

REGULAR RESEARCH ARTICLE

Prediction of Acute-Phase Treatment Outcomes by Adding a Single-Item Measure of Activity Impairment to Symptom Measurement: Development and Validation of an Interactive Calculator from the STAR*D and CO-MED Trials

Manish K. Jha,^o Charles South, Jay Trivedi, Abu Minhajuddin, A. John Rush, Madhukar H. Trivedi

Center for Depression Research and Clinical Care, UT Southwestern Medical Center, Dallas, Texas (Drs Jha, South, Minhajuddin, and Trivedi); Departments of Psychiatry and Neuroscience, Icahn School of Medicine at Mount Sinai, New York, New York (Dr Jha); Trinity Valley School, Fort Worth, Texas (Mr Trivedi); Prof Emeritus, Duke-National University of Singapore, Singapore (Dr Rush); Adjunct Professor, Dept. Psychiatry, Duke Medical School, Durham, North Carolina (Dr Rush); Adjunct Clinical Professor, Texas Tech University-Health Sciences Center, Permian Basin, Texas (Dr Rush).

Correspondence: Madhukar H. Trivedi, MD, Center for Depression Research and Clinical Care, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390-9119 (madhukar.trivedi@utsouthwestern.edu).

Abstract

Background: Day-to-day functioning is impaired in major depressive disorder. Yet there are no guidelines to systematically assess these functional changes. This report evaluates prognostic utility of changes in activity impairment to inform clinical decision-making at an individual level.

Methods: Mixed model analyses tested changes in activity impairment (sixth item of Work and Activity Impairment scale, rated 0–10) at mid-point (week 6) and end of step 1 (weeks 12–14) in the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial ($n=2697$) after controlling for depression severity [Quick Inventory of Depressive Symptomatology Self-Report (QIDS-SR)]. Interactive calculators for end of step 1 remission ($QIDS-SR \leq 5$) and no meaningful benefit ($<30\%$ QIDS-SR reduction from baseline) were developed for participants with complete data ($n=1476$) and independently replicated in the Combining Medications to Enhance Depression Outcomes trial ($n=399$).

Results: Activity impairment improved independently with acute-phase treatment in STAR*D ($F=7.27$; $df=2,2625$; $P<.001$). Baseline to mid-point activity impairment change significantly predicted remission ($P<.001$, model area under the curve = 0.823) and no meaningful benefit ($P<.001$, area under the curve = 0.821) in the STAR*D trial. Adding activity impairment variables to depression severity measures correctly reclassified 28.4% and 15.8% remitters and nonremitters (net reclassification improvement analysis, $P<.001$), and 11.4% and 16.8% of those with no meaningful benefit and meaningful benefit (net reclassification improvement analysis, $P<.001$). The STAR*D trial model estimates accurately predicted remission (area under the curve = 0.80) and no meaningful benefit (area under the curve = 0.82) in the Combining Medications to Enhance Depression Outcomes trial and was used to develop an interactive calculator.

Received: July 10, 2018; Revised: February 1, 2019; Accepted: March 21, 2019

© The Author(s) 2019. Published by Oxford University Press on behalf of CINP.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Significance Statement

While day-to-day functioning has been long recognized as an important secondary outcome, there are no guidelines or recommendations to incorporate functional changes in making treatment decisions for patients with major depressive disorder (MDD). In this report, we replicated previous findings that activity impairment reduced significantly with acute-phase antidepressant treatment even after controlling for changes in depression severity. Additionally, acute-phase remission rates were significantly lower in those with severe (15.5%) and moderate (34.3%) activity impairment than those with no/minimal impairment (66.7%) at week 6. We developed an interactive web-based calculator using baseline and week 6 depression severity and activity impairment scores to predict the likelihood of remission (Area Under the Curve [AUC]=0.82) and no meaningful benefit (<30% reduction, AUC=0.82) at the end of the acute phase (weeks 12–14) and demonstrated that inclusion of a single-item activity impairment measure significantly improved predictive accuracy of models compared with measuring depression severity only. We validated this calculator in an unrelated sample of MDD outpatients with comparable accuracy (remission AUC=0.80 and no meaningful benefit AUC=0.82).

Conclusion: A single-item self-report measure of activity impairment changes independently with antidepressant treatment. Baseline to week 6 changes in activity impairment and depression severity can be combined to predict acute-phase remission and no meaningful benefit at an individual level.

Keywords: activity impairment, remission, STAR*D, antidepressant treatment, response prediction, major depression, measurement-based care

Introduction

Major depressive disorder (MDD) is a widely prevalent (Kessler et al., 2003) and commonly disabling disorder (Vos et al., 2015) that is estimated to cost the United States over \$210 billion dollars per year (Greenberg et al., 2015). Only one-third of patients with MDD remit with initial antidepressant treatment, and over one-third do not respond to 2 or more antidepressant medications (Rush et al., 2006b). Additionally, measurement-based care protocols recommend evaluations every 2 to 3 weeks for at least 2 to 3 months to determine the therapeutic success of a given antidepressant (Guo et al., 2015). Either ineffective treatments are continued over prolonged periods in those who do not respond, or unnecessary visits and healthcare costs are incurred for those who remit with initial antidepressant treatment. Thus, patients with MDD who have a low likelihood of meaningful benefit can be switched to another antidepressant medication or psychotherapy, or be augmented with pharmacological, exercise, or brain stimulation (repetitive or deep transcranial magnetic stimulation) treatments (Rush et al., 2006c; Trivedi et al., 2006a, 2011; Gelenberg et al., 2010). Conversely, those at high likelihood of remitting with their initial treatment may be maintained on it with less frequent visits.

Previous efforts of predicting nonresponse to initial antidepressant treatment have been limited by their focus on core depressive symptoms with little additional information gained by addition of baseline features to the early changes in depressive symptoms (Kuk et al., 2010; Li et al., 2012; Perlis, 2013). Incorporating functional assessments can improve clinical prognostication as functional impairments persist even after symptomatic remission (Trivedi et al., 2009, 2013), and early improvements in functional measures predict better longer-term outcomes (Jha et al., 2016a, 2016b, 2016c).

One such measure in functional domain is impairment in day-to-day non-work-related activities (activity impairment). The sixth item of the Work Productivity and Activity Impairment (WPAI) scale is a widely used, well-validated measure that allows easy interpretability due to normative data available through the large-scale National Health and Wellness Survey (Reilly et al.,

1993; Gupta et al., 2012). It is also a more practical measure that is applicable to all patients with MDD than other measures such as work productivity, which apply only to employed patients. In a previous report, activity impairment improved early with antidepressant treatment (by week 6) and this early improvement predicted higher rates of remission at 3 and 7 months even after controlling for remission status at week 6 (Jha et al., 2017). However, guidelines for practically implementing these findings in clinical practice remain unclear. A publicly available web-based calculator to predict future outcomes incorporating measures of depressive symptoms and activity impairment may be used by patients and their clinicians to inform clinical decision-making.

