ORIGINAL ARTICLE

Incorporating Multidimensional Patient-Reported Outcomes of Symptom Severity, Functioning, and Quality of Life in the Individual Burden of Illness Index for Depression to Measure Treatment Impact and Recovery in MDD

Robert M. Cohen, PhD, MD; Jared M. Greenberg, MD; Waguih William IsHak, MD, FAPA

Context: The National Institute of Mental Health Affective Disorders Workgroup identified the assessment of an individual's burden of illness as an important need. The Individual Burden of Illness Index for Depression (IBI-D) metric was developed to meet this need.

Objective: To assess the use of the IBI-D for multidimensional assessment of treatment efficacy for depressed patients.

Design, Setting, and Patients: Complete data on depressive symptom severity, functioning, and quality of life (QOL) from depressed patients (N=2280) at entry and exit of level 1 of the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study (12-week citalopram treatment) were used as the basis for calculating IBI-D and self-rating scale changes.

Results: Principal component analysis of patient responses at the end of level 1 of STAR*D yielded a single principal component, IBI-D, with a nearly identical eigenvector to that previously reported. While changes in symptom severity (Quick Inventory of Depressive Symptomatology–Self Report) accounted for only 50% of the variance in changes in QOL (Quality of Life Enjoyment and Satisfaction Questionnaire–Short Form) and 47% of the variance in changes in functioning (Work and Social Adjustment Scale), changes in IBI-D captured 83% of the variance in changes in QOL and 80% in functioning, while also capturing 79% of the variance in change in symptom severity (Quick Inventory of Depressive Symptomatology–Self Report). Most importantly, the changes in IBI-D of the 36.6% of remitters who had abnormal QOL and/or functioning (mean [SD], 2.98 [0.35]) were significantly less than the changes in IBI-D of those who reported normal QOL and functioning (IBI-D=1.97; t=32.6; $P<10^{-8}$) with an effect size of a Cohen d of 2.58. In contrast, differences in symptom severity, while significant, had a Cohen d of only 0.78.

Conclusions: Remission in depressed patients, as defined by a reduction in symptom severity, does not denote normal QOL or functioning. By incorporating multidimensional patient-reported outcomes, the IBI-D provides a single measure that adequately captures the full burden of illness in depression both prior to and following treatment; therefore, it offers a more accurate metric of recovery.

Trial Registration: clinicaltrials.gov Identifier: NCT00021528

JAMA Psychiatry. 2013;70(3):343-350.
Published online January 2, 2013.
doi:10.1001/jamapsychiatry.2013.286

Author Affiliations:
Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine, University of California, Los Angeles, and Department of Psychiatry and Behavioral Neurosciences, Cedars-Sinai Medical Center, Los Angeles, California.

MAJOR DEPRESSIVE disorder (MDD) is an illness that carries enormous and well-established negative impact, and it is predicted to become the second-leading contributor to global burden of disease by the year 2020.[1] The National Institute of Mental Health (NIMH) Affective Disorders Workgroup identified the assessment of the burden of illness as one of the important needs in depression research.[2] We conceptualized the individual burden of illness to include suffering due to symptom severity (intensity, frequency, and duration), impairment in functioning (occupational, social, and leisure activities), and reduction in quality of life (QOL; satisfaction with health, occupational, social, and leisure activities), as illustrated in **Figure 1**.[3,4] Recently, we reported on the development and validation of an Individual Burden of Illness Index for Depression (IBI-D) as a useful single measure for capturing the multidimensional impact of depression on an individual,[4] whereas other existing measures focus on only 1 dimension such as

©2013 American Medical Association. All rights reserved.
Downloaded From: http://archpsyc.jamanetwork.com/ on 05/01/2016

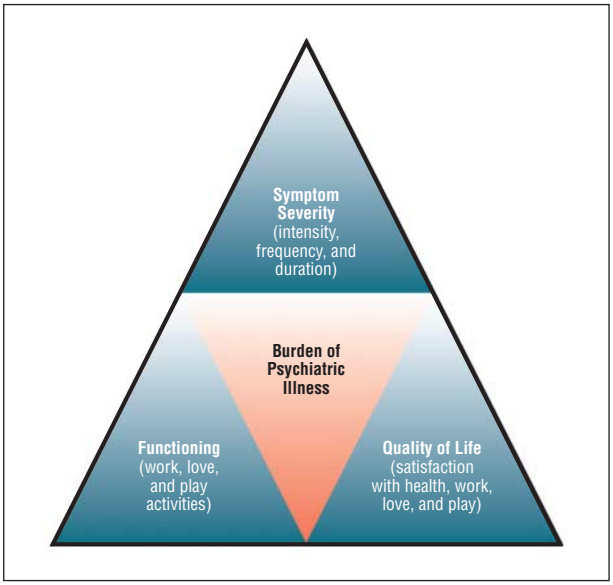

Figure 1. The components of the burden of illness: symptom severity, functioning, and quality of life.

symptom severity. The Individual Burden of Illness Index for Depression was the label given to the first and only significant principal component obtained from a principal component analysis (PCA) of 3 previously validated and reliable patient-reported scales: the Quick Inventory of Depressive Symptomatology–Self Report (QIDS-SR)[5] for depressive symptom severity, the Work and Social Adjustment Scale (WSAS)[6] for functioning, and the Quality of Life Enjoyment and Satisfaction Questionnaire–Short Form (Q-LES-Q)[7] for QOL. Enrollees in the Cedars-Sinai Psychiatric Treatment Outcome Registry[4] served as the development sample and enrollees in level 1 of the NIMH Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study[8] served as the validation sample for the PCA analyses.

