 57(4):375–380. [PubMed: 10768699]
- Kessler R, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, et al. The epidemiology of major depressive disorder: Results from the National Comorbidity Survey Replication (NCS-R). *JAMA*. 2003; 289(23):3095–3105. [PubMed: 12813115]
- Kruijshaar ME, Hoeymans N, Spijker J, Stouthard ME, Essink-Bot ML. Has the burden of depression been overestimated? *Bull World Health Organ*. 2005; 83(6):443–448. [PubMed: 15976895]

- McGuire T, Wells KB, Miranda J, Scheffler R, Durham M, Ford DE, et al. Wells KB, Miranda J, Scheffler R, Durham M, Ford DE, Lewis L. Burden of illness. *Mental Health Services Research*. 2002; 4(4):179–185. [PubMed: 12558002]
- Melse JM, Essink-Bot ML, Kramers PG, Hoeymans N. A national burden of disease calculation: Dutch disability-adjusted life-years. *American Journal of Public Health*. 2000; 90(8):1241–1247. [PubMed: 10937004]
- Molenaar PJ, Dekker J, Van R, Hendriksen M, Vink A, Schoevers RA. Does adding psychotherapy to pharmacotherapy improve social functioning in the treatment of outpatient depression? *Depress Anxiety*. 2007; 24(8):553–562. [PubMed: 17131302]
- Mundt JC, Marks IM, Shear MK, Greist JH. The work and social adjustment scale: A simple measure of impairment in functioning. *British Journal of Psychiatry*. 2002; 180:461–464. [PubMed: 11983645]
- Murray CJ. Quantifying the burden of disease: The technical basis for disability-adjusted life years. *Bull World Health Organ*. 1994; 72(3):429–445. [PubMed: 8062401]
- Murray CJ, Lopez AD. Global mortality, disability, and the contribution of risk factors: Global burden of disease study. *Lancet*. 1997; 349(9063):1436–1442. [PubMed: 9164317]
- Parikh SV, Lam RW. Clinical guidelines for the treatment of depressive disorders. Definitions, prevalence, and health burden. *Canadian Journal of Psychiatry*. 2001; 46(Suppl(1)):13S–20S. [PubMed: 11441768]
- Prüss-Üstün, A.; Mathers, C.; Corvalán, C.; Woodward, A. Introduction and methods: Assessing the environmental burden of disease at national and local levels (WHO Environmental Burden of Disease Series, No. 1). World Health Organization; Geneva: 2003.
- Rapaport MH, Clary C, Fayyad R, Endicott J. Quality-of-life impairment in depressive and anxiety disorders. *American Journal of Psychiatry*. 2005; 162:1171–1178. [PubMed: 15930066]
- Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, et al. The 16-item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating(QIDS-C), and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. *Biological Psychiatry*. 2003; 54(5):573–583. [PubMed: 12946886]
- Rush AJ, Fava M, Wisniewski SR, Lavori PW, Trivedi M, Sackeim HA, et al. Sequenced treatment alternatives to relieve depression (STAR*D): Rationale and design. *Controlled Clinical Trials*. 2004; 25(1):119–142. [PubMed: 15061154]
- Rush AJ, Kraemer HC, Sackeim HA, Fava M, Trivedi M, Frank E, et al. A report by the ACNP task force on response and remission in major depressive disorder. *Neuropsychopharmacology*. 2006; 31(9):1841–1853. [PubMed: 16794566]
- Sassi F. Calculating QALYs, comparing QALY and DALY calculations. *Health Policy and Planning*. 2006; 21(5):402–408. [PubMed: 16877455]
- Sheehan, DV. *The anxiety disease*. Scribner's; New York: 1983.
- Sheehan DV, Lecrubier Y, Sheehan KH, Amrim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry*. 1998; 59(Suppl(20)):22–33.
- Skärsäter I, Baigi A, Haglund L. Functional status and quality of life in patients with first-episode major depression. *Journal of Psychiatric and Mental Health Nursing*. 2006; 13(2):205–213. [PubMed: 16608476]
- Trivedi MH, Rush AJ, Wisniewski RS, Warden D, McKinney W, Downing M, et al. Factors associated with health-related quality of life among outpatients with major depressive disorder: A STAR*D report. *Journal of Clinical Psychiatry*. 2006; 67(2):185–195. [PubMed: 16566612]
- Waern M, Rubenwitz E, Runeson B, Skoog I, Wilhelmson K, Allebeck P. Burden of illness and suicide in elderly people: Case-control study. *BMJ*. 2002; 324(7350):1355–1358. [PubMed: 12052799]
- World Health Organization (WHO). [Dec 21, 2010] The global burden of disease, 2004. Oct. 2008 from http://www.who.int/healthinfo/global_burden_disease/GBD_report_2004update_full.pdf
- Zeckhauser RJ, Shepard DS. Where now for saving lives? *Law and Contemporary Problems*. 1976; 40(4):5–45.