There are 2 aims of this report. The first was to replicate previous findings that activity impairment improves early with antidepressant treatment and predicts longer-term clinical outcomes, independent of changes in depressive symptom severity, in a sample of treatment-seeking outpatients who were enrolled in step 1 of Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial (Trivedi et al., 2006b). The second was to develop an interactive calculator to predict remission and no meaningful benefit at an individual level in the STAR*D trial and replicate these predictions in a separate unrelated sample treatment of outpatients with MDD who participated in the Combining Medications to Enhance Depression Outcomes (CO-MED) trial (Rush et al., 2011).

Baseline activity/interest dimension obtained from measures of depressive symptom severity has been shown to predict acute-phase treatment outcomes in the STAR*D trial (Uher et al., 2012). However, the association of this depressive symptom dimension with patient-reported impairment in day-to-day activities has not been studied previously. This is the first report, to our knowledge, from the STAR*D trial to evaluate the changes in activity impairment and to demonstrate the improved prognostication of future clinical outcomes by adding assessment of this single item measure to the current practice of measuring depression severity.

Methods

Study Overview and Participants

STAR*D Trial

The data for this report were obtained from participants ($n=2697$) who completed the WPAI scale at baseline and at least once after baseline visit during step 1 of the STAR*D trial (NCT00021528). A subset of these participants who completed the activity impairment and depression rating scales at baseline and mid-point and the depression rating at the end of step 1 of STAR*D (referred to as complete data for this report) was used to develop the prognostic model for the interactive calculator ($n=1476$). The details of the STAR*D trial, including its rationale, methods, design, institutional review board approvals, and data safety monitoring board oversight, have been published elsewhere (Fava et al., 2003; Rush et al., 2004). Briefly, treatment-seeking outpatients with single or recurrent nonpsychotic MDD ($n=4041$) were recruited from 14 regional centers across the United States that included public or private psychiatric ($n=23$) and primary care ($n=18$) clinics. Broad inclusion and minimal exclusion criteria ensured that most patients qualified for the study. Eligibility criteria included age 18 to 75 years, baseline Hamilton Rating Scale of Depression 17-item (HRSD₁₇) (Hamilton, 1960) score ≥ 14 , and the clinician's determination that outpatient antidepressant medication was both safe and indicated. Outpatients who failed any of antidepressants used in the STAR*D trial were excluded. All study-related assessments and procedures were completed after obtaining informed consent.

CO-MED Trial

This report includes CO-MED trial (NCT00590863) participants who completed the activity impairment and depression rating scales at baseline and week 6 and the depression rating at week 12 ($n=399$). The details of the CO-MED trial, including the recruiting sites, inclusion and exclusion criteria, and institutional review board approvals, are described in detail by Rush et al. (Rush et al., 2011). The study was conducted by the Depression Treatment Network from March 2008 through February 2009 and enrolled participants from 6 primary and 9 psychiatric care sites after obtaining written informed consent from each participant prior to completing any study-related procedure or assessment (Rush et al., 2011). Inclusion in the CO-MED trial was restricted to 18- to 75-year-old MDD outpatients with nonpsychotic chronic (current episode exceeded 2 years) or recurrent depression with current episode ≥ 2 months and a baseline HRSD₁₇ ≥ 16 . At baseline visit, participants were randomly assigned to one of the following treatment arms in a 1:1:1 ratio after stratification by clinical sites: (1) escitalopram plus placebo (SSRI monotherapy), (2) sustained-release (SR) bupropion plus escitalopram (bupropion-SSRI combination), and (3) extended-release (XR) venlafaxine plus mirtazapine (venlafaxine-mirtazapine combination). Postrandomization visits were conducted at weeks 1, 2, 4, 6, 8, 10, and 12 for acute phase and weeks 16, 20, 24, and 28 for continuation phase.

Assessments

WPAI

The sixth item of this self-report measure, which has good construct validity and test-retest reliability, was used to measure non-work-related impairment in regular daily activities "such as work around the house, shopping, childcare, exercising, studying, etc." (or activity impairment) (Reilly et al., 1993). Activity impairment correlates significantly with depression severity and

functional impairment ($r=54-0.70$) (Jha et al., 2017). The activity impairment item (#6 of WPAI) is rated on a scale of 0 to 10, where 0 indicates no impairment and 10 indicates impairment that completely prevents participants from doing daily activities (Reilly et al., 1993). Using community norms from the large-scale National Health and Wellness Survey ($n=75000$) (Gupta et al., 2012), participants were categorized in the following 3 groups: no or minimal activity impairment (activity impairment score of 0–2), moderate activity impairment (activity impairment score of 3–6), and severe impairment (activity impairment score of 7–10).

Quick Inventory of Depressive Symptomatology Self-report (QIDS-SR)

The total score of the QIDS-SR (range of 0–27) is based on the 9 criteria symptom domains of MDD out of the 16 items, each of which is scored from 0 to 3 (Rush et al., 2003). The Pearson moment correlation between QIDS-SR and HRSD₁₇ was 0.86 and the Cronbach's α of QIDS-SR was 0.86 in a previous report (Rush et al., 2003). The QIDS-SR served as the measure of depressive symptoms in both the STAR*D and CO-MED trials.

Medications

Step 1 of STAR*D

Participants were started on citalopram 20 mg/d with a dose increase to 40 mg/d permitted by week 4 and to 60 mg/d (maximum dose) by week 6 using measurement-based care procedures with the aim of reaching symptom remission if side effects were tolerable. The protocol recommended treatment visits at 2, 4, 6, 9, and 12 weeks (with an optional week 14 visit if needed). Participants could discontinue citalopram before 12 weeks if (1) intolerable side effects required a medication change; (2) an optimal dose increase was not possible because of side effects or participant choice; or (3) significant symptoms (QIDS-C score ≥ 9) were present after 9 weeks at maximally tolerated doses. Participants could opt to move to the next treatment level if they had intolerable side effects or if the QIDS-C score was >5 after an adequate trial in terms of dose and duration.

CO-MED Trial

Participants in all 3 treatment arms received 2 types of pills in single-blind fashion where study personnel knew of both pill types, but participants knew only the first pill type. In the SSRI monotherapy arm, participants were started on escitalopram 10 mg/d with dose increase to 20 mg/d permitted at week 4; pill placebo was added as the second pill type at week 2. In the bupropion-SSRI treatment arm, bupropion SR was initiated at 150 mg/d and was increased to 300 mg/d at the week 1 visit, escitalopram was started at 10 mg/d as the second pill type at week 2, and dose increases of bupropion SR (up to 200 mg twice daily) and escitalopram (up to 20 mg/d) were permitted from weeks 4 to 8. In the venlafaxine-mirtazapine arm, venlafaxine XR was initiated at 37.5 mg/d and titrated to 150 mg/d by week 1, mirtazapine 15 mg/d was added as the second pill type at week 2, and dose increases of venlafaxine XR (up to 300 mg/d) and mirtazapine (up to 45 mg/d) were permitted from weeks 4 to 8.