The present article extends our previously referenced work by applying the IBI-D in a treatment context. To assess the potential use of the IBI-D in evaluating treatment efficacy, we used the IBI-D as a measure of the burden of illness in depressed patients at the end of level 1 (citalopram treatment) in the STAR*D trial. Moreover, as symptoms of depression are expected to contribute to functional impairment and whereas both depressive symptom severity and functional impairment are expected to alter subject perceptions of QOL, we expected that changes in all 3 rating scales in response to citalopram treatment would be correlated with each other just as they were at the time of subject entry, as was reported in our previous study. We hypothesized that changes in the 3 rating scales in response to treatment would be more closely correlated with changes in the IBI-D than to each other based on the previously mentioned theoretical considerations and the previous analysis of the empirical data obtained from patients prior to treatment demonstrating that the IBI-D served to adequately capture the shared variance among the 3 rating scales on which it was based. As one manifestation of the predicted higher correlations between the IBI-D and the QIDS-SR, WSAS, and Q-LES-Q, than, for example, between QIDS-SR and either Q-LES-Q or WSAS, we expected that the IBI-D would be of particular importance for discriminating among patients with remission as defined by a score of 5 or less on the QIDS-SR who also achieved normal QOL and function compared with those who did not. Empirical support for the previously mentioned hypotheses would suggest the benefit of using change in the IBI-D as a more complete indicator of the multidimensional impact of treatment on alleviating the burden of depression (ie, not just reducing symptoms, but also improving functioning) and QOL.[9] The focus of this study was to determine whether the multidimensional IBI-D would be as useful in evaluating the full impact of treatment as it is in capturing the burden of illness that patients present with prior to treatment.

METHODS

POPULATION

The STAR*D trial was a NIMH-funded study to evaluate the responses of patients with MDD to initial citalopram treatment, followed by additional treatment options for initially nonresponsive patients. The full details of the study are described elsewhere.[8,10] Briefly, STAR*D was conducted at 18 primary care settings and 23 psychiatric care settings in the United States from 2001 through 2007 and enrolled 4041 treatment-seeking outpatients aged 18 to 75 years who had a primary diagnosis of MDD.

For the purpose of this study, the patient sample was derived from the STAR*D trial. The authors obtained NIMH Data Use Certificate to use the STAR*D data set (STAR*D Pub Ver1). Subjects with complete data, collected by the interactive voice response system at entry and exit of level 1 of the study, were included in this analysis (N = 2280). The initial sample with entry data from 2967 subjects was reduced to 2324 owing to missing exit data, and then to 2280 after excluding patients who met criteria for remission at entry (QIDS-SR score ≤5).

OUTCOME MEASURES

The measures used to evaluate change in patients with MDD while participating in level 1 of the STAR*D trial were the QIDS-SR[5] for depressive symptom severity, the WSAS[6] for functioning, and the Q-LES-Q[7] for QOL. All 3 measures were obtained at time of entry into level 1 and at the close of level 1 (12 weeks), and the IBI-D was calculated at both points as defined next on the basis of these scales. Self-rated measures were chosen as the basis for our analyses as they offer ease of collection, while maintaining high reliability and validity with respect to clinician-rated measures. A detailed review of the measures is available in the table in the IBI-D development and validation study.[4]

CALCULATION OF IBI-D

Values for the means and standard deviations (SDs) of each rating scale were used to determine an individual patient's IBI-D. To obtain an individual's IBI-D, the individual's score on each scale was converted to a z score by first subtracting the scale's mean and dividing by the SD of the scale: For the QIDS, zQIDS-SR = (QIDS-SR–15.6)/5.1, and for the WSAS, zWSAS = (WSAS–23.9)/9.3.

For the Q-LES-Q, we corrected for the inversion of the scale by subtracting a score from 100 and inverting the sign to obtain the following formula for the z score: z Inverted Q-LES-Q (InvQ-LES-Q) = (41.4–Q-LES-Q)/15.3.

©2013 American Medical Association. All rights reserved.
Downloaded From: http://archpsyc.jamanetwork.com/ on 05/01/2016

Table 1. Demographic Characteristics of the STAR*D Sample

Characteristic	No. (%) (N = 2280)
Age, mean (SD) [range], y	42.6 (13.0) [18.1-75.6]
Women	1432 (62.8)
White	1846 (80.9)
Hispanic	239 (10.5)
College graduate	686 (30.1)
Employed	1301 (57.1)
Living with spouse/partner	1046 (45.9)

Abbreviation: STAR*D, Sequenced Treatment Alternatives to Relieve Depression.

The z scores were then substituted into the following formula, which uses the weightings (factor loadings) obtained from the previously reported PCA[4] to obtain the IBI-D and then dividing by the overall SD of the IBI-D obtained from our 2 previously studied sample populations: IBI-D = (0.57*[zQIDS-SR] + 0.58*[zWSAS] + 0.59*[zInvQ-LES-Q])/1.51.

STATISTICAL ANALYSIS

For confirmation of the validity of PCA for the end of treatment phase ratings: change scores on individual rating scales and on the IBI-D were calculated as z_B (at entry)–z_F (at end of level 1). Because all correlations were obtained from the same population sample (STAR*D), the test of significance of the difference between r's has to take into account dependence as formulated by Steiger[11] and Cohen and Cohen.[12] Analyses were performed using SAS software version 9.2 (SAS Institute Inc) and the open-source R programming language version 2.10.1 (The R Foundation for Statistical Computing).

As previously published, because the development of the IBI-D was based solely on data from depressed patients evaluated prior to treatment, it was important to confirm that at the end of the citalopram treatment phase (level 1) of the STAR*D trial, a PCA would support our initial findings. Support would come from finding a single component that essentially captured nearly all of the variance in the 3 scales (QIDS-SR, WSAS, and Q-LES-Q) that were the data input for the PCA, and that the loadings for each rating scale to the eigenvector, derived from the PCA, would be essentially the same as was initially derived.

RESULTS

The demographic characteristics of the STAR*D sample are displayed in **Table 1**. **Table 2** summarizes the results obtained from this new PCA along with the relevant values from our previous study to allow for ease of comparison. Significant changes between the baseline and exit analyses show that even higher variances are accounted for by principal component 1, the only significant principal component (**Figure 2** and **Figure 3**), labeled as IBI-D, in the exit analysis. Using the eigenvector obtained from the earlier analysis, we also found an even higher correlation between the IBI-D and each individual rating scale that served as input to the PCA at exit from level 1 than we obtained at entry. Presumably, the higher captured variance and greater correlations were likely the result of the greater ranges for each of the rating scales found at the end of level 1, as exit data include remitted individuals and others who remained severely depressed.