- Zimmerman M, McGlinchey JB, Posternak MA, Friedman M, Attiullah N, Boerescu D. How should remission from depression be defined? The depressed patient's perspective. *American Journal of Psychiatry*. 2006; 163(1):148–150. [PubMed: 16390903]
- Zimmerman M, McGlinchey JB, Posternak MA, Friedman M, Boerescu D, Attiullah N. Remission in depressed outpatients: More than just symptom resolution. *Journal of Psychiatric Research*. 2008; 42(10):797–801. [PubMed: 17986389]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

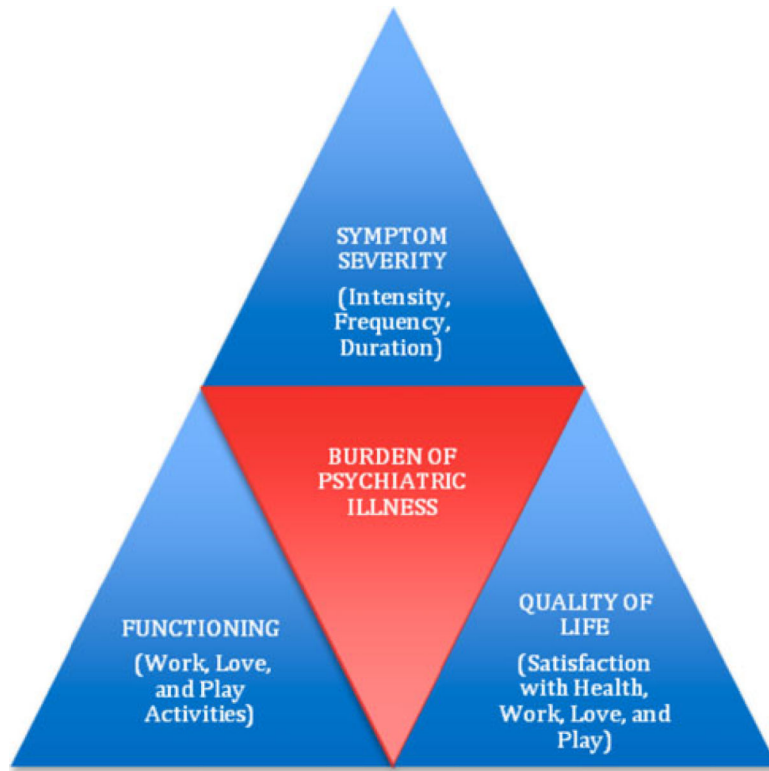


Fig. 1. The components of burden of illness: symptoms severity, functioning, and quality of life