Statistical Analysis Plan

Descriptive statistics were used to compare participants on clinical and sociodemographic variables based on their activity impairment level (no/minimal, moderate, or severe) at baseline of step 1 of the STAR*D trial ($n=2697$). Those who had complete

data and were included in the interactive calculator ($n=1476$) were also compared for these variables to participants who did not have complete data ($n=1221$). The activity-interest dimension was also computed at baseline (Uher et al., 2012) and its correlation with activity impairment was calculated.

Replication of Previous Findings of Changes in Activity Impairment

Separate repeated-measures mixed-model analyses with a random intercept assessed whether activity impairment improved from baseline to mid-point and end of step 1 of STAR*D trial ($n=2697$) using PROC MIXED in SAS, before and after controlling for changes in depression severity by including QIDS-SR as a time-varying covariate. To quantify the magnitude of change, effect sizes were estimated and proportion of participants at different activity impairment levels (no/minimal, moderate, and severe) were calculated. Descriptive statistics were also used to report the proportion of participants in remission at end of step 1 based on mid-point activity impairment levels and remission status.

Development of Predictive Models for Interactive Calculator in STAR*D Trial

For a parsimonious model incorporating activity impairment and depression severity, 2 separate logistic regression models were used to predict remission ($QIDS-SR \leq 5$) and no meaningful benefit (<30% decrease in QIDS-SR from baseline to end of step 1) (Rush et al., 2006a; South et al., 2017), respectively, at end of step 1 of STAR*D. These logistic regression models included baseline QIDS-SR, baseline activity impairment, change in QIDS-SR from baseline to mid-point [(mid-point QIDS-SR) – (baseline QIDS-SR)], and change in activity impairment from baseline to mid-point [(mid-point activity impairment) – (baseline activity impairment)]. Using the estimates obtained from these models, the probability of remission and no meaningful benefit were developed for an interactive calculator using the following equation:

$$\log\left(\frac{p}{1-p}\right) = b_0 + b_{\text{baseline QIDS-SR}} \cdot (\text{baseline QIDS-SR}) \\ + b_{\text{change in QIDS-SR}} \cdot (\text{change in QIDS-SR}) \\ + b_{\text{baseline activity impairment}} \cdot (\text{baseline activity impairment}) \\ + b_{\text{change in activity impairment}} \cdot (\text{change in activity impairment})$$

where p is the probability of the outcome variable, and b_i is the regression parameter for the i^{th} predictor. A receiver operating characteristics (ROC) curve was plotted to obtain the area under the curve (AUC) in the STAR*D trial, and calibration plots (Kuhn and Johnson, 2013) were used to evaluate the agreement between predicted probabilities and actual outcome. The initial versions of these models were based on percent changes in activity impairment and depression severity but excluded participants with an activity impairment score of 0 at baseline ($n=105$), thus reducing the generalizability. Hence, models using change scores were preferred and included in this report. To evaluate the internal replicability of these models (Morin & Davis, 2017), 10-fold cross-validation was conducted where the STAR*D sample was partitioned into 10 parts, and the logistic regression models with remission and no meaningful benefit were run on 9 parts of the sample. The β estimates for each variable from these models were then used to compute the probabilities of outcome in the tenth part of the sample, and accuracy was tested by comparing these predictions with observed occurrences using ROC plots. This process was repeated 10 times to obtain 10 AUC values each for predicted occurrence of remission and no meaningful benefit.

Improvement in Predictive Accuracy by Inclusion of Activity Impairment Variables

Separate net reclassification improvement analyses were conducted in the STAR*D trial for predictive models with remission and no meaningful benefit to evaluate whether addition of activity impairment (baseline and baseline to mid-point change) variables significantly improved the predictive accuracy compared with models including only depression severity (baseline and baseline to mid-point change) variables.

Validation of Predictive Models for Interactive Calculator in the CO-MED Trial

The β estimates obtained from the STAR*D trial for each variable were then used to compute probabilities (using the equation described above) of remission and no meaningful benefit for each participant in the CO-MED trial. Accuracy was tested by comparing these predictions with observed occurrence of remission or no meaningful benefit using ROC plots in the CO-MED trial.

Results

There were 3 analytic samples in this report. The first analytic sample included all participants in the STAR*D trial with an activity impairment score at baseline and at least one more score after baseline ($n=2697$). Of these 2697 participants, the second analytic sample (used for predictive model development) was limited to those with complete data (activity impairment and QIDS-SR scores at baseline, mid-point, and end of step 1; $n=1476$; also see [supplementary Figure 1](#)). Those who were excluded from the second sample ($n=1221$) were younger in age (mean=41.1 years, $SD=13.2$), had fewer years of education (mean=12.6, $SD=3.0$), and were more likely to be male (39.4%), uninsured (38.3%), and single/divorced (58.6%) ([supplementary Table 1](#)). The third analytic sample included CO-MED trial participants with activity impairment and QIDS-SR scores at baseline and 6 weeks and QIDS-SR at 12 weeks ($n=399$). The mean (SD) duration from baseline in weeks for mid-point and end of step 1 were 6.32 (0.65) and 12.38 (2.23), respectively, for participants with complete data ($n=1476$) in the STAR*D study.

Of the first analytic sample at baseline, 399 (14.79%) had no or minimal activity impairment, while 1155 (42.83%) had moderate and 1143 (42.38%) had severe activity impairment. Baseline clinical and sociodemographic differences based on activity impairment levels are reported in [Table 1](#). The correlation between participant-reported activity impairment and the activity-interest dimension as reported (Uher et al., 2012) was 0.37 ($n=2584$).