Table 2. Results of PC Analyses of STAR*D Data

	PC1	PC2	PC3
Eigenvalues of PCs			
Baseline	2.29	0.403	0.309
Exit	2.63	0.206	0.166
Proportion of variance explained			
Baseline	0.762	0.135	0.103
Exit	0.876	0.069	0.056
	Q-LES-Q	**QIDS-SR**	**WSAS**
Eigenvector for PC1			
Baseline	0.59	0.57	0.58
Exit	0.58	0.57	0.58
Correlations with PC1			
Baseline	0.891	0.896	0.847
Exit	0.941	0.928	0.939

Abbreviations: PC, principal component; QIDS-SR, Quick Inventory of Depressive Symptomatology–Self Report[5]; Q-LES-Q, Quality of Life Enjoyment and Satisfaction Questionnaire–Short Form[7]; STAR*D, Sequenced Treatment Alternatives to Relieve Depression; WSAS, Work and Social Adjustment Scale.[6]

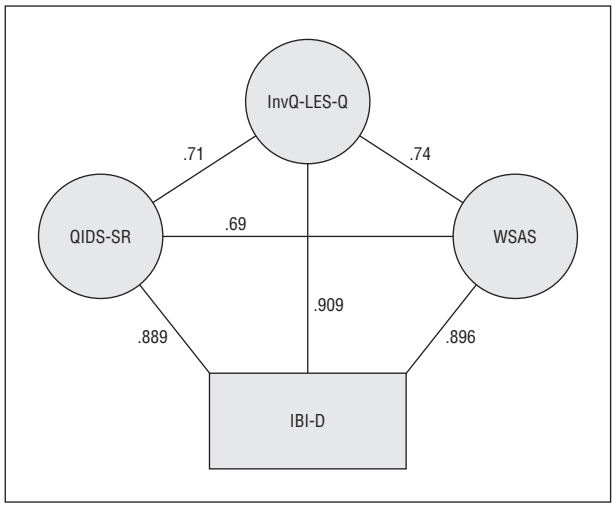

Figure 2. Correlations between changes on the Individual Burden of Illness Index for Depression (IBI-D) and the changes in the individual component rating scales—Quick Inventory of Depressive Symptomatology–Self Report (QIDS-SR),[5] Work and Social Adjustment Scale (WSAS),[6] and Inverted Quality of Life Enjoyment and Satisfaction Questionnaire–Short Form (InvQ-LES-Q)[7]—as well as between the component rating scales.

The previously mentioned analyses justified a more in-depth examination of the IBI-D as a tool for measuring antidepressant treatment effects. At exit from level 1, STAR*D trial subjects had a mean (SD) change from baseline of 1.181 (1.294) on the IBI-D, 1.197 (1.272) on the QIDS-SR z score, 0.986 (1.270) on the InvQ-LES-Q z score, and 0.895 (1.209) on the WSAS z score. Although all changes were of high statistical significance compared with improvement in depressive symptoms (QIDS-SR z score), patients had significantly lower z score changes on QOL and functioning as measured with the InvQ-LES-Q z score (t_{2280} = 10.344; $P < 10^{-15}$) and the WSAS z score (t_{2280} = 14.625; $P < 10^{-15}$), respectively. Furthermore, QOL measures (InvQ-LES-Q z score change) showed a somewhat greater improvement than functioning (WSAS z score change) as re-

©2013 American Medical Association. All rights reserved.
Downloaded From: http://archpsyc.jamanetwork.com/ on 05/01/2016

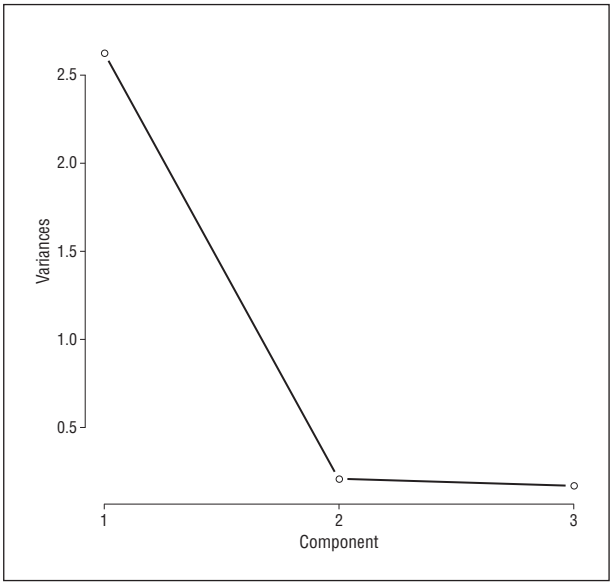

Figure 3. Scree plot of principal component analysis of patients at exit from level 1 of the Sequenced Treatment Alternatives to Relieve Depression trial.

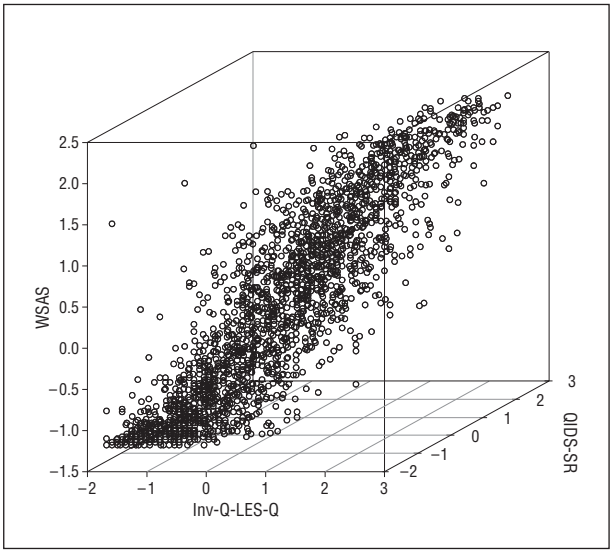

Figure 4. Scatterplot of patients' ratings on the Quick Inventory of Depressive Symptomatology–Self Report (QIDS-SR),[5] Work and Social Adjustment Scale (WSAS),[6] and Quality of Life Enjoyment and Satisfaction Questionnaire–Short Form (Q-LES-Q).[7] Plot reflects changed *z* scores.

flected in a paired t_{2280} value of 4.76 and a P value of 2×10^{-6}. The change in the IBI-D of 1.181 was somewhat smaller than the change in depressive symptoms of 1.197; however, the difference was not statistically significantly ($t_{2280} = 1.244$; $P = .21$).

