CanJPsychiatry 2015;60(1):9–13

Perspective

The STAR*D Trial: It Is Time to Reexamine the Clinical Beliefs That Guide the Treatment of Major Depression

H Edmund Pigott, PhD[1]

[1] Chief Clinical Officer, Positive Brain Training, LLC, Juno Beach, Florida; Psychologist, Guiding Light Wellness Center, Wellington, Florida.
Correspondence: 430 North Lyra Circle, Juno Beach, FL 33408; pathware@erols.com.

Key Words: depression, Sequenced Treatment Alternatives to Relieve Depression, STAR*D, selective outcome reporting

Received June 2014, revised, and accepted July 2014.

The STAR*D trial is the largest and most consequential antidepressant study ever conducted, with over 120 journal articles published by study investigators, innumerable citations of STAR*D's findings by other researchers, and extensive coverage in the media, thereby giving it an oversized impact on the treatment of depression, worldwide. Funded at a cost of 35-million US dollars, STAR*D enrolled 4041 patients who screened positive for major depression while seeking routine medical or psychiatric care. In contrast to most industry-funded trials, STAR*D included depressed patients with comorbid conditions, thereby increasing the generalizability of its findings and further, provided 12 months of free continuing care to monitor the durability of treatment effects.

STAR*D provided up to 4 treatment steps per patient and was designed to give guidance in selecting the best next-step treatment for the many patients who fail to get sufficient relief from their initial AD. Each step consisted of a 12-week, open-label trial, with an additional 2 weeks for patients deemed close to remission. ADs were administered using a system of measurement-based care that involved assessing symptoms and side effects at each visit to guide aggressive medication dosing to "ensure that the likelihood of achieving remission was maximized and that those who did not reach remission were truly resistant to the medication."[1, p 30]

STAR*D allowed patients to select treatment options for randomization in steps 2 to 4 "to empower patients, strengthen the therapeutic alliance, optimize treatment adherence, and improve outcome"[2, p 483] and evaluated the relative effectiveness of 11 pharmacologically distinct drug–drug combination treatments in 5 head-to-head comparisons. CT was also available as a switch or drug augmentation option in step 2, but too few patients included it as an acceptable treatment, resulting in only 101 contributing data after randomization. Therefore, CT was excluded from the step 2 switch and augmentation analyses.[3,4]

Patients who achieved remission during any step were encouraged to enter the 12 months of free follow-up care, as were responder patients who failed to attain remission but did not want to continue to the next treatment step as this would involve a change in medication. Remission was defined as a score of less than 8 on the HRSD, the study's prespecified primary outcome measure, and response as a 50% or greater reduction in depressive symptoms on it. The follow-up protocol "strongly recommended that participants continue the previously effective acute treatment medication(s) at the doses used in acute treatment" but treating physicians were allowed to make "any psychotherapy, medication, or medication dose change"[5, p 1908] they deemed necessary to sustain a positive outcome during follow-up, including scheduling additional visits if depressive symptoms returned and (or) intolerable side effects emerged.[6]

Prior Criticisms of Apparent Bias in the Reporting of STAR*D Findings

Pigott and colleagues[7,8] have previously criticized the investigators for the following:

1) not reporting in their summary article[5] remission and response rates using the prespecified HRSD but instead using the QIDS-SR, a nonblinded, clinic-administered assessment that was excluded from use as a research measure in the Research Protocol[9];
2) excluding from analysis patients who started on citalopram in their baseline visit and then dropped out without taking the exit HRSD despite the investigators' statement that "our primary analyses classified patients with missing exit HRSD scores as nonremitters a priori"[1, p 43] and these early dropout patients therefore should have been counted as treatment failures as prespecified;
3) asserting in their summary article a "theoretical cumulative remission rate" of 67% with the unrealistic provisos that "this estimate assumes no dropouts, and it assumes that those who exited the study would have had the same remission rates as those who stayed in the protocol"[5, p 1910]; and
4) other indicators of bias (see eTable 1).