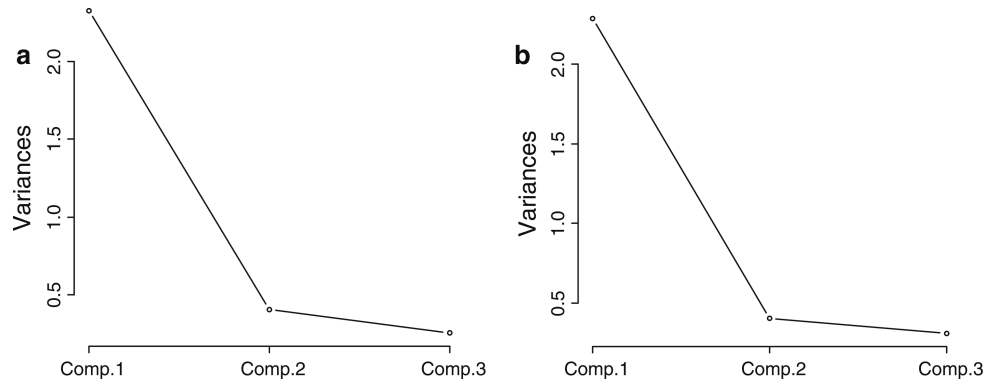
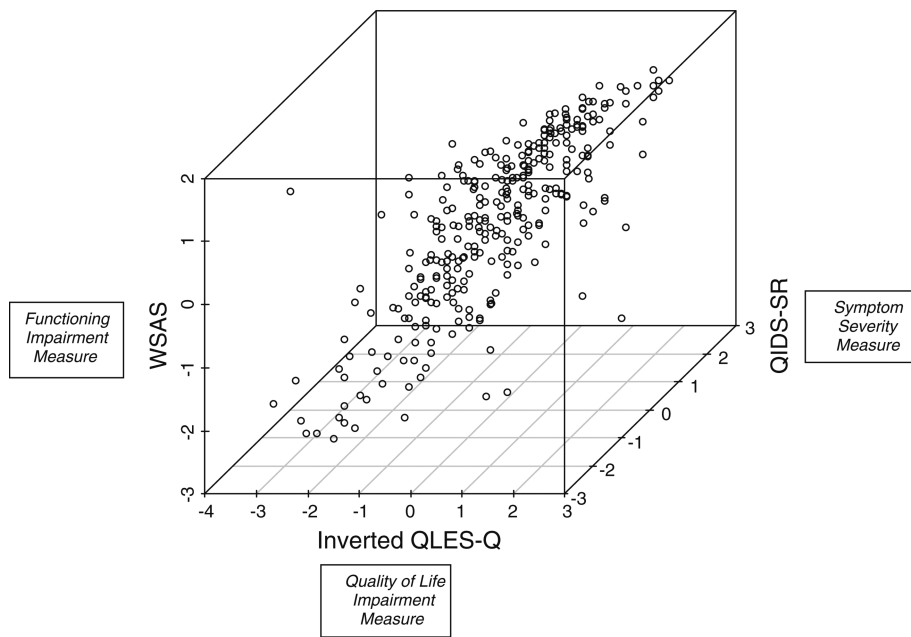


Fig. 2.
a Scree plot of PCA of CS-PTR data. **b** Scree plot of PCA of STAR*D data



QIDS-SR = Quick Inventory of Depressive Symptomatology-Self Report (Rush, et al., 2003).
 WSAS = Work and Social Adjustment Scale (Mundt, Marks, Shear, & Greist, 2002).
 Q-LES-Q = Quality of Life Enjoyment and Satisfaction Questionnaire – Short Form (Endicott, Nee, Harrison, & Blumenthal, 1993).

Fig. 3. Principal component analysis and the variance in each rating scale. *QIDS-SR* Quick Inventory of Depressive Symptomatology-Self Report (Rush et al. 2003). *WSAS* Work and Social Adjustment Scale (Mundt et al. 2002). *Q-LES-Q* Quality of Life Enjoyment and Satisfaction Questionnaire—short form (Endicott et al. 1993)

Table 1

Descriptions of measures (QIDS-SR, WSAS, and Q-LES-Q)

Measure (Reference)	Description	Validity and reliability
Quick Inventory of Depressive Symptomatology—Self Report or QIDS-SR (Rush et al. 2003)	A 16-item self-administered instrument that covers the nine DSM-IV criterion symptoms for major depressive disorder. It has been used in a variety of research and clinical settings, e.g., as an outcome measure in industry-sponsored randomized clinical trials, as well as in outpatient psychiatric and primary care clinics.	Good internal consistency and reliability has been reported for clinical populations with very high concurrent validity found between QIDS-SR and the HSRD in three large independent patient samples. It takes 5–10 min to complete, and has the advantage of providing a single numerical score for depression severity based on numerical score while having similar sensitivity to change as the HSRD in a clinical trial of chronic major depression.
Work and Social Adjustment Scale or WSAS (Mundt et al. 2002)	A five-item self-report scale of functional impairment attributable to an identified problem that is easy for patients to understand and complete.	Internal consistency is very good in depressed patients, with a Cronbach's α coefficient of 0.807 at baseline that increases with time up to 0.942 at week 30. There is also good convergent validity with HSRD-17, with a correlation coefficient of 0.76. Additionally, strong criterion validity (ability to stratify among different levels of severity) was established by association with HSRD severity strata.
Quality of Life Enjoyment and Satisfaction Questionnaire—Short Form or Q-LES-Q (Endicott et al., 1993)	A self-administered quality of life instrument that was designed to obtain sensitive information on the degree of enjoyment and satisfaction experienced by patients in various areas of daily functioning. The long form consists of 60 items and 5 subscales, whereas the short form consists of 16 items. The short form was used for the purpose of this study.	Q-LES-Q was shown to have a reliability coefficient of 0.90–0.96 for subscales and 0.90 for the summary scale, and a test-retest coefficient of 0.63–0.89 for subscales and 0.74 for summary scale. A correlation coefficient of -0.62 between the summary scale and the Clinical Global Impression (CGI) scale supports its convergent validity.