Importantly, while *z* score changes in all 3 scales were correlated with each other and of high statistical significance ($P < 10^{-15}$) as expected, these interscale correlations were significantly less than the correlation of each *z* score change with the IBI-D (**Figure 4**). Specifically, the correlation between subject changes in depressive symptoms (QIDS-SR *z* score) and changes in their perceptions of QOL (InvQ-LES-Q *z* score) were 0.706, whereas the IBI-D had a significantly higher correlation to the InvQ-LES-Q *z* score ($r = 0.909$; $t_{2278} = 57.9$; $P < 10^{-6}$). Although the change in depressive symptoms explained approximately 50% of the variance associated with changes in QOL perceptions, the IBI-D was able to account for approximately 83% of this variance. Similarly, whereas the change in QIDS-SR *z* score had a 0.687 correlation to change in functioning (WSAS *z* score), the change in functioning had a 0.889 correlation to the change in the IBI-D. The difference between the correlations was statistically significant at the $P < 10^{-6}$ level ($t_{2278} = 51.09$), with the change in depressive symptoms accounting for only about 47% of the variance in the change in functioning, whereas the IBI-D accounted for 79% of the variance in functioning change scores.

To be complete, we also examined whether the changes in functioning and QOL were more highly correlated with each other than to the change in depressive symptoms. The correlation between the change in WSAS *z* score and the change in InvQ-LES-Q *z* score was 0.735, which was significantly higher than the correlation of 0.706 between QIDS and InvQ-LES-Q ($t_{2278} = 2.77$; $P < .006$) and also higher than the correlation between the change in QIDS-SR *z* score and the change in WSAS *z* score ($r = 0.687$; $P < 10^{-5}$). Furthermore, we noted that the correlation between the change in depressive symptom severity (QIDS-SR *z* score) and change in perceived QOL (InvQ-LES-Q *z* score) of 0.706 did not statistically differ from that between the QIDS-SR and change in functioning (WSAS *z* score; $r = 0.687$; $t_{2278} = 1.86$; $P = .06$).

An alternative approach to viewing the potential importance of using the change in the IBI-D as a way of more accurately evaluating treatment efficacy is to examine the changes in QOL and functioning in remitters. To begin with, remitted patients differ from nonremitted patients on the QIDS-SR, WSAS, Q-LES-Q, and the IBI-D by substantial margins. The differences are highly statistically significant, although the size of the effect is greatest for the QIDS-SR and lowest for the Q-LES-Q, WSAS, and as we might expect for the IBI-D because it reflects changes in all 3 scales. Because remission is based solely on a patient having achieved a reduction in depression severity (ie, QIDS-SR score <5), these findings were not unexpected, but they do emphasize the potential problem with the sole use of the QIDS-SR as a measure of remission as it misses the very real variation among remitted patients with respect to functioning and QOL (**Table 3**).

Using the WSAS for functioning, patients with scores of less than 10 were considered within the normal range.[6] Using the Q-LES-Q for QOL, patients with scores within 10% of the community norms of 78.3 were considered within the normal range.[13-15] Of the 812 individuals who achieved remission in the STAR*D study, 158 (19.5%) had a WSAS score of 10 or greater, indicative of functional impairment. Similarly, 260 remitters (32.0%) had a Q-LES-Q score of less than 70.5 (InvQ-LES-Q>29.5), indicative of QOL impairment. Moreover, 121 remitters (14.9%) had both abnormal QOL and functioning measures. In total, 297 remitters (36.6%) had either an abnormal QOL or functioning or both. Furthermore, the data show that it is rare to have a QOL score in the normal range while reporting abnormal functioning: only 37 remitters (4.6%) had an abnormal score on functioning with a normal QOL score. In contrast, 139 remitters

©2013 American Medical Association. All rights reserved.
Downloaded From: http://archpsyc.jamanetwork.com/ on 05/01/2016

Table 3. Comparing Outcome Measures of Nonremitters and Remitters[a]

	Mean (SD)			
	Nonremitters (n = 1468)	Remitters (n = 812)	*t* Value	*P* Value
QIDS-SR	13.19 (5.23)	2.92 (1.58)	−69.7	$<10^{-8}$
Q-LES-Q	45.81 (17.55)	76.01 (14.4)	44.3	$<10^{-8}$
WSAS	21.34 (10.40)	4.93 (6.42)	−46.5	$<10^{-8}$
IBI-D	−0.397 (1.134)	−2.605 (0.626)	−59.9	$<10^{-8}$

Abbreviations: IBI-D, Individual Burden of Illness Index for Depression; QIDS-SR, Quick Inventory of Depressive Symptomatology–Self Report[5]; Q-LES-Q, Quality of Life Enjoyment and Satisfaction Questionnaire–Short Form[7]; WSAS, Work and Social Adjustment Scale.[6]

[a] Remission is defined as a score of 5 or less on the QIDS-SR.[5]

Table 4. Comparisons of Outcome Measures by QOL and Functioning in 812 Remitters[a]

	Mean (SD)			
	QIDS-SR	Q-LES-Q	WSAS	IBI-D
Abnormal functioning (n = 158)	3.77 (1.27)	61.15 (11.83)	15.79 (5.59)	−1.71 (0.44)
Normal QOL (n = 36)	3.53 (1.11)	75.06 (4.68)	13.58 (4.25)	−2.18 (0.21)
Abnormal QOL (n = 122)	3.84 (1.31)	57.05 (10.04)	16.45 (5.78)	−1.58 (0.39)
Normal Functioning (n = 654)	2.72 (1.58)	79.60 (12.55)	2.30 (2.85)	−2.82 (0.45)
Normal QOL (n = 515)	2.50 (1.55)	84.16 (9.29)	1.78 (2.50)	−2.98 (0.35)
Abnormal QOL (n = 139)	3.51 (1.42)	62.75 (7.80)	4.25 (3.22)	−2.25 (0.29)
Abnormal QOL (n = 260)	3.68 (1.37)	59.90 (9.02)	9.87 (7.52)	−1.94 (0.48)
Normal QOL (n = 552)	2.57 (1.55)	83.59 (9.40)	2.60 (4.13)	−2.92 (0.40)
	Statistical Comparison, *t* Value (*P* Value)			
Abnormal functioning vs normal functioning	8.9 ($<10^{-8}$)	17.4 ($<10^{-8}$)	−29.4 ($<10^{-8}$)	−28.3 ($<10^{-8}$)
Abnormal functioning + normal QOL vs abnormal functioning + abnormal QOL	−1.44 (NS)	15.0 ($<10^{-8}$)	−3.2 (<.002)	−12.13 ($<10^{-8}$)
Normal functioning + normal QOL vs normal functioning + abnormal QOL	−7.29 ($<10^{-8}$)	27.5 ($<10^{-8}$)	−8.4 ($<10^{-8}$)	−24.9 ($<10^{-8}$)
Abnormal QOL vs normal QOL	10.3 ($<10^{-8}$)	34.4 ($<10^{-8}$)	14.6($<10^{-8}$)	−28.9 ($<10^{-8}$)