As Pigott et al[7] document, the investigators' assumptions in calculating their theoretical remission rate of 67% are simply not true in the real world, and were certainly not true in STAR*D, as more patients dropped out in each step than remitted. Today, STAR*D investigators' provisos concerning their theoretical remission rate are sometimes dropped when portraying its findings. For example, a recent editorial in the *AJP* states STAR*D found, "after four optimized, well-delivered treatments, approximately 70% of patients achieve remission"[10, p 580] as though this is a factual statement of what occurred.

Figure 7 in STAR*D's Research Protocol has step-by-step predictions of patient drop out and the number of patients who would have a satisfactory response and enter follow-up.[9, p 35] The Protocol was obtained from NIMH in a Freedom of Information Act request. The predictions reflect the predicted aggregate outcomes for STAR*D's sequential treat-to-remission model of measurement-based care.

STAR*D's investigators found no significant group differences between any of the 11 drug–drug combination treatments. Further, no post hoc secondary analyses have reported significant predictors of outcomes between the pharmacologically distinct treatments. Therefore, STAR*D provides no next-step guidance to give hope for improving outcomes beyond that found in the study itself. As Barbui et al[11] note, AD study completion rates provide a "hard measure of treatment effectiveness and acceptability"[p 296] and STAR*D's investigators essentially ignore this fact when reporting study outcomes. Elsewhere, Pigott[8] has detailed the extensive efforts STAR*D made to keep patients in treatment and maximize their likelihood of achieving remission and maintaining it during the 12 months of follow-up care as well as completing the scheduled research assessments; thus it is essential to incorporate dropout patients into the evaluation of study outcomes.

Abbreviations

AD	antidepressant
AJP	*American Journal of Psychiatry*
CCJM	*Cleveland Clinic Journal of Medicine*
CT	cognitive therapy
HRSD	Hamilton Rating Scale for Depression
NIMH	National Institute of Mental Health
PCP	primary care physician
QIDS-SR	Quick Inventory of Depressive Symptoms—Self Report
RCT	randomized controlled trial
STAR*D	Sequenced Treatment Alternatives to Relieve Depression

Methods

The author abstracted from the Research Protocol and summary article's[5] published tables and figures to recalculate STAR*D's remission and relapse rates using 4041 as the denominator, as all patients were started on citalopram in their baseline visit. This approach seems more realistic than the summary article[5] that used 3671 as the denominator, thereby excluding from analysis the 370 early dropout patients who did not return for subsequent treatment visits.

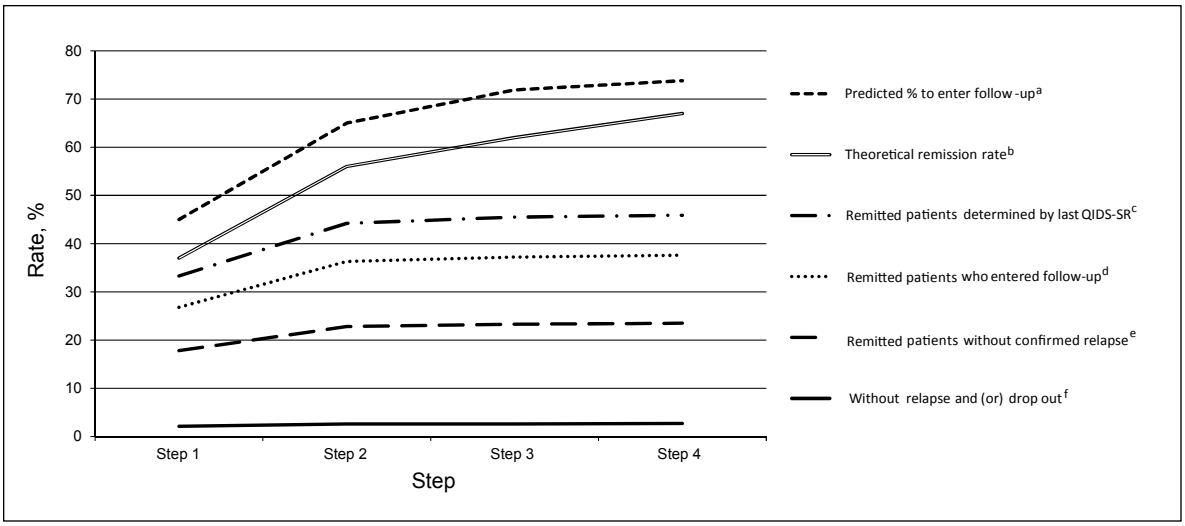
The Data

Figure 1 compares the cumulative step-by-step predicted per cent of patients who would be successfully treated and enter follow-up and STAR*D's theoretical remission rate to what occurred in the study.

In the summary article,[5] STAR*D's researchers used the nonblinded QIDS-SR to report acute and follow-up care outcomes instead of the protocol-specified HRSD. This assessment was administered at each clinic visit and monthly by telephone during follow-up. The QIDS-SR provided a more complete data set as 690 patients exited the study in step 1 alone, without taking the exit HRSD, 152 of whom had a QIDS-SR defined remission.[1, p 34] The cumulative per cent of patients who had a remission determined by their final QIDS-SR after up to 4 treatment trials was 45.9%. By step 4, the cumulative per cent of such patients who entered follow-up was only 37.6%. While both calculations used 4041 as the denominator, these findings improve little when using 3671 patients as the denominator (50.5% and 41.4%, respectively).

The data STAR*D investigators' provide for accessing the durability of treatment gains are even more discouraging. For step 1, only 17.8% of citalopram-treated patients had a remission as determined by their last clinic-administered

Figure 1 Comparison between predicted, theoretical, and actual step-by-step success rates



[a] Calculated from Figure 7 in STAR*D's Research Protocol predicting how many patients would have a satisfactory response and enter follow-up.

[b] Calculated by STAR*D's authors in the summary article.[5]

[c] Calculated from the summary article's Table 3 "Remission at each step" row but using 4041 as the denominator as all patients were started on citalopram in their baseline visit (that is, (1346 + 439 + 53 + 16)/4041 = 45.9% by step 4).

[d] Calculated from the summary article's Table 5, column 2, rows 3, 6, 9, and 12, by adding the number of remitted patients entering follow-up and dividing by 4041 (that is, (1085 + 383 + 35 + 15)/4041 = 37.6% by step 4).

[e] Calculated from the summary article's Table 5, column 6, rows 3, 6, 9, and 12. For each step, subtract the per cent relapse rate from 1.0 and then mulitply by the number of remitted patients entering follow-up (for example, step 1: 1.0 – 0.335 = 0.665 x 1085 = 722 remitted patients who did not have a confirmed relapse during follow-up). Repeat for steps 2 to 4. Add the remitted patients without confirmed relpase for steps 1 to 4 and divide by 4041.

[f] Calculated from the summary article Figure 3's table, column 5, by adding the number of surviving remitted patients without relapse and (or) drop out during months 10 to 12 and dividing by 4041 (that is, (84 + 20 + 2 + 2)/4041 = 2.7% by step 4).

QIDS-SR = Quick Inventory of Depressive Symptoms—Self Report; STAR*D = Sequenced Treatment Alternatives to Relieve Depression

QIDS-SR and during follow-up did not have a confirmed relapse on 1 or more of the 12 monthly administered telephonic QIDS-SR assessments. After up to 4 rounds of AD drug–drug combination treatments, the cumulative rate of patients who did not have a confirmed relapse improved to only 23.5%. When drop out is added, the durability of treatment effects is even paltrier. Only 2.7% of patients had a QIDS-SR determined remission after up to 4 rounds of AD drug care and neither relapsed nor dropped out as evidenced by taking at least 1 of the months 10-to-12 QIDS-SR telephonic assessments and not scoring as having relapsed in any of the 12 monthly administered assessments.

Figure 2 presents the rates of intolerable side effects and drop out by treatment step taken from the summary article's[5] Table 3 and Figure 1. It demonstrates how each change in treatment resulted in an increased rate of intolerable side effects from the newly prescribed medication, compared with the prior step, increasing from 16.3% in step 1 to 30.1% by step 4. This same trend is seen for study drop out, with it increasing from 28.1% in step 1 to 42.3% by step 3. Both measures demonstrate an increasing risk to patient care tied to each change in AD drug–drug combination treatment as patients progressed from step 1 through steps 2 to 4.

The Need for a Reexamination

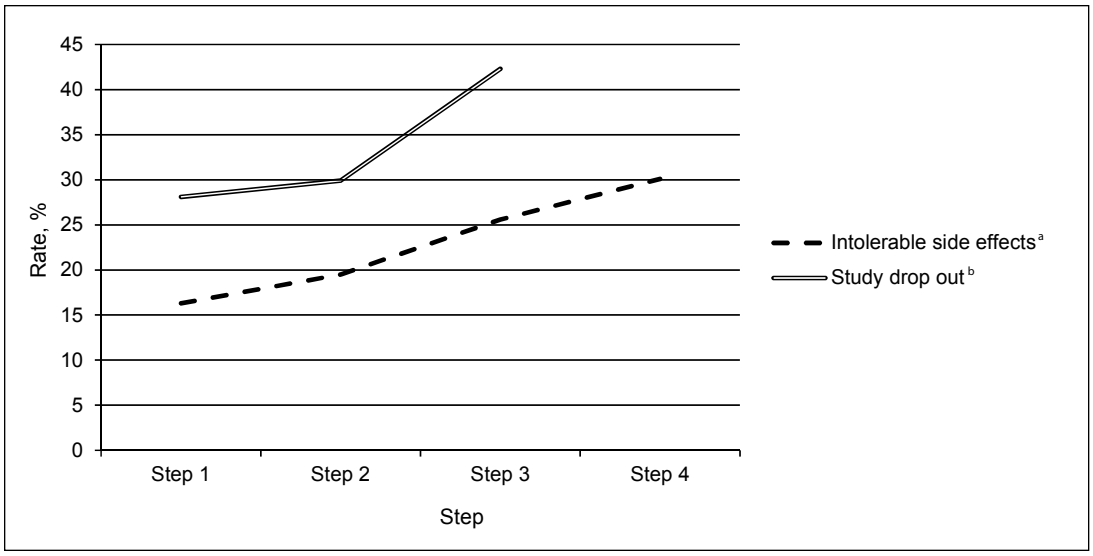
Following publication of STAR*D's steps 1 to 4 and summary results in 2006,[1–5,12–14] its investigators have continued publishing at a high rate of over 120 journal articles, including the recent *AJP* one which precipitated the editorial cited above.[15] In addition to not disclosing deviations from the prespecified criteria, often in these articles the investigators interpret what they do report in ways that lead to inaccurate conclusions and potentially harmful recommendations for patient care. For example, in their *CCJM* article's Key Points section, the investigators state:

> With persistent and vigorous treatment, most patients will enter remission: about 33% after one step, 50% after two steps, 60% after three steps, and 70% after four steps (assuming patients stay in treatment).[16, p 57]

The investigators then open this article stating:

> DEPRESSION can be treated successfully by primary care physicians under "real world" conditions. Furthermore, the particular drug or drugs used are not as important as following a rational plan: giving antidepressant medications in adequate

Figure 2 Rates of intolerable side effects and drop out by treatment step



[a] From the "Intolerable side effects" row, Table 3, of the summary article.[5]

[b] Calculated from Figure 1 by adding all patients who exited the study in steps 1 to 3 and dividing by the number enrolled in each step.

> doses, monitoring the patient's symptoms and side effects and adjusting the regimen accordingly, and switching drugs or adding new drugs to the regimen only after an adequate trial. These are among the lessons learned from the Sequenced Treatment Alternatives to Relieve Depression study, the largest prospective clinical trial of treatment of major depressive disorder ever conducted.[16, p 57]

There are several noteworthy points in this article's summary of STAR*D's lessons learned for PCPs who prescribe most ADs. First when summarizing their results, the highly theoretical 67% remission rate is rounded up to 70%. Second, the summary article's proviso that this rate "assumes that those who exited the study would have had the same remission rates as those who stayed in the protocol"[5, p 1910] is dropped, leaving only the assumption of no dropouts. Third, nowhere in this article nor elsewhere do the investigators acknowledge that through step 3, 43% of patients had in fact dropped out despite their best efforts in designing the study so as to minimize this "hard measure of treatment (in)effectiveness"[11, p 296] from occurring.

Fourth, while correctly stating that "the particular drug or drugs used are not as important"[16, p 57] as STAR*D provides no next-step guidance to improve outcomes, the investigators strongly endorse here as elsewhere their system of measurement-based care as the means to improve said outcomes. For example, the investigators state:

> *high quality of care* was delivered (*measurement-based care*) . . . Consequently, the outcomes in this report *may exceed* those that are presently obtained in daily practice wherein neither symptoms nor side-effects are consistently measured and wherein practitioners vary greatly in the timing and level of dosing [emphases added].[5, p 1914]

As Pigott et al[7] noted, STAR*D's sequential treat-to-remission model of measurement-based care may have been detrimental to patient care leading to worse outcomes than what occurs in daily practice.

> STAR*D encouraged all patients who did not achieve remission based on a number to enter the next trial despite the failure of the QIDS/HRSD to differentially weigh core depressive symptoms (e.g. mood, guilt, suicidal ideation, or anhedonia) and accessory ones (e.g. appetite, insomnia, or agitation) and patients' self-assessments of the relative importance of each.[p 277]

Fifth, STAR*D's investigators may have overstated the benefits of remission relative to patients with "merely substantial improvement" as both groups had high rates of confirmed relapse during follow-up and STAR*D's symptom-driven pursuit of more aggressive treatment was not shown to be better than usual care. More importantly, by encouraging PCPs whose patients achieve "merely substantial improvement" from one AD to "switching drugs or adding new drugs to the regimen"[16, p 57] in their pursuit of a less than 8 HRSD score, STAR*D's investigators fail to acknowledge the risks that came with each such change; that is, the step-by-step increasing rates of drug intolerance and study drop out.

The manner in which STAR*D's analyses were conducted and results communicated have created a narrative around its aggressive treat-to-remission model of measurement-based care that is not aligned with the actual results of the study.

STAR*D was a failed trial with results that were not shown to be superior to treatment-as-usual. Despite this lack of evidence though, some STAR*D investigators continue to advocate strongly for the superiority of their measurement-based model of care[17–19]—even arguing for the adoption of financial incentives and 9 other policy changes to make it the standard-of-care for psychiatric practice.[20]

The STAR*D data set are now available from NIMH for independent analysis by university-based researchers.[21] Researchers are urged to request this data set for independent analysis of the steps 1 to 4 and follow-up outcomes according to STAR*D's prespecified criteria. Only then will the lessons learned from this once-in-a-generation study be fully known and made available to inform clinical practice and guide subsequent research efforts to improve on the same.

Acknowledgements

In the past 3 years, Dr Pigott has consulted for Amen Clinics, CNS Response, and the International Society of Neurofeedback and Research. He also owns stock options in CNS Response, a company that markets a quantitative electroencephalogram database to predict medication response, and is a partner in Positive Brain Training, LLC, a neurofeedback practice.

He affirms that the manuscript is an honest, accurate, and transparent account of his analysis of what he has derived from the STAR*D investigators' published tables and figures reporting their findings and then comparing them to the Research Protocol he obtained from NIMH through the Freedom of Information Act.

Editor's Note

References 22 and 23 can be found in eTable 1.

References

1. Trivedi MH, Rush AJ, Wisniewski SR, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. Am J Psychiatry. 2006;163:28–40.
2. Fava M, Rush AJ, Trivedi MH, et al. Background and rationale for the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study. Psychiatr Clin North Am. 2003;26:457–494.
3. Rush AJ, Trivedi MH, Wisniewski SR, et al. Bupropion-SR, sertraline, or venlafaxine-XR after failure of SSRIs for depression. N Engl J Med. 2006;354:1231–1242.
4. Trivedi MH, Fava M, Wisniewski SR, et al. Medication augmentation after the failure of SSRIs for depression. N Engl J Med. 2006;354:1243–1252.
5. Rush AJ, Trivedi MH, Wisniewski SR, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. Am J Psychiatry. 2006;163:1905–1917.
6. Trivedi MH, Stegman D, Rush AJ, et al. STAR*D clinical procedures manual [Internet]. Pittsburgh (PA): University of Pittsburgh; 2002 Jul 31 [date cited unknown]. Available from: http://www.edc.gsph.pitt.edu/stard/public/study_manuals.html.
7. Pigott HE, Leventhal AM, Alter GS, et al. Efficacy and effectiveness of antidepressants: current status of research. Psychother Psychosom. 2010;79:267–279.
8. Pigott HE. STAR*D: a tale and trial of bias. Ethical Hum Psychol Psychiatry. 2011;13(1):6–28.
9. National Institute of Mental Health (NIMH). Sequenced Treatment Alternatives to Relieve Depression (STAR*D) research protocol. Revised ed. Washington (DC): NIMH; 2002. Received October 1, 2010, from NIMH following a request under the Freedom of Information Act. Also available from: http://www.madinamerica.com/wp-content/uploads/2011/12/STAR-D%20Protocol%20with%20Analytic%20Plan.pdf.
10. Greden JF. Workplace depression: personalize, partner, or pay the price. Am J Psychiatry. 2013;170:578–581.
11. Barbui C, Furukawa TA, Cipriani A. Effectiveness of paroxetine in the treatment of acute major depression in adults: a systematic re-examination of published and unpublished data from randomized trials. CMAJ. 2008;178:296–305.
12. Fava M, Rush AJ, Wisniewski SR, et al. A comparison of mirtazapine and nortriptyline following two consecutive failed medication treatments for depressed outpatients: a STAR*D report. Am J Psychiatry. 2006;163:1161–1172.
13. Nierenberg AA, Fava M, Trivedi MH, et al. A comparison of lithium and T3 augmentation following two failed medication treatments for depression: a STAR*D report. Am J Psychiatry. 2006;163:1519–1530.
14. McGrath PJ, Stewart JW, Fava M, et al. Tranylcypromine versus venlafaxine plus mirtazapine following three failed antidepressant medications trials for depression: a STAR*D report. Am J Psychiatry. 2006;163:1531–1541.
15. Trivedi MH, Morris DW, Wisniewski SR, et al. Increase in work productivity of depressed individuals with improvement in depressive symptom severity. Am J Psychiatry. 2013;170:633–641.
16. Gaynes BN, Rush AJ, Trivedi MH, et al. The STAR*D study: treating depression in the real world. Clev Clin J Med. 2008;75:57–66.
17. Morris DW, Trivedi MH. Measurement-based care for unipolar depression. Curr Psychiatry Rep. 2011;13(6):446–458.
18. Shelton RC, Trivedi MH. Using algorithms and computerized decision support systems to treat major depression. J Clin Psychiatry. 2011;72(12):e36.
19. Trivedi MH. Evaluating and monitoring treatment response in depression using measurement-based assessment and rating scales. J Clin Psychiatry. 2013;74(7):e14.
20. Harding KJ, Rush AJ, Arbuckle M, et al. Measurement-based care in psychiatric practice: a policy framework for implementation. J Clin Psychiatry. 2011;72(8):1136–1143.
21. National Institute of Mental Health (NIMH). Available limited access datasets from NIMH clinical trials: Sequenced Treatment Alternatives to Relieve Depression (STAR*D). Washington (DC): NIMH; [year of publication unknown] [cited 2013 Aug 26]. Available from: http://www.nimh.nih.gov/trials/datasets/imh-procedures-for-requesting-data-sets.shtml.
22. National Institute of Mental Health (NIMH). New strategies help depressed patients become symptom-free [Internet]. Washington (DC): NIMH; 2006 [cited 2013 Aug 26]. Available from: http://www.nimh.nih.gov/science-news/2006/new-strategieshelp-depressed-patients-become-symptom-free.shtml.
23. Hamilton M. Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol. 1967;6:278–296.