Replication of Previous Findings of Changes in Activity Impairment

Activity impairment decreased significantly from baseline to mid-point and end of step 1 of the STAR*D trial ($F=427.00$; $df=2,2183$; $P<.0001$, effect size=0.77). The estimated reductions from baseline to mid-point and from baseline to end of step 1 were 1.55 ($SE=0.07$) and 1.84 ($SE=0.07$) points, respectively. This improvement in activity impairment continued to be significant in mixed-model analyses that controlled for change in depression severity ($F=7.27$; $df=2,2625$; $P=.0007$, adjusted effect size=0.11) at each visit. Among the STAR*D participants with complete data ($n=1476$), the proportion of those with severe impairment reduced from 41.9% at baseline to 21.9% at mid-point and 20.1% at end of step 1, respectively ([Figure 1](#)). Participants

Table 1. Clinical and sociodemographic characteristics of STAR*D trial participants based on activity impairment categories prior to treatment initiation

	No/minimal activity impairment		Moderate activity impairment		Severe activity impairment		Test statistic	
Number in each category	399		1155		1143			
Categorical variables	n	%	n	%	n	%	χ^2	P value
Female gender	238	59.6	715	61.9	741	64.8	4.1	.128
Caucasian race	319	79.9	961	83.2	890	77.9	10.5	.005
Hispanic ethnicity	38	9.5	113	9.8	128	11.2	1.6	.454
Unemployed	137	34.3	481	41.6	620	54.2	61.9	<.001
Insured	263	69.2	773	68.3	696	63.0	8.8	.012
Married/cohabiting	170	42.6	543	47.0	475	41.6	7.3	.026
Lifetime suicide attempt	55	13.8	172	14.9	222	19.4	11.2	.004
Chronic depression	92	23.1	299	25.9	311	27.2	2.7	.263
Comorbid psychiatric disorders							83.1	<.001
0	194	49.5	494	43.5	389	35.0		
1	113	28.8	311	27.4	266	23.9		
2	50	12.8	172	15.1	187	16.8		
3	16	4.1	77	6.8	107	9.6		
4+	19	4.6	82	7.2	164	14.7		
Comorbid medical disorder							34.9	<.001
0	384	96.2	1067	92.4	999	87.4		
1	11	2.8	67	5.8	118	10.3		
2+	4	1.0	21	1.8	23	2.3		
Continuous variables	Mean	SD	Mean	SD	Mean	SD	F value	P value
Age, years	40.0	14.0	42.5	13.3	42.5	12.6	6.1	.002
Age at onset of 1st MDE	25.5	14.7	25.7	14.5	25.3	14.2	0.2	.833
Number of MDEs	4.8	7.4	6.0	11.0	5.4	8.2	2.3	.099
Education, years	13.0	3.1	13.0	3.2	12.5	3.1	9.2	<.001
Depression severity (QIDS-SR)	12.3	5.0	14.5	4.5	17.6	4.6	243.2	<.001
Activity impairment (WPAI #6)	0.8	0.9	4.7	1.0	8.1	1.0	8544.5	<.001
Psychosocial function (WSAS)	15.8	8.9	20.9	7.1	29.0	7.6	577.1	<.001
Quality of life (Q-LES-Q-SF)	42.8	7.8	39.8	6.7	33.0	7.5	375.5	<.001

Abbreviations: MDE, major depressive episode; QIDS-SR, Quick Inventory of Depressive Symptomatology Self-Report version; Q-LES-Q-SF, general activities summary scale of Quality of Life Enjoyment and Satisfaction Questionnaire; STAR*D, Sequenced Treatment Alternatives to Relieve Depression; WSAS, Work and Social Adjustment Scale.

Chronic depression was defined as index MDE ≥ 2 years. Activity impairment was measured with the sixth item (#6) of the Work Productivity and Activity Impairment Questionnaire (WPAI) and was categorized as following: no or minimal activity impairment (activity impairment score of 0–2), moderate activity impairment (activity impairment score of 3–6), and severe impairment (activity impairment score of 7–10).

with severe activity impairment at mid-point (week 6) of STAR*D were less likely to remit (15.5%, 50/323) at end of step 1 than those with moderate (34.3%, 204/595) and no/minimal impairment (66.7%, 372/558) at mid-point. These differences were maintained even after stratifying for remission status at mid-point (Figure 2).

Development of Predictive Models for Interactive Calculator in the STAR*D Trial

Of the 1476 participants with complete data in the STAR*D trial, 42.4% attained remission and 33.2% had no meaningful benefit. Similarly, of the 399 participants from the validation sample in CO-MED trial, 51.4% attained remission and 17.3% had no meaningful benefit.

In the STAR*D trial, greater baseline levels and smaller reductions from baseline to mid-point of depression severity and activity impairment predicted lower likelihood of remission and higher likelihood of no meaningful benefit at end of step 1 (Table 2). The AUC values in the STAR*D trial ($n=1476$) were 0.823 (95% DeLong CI=0.802, 0.844) for remission and 0.821 (95% DeLong CI=0.798, 0.843) for no meaningful benefit (Figure 3). Using

10-fold cross validation in the STAR*D study where β estimates from nine-tenth of the sample was used to predict outcomes in the remaining one-tenth of the sample, the mean AUC values for remission and no meaningful benefit were 0.819 (range = 0.764–0.901) (mean=0.819) and 0.818 (range = 0.773–0.873), respectively. Changes in activity impairment significantly predicted likelihood of remission and no meaningful benefit even when measured as percent change (excluding participants with baseline activity impairment score of 0) (supplementary Table 2).

Improvement in Predictive Accuracy by Inclusion of Activity Impairment Variables

Adding activity impairment variables to the model significantly improved the reclassification of both remission and no meaningful benefit. The net reclassification improvement for remission and no meaningful benefit were 0.44 (95% CI 0.342, 0.542; $P<.0001$) and 0.283 (95% CI 0.175, 0.390; $P<.0001$), respectively. With the inclusion of activity impairment variables in the remission model, 28.4% of the remitters were correctly reclassified and 15.8% of the nonremitters were correctly reclassified. In the no meaningful benefit model, 11.4% of those with no meaningful

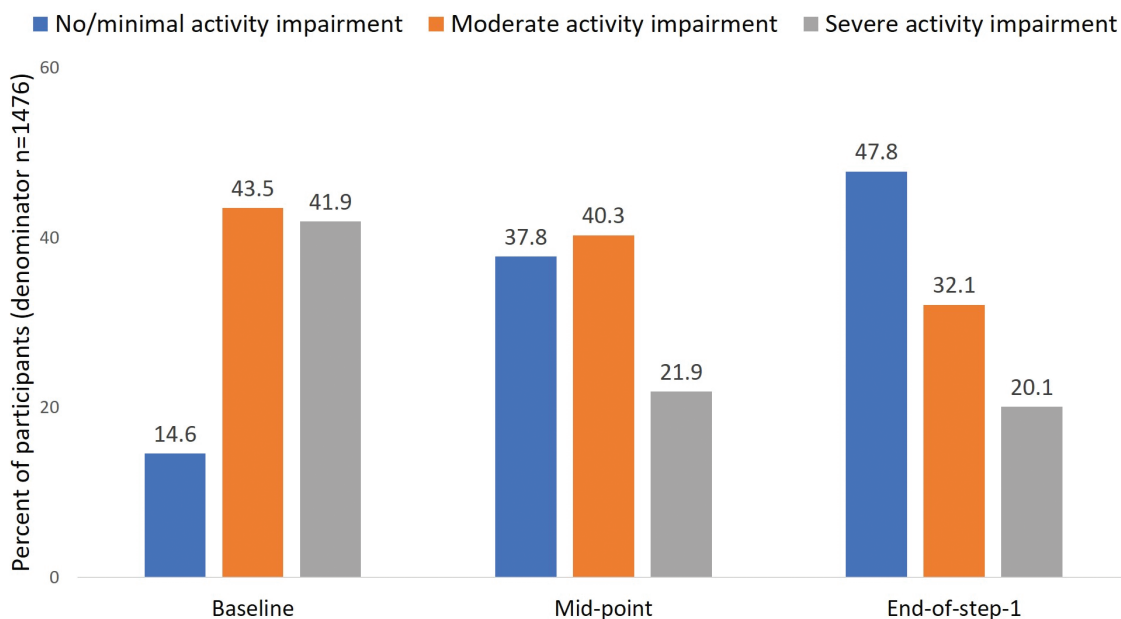


Figure 1. Changes in activity impairment from baseline to end-of-step 1 of the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial. Activity impairment was measured with the sixth item (#6) of the Work Productivity and Activity Impairment Questionnaire (WPAI) and was categorized as following: no or minimal activity impairment (activity impairment score of 0–2), moderate activity impairment (activity impairment score of 3–6), and severe impairment (activity impairment score of 7–10).

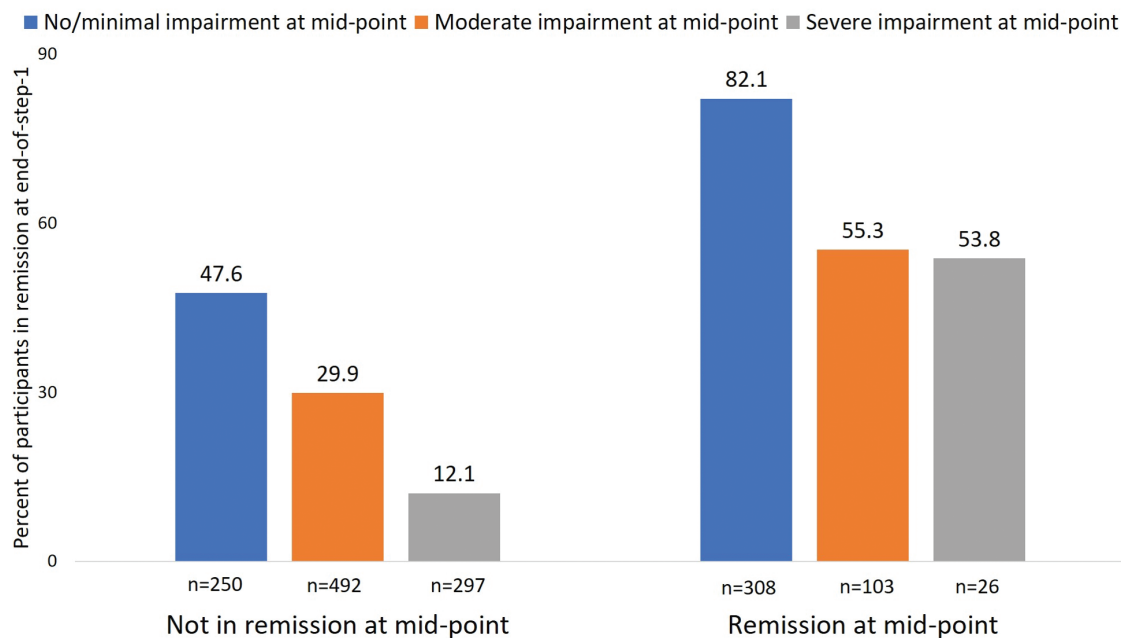


Figure 2. Remission rates at end of step 1 in the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial based on mid-point activity impairment categories. Remission was ascribed if end of step 1 Quick Inventory of Depressive Symptomatology Self-Report (QIDS-SR) was ≤ 5 , activity impairment was measured with the sixth item (#6) of the Work Productivity and Activity Impairment Questionnaire (WPAI) and was categorized as following: no or minimal activity impairment (activity impairment score of 0–2), moderate activity impairment (activity impairment score of 3–6), and severe impairment (activity impairment score of 7–10).

benefit were correctly reclassified, whereas 16.8% with meaningful-benefit were correctly reclassified. While the analyses for net reclassification improvement were done at an individual level, for descriptive purposes, probabilities of event (remission and no meaningful benefit) and nonevent (nonremission and meaningful-benefit) were grouped in tertiles (<33%, 33% to <67%, and $\geq 67\%$) and are presented in [supplementary Table 3](#) as comparison for models including QIDS-SR variables only and

models that included activity impairment variables along with QIDS-SR ones.

Validation of Predictive Models for Interactive Calculator in the CO-MED Trial

In the validation sample from the CO-MED trial ($n=399$), the AUC values were 0.798 (95% DeLong CI=0.7547, 0.8414) for

Table 2. Likelihood of remission and no meaningful benefit in the STAR*D trial (n=1476) based on changes in depressive symptoms and activity impairment.

	Remission			No meaningful benefit		
	OR	95% CI	P value	OR	95% CI	P value
Baseline QIDS-SR	0.784	0.753, 0.816	<.001	1.084	1.048, 1.121	<.001
Baseline activity impairment	0.989	0.983, 0.995	<.001	1.009	1.003, 1.016	.005
Change in QIDS-SR ^a	0.802	0.775, 0.830	<.001	1.288	1.244, 1.335	<.001
Change in activity impairment ^a	0.982	0.977, 0.988	<.001	1.011	1.005, 1.016	<.001

Abbreviations: CI, confidence interval; OR, odds ratio; QIDS-SR, Quick Inventory of Depressive Symptomatology Self-Report version; STAR*D, Sequenced Treatment Alternatives to Relieve Depression.

^aChanges in QIDS-SR and activity impairment were calculated by subtracting the score at baseline from that at mid-point, lower values denote improvement of symptoms and function, and activity impairment was measured with the sixth item (#6) of Work Productivity and Activity Impairment Questionnaire. Remission was ascribed if end of step 1 QIDS-SR was ≤ 5 . No meaningful benefit was ascribed if reduction in QIDS-SR from baseline to end of step 1 QIDS-SR was $<30\%$.

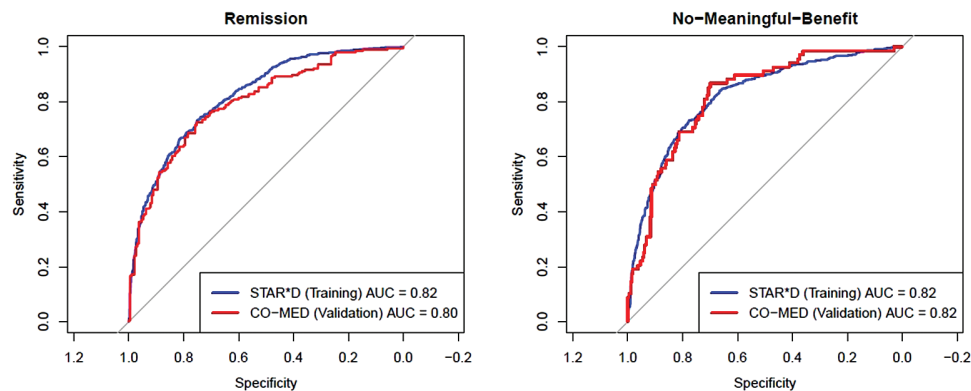


Figure 3. Receiver operating characteristics (ROC) curve for remission and no meaningful benefit in the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) and Combining Medications to Enhance Depression Outcomes (CO-MED) trials. Remission was ascribed if end of step 1 (around week 12) of the Quick Inventory of Depressive Symptomatology Self-Report (QIDS-SR) was ≤ 5 , and no meaningful benefit was ascribed if end of step 1 (around week 12) QIDS-SR was $\geq(0.70)$ * (baseline QIDS-SR).

remission and 0.822 (95% DeLong CI=0.7704, 0.874) for no meaningful benefit (Figure 3). Calibration plots (supplementary Figure 2) show that the predicted probabilities were well calibrated, aside from the tails.

Interactive Web-based Calculator

The predictive models for remission and no meaningful benefit that provide an individual-patient level probabilities for these outcomes were implemented as an interactive calculator using the Shiny package in R (<https://shiny.rstudio.com/>) and deployed on a server for universal use. With the interactive web-based calculator, users can specify the values of the 4 predictor variables (baseline and week 6 scores of both QIDS-SR and sixth item of WPAI [activity impairment]), view where the values lay according to the distributions in the STAR*D dataset, and obtain estimates for the probability of remission and probability of no meaningful benefit. A screenshot of the application is included as Figure 4.

Discussion

We found that activity impairment improves significantly with acute-phase antidepressant treatment, with about 50% reduction in the proportion of participants who report severe activity impairment from baseline to week 6 of acute-phase antidepressant treatment. This improvement predicts subsequent treatment outcomes. Smaller reductions in activity impairment were associated with lower likelihood of remission and higher

likelihood of no meaningful benefit at end of acute phase even after controlling for baseline depression severity and activity impairment as well as changes in depression severity. Combining changes in activity impairment (using a single-item measure) with depression severity significantly improves the accuracy of predicting remission and no meaningful benefit at an individual level compared with measuring changes in depression severity only.

The finding that activity impairment improves with antidepressant treatment is consistent with findings reported previously in the CO-MED trial and by Lam et al. (Lam et al., 2014; Jha et al., 2017). Previous reports have also found that persistent functional impairment after 6 weeks of antidepressant treatment was associated with poorer outcomes at 3 and 7 months when the same treatment was continued (Jha et al., 2016a, 2016c, 2017). Additionally, self-reported activity impairment (sixth item of WPAI) moderately correlates with the activity-interest dimension of depression severity (Uher et al., 2012) with $<15\%$ shared variance, suggesting these 2 are measuring different constructs of impairment associated with depression.

A clinical implication of these findings could be to try other treatments such as exercise or behavioral activation in those patients with severe activity impairment by week 6 of acute-phase treatment outcome. Arguably, the strongest clinical utility of this report is the availability of an interactive calculator that provides the predicted probabilities of an individual's remission and no meaningful benefit. These outcomes are clinically actionable (if high likelihood of remission then continue treatment, if

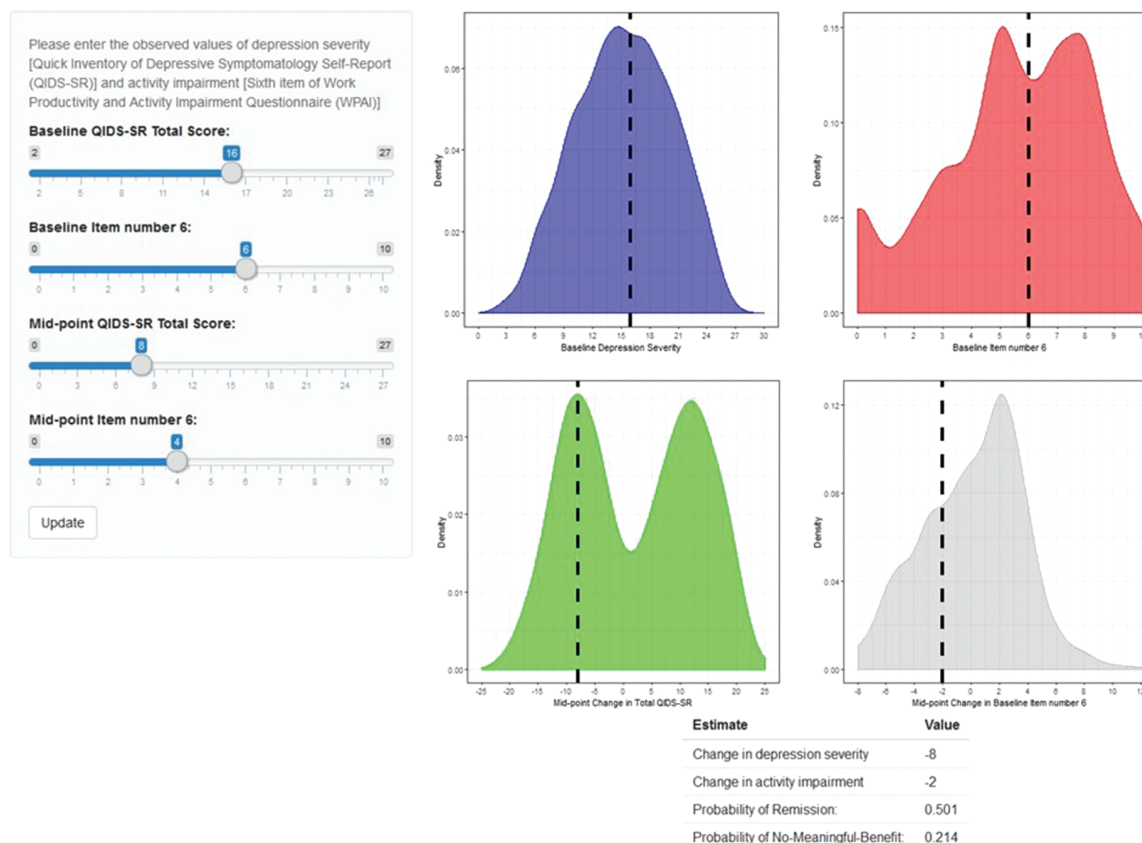


Figure 4. An individual level calculator of remission and no meaningful benefit using the estimates from Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial. Item number 6 (measure of activity impairment) is the sixth item of Work Productivity and Activity Impairment Questionnaire, and QIDS-SR is Quick Inventory of Depressive Symptomatology Self-Report version. Remission was ascribed if end of step 1 QIDS-SR was ≤ 5 . No meaningful benefit was ascribed if reduction in QIDS-SR from baseline to end of step 1 QIDS-SR was $<30\%$.

high likelihood of no meaningful benefit then modify treatment), require the addition of just a single-item self-report measure, and can be used by individual patients and their providers in clinical practice as well as research settings. A major strength of this report is the development and replication of a predictive model in 2 separate unrelated samples. The large sample size as well as recruitment of treatment-seeking outpatients from community practices with broad inclusion and minimal exclusion criteria increases the generalizability of these findings.

There are several limitations to this secondary analysis. The subjective nature of self-reported measure of activity impairment likely differs from objective measures, such as those collected as part of collateral information from relatives, friends, and family members, and does not fully assess impairment in different activities of life. The predictive models were developed and validated on participants who provided a complete (baseline, mid-point [around 6 weeks], and end of step 1 [around 12 weeks]) dataset at baseline. Hence, these findings could not be generalized to those who dropped out of care early or had an activity impairment score of 0 at baseline. Of note, those excluded had clinical features similar to those with complete data. Further, while week 6 or mid-point was selected as time-point due to its availability in the STAR*D, identification of predictors by week 2 or week 4 may be preferred as it may facilitate making clinical decisions earlier in course of treatment. In a large observational study of depressed patients started on pharmacotherapy using measurement-based care in primary care clinics ($n=2160$), there was an average 41.4-day delay from the initial

visit to the first follow-up visit (Jha et al., 2019). Thus, week 6 assessments may be especially useful in busy practices where patients may not be seen earlier. Additionally, the individual level calculator is restricted by the choice of QIDS-SR and WPAI as measures of depression severity and activity impairment. There is also evidence that the model is not well calibrated with probabilities above 0.6 in the no benefit model or above 0.8 in the remission model, so any estimates in that range should come with the understanding that they may not produce outcomes at the anticipated rate. To increase the generalizability of these findings, replication in a separate previously unreported dataset is recommended. Future studies are also needed to test the validity of this calculator with other measures of depression severity.

To conclude, the single-item self-report measure of activity impairment reflects improvement with antidepressant medication that is partly independent of change in depression severity and predicts longer-term clinical outcome. Assessment of activity impairment along with depression severity at baseline and mid-point (around 6 weeks) can be used to estimate an individual patient's likelihood of remission and no meaningful benefit at the end of acute-phase treatment.

Funding

This work was supported by the National Institute of Mental Health (NIMH), which funded the STAR*D trial (grant no. N01 MH-90003) and the CO-MED trial (grant no. N01 MH090003-02).

This work was also supported in part through the Center for Depression Research and Clinical Care at UT Southwestern (Principal Investigator: Madhukar H. Trivedi, MD), The Hersh Foundation, The Jordan Harris Foundation. The content of this publication does not necessarily reflect the views or policies of the U.S. Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US government. National Institute of Mental Health had no role in the drafting or review of the manuscript or in the collection or analysis of the data.

Acknowledgments

The authors thank the clinical staff at each clinical site of the STAR*D and CO-MED trials for their assistance with these projects and all of the study participants; and Jeremy A. Kee, MA for administrative support. Forest Pharmaceuticals, GlaxoSmithKline, Organon, and Wyeth Pharmaceuticals provided medications for CO-MED trial at no cost.

Statement of Interest

Drs. South and Minhajuddin as well as Mr. Trivedi did not have any conflicts of interest. Dr. Jha has received contract research support from Acadia Pharmaceuticals and Janssen Research and Development. Dr. Rush has received consulting fees from Akili, Brain Resource Inc., Compass Inc., Curbstone Consultant LLC, Eli Lilly, Emmes Corp, Liva-Nova, Mind Linc., Sunovion, Takeda USA, Taj Medical; speaking fees from Liva-Nova and Sing-Health; and royalties from Guilford Press and the University of Texas Southwestern Medical Center, Dallas, TX (for the Inventory of Depressive Symptoms and its derivatives). He is also named co-inventor on 2 patents: U.S. Patent No. 7,795,033: Methods to Predict the Outcome of Treatment with Antidepressant Medication, Inventors: McMahon FJ, Laje G, Manji H, Rush AJ, Paddock S, Wilson AS and U.S. Patent No. 7,906,283: Methods to Identify Patients at Risk of Developing Adverse Events During Treatment with Antidepressant Medication, Inventors: McMahon FJ, Laje G, Manji H, Rush AJ, Paddock S. Dr. Trivedi is or has been an advisor/consultant and received fees from: Alkermes, AstraZeneca, Cerecor, Eli Lilly & Company, Lundbeck, Naurex, Neuronetics, Otsuka Pharmaceuticals, PamLab, Pfizer Inc., SHIRE Development and Takeda. In addition, he has received grants/research support from National Institute of Mental Health and National Institute on Drug Abuse.

References

- Fava M, Rush AJ, Trivedi MH, Nierenberg AA, Thase ME, Sackeim HA, Quitkin FM, Wisniewski S, Lavori PW, Rosenbaum JF, Kupfer DJ (2003) Background and rationale for the sequenced treatment alternatives to relieve depression (STAR*D) study. *Psychiatr Clin North Am* 26:457–494.
- Gelenberg AJ, Freeman MP, Markowitz JC, Rosenbaum JF, Thase ME, Trivedi MH, Van Rhoads RS, Reus VI, J Raymond DePaulo M Jr, Fawcett JA (2010) Practice guideline for the treatment of patients with major depressive disorder third edition. *Am J Psychiatry* 167:1.
- Greenberg PE, Fournier AA, Sisitsky T, Pike CT, Kessler RC (2015) The economic burden of adults with major depressive disorder in the united states (2005 and 2010). *J Clin Psychiatry* 76:155–162.
- Guo T, Xiang YT, Xiao L, Hu CQ, Chiu HF, Ungvari GS, Correll CU, Lai KY, Feng L, Geng Y, Wang G (2015) Measurement-based care versus standard care for major depression: a randomized controlled trial with blind raters. *Am J Psychiatry* 172:1004–1013.
- Gupta S, Goren A, Phillips AL, Stewart M (2012) Self-reported burden among caregivers of patients with multiple sclerosis. *Int J MS Care* 14:179–187.
- Hamilton M (1960) A rating scale for depression. *J Neurol Neurosurg Psychiatry* 23:56–62.
- Jha MK, Grannemann B, Trombello JM, Clark EW, Levinson S, Lawson T, Greer TL, Rush AJ, Trivedi MH (2019) A structured approach for universal screening of depression and measurement-based treatment of major depressive disorder in primary care clinics. *Ann Fam Med*. In press.
- Jha MK, Greer TL, Grannemann BD, Carmody T, Rush AJ, Trivedi MH (2016b) Early normalization of quality of life predicts later remission in depression: findings from the CO-MED trial. *J Affect Disord* 206:17–22.
- Jha MK, Minhajuddin A, Greer TL, Carmody T, Rush AJ, Trivedi MH (2016a) Early improvement in psychosocial function predicts longer-term symptomatic remission in depressed patients. *Plos One* 11:e0167901.
- Jha MK, Minhajuddin A, Greer TL, Carmody T, Rush AJ, Trivedi MH (2016c) Early improvement in work productivity predicts future clinical course in depressed outpatients: findings from the CO-MED trial. *Am J Psychiatry* 173:1196–1204.
- Jha MK, Teer RB, Minhajuddin A, Greer TL, Rush AJ, Trivedi MH (2017) Daily activity level improvement with antidepressant medications predicts long-term clinical outcomes in outpatients with major depressive disorder. *Neuropsychiatr Dis Treat* 13:803–813.
- Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, Rush AJ, Walters EE, Wang PS, National Comorbidity Survey Replication (2003) The epidemiology of major depressive disorder: results from the national comorbidity survey replication (NCS-R). *Jama* 289:3095–3105.
- Kuhn M, Johnson K (2013) Applied predictive modeling. New York: Springer.
- Kuk AY, Li J, Rush AJ (2010) Recursive subsetting to identify patients in the STAR*D: a method to enhance the accuracy of early prediction of treatment outcome and to inform personalized care. *J Clin Psychiatry* 71:1502–1508.
- Lam RW, Endicott J, Hsu MA, Fayyad R, Guico-Pabia C, Boucher M (2014) Predictors of functional improvement in employed adults with major depressive disorder treated with desvenlafaxine. *Int Clin Psychopharmacol* 29:239–251.
- Li J, Kuk AY, Rush AJ (2012) A practical approach to the early identification of antidepressant medication non-responders. *Psychol Med* 42:309–316.
- Morin K, Davis JL (2017) Cross-validation: What is it and how is it used in regression?. *Commun Stat-Theor M* 46:5238–5251.
- Perlis RH (2013) A clinical risk stratification tool for predicting treatment resistance in major depressive disorder. *Biol Psychiatry* 74:7–14.
- Reilly MC, Zbrozek AS, Dukes EM (1993) The validity and reproducibility of a work productivity and activity impairment instrument. *Pharmacoeconomics* 4:353–365.
- 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 (2003) 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. *Biol Psychiatry* 54:573–583.
- 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 (2004) Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design. *Control Clin Trials* 25:119–142.
- Rush AJ, Kraemer HC, Sackeim HA, Fava M, Trivedi MH, Frank E, Ninan PT, Thase ME, Gelenberg AJ, Kupfer DJ, Regier DA, Rosenbaum JF, Ray O, Schatzberg AF, ACNP Task Force (2006a) Report by the ACNP task force on response and remission in major depressive disorder. *Neuropsychopharmacology* 31:1841–1853.
- Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, Niederehe G, Thase ME, Lavori PW, Lebowitz BD, McGrath PJ, Rosenbaum JF, Sackeim HA, Kupfer DJ, Luther J, Fava M (2006b) Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry* 163:1905–1917.
- Rush AJ, Trivedi MH, Wisniewski SR, Stewart JW, Nierenberg AA, Thase ME, Ritz L, Biggs MM, Warden D, Luther JF, Shores-Wilson K, Niederehe G, Fava M, STAR*D Study Team (2006c) Bupropion-SR, sertraline, or venlafaxine-XR after failure of ssris for depression. *N Engl J Med* 354:1231–1242.
- Rush AJ, Trivedi MH, Stewart JW, Nierenberg AA, Fava M, Kurian BT, Warden D, Morris DW, Luther JF, Husain MM, Cook IA, Shelton RC, Lesser IM, Kornstein SG, Wisniewski SR (2011) Combining medications to enhance depression outcomes (CO-MED): acute and long-term outcomes of a single-blind randomized study. *Am J Psychiatry* 168:689–701.
- South C, Rush AJ, Carmody TJ, Jha MK, Trivedi MH (2017) Accurately identifying patients who are excellent candidates or unsuitable for a medication: a novel approach. *Neuropsychiatr Dis Treat* 13:3001–3010.
- Trivedi MH, Fava M, Wisniewski SR, Thase ME, Quitkin F, Warden D, Ritz L, Nierenberg AA, Lebowitz BD, Biggs MM, Luther JF, Shores-Wilson K, Rush AJ, STAR*D Study Team (2006a) Medication augmentation after the failure of ssris for depression. *N Engl J Med* 354:1243–1252.
- Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, Ritz L, Norquist G, Howland RH, Lebowitz B, McGrath PJ, Shores-Wilson K, Biggs MM, Balasubramani GK, Fava M, STAR*D Study Team (2006b) Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry* 163:28–40.
- Trivedi MH, Corey-Lisle PK, Guo Z, Lennox RD, Pikalov A, Kim E (2009) Remission, response without remission, and nonresponse in major depressive disorder: impact on functioning. *Int Clin Psychopharmacol* 24:133–138.
- Trivedi MH, Greer TL, Church TS, Carmody TJ, Grannemann BD, Galper DI, Dunn AL, Earnest CP, Sunderajan P, Henley SS, Blair SN (2011) Exercise as an augmentation treatment for nonremitted major depressive disorder: a randomized, parallel dose comparison. *J Clin Psychiatry* 72:677–684.
- Trivedi MH, Morris DW, Wisniewski SR, Lesser I, Nierenberg AA, Daly E, Kurian BT, Gaynes BN, Balasubramani GK, Rush AJ (2013) Increase in work productivity of depressed individuals with improvement in depressive symptom severity. *Am J Psychiatry* 170:633–641.
- Uher R, Perlis RH, Henigsberg N, Zobel A, Rietschel M, Mors O, Hauser J, Dernovsek MZ, Souery D, Bajs M, Maier W, Aitchison KJ, Farmer A, McGuffin P (2012) Depression symptom dimensions as predictors of antidepressant treatment outcome: replicable evidence for interest-activity symptoms. *Psychol Med* 42:967–980.
- Vos T, Barber RM, Bell B, Bertozzi-Villa A, Biryukov S, Bolliger I, Charlson F, Davis A, Degenhardt L, Dicker D (2015) Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet* 386:743–800.