Abbreviations: IBI-D, Individual Burden of Illness Index for Depression; NS, not significant; QIDS-SR, Quick Inventory of Depressive Symptomatology–Self Report[5] (depressive symptom severity measure); Q-LES-Q, Quality of Life Enjoyment and Satisfaction Questionnaire–Short Form[7] (quality of life measure); QOL, quality of life; WSAS, Work and Social Adjustment Scale[6] (functioning measure).

[a] Remission is defined as a score of 5 or less on the QIDS-SR.[5] Using the WSAS for functioning, patients with scores of less than 10 are considered within the normal range.[6] Using the Q-LES-Q for QOL, patients with scores within 10% of the community norms of 78.3 are considered within the normal range.[13-15]

(17.1%) had abnormal QOL scores in the presence of normal functioning.

The size of differences among groups (ie, the ability to detect differences given equal sample sizes) is directly related to *t* values. For example, remitters with an abnormal WSAS score vs remitters with a normal WSAS score only differed by 1.05 on the QIDS-SR and had a *t* value of 8.9 compared with 29.4 for the WSAS. Similarly, remitters with normal functioning but abnormal QOL scores differed by 1.11 on the QIDS-SR and had a *t* value of 10.3, whereas the QOL *t* value was 34.4. However, the IBI-D is a powerful discriminator among these subsets of abnormal QOL ($t = 28.9$) and abnormal functioning ($t = 29.4$) compared with their normal QOL and functioning subsets of remitted patients (**Table 4**). Finally, remitted patients with either abnormal QOL and/or abnormal function (mean [SD] IBI-D, −1.97 [0.46]) are readily distinguished from remitters with normal QOL and normal WSAS (mean [SD] IBI-D, −2.98 [0.35]; $t = 32.6$; $P < 10^{-8}$). The effect size (Cohen *d*) for the IBI-D difference between these 2 subsets of remitted patients was 2.6, whereas the effect size for severity of symptoms was 0.78.

COMMENT

The current findings provide strong empirical support for the use of the IBI-D as a single composite measure of the full impact of treatment on depressive symptoms, QOL, and functioning. While outcome measurement in psychiatry has in the past primarily focused on symptom severity, the impact of depression on QOL and functioning in depressed patients and the importance of alleviating these effects with treatment has become increasingly recognized.[16-26] Although the IBI-D was developed using data obtained from patient ratings at their time of entry into treatment, the current study supports its validity for adequately describing patient ratings obtained at the end of treatment. By capturing the variance obtained on the 3 component rating scales for depressive symptoms, QOL, and functioning (QIDS-SR, Q-LES-Q, and WSAS, respectively) with a PCA at both time of entry and at the end of treatment, changes in the IBI-D provided a single measure for how effectively treatment alleviated that patient's individual burden of illness. We devote the rest of our commentary to illustra-

©2013 American Medical Association. All rights reserved.

Downloaded From: http://archpsyc.jamanetwork.com/ on 05/01/2016

tion of how the IBI-D performs in achieving this result in the evaluation of the effectiveness of citalopram treatment in level 1 of the STAR*D study.

Because of differences in scale properties, PCAs are based on *z* scores; for this same reason, throughout the current study, we examined changes in *z* scores on these same scales designed to measure depressive symptom severity, functioning, and QOL, and we compared and contrasted these changes with those of the IBI-D. Specifically, the 0.986 InvQ-LES-Q score change indicated that the average treated depressed patient at level 1 of the STAR*D trial had a perceived QOL that was better than all but 16.2% of the same group of patients when they entered the trial. Similarly, the average depressed treated patient was functioning at a level that was greater than all but 18.5% of the same group of depressed patients when they entered the study. These numbers are somewhat less favorable than the improvement in depressive symptom severity, in which only 11.6% of depressed patients would have had less depressive symptoms than the average depressed patient had after treatment. The change in the IBI-D of 1.18 indicated that following treatment, only 11.9% of depressed patients at time of entry had less of a burden of illness than the average patient at the end of treatment. The meaning of the smaller *z* score changes in the QOL (InvQ-LES-Q) and functioning scales (WSAS) compared with the depressive symptom severity (QIDS-SR) suggests that some portion of the QOL and functioning abnormalities at the time of entry may not be directly related to the depressive episode. Because we do not have information on premorbid QOL or functioning, one could postulate that individuals with low premorbid QOL or functioning impairments might be at risk for developing depression. Alternatively, QOL and functioning impairments may be more difficult to treat and/or are delayed in responding to antidepressant effects compared with depressive symptom severity. It is also possible that this delay may be related to the method of measuring QOL and functioning improvement (ie, self-report) as well as the depressed patients' negatively skewed ratings in comparison to observations by family, friends, and clinicians.

While these changes in *z* scores on all 3 rating scales and the IBI-D are similar for the average patient, the high coefficient of variation of each of the *z* score changes of more than 100% suggest the importance of examining the correlations across the scales (eg, how well an individual's change on symptom severity predicts his or her change in QOL or functioning). In examining the correlations between changes in *z* scores on the 3 rating scales, it is apparent that changes in symptom severity (QIDS-SR) accounted for only 50% of the variance in the change in QOL (Q-LES-Q) and only 47% of the variance in change in functioning (WSAS), whereas the IBI-D was able to capture 83% of the variance in the change in QOL and 80% of the variance in change in functioning. Moreover, in examining remitters, it is apparent that remitted patients can vary substantially with respect to both QOL and functioning that is not adequately explained by differences in their depression severity scores, whereas the IBI-D readily distinguishes between subjects that vary with respect to either QOL or WSAS or both. Furthermore, the data support the rarity of having a normal QOL despite remission if functioning lags behind (4.6% of remitters had an abnormal score on functioning with a normal QOL score). In contrast, QOL improvement can lag behind remission even in the presence of normal functioning (17.1% of remitters had abnormal QOL scores in the presence of normal functioning).

Given the variation in QOL and functioning even among remitters, it would seem prudent to include changes in function and QOL measurements in definitions of remission. As a single measure incorporating the 3 measures of symptom severity, QOL, and function that is able to discriminate among remitted patients that differ with respect to QOL and/or function, the IBI-D could potentially serve for defining remission in the clinic and the research arena. Which IBI-D number should define remission could theoretically be determined on the basis of likelihood of relapse or recurrence. The assumption we would make would be that true remission would be associated with a lower probability of relapse or recurrence. While we are currently working on this, absent this data, we favor a guideline based on the mean value and SD of the IBI-D of individuals remitted by the usual definition of a QIDS-SR score of 5 or less who also had normal Q-LES-Q and normal WSAS scores (ie, mean [SD], −2.98 [0.35]). The percentile interpretation of this value is that of all patients diagnosed as having depression and completing level 1 of the STAR*D study, only 0.14% (approximately 1 in a 1000) would have been expected to have an IBI-D this low. Allowing for 1 SD and rounding the number leads to using a guideline of an IBI-D score of −2.7 or less for defining patient remission. For example, in the clinic, the observation of a depressed patient treated with an antidepressant who has experienced good symptom reduction but remains with an IBI-D score of more than −2.7 would alert the clinician to the likelihood of the need to address issues related to QOL and/or function.

Metrics for recovery continue to be needed for mental disorders,[27] and they are of particular importance in chronic or cyclical disorders such as MDD. Although the 1991 consensus definitions of remission, recovery, relapse, and recurrence in MDD by Frank et al[28] set the stage to test ways of assessing recovery as a state of sustained remission over time, investigators have focused on the less complicated to quantify concept of remission. Resnick et al[29] showed in the schizophrenia literature that recovery could be described as remission over long duration (outcome) or as life satisfaction, hope, and empowerment (orientation). These 2 descriptions were referred to in Insel and Scolnick's[27] suggestion that recovery be defined as a complete and permanent remission. Our work supports the complete component of this definition.

This study and the IBI-D itself are of course not without their limitations. Although intended for use in both clinical and research settings, the steps necessary to calculate the IBI-D, not to mention the time required for a patient to complete the QIDS-SR, WSAS, and Q-LES-Q, may impede widespread adoption of IBI-D by practicing clinicians. It takes about 20 to 30 minutes to complete the 3 self-rated scales and about 5 to 10 minutes to calculate the IBI-D. At the same time, indices requiring multistep calculations, such as body mass index and estimated glomerular filtration rate, are widely used in clinical practice, with recent develop-

©2013 American Medical Association. All rights reserved.

Downloaded From: http://archpsyc.jamanetwork.com/ on 05/01/2016

ment of online resources to facilitate calculation. We are planning a website and a smart phone/tablet application, such as the ones available for body mass index, to calculate and inform users about the index parameters. Another barrier to administration/calculation is the need for validation of the IBI-D in different languages and cultures, although the 3 component scales have been validated in several languages. As a future step, we plan on using the Spanish versions of the scales to validate the index in Hispanic patients with MDD. Another concern relates to the self-rated measures on which the IBI-D is based: given that they are self-reported, they are open to a number of potential problems inherent in any self-rated scale, including various forms of bias such as social desirability bias, subjectivity, and state-dependent responses, all of which may affect validity. Because QOL, an important component of IBI-D, is primarily a self-reported construct, we deliberately chose for purposes of consistency self-rated scales for symptom severity and functional impairment. It is important to note that patients with depression could downrate their functioning level and QOL ratings owing to the innate manifestations of the illness. Another difficulty is the ongoing challenge of measuring functioning using self-report. Self-reports of functioning impairment carry the risk for attribution errors. Although the WSAS directs the patient to rate functioning deficits due to their psychiatric symptoms or mental illness, individuals could attribute/misattribute functioning impairments as a result of self-diagnosis, misperceptions, comorbid physical symptoms, fear of stigma, denial, and lack of insight. Conscious of these potential flaws, our approach was to use scales with well-established reliability and validity. Of course, a similar statistical approach could be taken to develop an alternative IBI-D to be used with clinician-rated measures, representing a possible future area of investigation. We fully appreciate that there are many scales available to measure symptom severity, QOL, and functioning. Because we were limited to choosing from scales that were part of the STAR*D trial, approaches based on using these alternative scales could not be evaluated. Therefore, although we find the IBI-D to meet criteria of validity and reliability, it is possible that creating scales based on these alternatives might offer more enhancements.

A second potential criticism of the IBI-D is that in coalescing the 3 domains of symptom severity, functioning, and QOL into a single value, one loses some information. While it is certainly possible to simply note the pretreatment to posttreatment changes in the QIDS-SR, Q-LES-Q, and WSAS to assess treatment efficacy with respect to symptom severity, QOL, and functioning, the IBI-D provides at least 3 advantages over such an approach: (1) It provides the ease and conceptual appeal of a single multidimensional index for tracking improvement. (2) By virtue of the statistical means used in its development and validation, the IBI-D allows for an individual to be compared with a large, validated data set of patients, lending a quantification of the expected level of improvement, whereas such an approach is not inherent in changes in the raw scores of the individual rating scales. (3) The IBI-D captures a shared variance that not only encompasses but also weighs the contributions of the 3 domains of interest to this variance prior to and following treatment. With only the changes in raw scores, one cannot achieve a sense either of the properly weighted contributions of each domain to overall burden or whether changes in QOL and functioning are secondary to other factors that are not related to the illness itself.

Additional strengths of the present study include a large sample size owing to the use of the STAR*D data set, and the generalizability owing to data collection from both primary care and specialty treatment settings. With more doubts being raised about the validity of MDD diagnosis in research-site samples and of self-reported measures of ad-recruited subjects,[30] the use of treatment-seeking patients in the STAR*D study rather than ad-recruited subjects is more likely to mirror actual clinical practice.

Lastly and not withstanding self-report limitations, our research is in line with the latest emphasis on patient-reported outcomes, an advance that is gaining significant traction in the field of medicine, locally and globally. Efforts to develop, test, and implement patient-reported outcome measures are expanding with initiatives such as the National Institutes of Health Patient-Reported Outcomes Measurement Information System[31]; the Patient-Centered Outcomes Research Institute[32]; the Federal Food and Drug Administration–supported initiative for patient-reported outcome (Critical Path Initiative)[33]; World Health Organization International Classification of Functioning, Disability and Health (ICF)[34]; and UK National Health Service Patient-Reported Outcome Measures.[35] Our work adds to some of these efforts, particularly in the area of MDD.

The Individual Burden of Illness Index for Depression is a simple multidimensional metric based on patient-reported outcomes to describe the complexity of depression as an illness including the burden it poses on the individual by incorporating depressive symptom severity, functioning, and QOL impairments. In this article, we have extended our previous work by demonstrating the suitability of the IBI-D in gauging treatment efficacy. These and previous data show that QOL and functioning remain impaired even following remission of symptoms. We hope that the IBI-D will help inform clinicians and researchers regarding outcomes before, during, and after treatment of depression in a manner that goes beyond the currently accepted focus on symptom resolution and more into recovery. The potential exists for the development of similar indices applicable to other disorders, which we hope will expand awareness of a broader conception of the burden of an illness at the individual level.

Submitted for Publication: April 2, 2012; final revision received June 13, 2012; accepted June 20, 2012.
Published Online: January 2, 2013. doi:10.1001/jamapsychiatry.2013.286
Correspondence: Waguih William IsHak, MD, FAPA, Cedars-Sinai Medical Center, Department of Psychiatry, 8730 Alden Dr, Thalians W-157, Los Angeles, CA 90048 (waguih.IsHak@cshs.org).
Conflict of Interest Disclosures: Dr IsHak received research support unrelated to the subject of this article from the National Alliance for Research on Schizophrenia and Depression (quality of life in major depression) and Pfizer (ziprasidone as monotherapy for major depression) that ended on December 31, 2011.

©2013 American Medical Association. All rights reserved.
Downloaded From: http://archpsyc.jamanetwork.com/ on 05/01/2016

Funding/Support: This study was supported by grant N01MH90003 from the National Institute of Mental Health to the University of Texas Southwestern Medical Center.
Disclaimer: This article reflects the views of the authors and may not reflect the opinions or views of the STAR*D Study Investigators or the National Institutes of Health.
Additional Information: The data used in the preparation of this article were obtained from the limited access data sets distributed from the National Institutes of Health–supported Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study, which focused on nonpsychotic major depressive disorder in adults seen in outpatient settings. The primary purpose of this research study was to determine which treatments worked best if the first treatment with medication did not produce an acceptable response.
Additional Contributions: Lev Gertsik, MD, Russell Poland, PhD, and Asbasia Mikhail, MD, contributed significantly to the original concepts highlighted in this article. Tammy Saah, MD, Hala Fakhry, MD, Shakiba Mobaraki, MD, A. John Rush, MD, Jennice Vilhauer, PhD, Mark H. Rapaport, MD, and Andrew Leon, PhD, provided valuable feedback about the concept implementation. Debates led by Peter Whybrow, MD, Ian Cook, MD, and James Spar, MD, contributed significantly to polishing the content of this article. We thank the editors and peer reviewers of the *American Journal of Psychiatry* and the *Administration and Policy in Mental Health and Mental Health Services Research,* and we express deep appreciation for the editor and peer reviewers of *JAMA Psychiatry* for their outstanding review and guidance.

REFERENCES

1. World Health Organization. The global burden of disease, 2004 update. http://www.who.int/healthinfo/global_burden_disease/GBD_report_2004update_full.pdf. Accessed March 1, 2012.
2. McGuire T, Wells KB, Bruce ML, Miranda J, Scheffler R, Durham M, Ford DE, Lewis L; NIMH Affective Disorders Workgroup. Burden of illness. *Ment Health Serv Res*. 2002;4(4):179-185.
3. IsHak WW, Burt T, Sederer LI. *Outcome Measurement in Psychiatry: A Critical Review*. Washington, DC: American Psychiatric Publishing Inc; 2002.
4. IsHak WW, Greenberg JM, Saah T, Mobaraki S, Fakhry H, Wu QV, Ngor E, Yu F, Cohen RM. Development and validation of the Individual Burden of Illness Index for Major Depressive Disorder (IBI-D) [published online October 4, 2011]. *Adm Policy Ment Health*. doi:10.1007/s10488-011-0376-6.
5. Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, Markowitz JC, Ninan PT, Kornstein S, Manber R, Thase ME, Kocsis JH, Keller MB. 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 [published correction appears in *Biol Psychiatry*. 2003;54(5):585]. *Biol Psychiatry*. 2003;54(5):573-583.
6. Mundt JC, Marks IM, Shear MK, Greist JH. The Work and Social Adjustment Scale: a simple measure of impairment in functioning. *Br J Psychiatry*. 2002;180:461-464.
7. Endicott J, Nee J, Harrison W, Blumenthal R. Quality of life enjoyment and satisfaction questionnaire: a new measure. *Psychopharmacol Bull*. 1993;29(2):321-326.
8. Rush AJ, Fava M, Wisniewski SR, Lavori PW, Trivedi MH, Sackeim HA, Thase ME, Nierenberg AA, Quitkin FM, Kashner TM, Kupfer DJ, Rosenbaum JF, Alpert J, Stewart JW, McGrath PJ, Biggs MM, Shores-Wilson K, Lebowitz BD, Ritz L, Niederehe G; STAR*D Investigators Group. Sequenced Treatment Alternatives to Relieve Depression (STAR*D): rationale and design. *Control Clin Trials*. 2004;25(1):119-142.
9. IsHak WW, Greenberg JM, Balayan K, Kapitanski N, Jeffrey J, Fathy H, Fakhry H, Rapaport MH. Quality of life: the ultimate outcome measure of interventions in major depressive disorder. *Harv Rev Psychiatry*. 2011;19(5):229-239.
10. Fava M, Rush AJ, Trivedi MH, Nierenberg AA, Thase ME, Sackeim HA, Quitkin FM, Wisniewski S, Lavori PW, Rosenbaum JF, Kupfer DJ. Background and rationale for the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study. *Psychiatr Clin North Am*. 2003;26(2):457-494, x.
11. Steiger JH. Tests for comparing elements of a correlation matrix. *Psychol Bull*. 1980;87(2):245-251. doi:10.1037/0033-2909.87.2.245.
12. Cohen J, Cohen P. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. 2nd ed. London, England: Lawrence Erlbaum; 1983.
13. Rapaport MH, Clary C, Fayyad R, Endicott J. Quality-of-life impairment in depressive and anxiety disorders. *Am J Psychiatry*. 2005;162(6):1171-1178.
14. Schechter D, Endicott J, Nee J. Quality of life of 'normal' controls: association with lifetime history of mental illness. *Psychiatry Res*. 2007;152(1):45-54.
15. Nierenberg AA, Rapaport MH, Schettler PJ, Howland RH, Smith JA, Edwards D, Schneider T, Mischoulon D. Deficits in psychological well-being and quality-of-life in minor depression: implications for *DSM-V*. *CNS Neurosci Ther*. 2010;16(4):208-216.
16. World Health Organization. Constitution of the World Health Organization. http://www.who.int/governance/eb/who_constitution_en.pdf. Accessed March 22, 2012.
17. Trivedi MH, Daly EJ. Measurement-based care for refractory depression: a clinical decision support model for clinical research and practice. *Drug Alcohol Depend*. 2007;88(suppl 2):S61-S71.
18. Pyne JM, Patterson TL, Kaplan RM, Gillin JC, Koch WL, Grant I. Assessment of the quality of life of patients with major depression. *Psychiatr Serv*. 1997;48(2):224-230.
19. De Almeida Fleck MP, Simon G, Herrman H, Bushnell D, Martin M, Patrick D; Longitudinal Investigation of Depression Outcomes Group. Major depression and its correlates in primary care settings in six countries: 9-month follow-up study. *Br J Psychiatry*. 2005;186:41-47.
20. Trivedi MH, Rush AJ, Wisniewski SR, Warden D, McKinney W, Downing M, Berman SR, Farabaugh A, Luther JF, Nierenberg AA, Callan JA, Sackeim HA. Factors associated with health-related quality of life among outpatients with major depressive disorder: a STAR*D report. *J Clin Psychiatry*. 2006;67(2):185-195.
21. Papakostas GI, Petersen T, Mahal Y, Mischoulon D, Nierenberg AA, Fava M. Quality of life assessments in major depressive disorder: a review of the literature. *Gen Hosp Psychiatry*. 2004;26(1):13-17.
22. Reed C, Monz BU, Perahia DG, Gandhi P, Bauer M, Dantchev N, Demyttenaere K, Garcia-Cebrian A, Grassi L, Quail D, Tylee A, Montejo AL. Quality of life outcomes among patients with depression after 6 months of starting treatment: results from FINDER. *J Affect Disord*. 2009;113(3):296-302.
23. Zimmerman M, McGlinchey JB, Posternak MA, Friedman M, Attiullah N, Boerescu D. How should remission from depression be defined? the depressed patient's perspective. *Am J Psychiatry*. 2006;163(1):148-150.
24. Zimmerman M, McGlinchey JB, Posternak MA, Friedman M, Boerescu D, Attiullah N. Remission in depressed outpatients: more than just symptom resolution? *J Psychiatr Res*. 2008;42(10):797-801.
25. Angermeyer MC, Holzinger A, Matschinger H, Stengler-Wenzke K. Depression and quality of life: results of a follow-up study. *Int J Soc Psychiatry*. 2002;48(3):189-199.
26. American Psychiatric Association. Practice guideline for the treatment of patients with major depressive disorder, 3rd Ed. http://psychiatryonline.org/content.aspx?bookid=28&sectionid=1667485. Accessed November 9, 2012.
27. Insel TR, Scolnick EM. Cure therapeutics and strategic prevention: raising the bar for mental health research. *Mol Psychiatry*. 2006;11(1):11-17.
28. Frank E, Prien RF, Jarrett RB, Keller MB, Kupfer DJ, Lavori PW, Rush AJ, Weissman MM. Conceptualization and rationale for consensus definitions of terms in major depressive disorder: remission, recovery, relapse, and recurrence. *Arch Gen Psychiatry*. 1991;48(9):851-855.
29. Resnick SG, Rosenheck RA, Lehman AF. An exploratory analysis of correlates of recovery. *Psychiatr Serv*. 2004;55(5):540-547.
30. Kobak KA, Kane JM, Thase ME, Nierenberg AA. Why do clinical trials fail? the problem of measurement error in clinical trials: time to test new paradigms? *J Clin Psychopharmacol*. 2007;27(1):1-5.
31. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, Ader D, Fries JF, Bruce B, Rose M; PROMIS Cooperative Group. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Med Care*. 2007;45(5)(suppl 1):S3-S11.
32. Methodology Committee of the Patient-Centered Outcomes Research Institute (PCORI). Methodological standards and patient-centeredness in comparative effectiveness research: the PCORI perspective. *JAMA*. 2012;307(15):1636-1640.
33. Critical Path Institute. Critical Path Institute project pipeline. http://www.c-path.org/project-pipeline.cfm. Accessed June 11, 2012.
34. World Health Organization. International Classification of Functioning, Disability and Health (ICF). http://www.who.int/classifications/icf/en/. Accessed June 11, 2012.
35. UK National Health Service (NHS) Information Centre for Health and Social Care. Patient Reported Outcome Measures (PROMS). http://www.ic.nhs.uk/statistics-and-data-collections/hospital-care/patient-reported-outcome-measures-proms. Accessed June 11, 2012.

©2013 American Medical Association. All rights reserved.

Downloaded From: http://archpsyc.jamanetwork.com/ on 05/01/2016