Table 2Demographic and clinical characteristics of the CS-PTR^a and STAR*D samples

	CS-PTR (<i>n</i> = 317)	STAR*D (<i>n</i> = 2,967)
Age, mean (SD), in years	43.7 (15.5)	41.2 (13.2)
Characteristics, <i>n</i> (%)		
Women	210 (66.2%)	1,869 (63.0%)
Race		
Caucasian	218 (69%)	2,009 (67.7%)
African american	47 (15%)	504 (17%)
Hispanic	20 (6%)	335 (11.3%)
Asian	12 (4%)	80 (2.7%)
Other	20 (6%)	39 (1.3%)
Employed	148 (46%)	1,377 (46.4%)
Primary diagnosis		
Major depression, single episode	73 (23%)	688 (23.2%)
Major depression, recurrent	244 (77%)	2,279 (76.8%)
Severity of depression (according to Rush et al. 2003)		
Remission (QIDS-SR = 0–5) ^b	15 (5%)	83 (2.8%)
Mild (QIDS-SR = 6–10)	41 (13%)	463 (15.6%)
Moderate (QIDS-SR = 11–15)	101 (32%)	914 (30.8%)
Severe (QIDS-SR = 16–20)	97 (30%)	961 (32.4%)
Very severe (QIDS-SR > 20)	63 (20%)	546 (18.4%)

^aCS-PTR Cedars-Sinai Psychiatric Treatment outcome Registry^bQIDS-SR Quick Inventory of Depressive Symptomatology—self report

Table 3

Comparison of individual rating scales between the CS-PTR and STAR*D samples

	<u>CS-PTR</u>		<u>STAR*D</u>		Effect size	<u>t-test</u>	
	Mean	SD	Mean	SD		t-value	P value
QIDS-SR	15.42	5.326	15.596	5.078	-0.034	-0.562	0.56
WSAS	24.9	10.624	23.703	9.163	0.128	1.93	0.054
InvQ-LES-Q	59.977	16.8	58.439	15.122	0.101	1.564	0.12

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4

Pearson product moment and partial correlations among depression rating scales

	<u>Correlations</u>		<u>Partial correlations</u>		KMO or MSA
	QIDS-SR	WSAS	QIDS-SR	WSAS	
CS-PTR Sample					
<i>Q-LES-Q</i>	0.742	0.634	0.574	0.33	0.685
<i>QIDS-SR</i>		0.621		0.291	0.693
<i>WSAS</i>					0.803
<i>Sum</i>					0.719
STAR*D Sample					
<i>Q-LES-Q</i>	0.642	0.685	0.394	0.487	0.692
<i>QIDS-SR</i>		0.603		0.293	0.763
<i>WSAS</i>					0.721
<i>Sum</i>					0.722

KMO Kaiser–Meyer–Olkin test. *MSA* measures of sampling adequacy

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 5

Principal component analysis

Samples	Eigenvalues of principal components			Proportion of variance explained		
	PC1	PC2	PC3	Q-LES-Q	QIDS-SR	WSAS
<i>CS-PTR</i>	2.33	0.409	0.258	0.778	0.136	0.086
<i>STAR*D</i>	2.29	0.403	0.309	0.762	0.135	0.103

Samples	Eigenvector for PC1			Correlations with PC1		
	Q-LES-Q	QIDS-SR	WSAS	Q-LES-Q	QIDS-SR	WSAS
<i>CS-PTR</i>	0.59	0.587	0.554	0.902	0.896	0.847
<i>STAR*D</i>	0.589	0.565	0.578	0.891	0.853	0.874

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript