

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

Restoring STAR*D: Treatment Remission, Response and Extent of Improvement Rates after up to Four Trials of Antidepressant Therapies

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2022-063095
Article Type:	Original research
Date Submitted by the Author:	31-Mar-2022
Complete List of Authors:	Pigott, Henry; None Kim, Thomas; University of Pennsylvania, Psychology Xu, Colin; University of Pennsylvania, Psychology Kirsch, Irving; Harvard Medical School Amsterdam, Jay; University of Pennsylvania Perelman School of Medicine, Psychiatry
Keywords:	Adult psychiatry < PSYCHIATRY, Depression & mood disorders < PSYCHIATRY, CLINICAL PHARMACOLOGY

SCHOLARONE™
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1
2
3 **Restoring STAR*D: Treatment Remission, Response and Extent of**
4 **Improvement Rates after up to Four Trials of Antidepressant Therapies**

5
6 **Authors:**

7 **H. Edmund Pigott, Thomas T. Kim, Colin Xu, Irving Kirsch, & Jay D. Amsterdam**
8
9

10 H. Edmund Pigott, Ph.D.

11 Position: Clinical Psychologist

12 Institution: none

13 Email: pathware@erols.com
14
15

16 Thomas T. Kim

17 Position: PhD candidate in Psychology

18 Institution: Department of Psychology, University of Pennsylvania
19
20

21 Colin Xu

22 Position: PhD candidate in Psychology

23 Institution: Department of Psychology, University of Pennsylvania
24
25

26 Irving Kirsch, Ph.D.

27 Position: Associate Director, Program in Placebo Studies

28 Institution: Harvard Medical School
29
30

31 Jay D. Amsterdam, MD

32 Position: Professor of Psychiatry (Emeritus), Depression Research Unit

33 Institution: Perelman School of Medicine, University of Pennsylvania
34
35

36 Corresponding author: H. Edmund Pigott, Ph.D.
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Abstract

Objective: Reanalyze the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial that examined the effectiveness of up to four optimized, and increasingly aggressive, antidepressant therapies in depressed adults.

Design: The trial was open label and semi-randomized.

Setting: 41 North American psychiatry and primary care treatment centers.

Participants: 4,041 adults screened positive for major depressive disorder. STAR*D enrolled patients seeking care (versus recruited) and included patients with a wide range of common co-morbid medical and psychiatric conditions to enhance the generalizability of findings to real-world clinical practice.

Interventions: STAR*D evaluated the relative effectiveness of 13 antidepressant therapies in levels 2-4 for depressed patients who failed to gain adequate benefit from their initial medication trial.

Main Outcome Measures: The primary outcome was remission, defined as a score <8 on the blinded Hamilton Rating Scale for Depression (HRSD). Response was a secondary outcome defined as ≥50% reduction in HRSD scores. STAR*D's protocol specifically excluded all non-blinded clinic-administered assessments from use as research outcome measures.

Results: We reanalyzed the STAR*D dataset with fidelity to the original research protocol wherever possible. Where the protocol was silent, we used other STAR*D publications to guide our analysis. STAR*D investigators did not use the protocol-stipulated HRSD to report cumulative remission and response rates in their summary article, and instead used a non-blinded clinic-administered assessment. This inflated their report of outcomes, as did their inclusion of 99 patients who scored as remitted on the HRSD at study outset as well as 125 who scored as remitted when initiating their next-level treatment. In contrast to the STAR*D-reported 67% cumulative remission rate after up to four antidepressant treatment trials, the actual rate was 35% when using the protocol-stipulated HRSD and inclusion in data analysis criteria.

Conclusion: STAR*D's cumulative remission rate was approximately 50% less than reported.

Strengths and limitations of this study

- We reanalyzed the largest ever antidepressant trial's patient-level dataset with fidelity to the original research protocol and related publications.
- The reanalysis was conducted under the guidelines of the *Restoring Invisible and Abandoned Trials* (RIAT) initiative.
- Treatment remission, response and extent of symptom improvement rates were calculated for 14 antidepressant therapies for those patients who met STAR*D's

1
2
3 inclusion in data analysis criteria as well as the overall cumulative remission rate after
4 up to four trials of antidepressant therapies.

- 5 • We calculated STAR*D's remission rate using the protocol-stipulated HRSD as well as the HRSD
6 remissions combined with a non-stipulated measure of remission for the 1,330 patients missing
7 an exit HRSD score.
- 8 • Finally, we compared STAR*D's outcomes to the protocol-predicted outcomes as well as to
9 those found in a meta-analysis of 7,030 patients enrolled in similar open-label antidepressant
10 comparator trials and therefore the appropriate data to compare STAR*D's outcomes to.
11
12
13

14 Introduction

15
16 The Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study is the largest and most
17 consequential antidepressant trial ever conducted, with over 100 journal articles published by study
18 investigators and innumerable citations of STAR*D's findings by other researchers giving it an oversized
19 impact on the treatment of depression world-wide.[1-7] Funded by the National Institute of Mental
20 Health (NIMH), STAR*D enrolled 4041 patients who screened positive for major depressive disorder
21 (MDD) while seeking routine medical or psychiatric care. In contrast to industry-funded trials, STAR*D
22 did not exclude patients with medical conditions and most comorbid psychiatric disorders, thereby
23 increasing the generalizability of its findings to real-world clinical practice.
24
25

26 The STAR*D study provided up to four treatment trials per patient and was designed to give guidance in
27 selecting the best next-level treatment option for the many people who fail to gain sufficient relief from
28 their first, and/or subsequent, antidepressant trial. To mimic clinical practice, STAR*D used an open-
29 label research design with no control group during any phase of the study.
30
31

32 The STAR*D investigators stated, "the primary outcome is depressive symptom severity, measured by
33 the 17-item Hamilton Rating Scale for Depression (HRSD)." [8, p. 120] STAR*D's prespecified primary
34 outcome was remission, defined as scoring <8 on the HRSD which was administered telephonically by
35 Research Outcome Assessors (ROAs) blind to patients' study status (level entry/exit/follow-up).
36 Response was a secondary outcome defined as ≥50% reduction in patients' HRSD scores. However,
37 despite its investigators numerous publications, neither change in HRSD depressive symptom severity
38 nor HRSD response rates have been reported for STAR*D's six trials[1-6] and summary article.[7]
39 Instead, response rates and change in symptom severity were reported using the clinic-administered
40 Quick Inventory of Depressive Symptomatology–Self Report (QIDS-SR), a measure developed by the
41 STAR*D principal investigators.[9] This occurred despite the fact that STAR*D's research protocol
42 specifically excluded all clinic-administered assessments, such as the QIDS-SR, from use as research
43 outcome measures since they were not blinded and instead, used to guide patient care. The protocol
44 states:
45
46

47
48 *Recall that the research outcomes assessments are distinguished from assessments*
49 *conducted at clinic visits. The latter are designed to collect information that guides*
50 *clinicians in the implementation of the treatment protocol. Research outcomes*
51 *assessments are **not** collected at the clinic visits. They are not collected by either*
52 *clinicians or Clinical Research Coordinators.[10,p.47-48; emphasis in the original]*
53

54 In their summary article, STAR*D investigators used the QIDS-SR as the sole measure to report
55 remission, response, and extent of symptom improvement. This article's Abstract states that "the
56
57
58
59

1
2
3 overall cumulative remission rate was 67%” with no qualifiers to this claim.[7, p.1905] Besides making
4 this claim based on an assessment the protocol specifically excluded from use as a research measure, it
5 is not until the article’s Results section that readers learn this high level of purported treatment success
6 did not occur. The STAR*D investigators’ claim was theoretical—an estimate based on the unrealistic
7 provisos of what would have happened if there were no study dropouts, and furthermore, “that those
8 who exited the study would have had the same remission rates as those who stayed in the protocol.”[7,
9 p.1910] As Pigott et al. document though, the investigators’ assumptions are not true in the real world,
10 since patients who drop out are more likely to be treatment non-responders, and more patients
11 dropped out than remitted in each treatment level of the STAR*D trial.[11]
12
13

14 Unfortunately, the STAR*D investigators’ claim of a 67% cumulative remission rate has become
15 accepted clinical wisdom, and the improbable provisions on which it is based are commonly not
16 referenced when portraying STAR*D’s findings. For example, an editorial in the *American Journal of*
17 *Psychiatry* states STAR*D found, “after four optimized, well-delivered treatments, approximately 70% of
18 patients achieve remission” as though this is a factual statement of what occurred.[12, p.580]
19
20

21 The first author has made published criticisms alleging protocol violations that appear to inflate
22 STAR*D’s findings and has called for the reanalysis of the dataset by independent investigators.[13] In
23 2018, the first and fourth authors collaborated with researchers from the University of Connecticut to
24 reanalyze STAR*D’s level 1 data obtained from NIMH.[14] This reanalysis found substantial inflation of
25 STAR*D’s reported remission and response rates. Furthermore, the reanalysis found that the extent of
26 HRSD improvement in STAR*D’s level 1 trial was approximately half that of the open-label
27 antidepressant comparator trials.
28
29

30 Our criticisms of STAR*D are as follows:[15]
31

- 32 • Using the QIDS-SR as the secondary outcome to report remission rates and sole measure to
33 report response rates in STAR*D’s level 1-4 articles without STAR*D investigators disclosing that
34 the protocol specifically excluded non-blinded/clinic-administered assessments such as the
35 QIDS-SR from use as outcome measures.[1-6] The QIDS-SR was also used as the sole measure to
36 report remission, response, and extent of improvement rates in their summary article.[7] The
37 primary outcome measure, the HRSD, should have been used to report the summary article’s
38 outcomes.
39
- 40 • Using data from the 931 patients deemed ineligible for analysis in STAR*D’s level 1 article
41 because these patients lacked a baseline ROA-administered HRSD score of ≥ 14 , in STAR*D’s
42 levels 2-4 and summary articles without clear disclosure. This included 99 patients who scored
43 < 8 on their baseline HRSD—indicating these patients met STAR*D’s remission criterion at study
44 outset and should not have been included in their report of outcomes.
45
- 46 • Excluding from analysis 370 patients who dropped out after starting on citalopram in their first
47 clinic visit without taking the exit HRSD despite STAR*D investigators stating, “our primary
48 analyses classified patients with missing exit HRSD scores as nonremitters a priori.”[1, p.34]
49 These 370 early dropout patients should have been counted as nonremitters as prespecified.
50
- 51 • Including in their analyses 125 patients who scored as remitted at entry into their next-level
52 treatment. This occurred despite STAR*D investigators prespecifying that, “patients who begin
53 a level with HRSD < 8 will be excluded from analyses.”[8, p.130]
54

55 This article reanalyzes STAR*D’s step-by-step treatment remission, response, and extent of
56 improvement after up to four trials of antidepressant therapies, using STAR*D’s primary outcome
57
58
59

1
2
3 measure, the ROA-administered HRSD. Subsequent efforts will focus on reanalyzing STAR*D's levels 2-4
4 semi-randomized comparator trials including 12-month follow-up outcomes tied to each compared
5 treatment.
6

7 8 **Method**

9 *RIAT Initiative*

10
11 The Restoring Invisible and Abandoned Trials (RIAT) initiative started in 2013 calling on funders
12 and investigators of abandoned (unpublished) or misreported trials to publish undisclosed
13 outcomes or correct misleading publications.[16] If investigators failed to correct a study
14 identified as misreported, independent investigators were encouraged to correct the record by
15 reanalyzing the study's patient-level dataset consistent with the research protocol and analytic
16 plan.
17
18

19
20 On March 6, 2019, the RIAT investigators published our response to a 'Call to Action' statement in the
21 *British Medical Journal*, in which we stated our intention to reanalyze the STAR*D dataset.[15] We then
22 notified STAR*D's principal investigators of our intention and requested they inform us whether they
23 would undertake a reanalysis of the dataset adhering to the research protocol. On March 22, 2019,
24 STAR*D investigators acknowledged our email notification, indicated the STAR*D data were in the public
25 domain, and stated they had no interest in undertaking a reanalysis.
26

27
28 In July 2019, we received a STAR*D Data Use Certificate, issued by the NIMH Data Archive Data Access
29 Committee, and gained access to the STAR*D levels 1-4 and follow-up subject-level dataset consisting of
30 26 text files, and limited supporting study documentation. In September 2019, we obtained funding
31 from the RIAT Support Center to reanalyze STAR*D.
32

33 *Patients*

34
35 STAR*D patients were 18 to 75 years of age, seeking care at 18 primary and 23 psychiatric care clinics.
36 Clinical research coordinators (CRCs) screened 4,790 patients for MDD. This screening included the
37 CRCs' administrating the HRSD, on which 4,041 patients scored ≥ 14 , met the other inclusion criteria, and
38 enrolled into the study. CRCs also gathered patients' psychiatric history, demographic information, and
39 administered both the Cumulative Illness Rating Scale and the Psychiatric Diagnostic Screening
40 Questionnaire to determine the extent of comorbid medical and psychiatric disorders.
41
42

43 *Acute Treatment*

44
45 STAR*D investigators sought to provide the highest quality of care to maximize the number of
46 remissions while minimizing dropouts (see appendix 1). Each treatment level consisted of 12 weeks of
47 antidepressant therapy, with an additional 2 weeks for patients deemed close to remission. Treatment
48 was administered using a system of measurement-based care that assessed symptoms and side effects
49 at each clinic visit. STAR*D investigators state, "*To enhance the quality and consistency of care,*
50 *physicians used the clinical decision support system that relied on the measurement of symptoms (QIDS-*
51 *C and QIDS-SR), side-effects, medication adherence, and clinical judgment based on patient progress.*"[1,
52 p.30] This system was used to guide aggressive medication dosing to "*ensure that the likelihood of*
53 *achieving remission was maximized and that those who did not reach remission were truly resistant to*
54 *the medication.*"[1, p.30]
55
56
57
58
59
60

1
2
3
4 STAR*D allowed patients to select acceptable treatment options for randomization in levels 2 to 4 “to
5 *empower patients, strengthen the therapeutic alliance, optimize treatment adherence, and improve*
6 *outcome*”[17, p.483] and evaluated the relative effectiveness of 11 pharmacologically distinct drug/drug
7 combinations as well as cognitive therapy in five head-to-head comparisons. Appendix 2 describes each
8 treatment.
9

10
11 During each treatment trial, patients who scored <6 on their last clinic-administered QIDS-Clinician
12 version (QIDS-C) were encouraged to enter follow-up. The QIDS-C was administered at every clinic visit
13 along with the QIDS-SR. Based on prior research, a QIDS score of <6 was estimated by STAR*D
14 investigators to correspond to a score of <8 on the HRSD, STAR*D’s prespecified primary outcome
15 measure for classifying patients as remitted.[9] Clinicians strongly encouraged patients who did not
16 obtain a QIDS-defined remission to enter the next-level treatment. Patients who failed to attain a QIDS-
17 defined remission, but did have a ≥50% reduction on the QIDS-C and did not want to be randomized to a
18 next-level treatment, were also encouraged to enter follow-up.
19
20

21 *Research Design*

22
23 STAR*D investigators developed a new research design for the study termed “equipoise-stratified” to
24 evaluate the relative efficacy of 13 antidepressant therapies in levels 2-4 for depressed patients who
25 failed to gain adequate benefit from their initial medication trial.[18] In level 1, all patients received
26 citalopram as their first treatment. In level 2, patients were informed regarding seven treatment
27 options to choose from: four switch options in which citalopram was stopped and the new treatment
28 initiated and three augmentation options in which citalopram was combined with a second
29 antidepressant treatment. In level 3, patients were informed regarding four treatment options to
30 choose from: two switch options and two augmentation options. Level 4 involved randomization to one
31 of two medication/medication combination switch options.
32
33

34 For most patients, their levels 3 and 4 treatments corresponded to treatment steps 3 and 4. For level 2
35 patients who failed to respond adequately to cognitive therapy alone or combined with citalopram and
36 chose to continue in the study, their third treatment step was designated level 2A and they were
37 randomized to one of two switch medications. For these patients, their level 2A treatment was their
38 third treatment step. For level 2A patients who did not adequately benefit from this medication trial,
39 they entered a fourth treatment step consisting of level 3 treatments. Similarly, the very few patients
40 who continued on to a fifth treatment step were randomized to level 4 treatments. STAR*D
41 investigators did not report on this group and neither do we.
42
43

44 *Analytic Plan*

45
46 We reanalyzed the STAR*D dataset with fidelity to the original research protocol wherever possible.
47 Where the protocol was silent, we used other STAR*D publications to guide our analysis. This occurred
48 four times. First, the protocol is silent regarding patients who entered the study without a ROA-
49 administered HRSD score of ≥14. In their level 1 article, STAR*D investigators deemed such patients
50 ineligible for analysis.[1] We do the same and extend this exclusion for such patients who continued on
51 to levels 2-4 because they did not meet this marker of depression severity. Second, the protocol is silent
52 on what to do with patients who met remission criteria on the HRSD at entry into their next-level
53 treatment. In STAR*D’s background article though, its investigators prespecify that “patients who begin
54 a level with HRSD <8 will be excluded from analyses.”[8, p.130] We therefore excluded 125 such
55
56
57
58
59
60

1
2
3 patients from our analyses of treatment levels 2-4. Third, the protocol is silent on how to analyze
4 patients who exit a treatment without taking the HRSD. STAR*D investigators state in their level 1
5 article, “our primary analyses classified patients with missing exit HRSD scores as nonremitters a
6 priori”[1, p.34] and repeat similar statements in their level 2-4 articles.[2-6] Therefore, we do likewise.
7

8
9 Finally, STAR*D had many patients with missing exit HRSD scores. In their level 2-4 articles, STAR*D
10 investigators mapped the final QIDS-SR score to the HRSD for patients missing their exit HRSD score to
11 assess the impact of their approach to counting such patients as “nonremitters a priori.” We therefore
12 calculated STAR*D’s remission rate both as prespecified based on an exit HRSD score of <8 as well as a
13 final QIDS-SR score of <6 for those patients missing an exit HRSD score.
14

15 All pre-processing and analyses were performed in R.[19] Authors 2 and 3 identified patients by their
16 subject key and used this variable to match information across datasets. Data on patients’ treatment
17 pathways, and when patients transitioned from one level to the next, were taken from the IVRA dataset
18 completed by CRCs, and verified against the data on patient level exits. Authors 2 and 3 then compared
19 the number of patients identified for all level 1-4 treatments to that reported in the STAR*D summary
20 article’s patient flowchart, and the Ns matched.[7, figure 1]
21
22

23 Next, authors 2 and 3 applied STAR*D’s inclusion for data analysis criterion to patients in treatment
24 levels 2-4 as well as excluded from analysis the 125 patients who scored <8 on the HRSD at entry into
25 their next-level treatment. We counted these 125 patients as remitted in the prior treatment level but
26 excluded them from the analyses of subsequent treatments. Appendix 3 presents the number of level 2-
27 4 patients excluded from our reanalysis, and the reasons for their exclusion. Appendix 4 is a table
28 identifying the number of patients with missing entry and/or exit HRSD scores for all level 1-4
29 treatments and used in our Last Observation Carried Forward (LOCF) analyses. As seen in Appendix 4,
30 1,330 patients were missing their exit HRSD score across all treatments.
31
32

33 We then compared STAR*D’s outcomes to those found in a meta-analysis of 7,030 patients enrolled in
34 antidepressant comparator trials. Similar to STAR*D, comparator trials typically are conducted open-
35 label without a control group and therefore are the appropriate data to compare STAR*D’s outcomes
36 to.[20] Continuous HRSD improvement means were provided by the first author of the meta-analysis.
37
38

39 Finally, we compared the STAR*D protocol’s step-by-step predictions of patient drop out and the
40 number of patients who would have a satisfactory treatment response and enter follow-up care to what
41 actually occurred.[10,figure 7] While these predictions’ purpose was to estimate the number of
42 continuing patients available for randomization in levels 2-4, at the meta-level these predictions are an
43 important hypothesis STAR*D tested by assessing how well its investigators could predict the aggregate
44 step-by-step successful treatment outcomes from their treat-to-remission model of care.
45
46

47 *Patient and Public Involvement*

48

49 Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans
50 of our research
51
52
53
54
55
56
57
58
59
60

Results

Figure 1 presents the overall flow of patients enrolled in the various protocol-defined treatment levels and places them in groups defined by the number of treatment steps. Of the 4,041 patients enrolled into STAR*D, 3,110 met the eligibility for data analysis criterion of having an ROA-administered HRSD score ≥ 14 . Figure 1 also identifies the number of patients who exited the study following each treatment step, the number who entered follow-up after each treatment step, and the number who were randomly assigned to a next-level treatment.

Appendix 5 describes the demographic and clinical features of the patients who entered treatment in steps 1-4 based on their initial baseline presentation when enrolling into the study. Summary statistics are presented as means and standard deviations for continuous variables and percentages for discrete variables. Note that 55.7% of STAR*D patients had 2 or more comorbid axis 1 disorders when first enrolled based on the Psychiatric Diagnostic Screening Questionnaire and averaged 2.5 comorbid medical conditions based on the Cumulative Illness Rating Scale. Furthermore, the average length of patients' current major depressive episode was 25.9 months. In a post hoc analysis, STAR*D investigators found that 77.8% of its enrolled patients would have been excluded from most antidepressant trials due to having two or more concurrent medical conditions, more than one comorbid psychiatric disorder, and/or a current depressive episode lasting > 2 years.[21]

Using LOCF, Table 1 presents the HRSD entry, exit, and mean change scores for patients by the specific treatment they received in steps 1-4 as well as the HRSD remission and response rates. Table 1 also provides the HRSD cumulative remission rate after up to 4 trials on antidepressant therapies as well as the combined HRSD plus QIDS-SR remission rate for the 1,330 patients missing an exit HRSD score.

Table 1: Outcomes Across All Treatments

Treatment Step	HRSD Score		HRSD Mean Change [95% confidence interval] (SD)	HRSD Remissions # (%)	# QIDS-SR Remissions for Patients with Missing HRSD	Combined HRSD Plus QIDS-SR Remissions # (%)	HRSD Response Rate # (%)
	Entry ^a Mean (SD)	Exit ^a Mean (SD)					
Step 1 (N=3,110)	21.87 (5.21)	15.46 (9.03)	6.41 [6.13, 6.70] (8.03)	794 (25.5%)	144	938 (30.2%)	1006 (32.3%)
Step 2 (N=1,134)	19.5 (6.16)	15.34 (8.46)	4.16 [3.75, 4.57] (6.97)	241 (21.3%)	43	284 (25%)	287 (25.3%)
Switch strategy (N=620)	20.44 (6.01)	16.17 (8.29)	4.27 [3.73, 4.81] (6.89)	113 (18.2%)	22	135 (21.8%)	152 (24.5%)
Bupropion(N=190)	20.75 (6.17)	17.03 (8.39)	3.72 [2.85, 4.58] (6.10)	31 (16.3%)	6	37 (19.5%)	41 (21.6%)
Sertraline (N=198)	20.60 (6.10)	16.30 (8.25)	4.30 [3.33, 5.26]	32 (16.2%)	4	36 (18.2%)	48 (24.2%)

			(6.93)				
Venlafaxine (N=192)	20.36 (5.91)	15.56 (8.21)	4.80 [3.77, 5.82] (7.26)	37 (19.3%)	8	45 (23.4%)	50 (26.0%)
Cognitive Therapy (N=40)	18.52 (5.06)	14.30 (8.20)	4.22 [1.63, 6.82] (8.36)	13 (32.5%)	4	17 (42.5%)	13 (32.5%)
Augmentation strategy (N=514)	18.37 (6.15)	14.34 (8.55)	4.03 [3.42, 4.64] (7.08)	128 (24.9%)	21	149 (29%)	135 (26.3%)
Bupropion (N=216)	17.94 (6.18)	13.71 (8.33)	4.22 [3.29, 5.15] (6.97)	54 (25.0%)	10	64 (29.6%)	61 (28.2%)
Buspirone (N=225)	18.74 (6.46)	14.85 (8.95)	3.89 [2.97, 4.82] (7.07)	58 (25.8%)	10	68 (30.2%)	56 (24.9%)
Cognitive Therapy (N=73)	18.52 (4.99)	14.64 (7.91)	3.88 [2.16, 5.59] (7.48)	16 (21.9%)	1	17 (23.3%)	18 (24.7%)
Step 3 (N=325)	19.86 (6.14)	17.29 (7.78)	2.57 [1.92, 3.22] (6.00)	43 (13.2%)	7	50 (15.4%)	52 (16.0%)
Level 2A (N=28)	20.71 (5.56)	17.75 (6.34)	2.96 [0.91, 5.02] (5.55)	3 (10.7%)	0	3 (10.7%)	4 (14.3%)
Bupropion (N=12)	19.50 (4.15)	17.75 (7.36)	1.75 [-1.65, 5.15] (6.00)	2 (16.7%)	0	2 (16.7%)	2 (16.7%)
Venlafaxine (N=16)	21.62 (6.41)	17.75 (5.71)	3.88 [1.33, 6.42] (5.20)	1 (6.3%)	0	1 (6.3%)	2 (12.5%)
Level 3 (N=297)	19.78 (6.19)	17.25 (7.91)	2.53 [1.84, 3.22] (6.05)	40 (13.5%)	7	47 (15.8%)	48 (16.2%)
Switch strategy (N=186)	20.28 (6.25)	17.80 (8.04)	2.49 [1.57, 3.41] (6.39)	23 (12.4%)	2	25 (13.4%)	27 (14.5%)
Nortriptyline (N=92)	20.04 (5.55)	17.87 (8.51)	2.17 [0.72, 3.63] (7.12)	15 (16.3%)	0	15 (16.3%)	15 (16.3%)
Mirtazapine (N=94)	20.52 (6.89)	17.72 (7.60)	2.80 [1.67, 3.93] (5.60)	8 (8.5%)	2	10 (10.6%)	12 (12.8%)
Augmentation strategy (N=111)	18.93 (6.01)	16.32 (7.64)	2.60 [1.59, 3.62] (5.46)	17 (15.3%)	5	22 (19.8)	21 (18.9%)
Lithium	19.22	16.98	2.24	7	2	9	7

(N=58)	(6.81)	(7.31)	[1.06, 3.42] (4.59)	(12.1%)		(15.5%)	(12.1%)
T3 (N=53)	18.60 (5.05)	15.60 (7.99)	3.00 [1.31, 4.69] (6.29)	10 (18.9%)	3	13 (24.5%)	14 (26.4%)
Step 4 (N=106)	20.84 (5.36)	17.44 (7.25)	3.40 [2.19, 4.60] (6.32)	11 (10.4%)	1	12 (11.3%)	17 (16.0%)
Level 3 (N=16)	20.56 (4.03)	17.62 (6.37)	2.94 [-0.12, 5.99] (6.23)	2 (12.5%)	0	2 (12.5%)	2 (12.5%)
Tranlycypromine (N=43)	21.40 (6.02)	17.67 (7.12)	3.72 [1.69, 5.75] (6.78)	3 (7.0%)	1	4 (9.3%)	8 (18.6%)
Venlafaxine XR/mirtazapine (N=47)	20.43 (5.17)	17.17 (7.78)	3.26 [1.53, 4.98] (6.03)	6 (12.8%)	0	6 (12.8%)	7 (14.9%)
Cumulative Remission Rate after up to four Antidepressant Therapies				1,089 (35.0%)		1,284 (41.3%)	

^a Used Last Observation Carried Forward for missing HRSD values.

Appendix 6 presents patients' aggregate HRSD status in terms of remission, response, and extent of symptomatic change at entry and exit for each treatment step as well as study dropout. In step 1, 25.5% of patients remitted. Steps 2-4 show a continuous decrease in remission rates from step 2's 21.3% to 13.2% for step 3 and 10.4% in step 4 with increasing rates of study dropout from step 1's 34.5% to step 3's 46.2%.

Appendixes 7 and 8 present figures comparing the HRSD remission, response and extent of symptom improvement rates for STAR*D patients in steps 1-4 to that found in a meta-analysis of 7,030 patients enrolled in non-blinded antidepressant comparator trials.[20] In step 1, these measures of improvement among STAR*D's patients were approximately half that found in comparator trials, and improvement grew progressively worse in each subsequent treatment episode.

Figure 2 compares the STAR*D protocol's predictions of patient dropout and the number of patients who would have a satisfactory treatment response and enter follow-up to what occurred. Cumulatively, STAR*D's investigators predicted that 73.8% of patients would have a successful treatment response and enter follow-up whereas in fact only 45.6% achieved this measure of treatment success. Furthermore, whereas its investigators predicted that over the course of up to four antidepressant therapies 20.7% of patients would dropout, in fact, 53.7% dropped out. On this measure of treatment failure, STAR*D's dropout rate was 2.6 times greater than predicted.

Figure 3 presents the step-by-step cumulative remission rate three ways. First, the 'theoretical' rate propagated by STAR*D investigators based on the provisos of what would have happened if there were no study dropouts and that those who did exit had the same QIDS-SR remission rates as those who stayed.[7] Next, the combined HRSD plus QIDS-SR remission rate based on either an exit HRSD score of

1
2
3 <8, OR a last clinic visit QIDS-SR score of <6 for the 1,330 patients missing an exit HRSD. Finally, our RIAT
4 reanalysis rate when using the protocol-specified exit HRSD score of <8 as the sole measure of remission
5 for the 3,110 patients who met STAR*D's inclusion for data analysis criteria. The cumulative remission
6 rate after up to four antidepressant therapies using the HRSD was 35% versus 41.3% when combined
7 with the QIDS-SR, both of which are substantially less than the 67% cumulative remission rate claimed in
8 the summary article's Abstract.
9

10 11 Discussion

12 13 *Principal findings and comparison with original STAR*D publication*

14
15 STAR*D's results highlight the discrepancy in likely outcomes between typical antidepressant clinical
16 trials with their exclusion criteria and the real-world patients for whom these medications are
17 commonly prescribed. Our RIAT reanalysis found poorer outcomes after up to four optimized, and
18 increasingly aggressive, antidepressant therapies than reported in the original publication. In contrast
19 to the 67% cumulative remission rate reported in the STAR*D summary article, the actual rate was 35%
20 when using the protocol-specified HRSD, and increased to 41.3% when combined with the QIDS-SR for
21 those patients missing an exit HRSD.
22
23

24 25 *Comparison with other studies*

26
27 Our reanalysis found that in step 1, STAR*D's remission, response, and extent of improvement rates
28 were only about half of those reported in other open-label antidepressant comparator trials and then
29 grew progressively worse in steps 2-4. Such trials typically exclude depressed patients with the range
30 and number of comorbid medical and/or psychiatric disorders that were included in STAR*D.
31

32 Regarding the protocol's predictions of treatment success and patient dropout, it states:

33
34
35 We arrived at these estimates using three experienced practitioners (Drs. Fava, Rush,
36 and Thase) who independently made estimates that were surprisingly close to each
37 other. Then, via teleconferencing, the final estimates were made. ***The underlying***
38 ***assumptions of these estimates come largely by inferences from results of***
39 ***published RCTs.***[10, p.31; emphasis added]
40

41
42 STAR*D's actual measures of treatment success and failure were significantly worse than predicted. As
43 Barbui et al. noted, antidepressant study dropout rates provide a "hard measure of treatment
44 effectiveness and acceptability"[22, p.296] and STAR*D's dropout rate was 2.6 times greater than that
45 predicted. This discrepancy further highlights the relative ineffectiveness of antidepressants in real-
46 world depressed patients, compared to those reported in conventional trials.
47

48
49 STAR*D's step 1 remission rate was 25.5% followed by a progressive decline in remission rates for those
50 patients receiving subsequent, and increasingly aggressive treatments, such that by step 4 it was only
51 10.4%. This decline in antidepressant medications' effectiveness essentially mirrors the findings from
52 randomized and naturalistic, prospective studies reporting a 20-30% loss of effectiveness with each
53 increase in the number of prior antidepressant trials.[23-28] Furthermore, several recent analyses
54 suggest that the sequential application of antidepressant medications for non-remitting depression may
55 in fact foster treatment resistance for many patients.[29-32]
56
57
58
59
60

Conclusion

STAR*D investigators did not find any significant group differences in the five level 2-4 comparisons of 11 pharmacologically-distinct drug/drug combination treatments.[2-6] Furthermore, no post hoc secondary analyses have reported significant predictors of outcomes between these pharmacologically-distinct treatments. Therefore, the largest and most expensive antidepressant trial conducted to date did not provide 'next-level' treatment guidance for improving outcomes. In this light, STAR*D was a failed trial both in terms of providing no guidance but also its disappointing outcomes when these optimized treatments were administered to depressed patients, many with a wide range of co-morbid medical and psychiatric conditions.

Our RIAT reanalysis documents scientific errors which inflated STAR*D investigators' report of positive outcomes. The STAR*D summary article's claim of a 67% remission rate from its try-try-try-and-try again approach to treating depression was published in 2006. If STAR*D's outcomes had been reported as prespecified, its sequential treat-to-remission model of care would likely have faced much stronger criticism 16 years ago and fueled a more vigorous search for evidence-based treatment alternatives.

Ethics Statement:

The study was overseen by NIMH; NCT00021528.

Data Availability Statement:

Data is available from the NIMH-supported National Database for Clinical Trials (NDCT).

Funding Statement:

Funding for this project was provided by The RIAT Support Center.

Competing Interests:

The authors have no relevant financial disclosures or other conflicts of interest to report.

Contributorship Statement:

HEP, JDA and IK contributed to the design of the study and secured funding. TK and CX conducted all of the data analyses. HEP wrote the manuscript with input from JDA, IK, TK, and CX.

Acknowledgements:

We thank Termeh Feinberg for her early efforts on this project, particularly her correspondence with the NIMH help desk to resolve issues with the 26 data files as well as the RIAT Support Center for funding this project. Data used in the preparation of this manuscript were obtained from the controlled access datasets distributed from the NIMH-supported National Database for Clinical Trials (NDCT). NDCT is a collaborative informatics system created by the National Institute of Mental Health to provide a national resource to support and accelerate discovery related to clinical trial research in mental health. The content of this publication does not necessarily reflect the views of the RIAT Support Center nor NIMH.

References:

1. 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. 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. 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. Bupropion-SR, sertraline, or venlafaxine-XR after failure of SSRIs for depression. *N Engl J Med*. 2006;354:1231–1242.
3. 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. Medication augmentation after the failure of SSRIs for depression. *N Engl J Med*. 2006;354:1243–1252.
4. Fava M, Rush AJ, Wisniewski SR, Nierenberg AA, Alpert JE, McGrath PJ, Thase ME, Warden D, Biggs MM, Luther JF, Niederehe G, Ritz L, Trivedi MH. 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.
5. Nierenberg AA, Fava M, Trivedi MH, Wisniewski SR, Thase ME, McGrath PJ, Alpert JE, Warden D, Luther JF, Niederehe G, Lebowitz BD, Shores-Wilson K, Rush AJ. 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.
6. McGrath PJ, Stewart JW, Fava M, Trivedi MH, Wisniewski SR, Nierenberg AA, Thase ME, Davis L, Biggs MM, Shores-Wilson K, Luther JF, Niederehe G, Warden D, Rush AJ. Tranylcypromine versus venlafaxine plus mirtazapine following three failed antidepressant medication trials for depression: a STAR*D report. *Am J Psychiatry*. 2006;163:1531–1541.
7. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, Niederehe G, Thase ME, Lavori PW, Lebowitz BD, McGrath PJ, Rosenbaum JF, Sackheim HA, Kupfer DJ, Luther J, Fava M. 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.
8. Rush AJ, Fava M, Wisniewski SR, Lavori PW, Trivedi MH, Sackheim 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. Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design. *Control Clin Trials*. 2004;25:119–142.
9. Rush AJ, Trivedi MH, Ibrahim HM, Carmody T J, Arnow B, Klein DN, ... Manber R. The 16-item Quick Inventory of Depressive Symptomatology (QIDS), Clinician Rating (QIDS-C), and Self-Report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biological Psychiatry*. 2003;54:573–583.
10. National Institute of Mental Health (NIMH). Sequenced Treatment Alternatives to Relieve Depression (STAR*D) Research Protocol. Washington (DC): NIMH; revised June 28, 2002.
11. Pigott HE, Leventhal AM, Alter GS, Boren JJ. Efficacy and effectiveness of antidepressants: current status of research. *Psychother Psychosom*. 2010;79:267–279.
12. Greden JF. Workplace depression: personalize, partner, or pay the price. *Am J Psychiatry*. 2013;170:578–581.
13. Pigott, HE. The STAR*D trial: it's time to reexamine the clinical beliefs which guide the treatment of major depression. *Canadian J Psychiatry*. 2015;60:9–13.
14. Kirsch I, Huedo-Medina TB, Pigott HE, Johnson B. Do outcomes of clinical trials resemble those “real world” patients? A re-analysis of STAR*D antidepressant data. *Psychology of Consciousness: Theory, Research, and Practice*. 2018 <https://doi.org/10.1037/cns0000164>

15. Pigott HE, Dubin, W, Kirsch I, Amsterdam J. Call to action: RIAT reanalysis of the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) Study. *BMJ*. March 6, 2019; <https://www.bmj.com/content/346/bmj.f2865/rr-10>
16. Doshi P, Dickersin K, Healy D, Vedula SS, Jefferson T. Restoring invisible and abandoned trials: a call for people to publish the findings. *BMJ* 2013;346:f2865.
17. 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.
18. Lavori PW, Rush AJ, Wisniewski SR, Alpert J, Fava M, Kupfer DJ, Nierenberg A, Quitkin FM, Sackeim HA, Thase ME, Trivedi M. Strengthening clinical effectiveness trials: equipoise-stratified randomization. *Biological Psychiatry*. 2001;50:792– 801.
19. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
20. Rutherford BR, Sneed JR, Roose SP. Does study design influence outcome? The effects of placebo control and treatment duration in antidepressant trials. *Psychotherapy and Psychosomatics*. 2009;78:172–181.
21. Wisniewski SR, Rush AJ, Nierenberg AA, Gaynes BN, Warden D, Luther JF, McGrath PJ, Lavori PW, Thase ME, Fava M, Trivedi MH: Can phase III trial results of antidepressant medications be generalized to clinical practice? A STAR * D report. *Am J Psychiatry* 2009;166:599–607.
22. 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.
23. Amsterdam JD, Maislin G. Fluoxetine efficacy in treatment resistant depression. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*. 1994;18:243-261.
24. Nierenberg A, Feighner JP, Rudolph R, Cole JO, Sullivan J. Venlafaxine for treatment-resistant unipolar depression. *J Clinical Psychopharmacology*. 1994;14:419-423.
25. Amsterdam JD, Shults J. MAOI safety and efficacy in advanced treatment-resistant depression: a retrospective analysis. *J Affective Disorders*. 2005; 89:183-188.
26. Amsterdam JD, Williams D, Michelson D, Adler LA, Dunner SL, Nierenberg A, Reimherr FW, Schatzberg AF. Tachyphylaxis after repeated antidepressant drug exposures in patients with recurrent major depressive disorder. *Neuropsychobiology*. 2009;59:227-233.
27. Leykin Y, Amsterdam JD, DeRubeis RJ, Shelton RC, Hollon SD. Progressive resistance to SSRI therapy but not to cognitive therapy in the treatment of major depression. *J Consulting & Clinical Psychology*. 2007;75:267-276.
28. Amsterdam JD, Lorenzo-Luaces L, DeRubeis RJ. Step-wise loss of antidepressant effectiveness after repeated antidepressant trials in bipolar II depression. *Bipolar Disorders*. 2016;18:563-570.
29. Fava GA, Offidani E. The mechanisms of tolerance in antidepressant action. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*. 2011;35:1593-602.
30. Amsterdam JD, Lorenzo-Luaces L. Increase in pharmacodynamic tolerance after repeated antidepressant trials in treatment-responsive bipolar II depressed subjects: An exploratory study. *Psychiatria Polska*. 2018;52:957–969.
31. Amsterdam JD, Kim TT. Increased risk of depressive relapse during maintenance therapy after repeated antidepressant trials in treatment-responsive subjects. *J Clinical Psychopharmacology*. 2019;39:344-350.
32. Andrews PW, Amsterdam JD. A hormetic approach to understanding antidepressant effectiveness and the development of antidepressant tolerance. *Psychiatria Polska*. 2020;54:1067–1090.

Figure 3 Footnote:

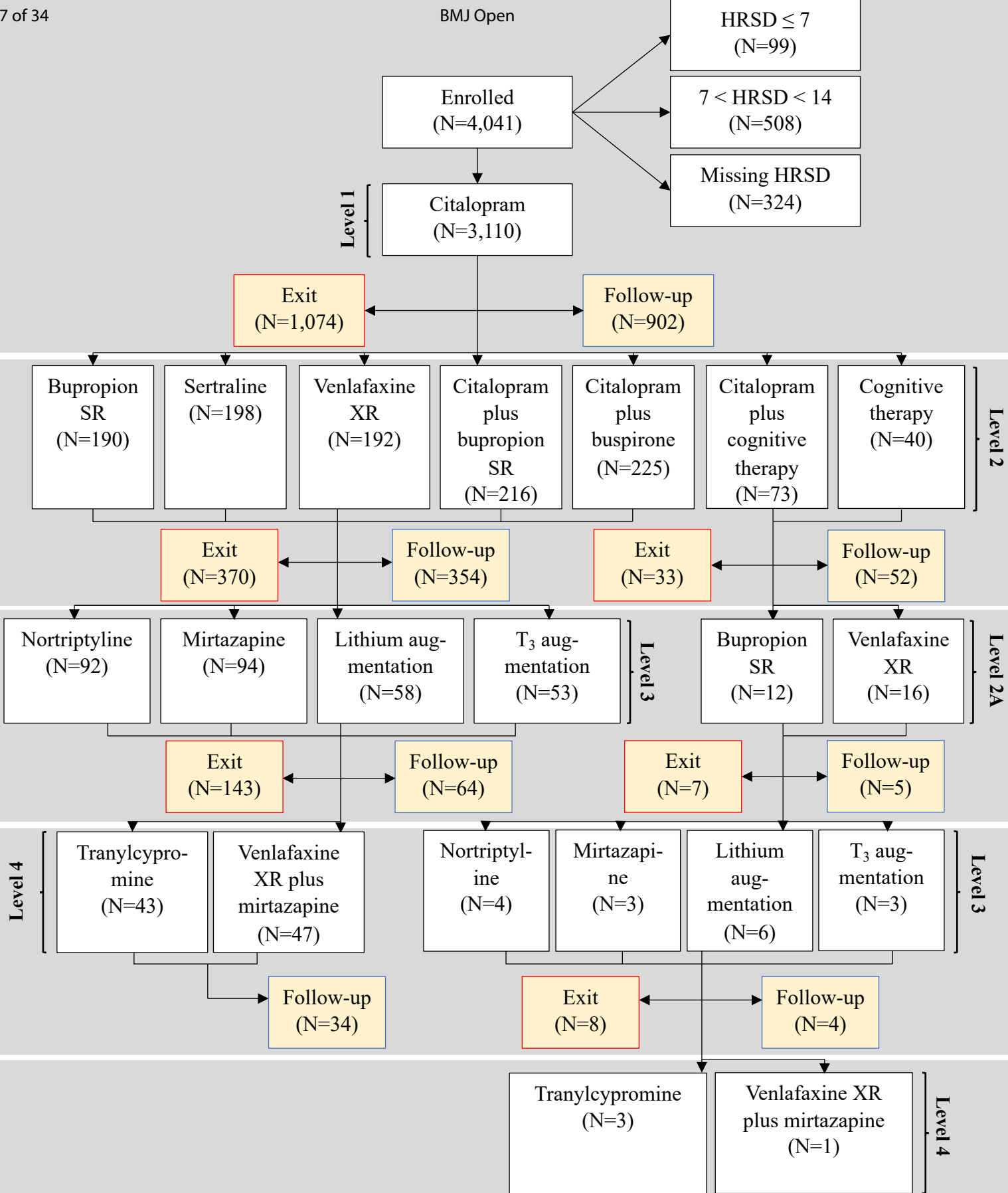
1
2
3
4 The step-by-step theoretical remission rates were obtained from the STAR*D summary article where it
5 states: "The theoretical cumulative remission rate is 67% (37+19+6+5)." [7, p.1910].
6

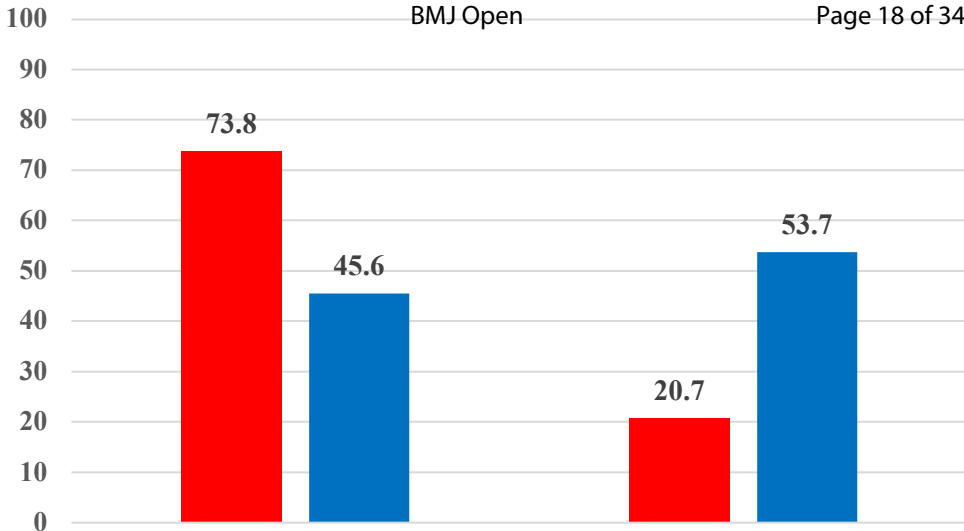
7
8 The HRSD + QIDS-SR cumulative remission rate was taken from Table 1. It combines the 1,089 patients
9 with an exit HRSD score of <8 with the 195 patients who were missing an exit HRSD score but had a final
10 clinic-visit QIDS-SR score of <6.
11

12 The RIAT Reanalysis cumulative remission rate is based on an exit HRSD score of <8 as the sole measure
13 of remission for the 3,110 patients who met STAR*D's inclusion for data analysis criteria.
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Step 1

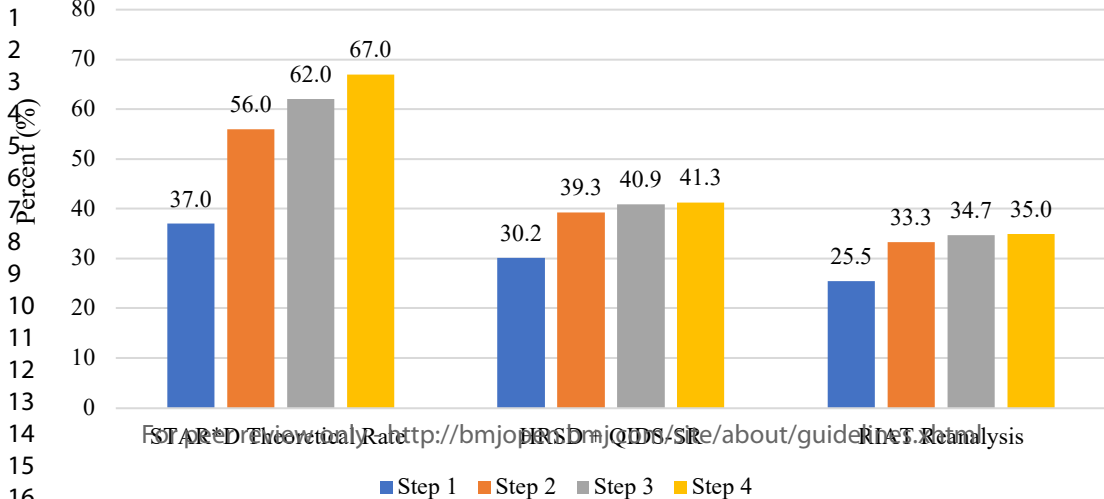
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18

For peer review only: <http://bmjopen.bmj.com/site/about/permissions.xhtml>

■ Star*D predicted ■ RIAT reanalysis



Appendix 1: Highest Quality of Acute and Continuing-Care to Maximize Remissions While Minimizing Relapse and Dropouts

Descriptor	Explanation
Optimized Sustained Study Participation to Minimize Dropouts ^{17, p. 473-474}	<ul style="list-style-type: none"> • Promoted patients' study affiliation via STAR*D-branded brochures, bimonthly newsletters, and an informational video emphasizing STAR*D's public health significance and the critical role played by patients; • Educated patients and families about depression and its treatment using a multi-step educational package; • Used a letter reminder system to alert patients before appointments in those clinics without such systems who had a >15% rate of missed appointments; • Ensured timely follow-up and rescheduling of missed appointments by calling patients on the day of the missed appointment, and again within 24 hours, if there was no response. Patient's physician sent letter within 48 hours if contact was not established; • Used a letter reminder system for all research outcome assessment calls during acute and continuing-care; • In every clinic visit, the Clinical Research Coordinator (CRC) discussed the research outcomes phone calls with the patient to ensure that the calls were completed on schedule and worked to resolve any problematic issues regarding said calls [Clinical Procedures Manual, page 75]; • Paid patients \$25.00 for participating in each telephonic research outcomes assessment; • Permitted patients to re-enter acute and/or continuing-care within four weeks after having dropped out [Clinical Procedures Manual, page 80]; • Recommended one-year of continuing-care for all patients who achieved a satisfactory clinical response with the essential goal of preventing relapse [Clinical Procedures Manual, page 15] and • Permitted continuing-care patients to remain in the study if they moved from the area [Clinical Procedures Manual, page 81].
Acute-Care Visits	Physicians met with patients on entry into each new step to initiate drug treatment with follow-up visits scheduled on weeks 2, 4, 6, 9, 12, with an optional week 14 visit.
Measurement-Based Care	Conducted structured evaluations of symptoms and side-effects at each visit and included a centralized treatment monitoring and physician feedback system to ensure consistent implementation of optimal care across research sites.
Aggressive Medication Dosing	Provided aggressive medication dosing with a fully adequate dose for a sufficient duration to <i>“ensure that the likelihood of achieving remission</i>

	<i>was maximized and that those who did not reach remission were truly resistant to the medication”</i> . ^{1, p.30}
Liberal Prescribing of Non-Study Medications	Physicians had great leeway in prescribing non-study medications to treat comorbid symptoms resulting in: <ul style="list-style-type: none"> • 17.2% taking Trazodone for sleep; • 11.9% taking an anti-anxiety medication; • 16.7% taking either a sedative or hypnotic medication; and • An undisclosed percent taking medications to address side-effects. ^{2, table 2}
Continuing-Care Visits	Patients saw their physician every 2 months and continued taking their treatment medication(s) at the same doses but their physicians were allowed to make any psychotherapy, medication, and/or medication dose changes to maximize the likelihood of maintaining patients' remission status. ^{7, p. 1908} Additional continuing-care visits were scheduled when patients began to experience a return of depressive symptoms and/or intolerable side-effects [Clinical Procedures Manual. page 78].
Clinical Research Coordinator (CRC)	Each site had a CRC who: ^{1, p. 30} <ul style="list-style-type: none"> • Saw patients before each visit administering multiple measures to them including the QIDS-SR during each acute-care visit; • Assisted physicians in protocol implementation; and • Provided patients support and encouragement in protocol implementation.
Treatment Designed to Enhance Subject Retention	Treatment was designed to minimize drop-outs and/or non-compliance including: <ul style="list-style-type: none"> • Open label prescribing during acute and continuing-care with no placebo control condition during any study phase; • Patients chose their acceptable treatment assignments for steps two and three to eliminate any concerns they might have about receiving an unacceptable assignment. This resulted in only 21 of 1,439 (1.5%) Step-2 patients making themselves available for random assignment to all treatment options ^{2, p. 1235} while only 29 of 377 (7.7%) did so in Step-3. ^{5, p. 1521} • During each step, patients could enroll immediately into the next step if they had intolerable side-effects or had maximized their current medication(s)' dosing without achieving a remission; and • During any step, patients could enter continuing-care directly on their current medication(s) if they were treatment responders even if they had not achieved remission. This was done to minimize responders from dropping out in order to avoid having to discontinue their current medication(s) and start a new drug regimen.

***Trivedi MH, Stegman D, Rush AJ, Wisniewski SR, Nierenberg AA: STAR * D clinical procedures manual. July 31, 2002. www.edc.pitt.edu/stard/public/study_manuals.html

Appendix 2: Description of Levels 1-4 Treatments

Level 1:

STAR*D investigators report that Citalopram (Celexa) was chosen as the first-line SSRI treatment because (1) absence of discontinuation symptoms; (2) demonstrated safety in elderly and medically fragile patients; (3) easy once-a-day dosing with few dose adjustments; and (4) favorable drug–drug interaction profile.¹ Citalopram was started at 20 mg/day and then raised to 40 mg/day by day 28 and up to 60 mg/day by day 43 and onward. Dose adjustments were based on how long a patient had received a particular dose, symptom changes, and side effect burden.

Level 2 switch treatments:

Citalopram was discontinued without a tapering at the initiation of each level 2 switch treatment. STAR*D investigators chose pharmacologically distinct switch medications.² The level 2 treatments were:

- Sertraline (Zoloft), an SSRI with the same pharmacological profile as citalopram. Sertraline was started at a daily dose of 50 mg and increased to 100 mg at day 8, to 150 mg at day 28, and to 200 mg at day 63 and onward.
- Sustained-release bupropion (Wellbutrin SR), an “out-of-class” agent whose neurochemical action mechanisms are unknown; other than that, it does not inhibit serotonin reuptake and is believed to produce antidepressant effects by blocking the reuptake of dopamine and norepinephrine. The daily dose of sustained-release bupropion was 150 mg for week 1, 200 mg from day 8 to 27, 300 mg from day 28 to 41, and 400 mg from day 42 onward.
- Extended-release venlafaxine (Effexor), a “dual-action” agent that inhibits the reuptake of both serotonin and norepinephrine. The starting daily dose of extended-release venlafaxine was 37.5 mg for week 1 and increased to 75 mg from day 8 to 14, to 150 mg from day 15 to 27, to 225 mg from day 28 to 41, to 300 mg from day 42 to 62, and to 375 mg from day 63 onward.
- Cognitive therapy was provided by a trained psychotherapist and scheduled twice weekly for the first four weeks, then once weekly for the remaining 8 weeks (16 sessions total).

Level 2 Citalopram augmentation treatments:

During the augmentation trial, the citalopram dose was kept constant but reduced if side effects developed. The level 2 augmentation treatments were:

- Buspirone (Buspar), a partial agonist at the postsynaptic 5-hydroxytryptamine 1A (5-HT_{1A}) receptor that is believed to enhance the activity of SSRIs through the 5HT_{1A} receptors. The starting dose was 15 mg per day week 1, increasing to 30 mg per day week 2, and then to 45 mg per day for weeks 3 through 5, and a final, maximum dose of 60 mg per day week 6 and onward.
- Sustained-release bupropion (Wellbutrin SR) whose neurochemical action mechanisms are unknown but is believed to produce antidepressant effects by blocking the reuptake of dopamine and norepinephrine. The initial dose was 200 mg per day during weeks 1 and 2, increasing to 300 mg per day by week 4 and to 400 mg per day week 6 and onward.
- Cognitive therapy was provided by a trained psychotherapist and scheduled twice weekly for the first four weeks, then once weekly for the remaining 8 weeks (16 sessions total).

Level 3 switch treatments:

At entry into the Level 3 switch trial, all level 2 medications were discontinued without tapering at the initial Level 3 treatment visit. The level 3 switch treatments were:

- Nortriptyline (Pamelor), a tricyclic antidepressant. Recommended doses were 25 mg/day for 3 days, 50 mg/day for 4 days, and then 75 mg/day by day 8, 100 mg/day by day 28, and, if necessary, 150 mg/day by day 42 and onward.
- Mirtazapine (Remeron), a tetracyclic antidepressant that blocks inhibitory α -adrenoceptors on norepinephrine and serotonin neurons to enhance both norepinephrine and serotonin neurotransmission. Recommended mirtazapine doses were 15 mg/day for the first 7 days, 30 mg/day by day 8, 45 mg/day by day 28, and, if necessary, 60 mg/day by day 42 and onward.

Level 3 augmentation treatments of level 2 medications:

The two medication augmentation options used in level 2, buspirone and sustained-release bupropion, were discontinued without tapering in the initial level 3 visit. The two medication augmentation treatments in level 3 were added to ongoing treatment with citalopram, sertraline, sustained-release bupropion, or extended-release venlafaxine. The level 3 augmentation treatments were:

- Lithium started at 450 mg/day, and at week 2 it was increased to the recommended dose of 900 mg/day. If participants could not tolerate the initial dose, it could be reduced to 225 mg/day for 1 week then increased to 450 mg/day. There was no monitoring of lithium levels.
- Triiodothyronine (T₃), a thyroid hormone, started at 25 μ g/day for 1 week and then increased to the recommended dose of 50 μ g/day. There was no pretreatment assessment, nor ongoing monitoring, of thyroid functioning.

Level 4 switch treatments:

The level 4 switch treatments were:

- Tranylcypromine (Parnate), a monoamine oxidase inhibitor. A 2-week washout period of Level 3 medications was required for patients assigned to the tranylcypromine group. The recommended dosing for tranylcypromine was 10 mg/day for the first 2 weeks, followed by weekly increases of 10 mg/day until a maximum of 60 mg/day.
- Co-administered venlafaxine (Effexor) and mirtazapine (Remeron) to inhibit the reuptake of both serotonin and norepinephrine and block inhibitory α -2-adrenoceptors on both norepinephrine and serotonin neurons to enhance both norepinephrine and serotonin neurotransmission. Level 3 medications were discontinued without tapering for patients assigned to this group. The dosage of extended-release venlafaxine was 37.5 mg/day for the first week, 75 mg/day for the second week, 150 mg/day for weeks 3–5, 225 mg/day for weeks 6–8, and 300 mg/day onward. Mirtazapine was started at 15 mg/day for the first 3 weeks, 30 mg/day for weeks 4 to 8, and then 45 mg/day onward.

Appendix 3: Number of Level 2-4 Participants Excluded from our RIAT Reanalysis, and the Reasons for their Exclusion, yet Included in STAR*D

Level 2 Treatments

Number of Level 2 Participants Excluded from our Reanalysis but Included in STAR*D	Bup	Sert	Ven	CT	Cit + BUP	Cit + Busp	Cit + CT	Total
Scored as Remitted at ENTRY into Level 2 yet still included in STAR*D's Level 2 analyses	22	8	14	7	30	24	4	109
Scored as only mildly depressed (HRSD >7 & <14) at entry into Level 1, and therefore excluded from STAR*D's data analysis, yet still treated in Level 1, progressed to Level 2, and then included in STAR*D's Level 2 data analyses	21	15	25	15	20	30	7	133
Scored as Remitted at entry into Level 1 (HRSD ≤ 7), and therefore excluded from STAR*D's data analysis, yet still treated in Level 1 and progressed to Level 2 and then included in STAR*D's Level 2 data analyses	6	1	4	2	2	2	2	19
Missing baseline HRSD at entry into Level 1, and therefore excluded from STAR*D's data analysis, yet still treated in Level 1, and progressed to Level 2, and then included in STAR*D's Level 2 data analyses	12	18	22	4	16	13	1	86
Number meeting 2 exclusion criterions	12	2	7	6	5	8	2	42

Bup=Sustained-release Bupropion; Sert= Sertraline; Ven= Extended-release Venlafaxine; CT=Cognitive Therapy; Cit+BUP= Citalopram + Sustained-release Bupropion; Cit+Busp=Citalopram + Buspirone; Cit+CT= Citalopram + Cognitive Therapy

Level 3 Treatments

	Nortriptyline	Mirtazapine	Lithium Augmentation	Triiodothyronine Augmentation	Total
Scored as Remitted at ENTRY into Level 3 yet still included in STAR*D's Level 3 analyses	4	0	1	5	10
Scored as only mildly depressed (HRSD >7 &	8	5	3	4	20

<14) at entry into Level 1, and therefore excluded from STAR*D's data analysis, yet still treated in Level 1, progressed to Level 2, and then 3 and included in STAR*D's Level 3 data analyses					
Scored as Remitted at entry into Level 1 (HRSD \leq 7), and therefore excluded from STAR*D's data analysis, yet still treated in Level 1 and progressed to Level 2 and then 3 and included in STAR*D's Level 3 data analyses	2	1	0	1	4
Missing baseline HRSD at entry into Level 1, and therefore excluded from STAR*D's data analysis, yet still treated in Level 1, and progressed to Level 2, and then level 3 and included in STAR*D's Level 3 data analyses	7	8	1	7	23

Level 4 Treatments

	Tranlycypromine	Venlafaxine + Mirtazapine	Total
Scored as Remitted at ENTRY into Level 4 yet still included in STAR*D's Level 4 analyses	5	1	6
Scored as only mildly depressed (HRSD $>$ 7 & $<$ 14) at entry into Level 1, and therefore excluded from STAR*D's data analysis, yet still treated in Level 1, progressed to Level 2, 3 and then 4 and included in STAR*D's Level 4 data analyses	3	1	4
Scored as Remitted at entry into Level 1 (HRSD \leq 7), and therefore excluded from STAR*D's data analysis, yet still treated in Level 1 and progressed	0	0	0

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

to Level 2, then Level 3 and included in STAR*D's Level 4 data analyses			
Missing baseline HRSD at entry into Level 1, and therefore excluded from STAR*D's data analysis, yet still treated in Level 1, and progressed to Level 2, and then Level 3 and 4 included in STAR*D's Level 4 data analyses	5	1	6

For peer review only

Appendix 4: Number and Percent of Participants Missing Entry and/or Exit HRSD Used for Last Observation Carried Forward Analyses

	#/(%) with Missing Entry HRSD	#/(%) with Missing Exit HRSD
Step 1 (N=3,110)	0 (0%)	926 (29.8%)
Step 2 (N=1,134)	168 (14.8%)	304 (26.8%)
Switch strategy (N=620)	90 (14.5%)	183 (29.5%)
Bupropion (N=190)	34	58
Sertraline (N=198)	27	56
Venlafaxine (N=192)	24	56
Cognitive Therapy (N=40)	5	13
Augmentation strategy (N=514)	78 (15.2%)	121 (23.5%)
Bupropion (N=216)	35	58
Bupirone (N=225)	37	52
Cognitive Therapy (N=73)	6	11
Step 3 (N=325)	42 (12.9%)	78 (24%)
Level 2A (N=28)	3	6
Bupropion (N=12)	3	2
Venlafaxine (N=16)	0	4
Level 3 (N=297)	39	72
Switch strategy (N=186)	26	49
Nortriptyline (N=92)	11	23
Mirtazapine (N=94)	15	26
Augmentation strategy (N=111)	13	23
Lithium (N=58)	9	13
Bupropion SR (N=17)	3	2
Citalopram (N=22)	5	6
Sertraline (N=11)	1	3
Venlafaxine XR (N=8)	0	2
T3 (N=53)	4	10
Bupropion SR (N=6)	1	0
Citalopram (N=26)	1	7
Sertraline (N=8)	1	1
Venlafaxine XR (N=13)	1	2
Step 4 (N=106)	15 (14.2%)	22 (20.8%)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Level 3 (N=16)	3	3
Tranlycypromine (N=43)	7	10
Venlafaxine XR/mirtazapine (N=47)	5	9
Total Across Treatment Steps	225	1,330

For peer review only

Appendix 5: Baseline Demographic and Clinical Features by Treatment Step

	Treatment Step ^a							
	Step 1 (N=3,110)		Step 2 (N=1,134)		Step 3 (N=325)		Step 4 (N=106)	
Demographic Features								
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age	41.0	13.0	42.0	12.6	44.1	12.0	46.9	11.0
Education (years)	13.6	3.2	13.2	3.3	12.8	3.1	12.6	2.3
Monthly household income	2,289	2,732	1,744	1,539	1,470	1,383	1,003	887
	N	%	N	%	N	%	N	%
Female	1,469	74.6	502	73.1	113	65.3	34	65.4
Race								
White	2,328	74.9	870	76.7	259	79.7	86	81.1
Black	333	10.7	115	10.1	29	8.9	7	6.6
Other	449	14.4	149	13.1	37	11.4	13	12.3
Hispanic	402	12.9	139	12.3	45	13.8	16	15.1
Employment status								
Employed	975	58.7	314	54.2	69	46.9	19	43.2
Unemployed	612	36.9	243	42.0	72	49.0	24	54.5
Retired	73	4.4	22	3.8	6	4.1	1	2.3
Medical insurance								
Private	848	52.2	254	44.5	52	36.6	14	31.8
Public	282	17.4	109	19.2	30	21.4	10	23.3
None	534	33.2	223	39.3	60	43.2	20	46.5
Marital status								
Single	475	28.6	171	29.5	40	27.2	10	22.7
Married/cohabiting	716	43.1	238	41.0	61	41.5	18	40.9
Divorce/separated	429	25.8	155	26.7	42	28.6	14	31.8
Widowed	41	2.5	16	2.8	4	2.7	2	4.5
Clinical Features	N	%	N	%	N	%	N	%
First episode occurrence before age 18	1,200	39.0	436	38.8	120	37.0	41	38.7
Recurrent depression	1,940	66.8	718	68.3	188	63.3	59	60.8
Family history of depression	1,694	55.4	609	54.9	165	51.7	58	54.7

Duration of current episode \geq 2 years	787	25.6	311	27.7	88	27.2	34	32.1
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age at first episode (years)	24.9	14.5	24.7	14.2	25.9	14.6	25.9	14.3
Illness duration (years)	16.1	13.5	17.2	13.7	18.2	14.1	21.0	14.9
Number of episodes	4.4	9.7	4.9	11.1	4.4	10.3	5.0	12.1
Duration of current episode (months)	25.9	52.0	28.1	58.8	32.1	68.5	45.9	82.0
Median duration of current episode (months)	8.3		8.7		9.5		10.1	
Quality of Life and Enjoyment Satisfaction Questionnaire score ^b	39.1	14.3	36.5	13.6	33.7	13.5	31.6	12.8
SF-12 Mental ^c	25.6	8.1	25.0	7.7	24.4	7.7	24.0	7.5
SF-12 Physical ^c	48.6	12.1	47.0	12.4	44.5	12.1	43.8	12.3
Work and Social Adjustment Scale score ^d	25.0	8.7	26.3	8.2	28.3	7.7	29.4	7.0
HRSD ₁₇ score	21.9	5.2	22.5	5.2	23.4	5.2	23.9	5.4
IDS-C ₃₀ score ^e	39.1	9.6	40.6	9.7	42.6	9.4	43.6	9.8
QIDS-IVR score ^f	16.9	3.3	17.3	3.3	17.9	3.0	18.3	3.1
Cumulative Illness Rating Scale								
Categories endorsed	2.5	1.5	2.6	1.6	2.8	1.6	3.1	1.6
Total score	4.7	3.9	5.1	4.0	5.8	4.5	6.2	4.3
Severity score	1.8	0.8	1.8	0.8	2.0	0.9	2.0	0.7
	N	%	N	%	N	%	N	%
Psychiatric Diagnostic Screening Questionnaire								
Agoraphobia	559	18.2	240	21.4	89	27.5	32	30.2
Alcohol abuse/dependence	371	12.0	136	12.1	36	11.1	8	7.5
Bulimia	607	19.7	232	20.6	67	20.7	20	18.9

Drug abuse/dependence	234	7.6	80	7.1	21	6.5	7	6.6
Generalized anxiety disorder	736	23.9	290	25.8	94	29.0	36	34.0
Hypochondriasis	336	10.9	139	12.4	45	13.9	14	13.2
Obsessive-compulsive disorder	723	23.5	265	23.6	97	29.9	31	29.2
Panic disorder	422	13.7	183	16.3	65	20.1	21	19.8
Posttraumatic stress disorder	387	12.6	172	15.3	55	17.0	16	15.1
Social phobia	963	31.3	379	33.7	117	36.1	35	33.0
Somatoform disorder	284	9.2	105	9.3	35	10.8	9	8.5
Number of axis I comorbid psychiatric disorders								
0	606	19.7	190	16.9	48	14.8	12	11.3
1	740	24.0	257	22.9	68	21.0	23	21.7
2	577	18.7	217	19.3	62	19.1	25	23.6
3	363	11.8	139	12.4	42	13.0	14	13.2
4+	793	25.8	321	28.6	104	32.1	32	30.2

^a Sums do not always equal N due to missing values. Percentages are based on available data.

^b Integrated voice response (IVR) administered version of the Quality-of-Life Enjoyment and Satisfaction Questionnaire assessing participants' global rate of satisfaction. Higher scores (range=0–100) represent greater life enjoyment and satisfaction.

^c IVR-administered version of the SF-12 assessing perceived mental and physical health status. Two subscales (physical health factor and mental health) range from 0 to 100— higher scores indicate better functioning with a population norm for each score of 50.

^d IVR-administered version of the Work and Social Adjustment Scale. Scores between 10 and 20 are associated with significant functional impairment while scores above 20 suggest moderate to severe functional impairment.

^e Inventory of Depressive Symptomatology administered telephonically.

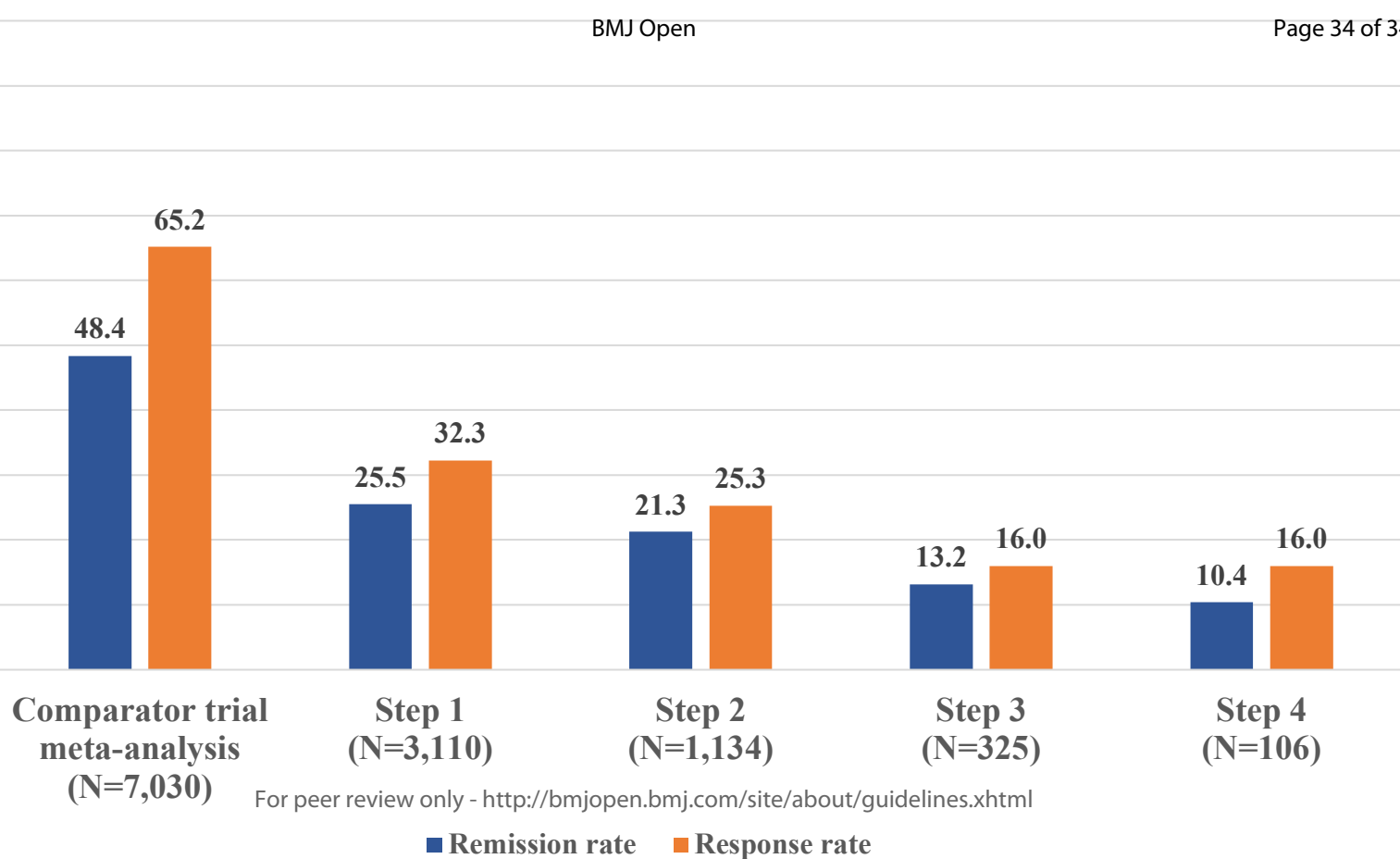
^f IVR-administered version of the QIDS.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

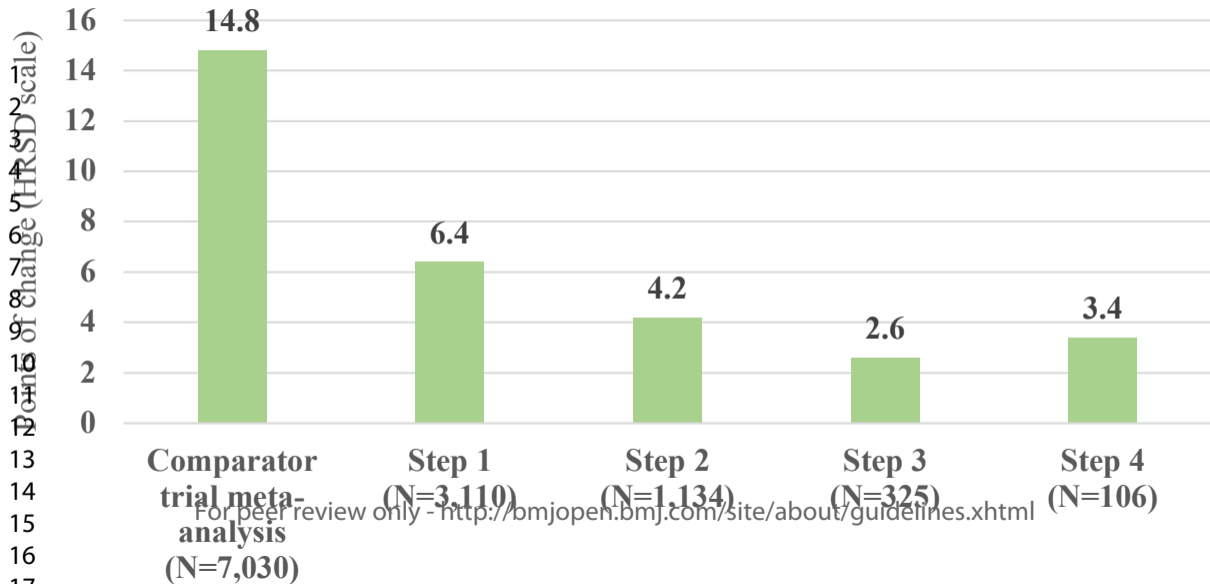
Appendix 6: Outcomes by Treatment Step

	Step 1 (N=3,110)		Step 2 (N=1,134)		Step 3 (N=325)		Step 4 (N=106)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
HRSD score at entry into step*	21.87	5.21	19.5	6.16	19.86	6.14	20.84	5.36
HRSD score at exit from step*	15.46	9.03	15.34	8.46	17.29	7.78	17.44	7.25
HRSD Mean Change	6.41	8.03	4.16	6.97	2.57	6.0	3.40	6.32
	N	%	N	%	N	%	N	%
Remission at each step exit	794	25.5%	241	21.3%	43	13.2%	11	10.4%
Response at each step exit	1006	32.3%	287	25.3%	52	16.0%	17	16.0%
Entered Follow-up	902	29.0%	406	35.8%	69	21.2%	38	35.9%
Study Exit/Dropout	1,074	34.5%	403	35.5%	150	46.2%		

* Last observation carried forward

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27For peer review only - <http://bmjopen.bmj.com/site/about/guidelines.xhtml>

■ Remission rate ■ Response rate



BMJ Open

What are the Treatment Remission, Response, and Extent of Improvement Rates after up to Four Trials of Antidepressant Therapies in Real-World Depressed Patients? A Reanalysis of the STAR*D Study with Fidelity to the Original Research Protocol

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2022-063095.R1
Article Type:	Original research
Date Submitted by the Author:	10-Nov-2022
Complete List of Authors:	Pigott, H.; None Kim, Thomas; University of Pennsylvania, Psychology Xu, Colin; University of Pennsylvania, Psychology Kirsch, Irving; Harvard Medical School Amsterdam, Jay; University of Pennsylvania Perelman School of Medicine, Psychiatry
Primary Subject Heading:	Mental health
Secondary Subject Heading:	Pharmacology and therapeutics
Keywords:	Adult psychiatry < PSYCHIATRY, Depression & mood disorders < PSYCHIATRY, CLINICAL PHARMACOLOGY

SCHOLARONE™
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1
2
3 **What are the Treatment Remission, Response, and Extent of**
4 **Improvement Rates after up to Four Trials of Antidepressant Therapies in Real-World**
5 **Depressed Patients? A Reanalysis of the STAR*D Study with Fidelity to the Original Research**
6 **Protocol**

7 **Authors:**

8
9 H. Edmund Pigott, Thomas T. Kim, Colin Xu, Irving Kirsch, & Jay D. Amsterdam
10
11
12

13 H. Edmund Pigott, Ph.D.
14 Position: Clinical Psychologist
15 Institution: none
16 Email: pathware@erols.com
17
18

19 Thomas T. Kim
20 Position: PhD candidate in Psychology
21 Institution: Department of Psychology, University of Pennsylvania
22
23

24 Colin Xu
25 Position: PhD candidate in Psychology
26 Institution: Department of Psychology, University of Pennsylvania
27
28

29 Irving Kirsch, Ph.D.
30 Position: Associate Director, Program in Placebo Studies
31 Institution: Harvard Medical School
32
33

34 Jay D. Amsterdam, MD
35 Position: Professor of Psychiatry (Emeritus), Depression Research Unit
36 Institution: Perelman School of Medicine, University of Pennsylvania
37
38

39 Corresponding author: H. Edmund Pigott, Ph.D.
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Abstract

Objective: Reanalyze the patient-level dataset of the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study with fidelity to the original research protocol and related publications.

Design: The study was open label and semi-randomized examining the effectiveness of up to four optimized, and increasingly aggressive, antidepressant therapies in depressed adults. Patients who failed to gain adequate relief from their level 1 trial on the SSRI citalopram could receive up to three additional treatment trials in levels 2-4.

Setting: 41 North American psychiatry and primary care treatment centers.

Participants: 4,041 adults screened positive for major depressive disorder. STAR*D enrolled patients seeking care (versus recruited) and included patients with a wide range of common co-morbid medical and psychiatric conditions to enhance the generalizability of findings to real-world clinical practice.

Interventions: STAR*D evaluated the relative effectiveness of 13 antidepressant therapies in treatment levels 2-4 for depressed patients who failed to gain adequate benefit from their level 1 medication trial.

Main Outcome Measures: The primary outcome was remission, defined as a score <8 on the blinded Hamilton Rating Scale for Depression (HRSD). Response was a secondary outcome defined as ≥50% reduction in HRSD scores. STAR*D's protocol specifically excluded all non-blinded clinic-administered assessments from use as research outcome measures.

Results: STAR*D investigators did not use the protocol-stipulated HRSD to report cumulative remission and response rates in their summary article, and instead used a non-blinded clinic-administered assessment. This inflated their report of outcomes, as did their inclusion of 99 patients who scored as remitted on the HRSD at study outset as well as 125 who scored as remitted when initiating their next-level treatment. These patients should have been excluded from data analysis. In contrast to the STAR*D-reported 67% cumulative remission rate after up to four antidepressant treatment trials, the rate was 35% when using the protocol-stipulated HRSD and inclusion in data analysis criteria.

Conclusion: STAR*D's cumulative remission rate was approximately half than that reported.

Strengths and limitations of this study

- We reanalyzed the largest ever prospective antidepressant trial's patient-level dataset with fidelity to the original research protocol and related publications.
- The reanalysis was conducted under the guidelines of the *Restoring Invisible and Abandoned Trials* (RIAT) initiative.

- Treatment remission, response, and extent of symptom improvement rates were calculated for 14 antidepressant therapies for those patients who met STAR*D's inclusion in data analysis criteria as well as the overall cumulative remission rate after up to four trials of antidepressant therapies.
- We calculated STAR*D's remission rate using the protocol-stipulated HRSD as well as combining the HRSD remissions with those from a non-stipulated measure of remission for patients missing an exit HRSD score.
- Finally, we compared STAR*D's outcomes to the protocol-predicted outcomes as well as to those found in a meta-analysis of 7,030 patients enrolled in similar open-label antidepressant comparator trials. Comparator trials are the appropriate comparison data for STAR*D's outcomes.

Introduction

The 35-million US dollar Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study is the largest prospective antidepressant study ever conducted, with over 100 journal articles published by study investigators and innumerable citations of STAR*D's findings by other researchers, giving it an oversized impact on the treatment of depression world-wide.[1-7] Funded by the National Institute of Mental Health (NIMH), STAR*D enrolled 4041 patients who screened positive for major depressive disorder (MDD) while seeking routine medical or psychiatric care. In contrast to industry-funded studies, STAR*D did not exclude patients with medical conditions and most comorbid psychiatric disorders, thereby increasing the generalizability of its findings to real-world clinical practice.

The STAR*D study provided up to four treatment trials per patient and was designed to give guidance in selecting the best next-level treatment option for the many people who fail to gain sufficient relief from their first, and/or subsequent, antidepressant trial. To mimic clinical practice, STAR*D used an open-label research design with no control group during any phase of the study.

Our STAR*D reanalysis examines key methodological deviations from its research protocol and related publications, and these deviations' impact on its investigators report of outcomes. In STAR*D's Rationale and Research Design article, and repeated in the level 1-4 published study outcomes, STAR*D investigators stated, "*the primary outcome is depressive symptom severity, measured by the 17-item Hamilton Rating Scale for Depression (HRSD).*"[8, p. 120]. STAR*D's prespecified primary outcome was remission, defined as scoring <8 on the HRSD which was administered telephonically by Research Outcome Assessors (ROAs) blind to patients' study status (treatment level entry/exit/follow-up). Response was a secondary outcome defined as a $\geq 50\%$ reduction in patients' HRSD scores. However, despite its investigators numerous publications, neither change in HRSD depressive symptom severity nor HRSD response rates have been reported for STAR*D's six primary studies [1-6] and summary article.[7] Instead, response rates and change in symptom severity were reported using the clinic-administered Quick Inventory of Depressive Symptomatology–Self Report (QIDS-SR), a measure developed by the STAR*D principal investigators.[9] This occurred despite the fact that STAR*D's research protocol specifically excluded all clinic-administered assessments, such as the QIDS-SR, from use as research outcome measures since they were not blinded and instead, used to guide patient care. The protocol states:

Recall that the research outcomes assessments are distinguished from assessments conducted at clinic visits. The latter are designed to collect information that guides

1
2
3 *clinicians in the implementation of the treatment protocol. Research outcomes*
4 *assessments are **not** collected at the clinic visits. They are not collected by either*
5 *clinicians or Clinical Research Coordinators.[10,p.47-48; emphasis in the original]*
6

7
8 In their summary article, STAR*D investigators used the QIDS-SR as the sole measure to report
9 remission, response, and extent of symptom improvement. This article's Abstract states that "the
10 overall cumulative remission rate was 67%" with no qualifiers to this claim.[7, p.1905] Besides making
11 this claim based on an assessment the protocol specifically excluded from use as a research measure, it
12 is not until the article's Results section that readers learn this high level of purported treatment success
13 did not occur. The STAR*D investigators' claim was theoretical—an estimate based on the unrealistic
14 provisos of what would have happened if there were no study dropouts, and furthermore, "that those
15 who exited the study would have had the same remission rates as those who stayed in the protocol." [7,
16 p.1910] As Pigott et al. document though, the investigators' assumptions are not true in the real world
17 since more patients dropped out than remitted in each STAR*D treatment level,[11] and patients who
18 drop out are more likely to have had adverse treatment side effects and/or emergent suicidality.[12]
19

20
21 Unfortunately, the STAR*D investigators' claim of a 67% cumulative remission rate has become
22 accepted clinical wisdom, and the improbable provisions on which it is based are commonly not
23 referenced when portraying STAR*D's findings. For example, in 2009 NIMH's Director Dr. Thomas Insel
24 claimed STAR*D found "at the end of 12 months, with up to four treatment steps, roughly 70% of
25 participants were in remission." [13, p.1466]. Similarly in 2013, an editorial in the *American Journal of*
26 *Psychiatry (AJP)* claimed STAR*D found "after four optimized, well-delivered treatments, approximately
27 70% of patients achieve remission." [14, p.580]. These are not factual statements of STAR*D's findings.
28

29
30 The first author has made published criticisms alleging protocol violations that appear to inflate
31 STAR*D's findings and called for the reanalysis of the dataset by independent investigators.[15] In 2018,
32 the first and fourth authors collaborated with researchers from the University of Connecticut to
33 reanalyze STAR*D's level 1 data obtained from NIMH.[16] This reanalysis found substantial inflation of
34 STAR*D's reported remission and response rates. Furthermore, the reanalysis found that the extent of
35 HRSD improvement in STAR*D's level 1 trial was approximately half that of open-label antidepressant
36 comparator trials.
37

38
39 Our published criticisms of STAR*D investigators' report of outcomes are as follows:[17]
40

- 41 • Using the QIDS-SR as the secondary outcome to report remission rates and sole measure to
42 report response rates in STAR*D's level 1-4 articles without the investigators disclosing that the
43 protocol specifically excluded non-blinded/clinic-administered assessments such as the QIDS-SR
44 from use as outcome measures.[1-6] While STAR*D investigators used the used HRSD to report
45 remission rates in their levels 1-4 articles,[1-6] the QIDS-SR was used as the sole measure to
46 report remission, response, and extent of improvement rates in their summary article.[7] The
47 primary outcome measure, the HRSD, should have been used to report the summary article's
48 outcomes.
49
- 50 • Using data from the 931 patients deemed ineligible for analysis in STAR*D's level 1 article
51 because these patients lacked a baseline ROA-administered HRSD score of ≥ 14 , in STAR*D's
52 levels 2-4 and summary articles without clear disclosure. This included 99 patients who scored
53 < 8 on their baseline HRSD—indicating these patients met STAR*D's remission criterion at study
54 outset and should not have been included in their report of outcomes.
55

- Excluding from analysis 370 patients who dropped out after starting on citalopram in their first clinic visit without taking the exit HRSD despite STAR*D investigators stating, “our primary analyses classified patients with missing exit HRSD scores as nonremitters a priori.”[1, p.34] These 370 early dropout patients should have been counted as nonremitters as prespecified.
- Including in their analyses 125 patients who scored as remitted at entry into their next-level treatment. This occurred despite STAR*D investigators prespecifying that, “patients who begin a level with HRSD <8 will be excluded from analyses.”[8, p.130]

This article reanalyzes STAR*D’s treatment remission, response, and extent of improvement after up to four trials of antidepressant therapies, using STAR*D’s protocol-specified primary outcome measure, the ROA-administered HRSD. This effort builds on Pigott et al’s 2010 article [11] that focused on deconstructing STAR*D investigators’ levels 1-4 and summary articles,[1-7] by analyzing STAR*D’s patient-level dataset obtained from NIMH in 2019 with fidelity to the original research protocol and inclusion in data analysis criteria. Subsequent efforts will focus on reanalyzing STAR*D’s levels 2-4 semi-randomized comparator trials including the extent of emergent suicidal ideation and 12-month follow-up outcomes tied to each compared treatment.

Method

RIAT Initiative

The Restoring Invisible and Abandoned Trials (RIAT) initiative started in 2013 calling on funders and investigators of abandoned (unpublished) or misreported studies to publish undisclosed outcomes or correct misleading publications.[18] If investigators failed to correct a study identified as misreported, independent investigators were encouraged to correct the record by reanalyzing the study’s patient-level dataset consistent with the research protocol and analytic plan.

On March 6, 2019, the RIAT investigators published our response to a ‘Call to Action’ statement in the *British Medical Journal*, in which we stated our intention to reanalyze the STAR*D dataset.[17] We then notified STAR*D’s principal investigators of our intention and requested they inform us whether they would undertake a reanalysis of the dataset adhering to the research protocol. On March 22, 2019, STAR*D investigators acknowledged our email notification, indicated the STAR*D data were in the public domain, and stated they had no interest in undertaking a reanalysis.

In July 2019, we received a STAR*D Data Use Certificate, issued by the NIMH Data Archive Data Access Committee, and gained access to the STAR*D levels 1-4 and follow-up patient-level dataset consisting of 26 text files, and limited supporting study documentation. In September 2019, we obtained funding from the RIAT Support Center to reanalyze STAR*D.

Patients

STAR*D patients were 18 to 75 years of age, seeking care at 18 primary and 23 psychiatric care clinics. Clinical research coordinators (CRCs) screened 4,790 patients for MDD. This screening included the CRCs’ administering the HRSD, on which 4,041 patients scored ≥ 14 , met the other inclusion criteria, and enrolled into the study. CRCs also gathered patients’ psychiatric history, demographic information, and administered both the Cumulative Illness Rating Scale and the Psychiatric Diagnostic Screening Questionnaire to determine the extent of comorbid medical and psychiatric disorders.

Levels/Steps of Acute Treatment

STAR*D investigators sought to provide the highest quality of care to maximize the number of remissions while minimizing dropouts (see appendix 1). Appendix 2 describes the antidepressant therapies available in treatment levels 1-4 while steps refer to the numeric order of treatments. As seen in figure 1, treatment steps 1 and 2 correspond to level 1 and 2 treatments. Similarly, for most patients their level 3 and 4 treatments correspond to treatment steps 3 and 4. For level/step 2 patients though who failed to respond adequately to cognitive therapy alone or combined with citalopram and chose to continue in the study, their third treatment step was designated level 2A and they were randomized to one of two level 2 switch medications. For these patients, their level 2A treatment was their third treatment step. For level 2A patients who did not adequately benefit from this medication trial and chose to continue in the study, they entered a fourth treatment step consisting of level 3 treatments.

All patients were administered the SSRI citalopram for their level 1 treatment. Each treatment level consisted of 12 weeks of antidepressant therapy, with an additional 2 weeks for patients deemed close to remission. Treatment was administered using a system of measurement-based care that assessed symptoms and side effects at each clinic visit. STAR*D investigators state, *"To enhance the quality and consistency of care, physicians used the clinical decision support system that relied on the measurement of symptoms (QIDS-C and QIDS-SR), side-effects, medication adherence, and clinical judgment based on patient progress."*[1, p.30] This system was used to guide medication management of a fully adequate dose for a sufficient time 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]

For those patients who failed to gain an adequate response from citalopram, STAR*D allowed them to select acceptable treatment options for randomization in levels 2 to 4 *"to empower patients, strengthen the therapeutic alliance, optimize treatment adherence, and improve outcome"*[19, p.483]. The treatment options available for randomization involved either switching to a new treatment or augmenting the patient's current treatment. Treatment levels 2 to 4 evaluated the relative effectiveness of 11 pharmacologically distinct drug/drug combination treatments. Cognitive therapy was also available as either a switch or citalopram augmentation option in level 2.

*STAR*D Follow Up Phase*

In each treatment trial for levels 1-4, patients who scored <6 on their last QIDS-Clinician version (QIDS-C) were considered clinician-rated remissions and encouraged to enter the 12-month follow-up phase. During follow-up, patients continued their "previously effective acute treatment medication(s) at the doses used in acute treatment but that any psychotherapy, medication, or medication dose change could be used." [7, p.1908] Based on prior research, a QIDS score of <6 was estimated by STAR*D investigators to correspond to a score of <8 on the HRSD, STAR*D's prespecified primary outcome measure for classifying patients as remitted.[9] Clinicians strongly encouraged patients who did not obtain a QIDS-defined remission to enter the next-level treatment. Patients who failed to attain a QIDS-defined remission, but did have a $\geq 50\%$ reduction on the QIDS-C and did not want to be randomized to a next-level treatment, were also encouraged to enter follow-up.

*Research Design of the STAR*D Study*

1
2
3 STAR*D investigators developed a new research design for the study termed “equipoise-stratified” to
4 evaluate the relative efficacy of 13 antidepressant therapies in levels 2-4 for depressed patients who
5 failed to gain adequate benefit from their level 1 medication trial.[20]. In level 1, all patients received
6 citalopram as their first treatment. In level 2, patients were informed regarding seven treatment
7 options to choose from: four switch options in which citalopram was stopped and the new treatment
8 initiated and three augmentation options in which citalopram was combined with a second
9 antidepressant treatment. In level 3, patients were informed regarding four treatment options to
10 choose from: two switch options and two augmentation options. Level 4 involved randomization to one
11 of two medication/medication combination switch options.
12
13

14 *Analytic Plan of the RIAT Reanalysis*

15

16 We reanalyzed the STAR*D dataset with fidelity to the original research protocol wherever possible.
17 Where the protocol was silent, we used other STAR*D publications to guide our analysis. This occurred
18 four times. First, the protocol is silent regarding patients who entered the study without a baseline
19 ROA-administered HRSD score of ≥ 14 . In their level 1 article, STAR*D investigators deemed the 931 such
20 patients who lacked this marker of depression severity ineligible for inclusion in data analysis.[1] We do
21 the same and extend this exclusion for such patients who continued on to levels 2-4 because their
22 extent of depression severity at study outset is not known. Second, the protocol is silent on what to do
23 with patients who met the remission criteria on the HRSD at entry into their next-level treatment. In
24 STAR*D’s Rationale and Research Design article though, its investigators prespecify that “patients who
25 begin a level with HRSD < 8 will be excluded from analyses.”[8, p.130] We therefore excluded 125 such
26 patients from our analyses of treatment levels 2-4. Third, the protocol is silent on how to analyze
27 patients who exit a treatment without taking the HRSD. STAR*D investigators state in their level 1
28 article, “our primary analyses classified patients with missing exit HRSD scores as nonremitters a
29 priori”[1, p.34] and repeat similar statements in their level 2-4 articles.[2-6] Therefore, we do likewise.
30
31
32

33 Finally, STAR*D had many patients with missing exit HRSD scores. In their level 2-4 articles, STAR*D
34 investigators mapped the final QIDS-SR score to the HRSD for patients missing their exit HRSD score to
35 assess the impact of their approach to counting such patients as “nonremitters a priori.” We therefore
36 calculated STAR*D’s remission rate both as prespecified based on an exit HRSD score of < 8 as well as a
37 final QIDS-SR score of < 6 for those patients missing an exit HRSD score.
38
39

40 All pre-processing and analyses were performed in R.[21] Authors 2 and 3 identified patients by their
41 subject key and used this variable to match information across datasets. Data on patients’ treatment
42 pathways, and when patients transitioned from one level to the next, were taken from the IVRA dataset
43 completed by CRCs, and verified against the data on patient level exits. Authors 2 and 3 then compared
44 the number of patients identified for all level 1-4 treatments to that reported in the STAR*D summary
45 article’s patient flowchart, and the Ns matched.[7, figure 1]
46
47

48 Next, authors 2 and 3 applied STAR*D’s level 1 inclusion in data analysis criterion to patients in
49 treatment levels 2-4 as well as excluded from analysis the 125 patients who scored < 8 on the HRSD at
50 entry into their next-level treatment. We counted these 125 patients as remitted in the prior treatment
51 level but excluded them from the analyses of subsequent treatments. Appendix 3 presents the number
52 of level 2-4 patients excluded from our reanalysis, and the reasons for their exclusion. Appendix 4 is a
53 table identifying the number of patients with missing entry and/or exit HRSD scores for all level 1-4
54 treatments and used in our Last Observation Carried Forward (LOCF) analyses. As seen in Appendix 4,
55 1,330 patients were missing their exit HRSD score across all treatments.
56
57
58
59
60

We then compared STAR*D's outcomes to those found in a meta-analysis of 7,030 patients enrolled in antidepressant comparator trials.[22] Similar to STAR*D, comparator trials typically are conducted open-label without a control group and therefore are the appropriate comparison data for STAR*D's outcomes. Continuous HRSD improvement means were provided by the first author of the meta-analysis.[22]

Finally, we compared the STAR*D protocol's step-by-step predictions of patient drop out and the number of patients who would have a satisfactory treatment response and enter follow-up care to what actually occurred.[10] While these predictions' purpose was to estimate the number of continuing patients available for randomization in treatment levels 2-4, at the meta-level these predictions are an important hypothesis STAR*D tested by assessing how well its investigators could predict the aggregate step-by-step successful treatment outcomes from their treat-to-remission model of care.

Patient and Public Involvement

Neither patients nor the public were involved in the design, conduct, reporting, or dissemination plans of our research.

Results

Figure 1 presents the overall flow of patients enrolled in the various protocol-defined treatment levels and places them in groups defined by the number of treatment steps. Of the 4,041 patients enrolled into STAR*D, 3,110 met the eligibility for data analysis criterion of having a ROA-administered HRSD score ≥ 14 at study outset. Figure 1 also identifies the number of patients who exited the study following each treatment step, the number who entered follow-up after each treatment step, and the number who were randomly assigned to a next-level treatment.

Appendix 5 describes the demographic and clinical features of the patients who entered treatment in steps 1-4 based on their level 1 baseline presentation when enrolling into the study. Summary statistics are presented as means and standard deviations for continuous variables and percentages for discrete variables. Note that 55.7% of STAR*D patients had 2 or more comorbid axis 1 disorders when first enrolled based on the Psychiatric Diagnostic Screening Questionnaire and averaged 2.5 comorbid medical conditions based on the Cumulative Illness Rating Scale. Furthermore, the average length of patients' current depressive episode was 25.9 months. In a post hoc analysis, STAR*D investigators found that 77.8% of its enrolled patients would have been excluded from most antidepressant trials due to having two or more concurrent medical conditions, more than one comorbid psychiatric disorder, and/or a current depressive episode lasting > 2 years.[23]

Using LOCF, Table 1 presents the HRSD entry, exit, and mean change scores for patients by the specific treatment they received in steps 1-4 as well as the HRSD remission and response rates. Table 1 also provides the HRSD cumulative remission rate after up to 4 trials on antidepressant therapies as well as the combined HRSD plus QIDS-SR remission rate for the 1,330 patients missing an exit HRSD score.

Table 1: Outcomes Across All Treatments

Treatment Step	HRSD Score		HRSD Mean Change [95% confidence interval] (SD)	HRSD Remissions # (%)	# QIDS-SR Remissions for Patients with Missing HRSD	Combined HRSD Plus QIDS-SR Remissions # (%)	HRSD Response Rate # (%)
	Entry ^a Mean (SD)	Exit ^a Mean (SD)					
Step 1 (N=3,110)	21.87 (5.21)	15.46 (9.03)	6.41 [6.13, 6.70] (8.03)	794 (25.5%)	144	938 (30.2%)	1006 (32.3%)
Step 2 (N=1,134)	19.5 (6.16)	15.34 (8.46)	4.16 [3.75, 4.57] (6.97)	241 (21.3%)	43	284 (25%)	287 (25.3%)
Switch strategy (N=620)	20.44 (6.01)	16.17 (8.29)	4.27 [3.73, 4.81] (6.89)	113 (18.2%)	22	135 (21.8%)	152 (24.5%)
Bupropion(N=190)	20.75 (6.17)	17.03 (8.39)	3.72 [2.85, 4.58] (6.10)	31 (16.3%)	6	37 (19.5%)	41 (21.6%)
Sertraline (N=198)	20.60 (6.10)	16.30 (8.25)	4.30 [3.33, 5.26] (6.93)	32 (16.2%)	4	36 (18.2%)	48 (24.2%)
Venlafaxine (N=192)	20.36 (5.91)	15.56 (8.21)	4.80 [3.77, 5.82] (7.26)	37 (19.3%)	8	45 (23.4%)	50 (26.0%)
Cognitive Therapy (N=40)	18.52 (5.06)	14.30 (8.20)	4.22 [1.63, 6.82] (8.36)	13 (32.5%)	4	17 (42.5%)	13 (32.5%)
Augmentation strategy (N=514)	18.37 (6.15)	14.34 (8.55)	4.03 [3.42, 4.64] (7.08)	128 (24.9%)	21	149 (29%)	135 (26.3%)
Bupropion (N=216)	17.94 (6.18)	13.71 (8.33)	4.22 [3.29, 5.15] (6.97)	54 (25.0%)	10	64 (29.6%)	61 (28.2%)
Bupirone (N=225)	18.74 (6.46)	14.85 (8.95)	3.89 [2.97, 4.82] (7.07)	58 (25.8%)	10	68 (30.2%)	56 (24.9%)
Cognitive Therapy (N=73)	18.52 (4.99)	14.64 (7.91)	3.88 [2.16, 5.59] (7.48)	16 (21.9%)	1	17 (23.3%)	18 (24.7%)
Step 3 (N=325)	19.86 (6.14)	17.29 (7.78)	2.57 [1.92, 3.22] (6.00)	43 (13.2%)	7	50 (15.4%)	52 (16.0%)
Level 2A (N=28)	20.71 (5.56)	17.75 (6.34)	2.96 [0.91, 5.02] (5.55)	3 (10.7%)	0	3 (10.7%)	4 (14.3%)
Bupropion (N=12)	19.50 (4.15)	17.75 (7.36)	1.75 [-1.65, 5.15]	2 (16.7%)	0	2 (16.7%)	2 (16.7%)

			(6.00)				
Venlafaxine (N=16)	21.62 (6.41)	17.75 (5.71)	3.88 [1.33, 6.42] (5.20)	1 (6.3%)	0	1 (6.3%)	2 (12.5%)
Level 3 (N=297)	19.78 (6.19)	17.25 (7.91)	2.53 [1.84, 3.22] (6.05)	40 (13.5%)	7	47 (15.8%)	48 (16.2%)
Switch strategy (N=186)	20.28 (6.25)	17.80 (8.04)	2.49 [1.57, 3.41] (6.39)	23 (12.4%)	2	25 (13.4%)	27 (14.5%)
Nortriptyline (N=92)	20.04 (5.55)	17.87 (8.51)	2.17 [0.72, 3.63] (7.12)	15 (16.3%)	0	15 (16.3%)	15 (16.3%)
Mirtazapine (N=94)	20.52 (6.89)	17.72 (7.60)	2.80 [1.67, 3.93] (5.60)	8 (8.5%)	2	10 (10.6%)	12 (12.8%)
Augmentation strategy (N=111)	18.93 (6.01)	16.32 (7.64)	2.60 [1.59, 3.62] (5.46)	17 (15.3%)	5	22 (19.8)	21 (18.9%)
Lithium (N=58)	19.22 (6.81)	16.98 (7.31)	2.24 [1.06, 3.42] (4.59)	7 (12.1%)	2	9 (15.5%)	7 (12.1%)
T3 (N=53)	18.60 (5.05)	15.60 (7.99)	3.00 [1.31, 4.69] (6.29)	10 (18.9%)	3	13 (24.5%)	14 (26.4%)
Step 4 (N=106)	20.84 (5.36)	17.44 (7.25)	3.40 [2.19, 4.60] (6.32)	11 (10.4%)	1	12 (11.3%)	17 (16.0%)
Level 3 (N=16)	20.56 (4.03)	17.62 (6.37)	2.94 [-0.12, 5.99] (6.23)	2 (12.5%)	0	2 (12.5%)	2 (12.5%)
Tranlycypromine (N=43)	21.40 (6.02)	17.67 (7.12)	3.72 [1.69, 5.75] (6.78)	3 (7.0%)	1	4 (9.3%)	8 (18.6%)
Venlafaxine XR/mirtazapine (N=47)	20.43 (5.17)	17.17 (7.78)	3.26 [1.53, 4.98] (6.03)	6 (12.8%)	0	6 (12.8%)	7 (14.9%)
Cumulative Remission Rate after up to four Antidepressant Therapies				1,089 (35.0%)		1,284 (41.3%)	

^a Used Last Observation Carried Forward for missing HRSD values.

Table 2 presents patients' aggregate HRSD status in terms of remission, response, and extent of symptomatic change at entry and exit for each treatment step as well as study dropout. In step 1, 25.5% of patients remitted. Steps 2-4 show a continuous decrease in remission rates from step 2's 21.3% to 13.2% for step 3 and 10.4% in step 4 with increasing rates of study dropout from step 1's 34.5% to step 3's 46.2%.

Table 2: Outcomes by Treatment Step

	Step 1 (N=3,110)		Step 2 (N=1,134)		Step 3 (N=325)		Step 4 (N=106)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
HRSD score at entry into step*	21.87	5.21	19.5	6.16	19.86	6.14	20.84	5.36
HRSD score at exit from step*	15.46	9.03	15.34	8.46	17.29	7.78	17.44	7.25
HRSD Mean Change	6.41	8.03	4.16	6.97	2.57	6.0	3.40	6.32
	N	%	N	%	N	%	N	%
Remission at each step exit	794	25.5%	241	21.3%	43	13.2%	11	10.4%
Response at each step exit	1006	32.3%	287	25.3%	52	16.0%	17	16.0%
Entered Follow-up	902	29.0%	406	35.8%	69	21.2%	38	35.9%
Study Exit/Dropout	1,074	34.5%	403	35.5%	150	46.2%		

* Last observation carried forward

Figures 2 and 3 compare the HRSD remission, response, and extent of symptom improvement rates for STAR*D patients in steps 1-4 to that found in a meta-analysis of 7,030 patients enrolled in non-blinded antidepressant comparator trials.[22] In step 1, these measures of improvement among STAR*D's patients were approximately half that found in comparator trials, and improvement grew progressively worse in each subsequent treatment episode.

Appendix 6 presents a figure comparing the STAR*D protocol's predictions of patient dropout and the number of patients who would have a satisfactory treatment response and enter follow-up to what occurred. Cumulatively, STAR*D's investigators predicted that 73.8% of patients would have a successful treatment response and enter follow-up whereas in fact only 45.6% achieved this measure of treatment success. Furthermore, whereas its investigators predicted that over the course of up to four antidepressant therapies 20.7% of patients would dropout, in fact, 53.7% dropped out. On this measure of treatment failure, STAR*D's dropout rate was 2.6 times greater than predicted.

Figure 4 presents the step-by-step cumulative remission rate three ways. First, the 'theoretical' rate propagated by STAR*D investigators based on the provisos of what would have happened if there were no study dropouts and that those who did exit had the same QIDS-SR remission rates as those who stayed.[7] Next, the combined HRSD plus QIDS-SR remission rate based on either an exit HRSD score of <8, OR a last clinic visit QIDS-SR score of <6 for the 1,330 patients missing an exit HRSD. Finally, our RIAT reanalysis rate when using the protocol-specified exit HRSD score of <8 as the sole measure of remission for the 3,110 patients who met STAR*D's inclusion in data analysis criteria. The cumulative remission rate after up to four antidepressant therapies using the HRSD was 35% versus 41.3% when combined with the QIDS-SR, both of which are substantially less than the 67% cumulative remission rate claimed in the summary article's Abstract.

Discussion

*Principal findings and comparison with original STAR*D publication*

1
2
3 STAR*D's results highlight the discrepancy in likely outcomes between typical antidepressant clinical
4 trials with their exclusion criteria and the real-world patients for whom these medications are
5 commonly prescribed. Our RIAT reanalysis found poorer outcomes after up to four optimized, and
6 increasingly aggressive, antidepressant therapies than reported in STAR*D's summary article published
7 in *AJP*.^[7] In contrast to the 67% cumulative remission rate reported in *AJP*, the actual rate was 35%
8 when using the protocol-specified HRSD, and increased to 41.3% when combined with a final clinic-visit
9 QIDS-SR score of <6 for patients missing exit HRSD scores in treatment steps 1-4. The 41.3% cumulative
10 remission rate should be viewed as the "best case scenario" since it added an additional 195 QIDS-
11 defined remissions (a remission measure not specified in the protocol) from the 1,330 patients with
12 missing exit HRSD scores. As there was neither a placebo nor waitlist control group during any phase of
13 the STAR*D study, it is impossible to know if even the meager results that were observed were due to
14 the pharmacologic effects of the prescribed medications, placebo effects, or the mere passage of time.
15
16

17
18 Our reanalysis did not assess the durability of treatment effects during the 12-month follow up phase.
19 In their summary article though, STAR*D authors reported an overall relapse rate of 46.1% for the 1,729
20 patients for whom they had at least one assessment (of up to 12 scheduled) during follow up using a
21 telephonic-administered version of the QIDS [7] whereas Pigott et al found a far lower sustained
22 recovery rate when incorporating patient dropout in the analysis.^[11]
23

24 *Comparison with other studies*

25
26 Our reanalysis found that in step 1, STAR*D's remission, response, and extent of improvement rates
27 were only about half of those reported in other open-label antidepressant comparator trials and then
28 grew progressively worse in steps 2-4.^[22] Such studies typically exclude depressed patients with the
29 range and number of comorbid medical and/or psychiatric disorders that were included in STAR*D.
30
31

32 STAR*D's step 1 remission rate was 25.5% followed by a progressive decline in remission rates for those
33 patients receiving subsequent, and increasingly aggressive treatments, such that by step 4 it was only
34 10.4%. This decline in antidepressant medications' effectiveness essentially mirrors the findings from
35 randomized and naturalistic, prospective studies reporting a 20-30% loss of effectiveness with each
36 increase in the number of prior antidepressant trials.^[24-29] Furthermore, several recent analyses
37 suggest that the sequential application of antidepressant medications for non-remitting depression may
38 in fact foster treatment resistance for many patients.^[30-33]
39
40

41 Regarding the protocol's predictions of treatment success and patient dropout, it states:

42
43 We arrived at these estimates using three experienced practitioners who
44 independently made estimates that were surprisingly close to each other. Then, via
45 teleconferencing, the final estimates were made. ***The underlying assumptions of***
46 ***these estimates come largely by inferences from results of published RCTs.***^{[10, p.31;}
47 ***emphasis added]***
48
49

50 STAR*D's actual measures of treatment success and failure were significantly worse than predicted. As
51 Barbui et al. noted, antidepressant study dropout rates provide a "hard measure of treatment
52 effectiveness and acceptability"^[12, p.296] and STAR*D's dropout rate was 2.6 times greater than
53 predicted. This discrepancy further highlights the relative ineffectiveness of antidepressants in real-
54 world depressed patients, compared to those reported in conventional studies. Consequently, the claim
55
56
57
58
59
60

1
2
3 that antidepressants, with their suspect efficacy,[34,35] work better in real-world clinical practice is not
4 supported by the STAR*D study when its patient-level data are analyzed as per protocol.
5

6 *Conclusion*

7
8 Bias in the clinical literature is commonly associated with industry-funded RCTs, not publicly funded
9 ones.[36] Our RIAT reanalysis though documents numerous scientific errors in this NIMH-funded study.
10 These errors inflated STAR*D investigators' report of positive outcomes, taking a failed study and
11 portraying it as positive.
12
13

14 The STAR*D summary article's claim of a 67% cumulative remission rate was published in 2006. If
15 STAR*D's outcomes had been reported as prespecified, its measurement-based treat-to-remission
16 model of care would likely have faced much stronger criticism 16 years ago and fueled a more vigorous
17 search for evidence-based treatment alternatives.
18
19

20 **Ethics Statement:**

21
22 The study was overseen by NIMH; NCT00021528.
23

24 **Data Availability Statement:**

25
26 Data is available from the NIMH-supported National Database for Clinical Trials (NDCT).
27
28

29 **Funding Statement:**

30
31 Funding for this project was provided by The RIAT Support Center.
32
33

34 **Competing Interests:**

35
36 The authors have no relevant financial disclosures or other conflicts of interest to report.
37
38

39 **Contributorship Statement:**

40
41 HEP, JDA and IK contributed to the design of the study and secured funding. TK and CX conducted all of
42 the data analyses. HEP wrote the manuscript with input from JDA, IK, TK, and CX.
43
44

45 **Acknowledgements:**

46
47 We thank Termeh Feinberg for her early efforts on this project, particularly her correspondence with the
48 NIMH help desk to resolve issues with the 26 data files as well as the RIAT Support Center for funding
49 this project. Data used in the preparation of this manuscript were obtained from the controlled access
50 datasets distributed from the NIMH-supported National Database for Clinical Trials (NDCT). NDCT is a
51 collaborative informatics system created by the National Institute of Mental Health to provide a national
52 resource to support and accelerate discovery related to clinical trial research in mental health. The
53 content of this publication does not necessarily reflect the views of the RIAT Support Center nor NIMH.
54

55 **References:**

1. Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, Ritz L, Norquist G, Howland RH, Lebowitz B, McGarth PJ, Shores-Wilson K, Biggs MM, Balasubramani GK, Fava M. 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. 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. Bupropion-SR, sertraline, or venlafaxine-XR after failure of SSRIs for depression. *N Engl J Med*. 2006;354:1231–1242.
3. 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. Medication augmentation after the failure of SSRIs for depression. *N Engl J Med*. 2006;354:1243–1252.
4. Fava M, Rush AJ, Wisniewski SR, Nierenberg AA, Alpert JE, McGrath PJ, Thase ME, Warden D, Biggs MM, Luther JF, Niederehe G, Ritz L, Trivedi MH. 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.
5. Nierenberg AA, Fava M, Trivedi MH, Wisniewski SR, Thase ME, McGrath PJ, Alpert JE, Warden D, Luther JF, Niederehe G, Lebowitz BD, Shores-Wilson K, Rush AJ. 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.
6. McGrath PJ, Stewart JW, Fava M, Trivedi MH, Wisniewski SR, Nierenberg AA, Thase ME, Davis L, Biggs MM, Shores-Wilson K, Luther JF, Niederehe G, Warden D, Rush AJ. Tranylcypromine versus venlafaxine plus mirtazapine following three failed antidepressant medication trials for depression: a STAR*D report. *Am J Psychiatry*. 2006;163:1531–1541.
7. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, Niederehe G, Thase ME, Lavori PW, Lebowitz BD, McGrath PJ, Rosenbaum JF, Sackheim HA, Kupfer DJ, Luther J, Fava M. 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.
8. Rush AJ, Fava M, Wisniewski SR, Lavori PW, Trivedi MH, Sackheim 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. Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design. *Control Clin Trials*. 2004;25:119–142.
9. Rush AJ, Trivedi MH, Ibrahim HM, Carmody T J, Arnow B, Klein DN, ... Manber R. The 16-item Quick Inventory of Depressive Symptomatology (QIDS), Clinician Rating (QIDS-C), and Self-Report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biological Psychiatry*. 2003;54:573–583.
10. National Institute of Mental Health (NIMH). Sequenced Treatment Alternatives to Relieve Depression (STAR*D) Research Protocol. Washington (DC): NIMH; revised June 28, 2002.
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. Pigott HE, Leventhal AM, Alter GS, Boren JJ. Efficacy and effectiveness of antidepressants: current status of research. *Psychother Psychosom*. 2010;79:267–279.
13. Insel TR, Wang PS. The STAR*D trial: revealing the need for better treatments. *Psychiatric Services*. 2009;60:1466–1467.
14. Greden JF. Workplace depression: personalize, partner, or pay the price. *Am J Psychiatry*. 2013;170:578–581.

15. Pigott HE. The STAR*D trial: it's time to reexamine the clinical beliefs which guide the treatment of major depression. *Canadian J Psychiatry*. 2015;60:9–13.
16. Kirsch I, Huedo-Medina TB, Pigott HE, Johnson B. Do outcomes of clinical trials resemble those “real world” patients? A re-analysis of STAR*D antidepressant data. *Psychology of Consciousness: Theory, Research, and Practice*. 2018 <https://doi.org/10.1037/cns0000164>
17. Pigott HE, Dubin, W, Kirsch I, Amsterdam J. Call to action: RIAT reanalysis of the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) Study. *BMJ*. March 6, 2019; <https://www.bmj.com/content/346/bmj.f2865/rr-10>
18. Doshi P, Dickersin K, Healy D, Vedula SS, Jefferson T. Restoring invisible and abandoned trials: a call for people to publish the findings. *BMJ* 2013;346:f2865.
19. 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.
20. Lavori PW, Rush AJ, Wisniewski SR, Alpert J, Fava M, Kupfer DJ, Nierenberg A, Quitkin FM, Sackeim HA, Thase ME, Trivedi M. Strengthening clinical effectiveness trials: equipose-stratified randomization. *Biological Psychiatry*. 2001;50:792– 801.
21. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
22. Rutherford BR, Sneed JR, Roose SP. Does study design influence outcome? The effects of placebo control and treatment duration in antidepressant trials. *Psychotherapy and Psychosomatics*. 2009;78:172–181.
23. Wisniewski SR, Rush AJ, Nierenberg AA, Gaynes BN, Warden D, Luther JF, McGrath PJ, Lavori PW, Thase ME, Fava M, Trivedi MH: Can phase III trial results of antidepressant medications be generalized to clinical practice? A STAR * D report. *Am J Psychiatry* 2009;166:599–607.
24. Amsterdam JD, Maislin G. Fluoxetine efficacy in treatment resistant depression. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*. 1994;18:243-261.
25. Nierenberg A, Feighner JP, Rudolph R, Cole JO, Sullivan J. Venlafaxine for treatment-resistant unipolar depression. *J Clinical Psychopharmacology*. 1994;14:419-423.
26. Amsterdam JD, Shults J. MAOI safety and efficacy in advanced treatment-resistant depression: a retrospective analysis. *J Affective Disorders*. 2005; 89:183-188.
27. Amsterdam JD, Williams D, Michelson D, Adler LA, Dunner SL, Nierenberg A, Reimherr FW, Schatzberg AF. Tachyphylaxis after repeated antidepressant drug exposures in patients with recurrent major depressive disorder. *Neuropsychobiology*. 2009;59:227-233.
28. Leykin Y, Amsterdam JD, DeRubeis RJ, Shelton RC, Hollon SD. Progressive resistance to SSRI therapy but not to cognitive therapy in the treatment of major depression. *J Consulting & Clinical Psychology*. 2007;75:267-276.
29. Amsterdam JD, Lorenzo-Luaces L, DeRubeis RJ. Step-wise loss of antidepressant effectiveness after repeated antidepressant trials in bipolar II depression. *Bipolar Disorders*. 2016;18:563-570.
30. Fava GA, Offidani E. The mechanisms of tolerance in antidepressant action. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*. 2011;35:1593-602.
31. Amsterdam JD, Lorenzo-Luaces L. Increase in pharmacodynamic tolerance after repeated antidepressant trials in treatment-responsive bipolar II depressed subjects: An exploratory study. *Psychiatria Polska*. 2018;52:957–969.
32. Amsterdam JD, Kim TT. Increased risk of depressive relapse during maintenance therapy after repeated antidepressant trials in treatment-responsive subjects. *J Clinical Psychopharmacology*. 2019;39:344-350.
33. Andrews PW, Amsterdam JD. A hormetic approach to understanding antidepressant effectiveness and the development of antidepressant tolerance. *Psychiatria Polska*. 2020;54:1067–1090.

- 1
2
3 34. Munkholm K, Paludan-Müller AS, Boesen K. Considering the methodological limitations in the
4 evidence base of antidepressants for depression: a reanalysis of a network meta-analysis. *BMJ*
5 *Open* 2019;9:e024886. doi:10.1136/bmjopen-2018-024886
6
7 35. Stone MB, Yaseen ZS, Miller BJ, Richardville K, Kalaria SN, Kirsch I: Response to acute
8 monotherapy for major depressive disorder in randomized, placebo controlled trials submitted
9 to the US Food and Drug Administration: individual participant data analysis. *BMJ*.
10 2022;378:e067606 <http://dx.doi.org/10.1136/bmj-2021-067606>
11
12 36. Amsterdam JD, McHenry LB, Jureidini JN. Industry-corrupted psychiatric trials. *Psychiatria*
13 *Polska*. 2017;51:993–1008.
14
15

Figure 1 Caption: Patient Flowchart

Figure 1 Footnote:

16
17
18
19
20 * In level 2, 580 patients were randomized to switch medications, 441 to medication augmentation, and
21 113 to Cognitive Therapy as either a switch or medication augmentation treatment. In level 2A, 28
22 patients were randomized to one of two level 2 switch medications. For step 3/level 3 patients, 186
23 were randomized to medication switch and 111 to medication augmentation. For step 4/level 3
24 patients, 7 were randomized to medication switch and 9 to medication augmentation. For step 4/level 4
25 patients, 90 were randomized to one of two medication/medication combination switch options.
26
27 ** Exit refers to the number of patients who exit the study and do not proceed either to the next
28 treatment level nor enter follow-up.
29
30 *** Follow-up refers to the number of patients who exit a treatment and enter the 12-month follow-up
31 phase.
32

Figure 2 Caption: Remission and Response Rate Comparison between STAR*D and Antidepressant Comparator Trials

Figure 2 Footnote:

33
34
35
36
37
38 The step-by-step theoretical remission rates were obtained from the STAR*D summary article where it
39 states: "The theoretical cumulative remission rate is 67% (37+19+6+5)." [7, p.1910].
40

41
42 The HRSD + QIDS-SR cumulative remission rate was taken from Table 1. It combines the 1,089 patients
43 with an exit HRSD score of <8 with the 195 patients who were missing an exit HRSD score but had a final
44 clinic-visit QIDS-SR score of <6.
45

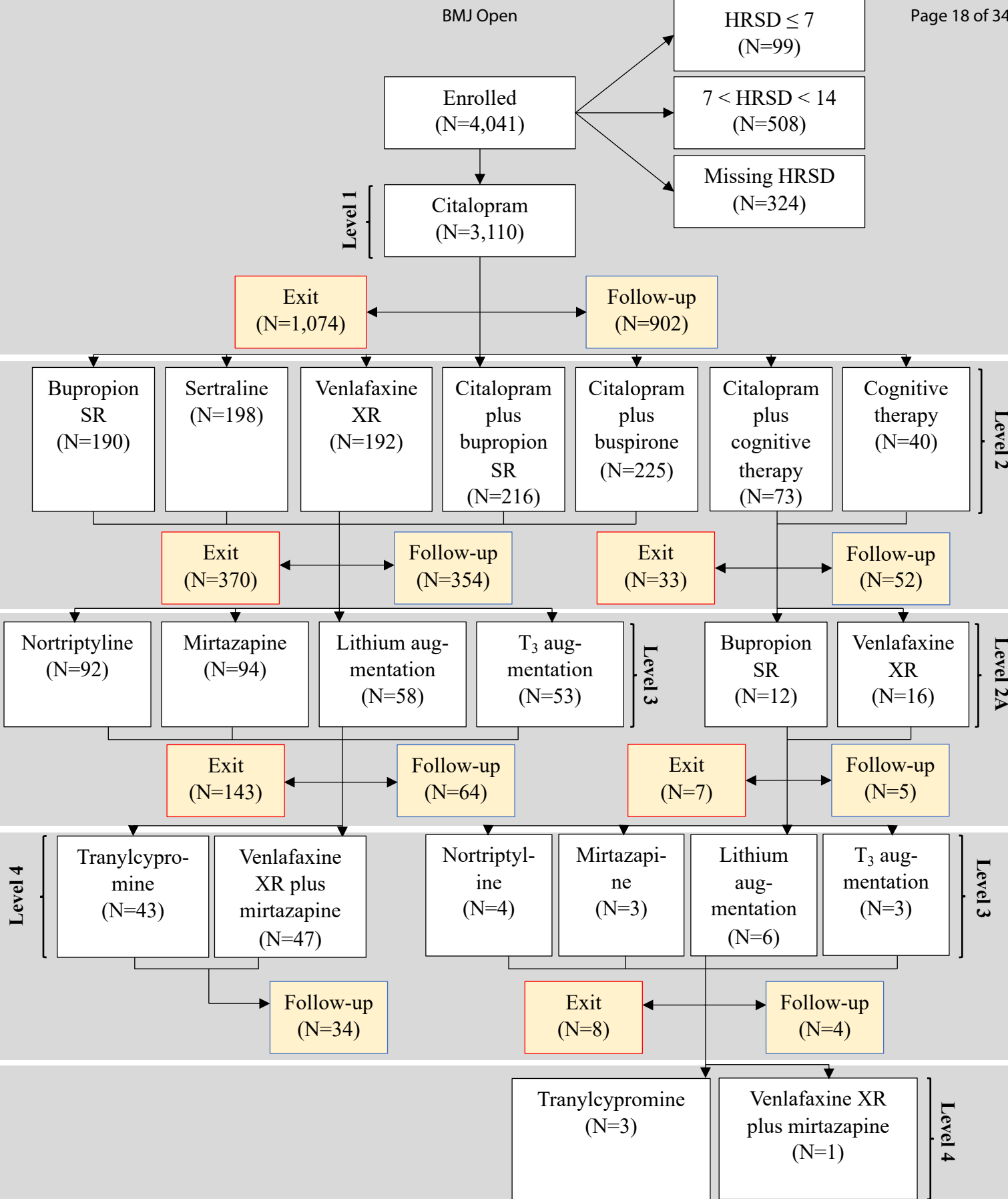
46
47 The RIAT Reanalysis cumulative remission rate is based on an exit HRSD score of <8 as the sole measure
48 of remission for the 3,110 patients who met STAR*D's inclusion for data analysis criteria.
49

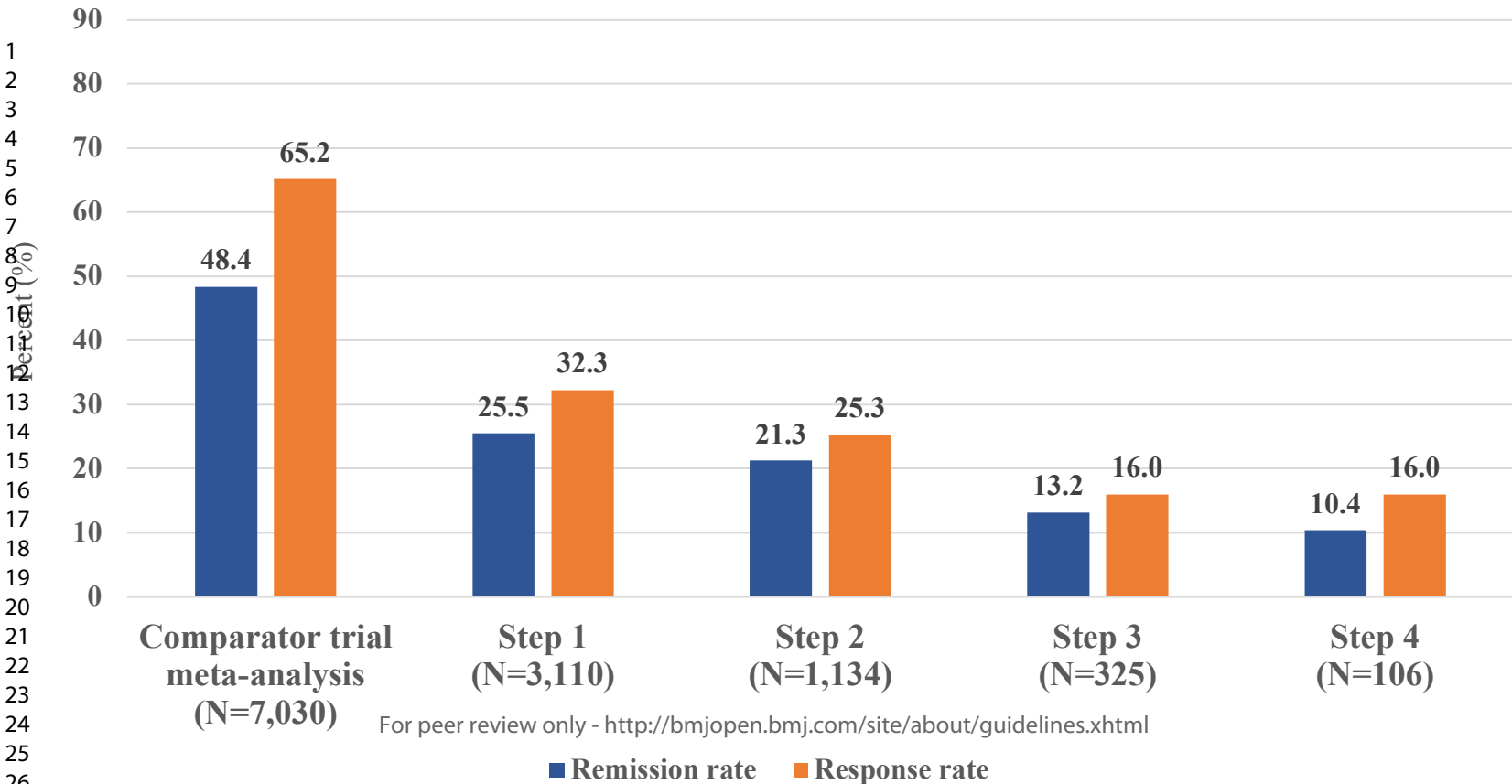
Figure 3 Caption: Comparison in Mean Change HRSD scores between STAR*D and Antidepressant Comparator Trials

Figure 4 Caption: Cumulative Remission Rate Presented Three Ways

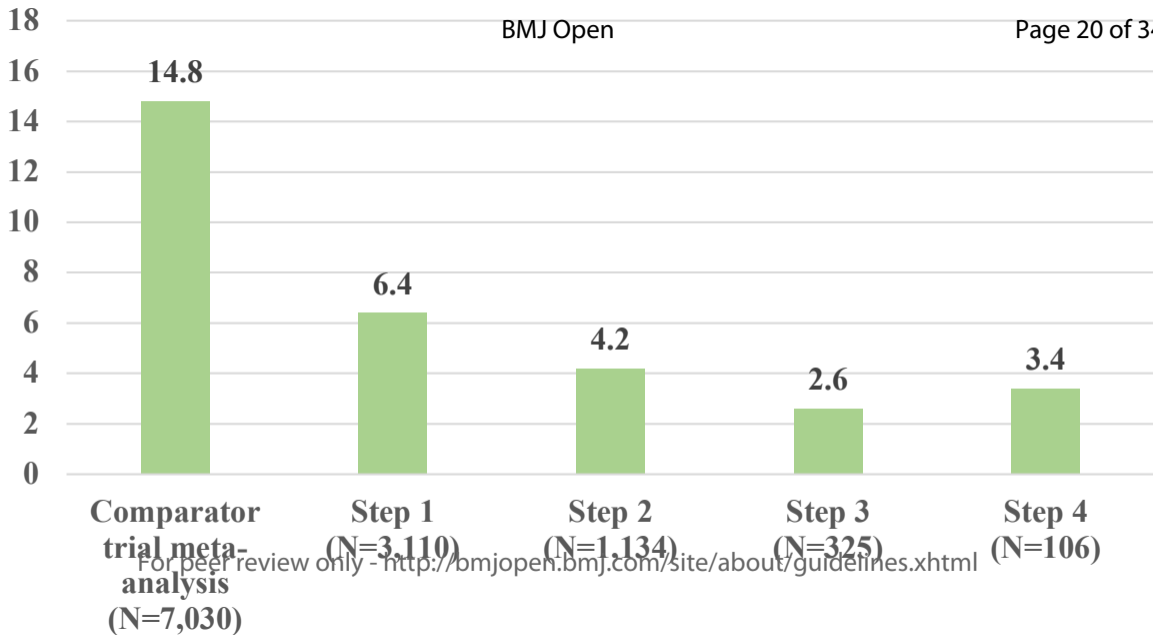
Step 1

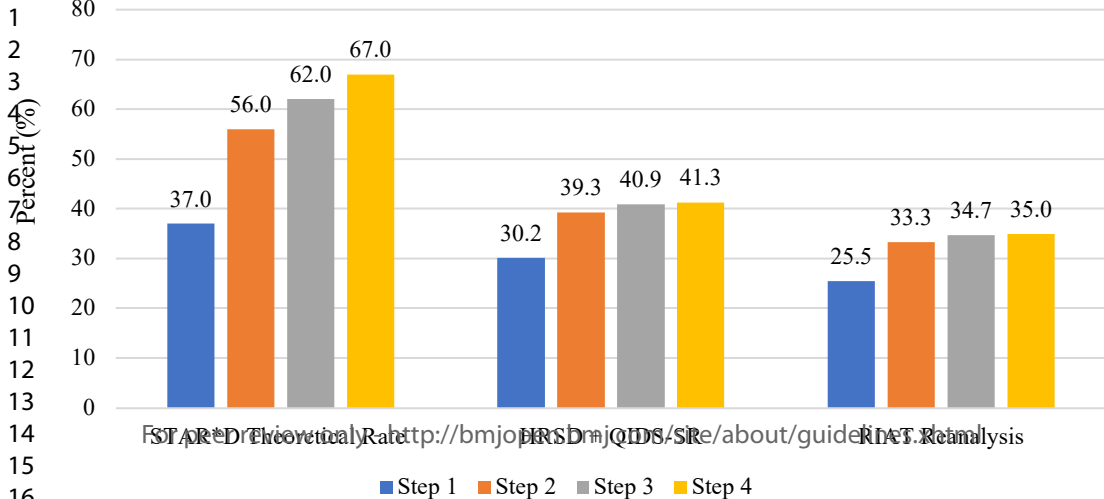
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56





For peer review only - <http://bmjopen.bmj.com/site/about/guidelines.xhtml>

Points of Change (HRSD² scale)For peer review only - <http://bmjopen.bmj.com/site/about/guidelines.xhtml>



Appendix 1: Highest Quality of Acute and Continuing-Care to Maximize Remissions While Minimizing Relapse and Dropouts

Descriptor	Explanation
Optimized Sustained Study Participation to Minimize Dropouts ^{17, p. 473-474}	<ul style="list-style-type: none"> • Promoted patients' study affiliation via STAR*D-branded brochures, bimonthly newsletters, and an informational video emphasizing STAR*D's public health significance and the critical role played by patients; • Educated patients and families about depression and its treatment using a multi-step educational package; • Used a letter reminder system to alert patients before appointments in those clinics without such systems who had a >15% rate of missed appointments; • Ensured timely follow-up and rescheduling of missed appointments by calling patients on the day of the missed appointment, and again within 24 hours, if there was no response. Patient's physician sent letter within 48 hours if contact was not established; • Used a letter reminder system for all research outcome assessment calls during acute and continuing-care; • In every clinic visit, the Clinical Research Coordinator (CRC) discussed the research outcomes phone calls with the patient to ensure that the calls were completed on schedule and worked to resolve any problematic issues regarding said calls [Clinical Procedures Manual, page 75]; • Paid patients \$25.00 for participating in each telephonic research outcomes assessment; • Permitted patients to re-enter acute and/or continuing-care within four weeks after having dropped out [Clinical Procedures Manual, page 80]; • Recommended one-year of continuing-care for all patients who achieved a satisfactory clinical response with the essential goal of preventing relapse [Clinical Procedures Manual, page 15] and • Permitted continuing-care patients to remain in the study if they moved from the area [Clinical Procedures Manual, page 81].
Acute-Care Visits	Physicians met with patients on entry into each new step to initiate drug treatment with follow-up visits scheduled on weeks 2, 4, 6, 9, 12, with an optional week 14 visit.
Measurement-Based Care	Conducted structured evaluations of symptoms and side-effects at each visit and included a centralized treatment monitoring and physician feedback system to ensure consistent implementation of optimal care across research sites.
Aggressive Medication Dosing	Provided aggressive medication dosing with a fully adequate dose for a sufficient duration to <i>“ensure that the likelihood of achieving remission</i>

	<i>was maximized and that those who did not reach remission were truly resistant to the medication”</i> . ^{1, p.30}
Liberal Prescribing of Non-Study Medications	Physicians had great leeway in prescribing non-study medications to treat comorbid symptoms resulting in: <ul style="list-style-type: none"> • 17.2% taking Trazodone for sleep; • 11.9% taking an anti-anxiety medication; • 16.7% taking either a sedative or hypnotic medication; and • An undisclosed percent taking medications to address side-effects.^{2, table 2}
Continuing-Care Visits	Patients saw their physician every 2 months and continued taking their treatment medication(s) at the same doses but their physicians were allowed to make any psychotherapy, medication, and/or medication dose changes to maximize the likelihood of maintaining patients' remission status. ^{7, p. 1908} Additional continuing-care visits were scheduled when patients began to experience a return of depressive symptoms and/or intolerable side-effects [Clinical Procedures Manual. page 78].
Clinical Research Coordinator (CRC)	Each site had a CRC who: ^{1, p. 30} <ul style="list-style-type: none"> • Saw patients before each visit administering multiple measures to them including the QIDS-SR during each acute-care visit; • Assisted physicians in protocol implementation; and • Provided patients support and encouragement in protocol implementation.
Treatment Designed to Enhance Subject Retention	Treatment was designed to minimize drop-outs and/or non-compliance including: <ul style="list-style-type: none"> • Open label prescribing during acute and continuing-care with no placebo control condition during any study phase; • Patients chose their acceptable treatment assignments for steps two and three to eliminate any concerns they might have about receiving an unacceptable assignment. This resulted in only 21 of 1,439 (1.5%) Step-2 patients making themselves available for random assignment to all treatment options^{2, p. 1235} while only 29 of 377 (7.7%) did so in Step-3.^{5, p. 1521} • During each step, patients could enroll immediately into the next step if they had intolerable side-effects or had maximized their current medication(s)' dosing without achieving a remission; and • During any step, patients could enter continuing-care directly on their current medication(s) if they were treatment responders even if they had not achieved remission. This was done to minimize responders from dropping out in order to avoid having to discontinue their current medication(s) and start a new drug regimen.

***Trivedi MH, Stegman D, Rush AJ, Wisniewski SR, Nierenberg AA: STAR * D clinical procedures manual. July 31, 2002. www.edc.pitt.edu/stard/public/study_manuals.html

Appendix 2: Description of Levels 1-4 Treatments

Level 1:

STAR*D investigators report that Citalopram (Celexa) was chosen as the first-line SSRI treatment because (1) absence of discontinuation symptoms; (2) demonstrated safety in elderly and medically fragile patients; (3) easy once-a-day dosing with few dose adjustments; and (4) favorable drug–drug interaction profile.¹ Citalopram was started at 20 mg/day and then raised to 40 mg/day by day 28 and up to 60 mg/day by day 43 and onward. Dose adjustments were based on how long a patient had received a particular dose, symptom changes, and side effect burden.

Level 2 switch treatments:

Citalopram was discontinued without a tapering at the initiation of each level 2 switch treatment. STAR*D investigators chose pharmacologically distinct switch medications.² The level 2 treatments were:

- Sertraline (Zoloft), an SSRI with the same pharmacological profile as citalopram. Sertraline was started at a daily dose of 50 mg and increased to 100 mg at day 8, to 150 mg at day 28, and to 200 mg at day 63 and onward.
- Sustained-release bupropion (Wellbutrin SR), an “out-of-class” agent whose neurochemical action mechanisms are unknown; other than that, it does not inhibit serotonin reuptake and is believed to produce antidepressant effects by blocking the reuptake of dopamine and norepinephrine. The daily dose of sustained-release bupropion was 150 mg for week 1, 200 mg from day 8 to 27, 300 mg from day 28 to 41, and 400 mg from day 42 onward.
- Extended-release venlafaxine (Effexor), a “dual-action” agent that inhibits the reuptake of both serotonin and norepinephrine. The starting daily dose of extended-release venlafaxine was 37.5 mg for week 1 and increased to 75 mg from day 8 to 14, to 150 mg from day 15 to 27, to 225 mg from day 28 to 41, to 300 mg from day 42 to 62, and to 375 mg from day 63 onward.
- Cognitive therapy was provided by a trained psychotherapist and scheduled twice weekly for the first four weeks, then once weekly for the remaining 8 weeks (16 sessions total).

Level 2 Citalopram augmentation treatments:

During the augmentation trial, the citalopram dose was kept constant but reduced if side effects developed. The level 2 augmentation treatments were:

- Buspirone (Buspar), a partial agonist at the postsynaptic 5-hydroxytryptamine 1A (5-HT_{1A}) receptor that is believed to enhance the activity of SSRIs through the 5HT_{1A} receptors. The starting dose was 15 mg per day week 1, increasing to 30 mg per day week 2, and then to 45 mg per day for weeks 3 through 5, and a final, maximum dose of 60 mg per day week 6 and onward.
- Sustained-release bupropion (Wellbutrin SR) whose neurochemical action mechanisms are unknown but is believed to produce antidepressant effects by blocking the reuptake of dopamine and norepinephrine. The initial dose was 200 mg per day during weeks 1 and 2, increasing to 300 mg per day by week 4 and to 400 mg per day week 6 and onward.
- Cognitive therapy was provided by a trained psychotherapist and scheduled twice weekly for the first four weeks, then once weekly for the remaining 8 weeks (16 sessions total).

Level 3 switch treatments:

At entry into the Level 3 switch trial, all level 2 medications were discontinued without tapering at the initial Level 3 treatment visit. The level 3 switch treatments were:

- Nortriptyline (Pamelor), a tricyclic antidepressant. Recommended doses were 25 mg/day for 3 days, 50 mg/day for 4 days, and then 75 mg/day by day 8, 100 mg/day by day 28, and, if necessary, 150 mg/day by day 42 and onward.
- Mirtazapine (Remeron), a tetracyclic antidepressant that blocks inhibitory α_2 -adrenoceptors on norepinephrine and serotonin neurons to enhance both norepinephrine and serotonin neurotransmission. Recommended mirtazapine doses were 15 mg/day for the first 7 days, 30 mg/day by day 8, 45 mg/day by day 28, and, if necessary, 60 mg/day by day 42 and onward.

Level 3 augmentation treatments of level 2 medications:

The two medication augmentation options used in level 2, buspirone and sustained-release bupropion, were discontinued without tapering in the initial level 3 visit. The two medication augmentation treatments in level 3 were added to ongoing treatment with citalopram, sertraline, sustained-release bupropion, or extended-release venlafaxine. The level 3 augmentation treatments were:

- Lithium started at 450 mg/day, and at week 2 it was increased to the recommended dose of 900 mg/day. If participants could not tolerate the initial dose, it could be reduced to 225 mg/day for 1 week then increased to 450 mg/day. There was no monitoring of lithium levels.
- Triiodothyronine (T₃), a thyroid hormone, started at 25 µg/day for 1 week and then increased to the recommended dose of 50 µg/day. There was no pretreatment assessment, nor ongoing monitoring, of thyroid functioning.

Level 4 switch treatments:

The level 4 switch treatments were:

- Tranylcypromine (Parnate), a monoamine oxidase inhibitor. A 2-week washout period of Level 3 medications was required for patients assigned to the tranylcypromine group. The recommended dosing for tranylcypromine was 10 mg/day for the first 2 weeks, followed by weekly increases of 10 mg/day until a maximum of 60 mg/day.
- Co-administered venlafaxine (Effexor) and mirtazapine (Remeron) to inhibit the reuptake of both serotonin and norepinephrine and block inhibitory α -2-adrenoceptors on both norepinephrine and serotonin neurons to enhance both norepinephrine and serotonin neurotransmission. Level 3 medications were discontinued without tapering for patients assigned to this group. The dosage of extended-release venlafaxine was 37.5 mg/day for the first week, 75 mg/day for the second week, 150 mg/day for weeks 3–5, 225 mg/day for weeks 6–8, and 300 mg/day onward. Mirtazapine was started at 15 mg/day for the first 3 weeks, 30 mg/day for weeks 4 to 8, and then 45 mg/day onward.

Appendix 3: Number of Level 2-4 Participants Excluded from our RIAT Reanalysis, and the Reasons for their Exclusion, yet Included in STAR*D

Level 2 Treatments

Number of Level 2 Participants Excluded from our Reanalysis but Included in STAR*D	Bup	Sert	Ven	CT	Cit + BUP	Cit + Busp	Cit + CT	Total
Scored as Remitted at ENTRY into Level 2 yet still included in STAR*D's Level 2 analyses	22	8	14	7	30	24	4	109
Scored as only mildly depressed (HRSD >7 & <14) at entry into Level 1, and therefore excluded from STAR*D's data analysis, yet still treated in Level 1, progressed to Level 2, and then included in STAR*D's Level 2 data analyses	21	15	25	15	20	30	7	133
Scored as Remitted at entry into Level 1 (HRSD ≤ 7), and therefore excluded from STAR*D's data analysis, yet still treated in Level 1 and progressed to Level 2 and then included in STAR*D's Level 2 data analyses	6	1	4	2	2	2	2	19
Missing baseline HRSD at entry into Level 1, and therefore excluded from STAR*D's data analysis, yet still treated in Level 1, and progressed to Level 2, and then included in STAR*D's Level 2 data analyses	12	18	22	4	16	13	1	86
Number meeting 2 exclusion criterions	12	2	7	6	5	8	2	42

Bup=Sustained-release Bupropion; Sert= Sertraline; Ven= Extended-release Venlafaxine; CT=Cognitive Therapy; Cit+BUP= Citalopram + Sustained-release Bupropion; Cit+Busp=Citalopram + Buspirone; Cit+CT= Citalopram + Cognitive Therapy

Level 3 Treatments

	Nortriptyline	Mirtazapine	Lithium Augmentation	Triiodothyronine Augmentation	Total
Scored as Remitted at ENTRY into Level 3 yet still included in STAR*D's Level 3 analyses	4	0	1	5	10
Scored as only mildly depressed (HRSD >7 &	8	5	3	4	20

<14) at entry into Level 1, and therefore excluded from STAR*D's data analysis, yet still treated in Level 1, progressed to Level 2, and then 3 and included in STAR*D's Level 3 data analyses					
Scored as Remitted at entry into Level 1 (HRSD \leq 7), and therefore excluded from STAR*D's data analysis, yet still treated in Level 1 and progressed to Level 2 and then 3 and included in STAR*D's Level 3 data analyses	2	1	0	1	4
Missing baseline HRSD at entry into Level 1, and therefore excluded from STAR*D's data analysis, yet still treated in Level 1, and progressed to Level 2, and then level 3 and included in STAR*D's Level 3 data analyses	7	8	1	7	23

Level 4 Treatments

	Tranlycypromine	Venlafaxine + Mirtazapine	Total
Scored as Remitted at ENTRY into Level 4 yet still included in STAR*D's Level 4 analyses	5	1	6
Scored as only mildly depressed (HRSD $>$ 7 & $<$ 14) at entry into Level 1, and therefore excluded from STAR*D's data analysis, yet still treated in Level 1, progressed to Level 2, 3 and then 4 and included in STAR*D's Level 4 data analyses	3	1	4
Scored as Remitted at entry into Level 1 (HRSD \leq 7), and therefore excluded from STAR*D's data analysis, yet still treated in Level 1 and progressed	0	0	0

to Level 2, then Level 3 and included in STAR*D's Level 4 data analyses			
Missing baseline HRSD at entry into Level 1, and therefore excluded from STAR*D's data analysis, yet still treated in Level 1, and progressed to Level 2, and then Level 3 and 4 included in STAR*D's Level 4 data analyses	5	1	6

For peer review only

Appendix 4: Number and Percent of Participants Missing Entry and/or Exit HRSD Used for Last Observation Carried Forward Analyses

	#/(%) with Missing Entry HRSD	#/(%) with Missing Exit HRSD
Step 1 (N=3,110)	0 (0%)	926 (29.8%)
Step 2 (N=1,134)	168 (14.8%)	304 (26.8%)
Switch strategy (N=620)	90 (14.5%)	183 (29.5%)
Bupropion (N=190)	34	58
Sertraline (N=198)	27	56
Venlafaxine (N=192)	24	56
Cognitive Therapy (N=40)	5	13
Augmentation strategy (N=514)	78 (15.2%)	121 (23.5%)
Bupropion (N=216)	35	58
Bupirone (N=225)	37	52
Cognitive Therapy (N=73)	6	11
Step 3 (N=325)	42 (12.9%)	78 (24%)
Level 2A (N=28)	3	6
Bupropion (N=12)	3	2
Venlafaxine (N=16)	0	4
Level 3 (N=297)	39	72
Switch strategy (N=186)	26	49
Nortriptyline (N=92)	11	23
Mirtazapine (N=94)	15	26
Augmentation strategy (N=111)	13	23
Lithium (N=58)	9	13
Bupropion SR (N=17)	3	2
Citalopram (N=22)	5	6
Sertraline (N=11)	1	3
Venlafaxine XR (N=8)	0	2
T3 (N=53)	4	10
Bupropion SR (N=6)	1	0
Citalopram (N=26)	1	7
Sertraline (N=8)	1	1
Venlafaxine XR (N=13)	1	2
Step 4 (N=106)	15 (14.2%)	22 (20.8%)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Level 3 (N=16)	3	3
Tranlycypromine (N=43)	7	10
Venlafaxine XR/mirtazapine (N=47)	5	9
Total Across Treatment Steps	225	1,330

For peer review only

Appendix 5: Baseline Demographic and Clinical Features by Treatment Step

	Treatment Step ^a							
	Step 1 (N=3,110)		Step 2 (N=1,134)		Step 3 (N=325)		Step 4 (N=106)	
Demographic Features								
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age	41.0	13.0	42.0	12.6	44.1	12.0	46.9	11.0
Education (years)	13.6	3.2	13.2	3.3	12.8	3.1	12.6	2.3
Monthly household income	2,289	2,732	1,744	1,539	1,470	1,383	1,003	887
	N	%	N	%	N	%	N	%
Female	1,469	74.6	502	73.1	113	65.3	34	65.4
Race								
White	2,328	74.9	870	76.7	259	79.7	86	81.1
Black	333	10.7	115	10.1	29	8.9	7	6.6
Other	449	14.4	149	13.1	37	11.4	13	12.3
Hispanic	402	12.9	139	12.3	45	13.8	16	15.1
Employment status								
Employed	975	58.7	314	54.2	69	46.9	19	43.2
Unemployed	612	36.9	243	42.0	72	49.0	24	54.5
Retired	73	4.4	22	3.8	6	4.1	1	2.3
Medical insurance								
Private	848	52.2	254	44.5	52	36.6	14	31.8
Public	282	17.4	109	19.2	30	21.4	10	23.3
None	534	33.2	223	39.3	60	43.2	20	46.5
Marital status								
Single	475	28.6	171	29.5	40	27.2	10	22.7
Married/cohabiting	716	43.1	238	41.0	61	41.5	18	40.9
Divorce/separated	429	25.8	155	26.7	42	28.6	14	31.8
Widowed	41	2.5	16	2.8	4	2.7	2	4.5
Clinical Features	N	%	N	%	N	%	N	%
First episode occurrence before age 18	1,200	39.0	436	38.8	120	37.0	41	38.7
Recurrent depression	1,940	66.8	718	68.3	188	63.3	59	60.8
Family history of depression	1,694	55.4	609	54.9	165	51.7	58	54.7

Duration of current episode \geq 2 years	787	25.6	311	27.7	88	27.2	34	32.1
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age at first episode (years)	24.9	14.5	24.7	14.2	25.9	14.6	25.9	14.3
Illness duration (years)	16.1	13.5	17.2	13.7	18.2	14.1	21.0	14.9
Number of episodes	4.4	9.7	4.9	11.1	4.4	10.3	5.0	12.1
Duration of current episode (months)	25.9	52.0	28.1	58.8	32.1	68.5	45.9	82.0
Median duration of current episode (months)	8.3		8.7		9.5		10.1	
Quality of Life and Enjoyment Satisfaction Questionnaire score ^b	39.1	14.3	36.5	13.6	33.7	13.5	31.6	12.8
SF-12 Mental ^c	25.6	8.1	25.0	7.7	24.4	7.7	24.0	7.5
SF-12 Physical ^c	48.6	12.1	47.0	12.4	44.5	12.1	43.8	12.3
Work and Social Adjustment Scale score ^d	25.0	8.7	26.3	8.2	28.3	7.7	29.4	7.0
HRSD ₁₇ score	21.9	5.2	22.5	5.2	23.4	5.2	23.9	5.4
IDS-C ₃₀ score ^e	39.1	9.6	40.6	9.7	42.6	9.4	43.6	9.8
QIDS-IVR score ^f	16.9	3.3	17.3	3.3	17.9	3.0	18.3	3.1
Cumulative Illness Rating Scale								
Categories endorsed	2.5	1.5	2.6	1.6	2.8	1.6	3.1	1.6
Total score	4.7	3.9	5.1	4.0	5.8	4.5	6.2	4.3
Severity score	1.8	0.8	1.8	0.8	2.0	0.9	2.0	0.7
	N	%	N	%	N	%	N	%
Psychiatric Diagnostic Screening Questionnaire								
Agoraphobia	559	18.2	240	21.4	89	27.5	32	30.2
Alcohol abuse/dependence	371	12.0	136	12.1	36	11.1	8	7.5
Bulimia	607	19.7	232	20.6	67	20.7	20	18.9

Drug abuse/dependence	234	7.6	80	7.1	21	6.5	7	6.6
Generalized anxiety disorder	736	23.9	290	25.8	94	29.0	36	34.0
Hypochondriasis	336	10.9	139	12.4	45	13.9	14	13.2
Obsessive-compulsive disorder	723	23.5	265	23.6	97	29.9	31	29.2
Panic disorder	422	13.7	183	16.3	65	20.1	21	19.8
Posttraumatic stress disorder	387	12.6	172	15.3	55	17.0	16	15.1
Social phobia	963	31.3	379	33.7	117	36.1	35	33.0
Somatoform disorder	284	9.2	105	9.3	35	10.8	9	8.5
Number of axis I comorbid psychiatric disorders								
0	606	19.7	190	16.9	48	14.8	12	11.3
1	740	24.0	257	22.9	68	21.0	23	21.7
2	577	18.7	217	19.3	62	19.1	25	23.6
3	363	11.8	139	12.4	42	13.0	14	13.2
4+	793	25.8	321	28.6	104	32.1	32	30.2

^a Sums do not always equal N due to missing values. Percentages are based on available data.

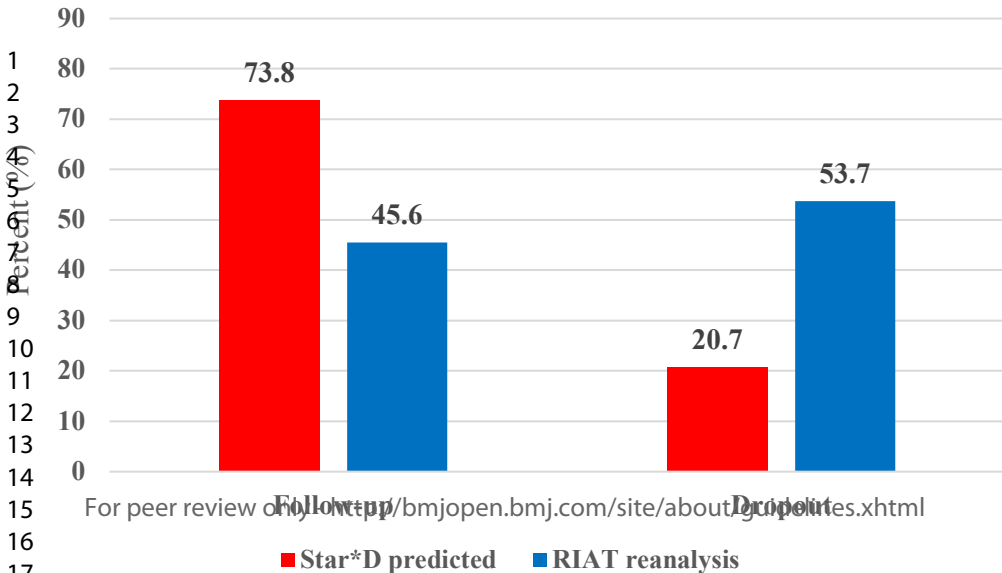
^b Integrated voice response (IVR) administered version of the Quality-of-Life Enjoyment and Satisfaction Questionnaire assessing participants' global rate of satisfaction. Higher scores (range=0–100) represent greater life enjoyment and satisfaction.

^c IVR-administered version of the SF-12 assessing perceived mental and physical health status. Two subscales (physical health factor and mental health) range from 0 to 100— higher scores indicate better functioning with a population norm for each score of 50.

^d IVR-administered version of the Work and Social Adjustment Scale. Scores between 10 and 20 are associated with significant functional impairment while scores above 20 suggest moderate to severe functional impairment.

^e Inventory of Depressive Symptomatology administered telephonically.

^f IVR-administered version of the QIDS.



BMJ Open

What are the Treatment Remission, Response, and Extent of Improvement Rates after up to Four Trials of Antidepressant Therapies in Real-World Depressed Patients? A Reanalysis of the STAR*D Study's Patient-Level Data with Fidelity to the Original Research Protocol

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2022-063095.R2
Article Type:	Original research
Date Submitted by the Author:	14-Apr-2023
Complete List of Authors:	Pigott, H.; None Kim, Thomas; University of Pennsylvania, Psychology Xu, Colin; University of Pennsylvania, Psychology Kirsch, Irving; Harvard Medical School Amsterdam, Jay; University of Pennsylvania Perelman School of Medicine, Psychiatry
Primary Subject Heading:	Mental health
Secondary Subject Heading:	Pharmacology and therapeutics
Keywords:	Adult psychiatry < PSYCHIATRY, Depression & mood disorders < PSYCHIATRY, CLINICAL PHARMACOLOGY

SCHOLARONE™
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1
2
3
4
5
6
7
8
9
10
11
12

**What are the Treatment Remission, Response, and Extent of
Improvement Rates after up to Four Trials of Antidepressant Therapies in Real-World
Depressed Patients? A Reanalysis of the STAR*D Study's Patient-Level Data
with Fidelity to the Original Research Protocol**

Authors:

H. Edmund Pigott, Thomas T. Kim, Colin Xu, Irving Kirsch, & Jay D. Amsterdam

13
14
15
16
17
18

H. Edmund Pigott, Ph.D.

Position: Clinical Psychologist

Institution: none

Email: pathware@erols.com

19
20
21
22
23

Thomas T. Kim

Position: PhD candidate in Psychology

Institution: Department of Psychology, University of Pennsylvania

24
25
26
27
28

Colin Xu

Position: PhD candidate in Psychology

Institution: Department of Psychology, University of Pennsylvania

29
30
31
32
33

Irving Kirsch, Ph.D.

Position: Associate Director, Program in Placebo Studies

Institution: Harvard Medical School

34
35
36
37
38

Jay D. Amsterdam, MD

Position: Professor of Psychiatry (Emeritus), Depression Research Unit

Institution: Perelman School of Medicine, University of Pennsylvania

39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Corresponding author: H. Edmund Pigott, Ph.D.

Abstract

Objective: Reanalyze the patient-level dataset of the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study with fidelity to the original research protocol and related publications.

Design: The study was open label and semi-randomized examining the effectiveness of up to four optimized, and increasingly aggressive, antidepressant therapies in depressed adults. Patients who failed to gain adequate relief from their level 1 trial on the SSRI citalopram could receive up to three additional treatment trials in levels 2-4.

Setting: 41 North American psychiatry and primary care treatment centers.

Participants: 4,041 adults screened positive for major depressive disorder. In contrast to most clinical trials, STAR*D enrolled patients seeking care (versus recruited) and included patients with a wide range of common co-morbid medical and psychiatric conditions to enhance the generalizability of findings to real-world clinical practice.

Interventions: STAR*D evaluated the relative effectiveness of 13 antidepressant therapies in treatment levels 2-4 for depressed patients who failed to gain adequate benefit from their level 1 medication trial.

Main Outcome Measures: According to the STAR*D protocol, the primary outcome was remission, defined as a score <8 on the blinded Hamilton Rating Scale for Depression (HRSD). Response was a secondary outcome defined as ≥50% reduction in HRSD scores. STAR*D's protocol specifically excluded all non-blinded clinic-administered assessments from use as research outcome measures.

Results: STAR*D investigators did not use the protocol-stipulated HRSD to report cumulative remission and response rates in their summary article, and instead used a non-blinded clinic-administered assessment. This inflated their report of outcomes, as did their inclusion of 99 patients who scored as remitted on the HRSD at study outset as well as 125 who scored as remitted when initiating their next-level treatment. These patients should have been excluded from data analysis. In contrast to the STAR*D-reported 67% cumulative remission rate after up to four antidepressant treatment trials, the rate was 35.0% when using the protocol-stipulated HRSD and inclusion in data analysis criteria.

Conclusion: STAR*D's cumulative remission rate was approximately half of that reported.

Strengths and limitations of this study

- We reanalyzed the largest ever prospective antidepressant trial's patient-level dataset with fidelity to the original research protocol and related publications.
- The reanalysis was conducted under the guidelines of the *Restoring Invisible and Abandoned Trials* (RIAT) initiative.

- Treatment remission, response, and extent of symptom improvement rates were calculated for 14 antidepressant therapies for those patients who met STAR*D's inclusion in data analysis criteria as well as the overall cumulative remission rate after up to four trials of antidepressant therapies.
- We calculated STAR*D's remission rate using the protocol-stipulated HRSD as well as combining the HRSD remissions with those from a non-stipulated measure of remission for patients missing an exit HRSD score. Combining STAR*D's HRSD-defined remissions with those from the non-stipulated measure increased its cumulative remission rate from 35.0% to 41.3%.
- Finally, we compared STAR*D's outcomes to those found in a meta-analysis of 7,030 patients enrolled in similar open-label antidepressant comparator trials. Whereas the treatment remission and response rates in comparator trials averaged 48.4% and 65.2% respectively, they were only 25.5% and 40.5% for STAR*D's level 1 patients and worse in treatment levels 2-4. Similarly, comparator trials patients' mean change on the HRSD was 14.8 points versus 8.4 points for STAR*D's level 1 patients and worse for patients in treatment levels 2-4.

Introduction

At a cost of 35 million US dollars, the National Institute of Mental Health (NIMH) funded Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study is the largest and most expensive prospective antidepressant trial ever conducted with over 100 journal articles published by study investigators.[1-7] In contrast to most clinical trials that enroll symptomatic volunteers (typically recruited through advertising), STAR*D enrolled 4041 patients who screened positive for major depressive disorder (MDD) while seeking routine medical or psychiatric care. STAR*D did not exclude patients with medical conditions and most comorbid psychiatric disorders, thereby increasing the generalizability of its findings to real-world clinical practice.

The STAR*D study provided up to four treatment trials per patient and was designed to give guidance in selecting the best next-level treatment option for the many patients who fail to gain sufficient relief from their first, and/or subsequent, antidepressant trial. To mimic clinical practice, STAR*D used an open-label research design with no control group during any phase of the study.

Our STAR*D reanalysis examines key methodological deviations from its research protocol and related publications, and these deviations' impact on its investigators' report of outcomes. In STAR*D's Rationale and Research Design article, and repeated in the level 1-4 published study outcomes, STAR*D investigators stated, "*the primary outcome is depressive symptom severity, measured by the 17-item Hamilton Rating Scale for Depression (HRSD)*." [8, p. 120]. STAR*D's prespecified primary outcome was remission, defined as scoring <8 on the HRSD which was administered telephonically by Research Outcome Assessors (ROAs) blind to patients' study status (treatment level entry/exit/follow-up). Response was a secondary outcome defined as a ≥50% reduction in patients' HRSD scores. Remission as defined by the HRSD was not presented in STAR*D's summary article.[7] Furthermore, despite its investigators' numerous publications, neither change in HRSD depressive symptom severity nor HRSD response rates have been reported for STAR*D's six primary studies [1-6] and summary article.[7] Instead, response rates and change in symptom severity were reported using the clinic-administered Quick Inventory of Depressive Symptomatology–Self Report (QIDS-SR), a measure developed by the STAR*D principal investigators.[9] This occurred despite the fact that STAR*D's research protocol specifically excluded all clinic-administered assessments, such as the QIDS-SR, from

1
2
3 use as research outcome measures since they were not blinded and instead, used to guide patient care.
4 The protocol states:
5

6
7 *Recall that the research outcomes assessments are distinguished from assessments*
8 *conducted at clinic visits. The latter are designed to collect information that guides*
9 *clinicians in the implementation of the treatment protocol. Research outcomes*
10 *assessments are **not** collected at the clinic visits. They are not collected by either*
11 *clinicians or Clinical Research Coordinators.*[10,p.47-48; emphasis in the original]
12

13 In their summary article, STAR*D investigators used the QIDS-SR as the sole measure to report
14 remission, response, and extent of symptom improvement. This article's Abstract states that "the
15 overall cumulative remission rate was 67%" with no qualifiers to this claim.[7, p.1905] Besides making
16 this claim based on an assessment the protocol specifically excluded from use as a research measure, it
17 is not until the article's Results section that readers learn this high level of treatment success did not
18 occur. The STAR*D investigators' claim was theoretical—an estimate based on the provisos of what
19 would have happened if there were no study dropouts, and furthermore, "that those who exited the
20 study would have had the same remission rates as those who stayed in the protocol." [7, p.1910] As
21 Pigott et al. documented though, the investigators' assumptions are not true in the real world since
22 more patients dropped out than remitted in each STAR*D treatment level,[11] and furthermore, it has
23 been found in placebo-controlled trials that patients who drop out are more likely to have had adverse
24 treatment side effects and/or emergent suicidality.[12]
25
26

27
28 Unfortunately, the STAR*D investigators' claim of a 67% cumulative remission rate has
29 become accepted clinical wisdom, and the provisions on which it is based are commonly not
30 referenced when portraying STAR*D's findings. For example, in 2009 NIMH's Director Dr.
31 Thomas Insel claimed STAR*D found "at the end of 12 months, with up to four treatment
32 steps, roughly 70% of participants were in remission." [13, p.1466] Similarly in 2013, an
33 editorial in the *American Journal of Psychiatry (AJP)* claimed STAR*D found "after four
34 optimized, well-delivered treatments, approximately 70% of patients achieve remission." [14,
35 p.580]. More recently (2022), a New York Times' article claimed that half of STAR*D's
36 participants "had significantly improved after using either the first or second medication, and
37 nearly 70 percent of people had become symptom-free by the fourth antidepressant." [15]
38 These are not factual statements of STAR*D's findings.
39
40

41 The first author has made published criticisms alleging protocol violations that appear to inflate
42 STAR*D's findings and called for the reanalysis of the dataset by independent investigators.[16] In 2018,
43 the first and fourth authors collaborated with researchers from the University of Connecticut to
44 reanalyze STAR*D's level 1 data obtained from NIMH.[17] This reanalysis found substantial inflation of
45 STAR*D's reported remission and response rates. Furthermore, the reanalysis found that the extent of
46 HRSD improvement in STAR*D's level 1 trial was approximately half that of open-label antidepressant
47 comparator trials.
48
49

50 Our published criticisms of STAR*D investigators' report of outcomes are as follows:[18]
51

- 52 • While STAR*D investigators used the HRSD to report remission rates in their levels 1-4
53 articles,[1-6] the QIDS-SR was used as the sole measure to report remission, response, and
54 extent of improvement rates in their summary article[7] without disclosing that the protocol
55 specifically excluded all non-blinded/clinic-administered assessments such as the QIDS-SR from
56
57
58
59
60

1
2
3 use as outcome measures. The primary outcome measure, the HRSD, should have been used to
4 report the summary article's outcomes.

- 5 • Using data from the 931 patients deemed ineligible for analysis in STAR*D's level 1 article
6 because these patients lacked a baseline ROA-administered HRSD score of ≥ 14 , in STAR*D's
7 levels 2-4 and summary articles without clear disclosure. This included 99 patients who scored
8 < 8 on their baseline HRSD—indicating these patients met STAR*D's remission criterion at study
9 outset and should not have been included in their report of outcomes.
- 10 • Excluding from analysis 370 patients who dropped out after starting on citalopram in their first
11 clinic visit without taking the exit HRSD despite STAR*D investigators stating, "our primary
12 analyses classified patients with missing exit HRSD scores as nonremitters a priori." [1, p.34]
13 These 370 early dropout patients should have been counted as nonremitters as prespecified.
- 14 • Including in their analyses 125 patients who scored as remitted at entry into their next-level
15 treatment. This occurred despite STAR*D investigators prespecifying that, "patients who begin
16 a level with HRSD < 8 will be excluded from analyses." [8, p.130]

17
18
19
20 This article reanalyzes STAR*D's treatment remission, response, and extent of improvement after up to
21 four trials of antidepressant therapies, using STAR*D's protocol-specified primary outcome measure, the
22 ROA-administered HRSD. This effort builds on Pigott et al's 2010 article [11] that focused on
23 deconstructing STAR*D investigators' levels 1-4 and summary articles, [1-7] by reanalyzing STAR*D's
24 patient-level dataset obtained from NIMH in 2019 with fidelity to the original research protocol and
25 inclusion in data analysis criteria. Future efforts will focus on reanalyzing STAR*D's levels 2-4 semi-
26 randomized comparator trials, including the extent of emergent suicidal ideation and 12-month follow-
27 up outcomes tied to each compared treatment.

30 Method

31 *RIAT Initiative*

32
33
34 The Restoring Invisible and Abandoned Trials (RIAT) initiative started in 2013 calling on funders
35 and investigators of abandoned (unpublished) or misreported studies to publish undisclosed
36 outcomes or correct misleading publications. [19] If investigators failed to correct a study
37 identified as misreported, independent investigators were encouraged to correct the record by
38 reanalyzing the study's patient-level dataset consistent with the research protocol and analytic
39 plan.

40
41
42 On March 6, 2019, the RIAT investigators published our response to a 'Call to Action' statement in the
43 *British Medical Journal*, in which we stated our intention to reanalyze the STAR*D dataset. [18] We then
44 notified STAR*D's principal investigators of our intention and requested they inform us whether they
45 would undertake a reanalysis of the dataset adhering to the research protocol. On March 22, 2019,
46 STAR*D investigators acknowledged our email notification, indicated the STAR*D data were in the public
47 domain, and stated they had no interest in undertaking a reanalysis.

48
49
50 In July 2019, we received a STAR*D Data Use Certificate, issued by the NIMH Data Archive Data Access
51 Committee, and gained access to the STAR*D levels 1-4 and follow-up patient-level dataset consisting of
52 26 text files, and limited supporting study documentation. In September 2019, we obtained funding
53 from the RIAT Support Center to reanalyze STAR*D.

54 *Patients*

1
2
3
4 STAR*D patients were 18 to 75 years of age, seeking care at 18 primary and 23 psychiatric care clinics.
5 Clinical research coordinators (CRCs) screened 4,790 patients for MDD. This screening included the
6 CRCs' administering the HRSD, on which 4,041 patients scored ≥ 14 , met the other inclusion criteria, and
7 enrolled into the study. CRCs also gathered patients' psychiatric history, demographic information, and
8 administered both the Cumulative Illness Rating Scale and the Psychiatric Diagnostic Screening
9 Questionnaire to determine the extent of comorbid medical and psychiatric disorders.
10

11 12 *Levels/Steps of Acute Treatment*

13
14 STAR*D investigators sought to provide the highest quality of care to maximize the number of
15 remissions while minimizing dropouts (see Supplemental Table 1). Supplemental Table 2 describes the
16 antidepressant therapies available in treatment levels 1-4 while steps refer to the numeric order of
17 treatments. As seen in Figure 1, treatment steps 1 and 2 correspond to levels 1 and 2 treatments.
18 Similarly, for most patients their levels 3 and 4 treatments correspond to treatment steps 3 and 4. For
19 level/step 2 patients though who failed to respond adequately to cognitive therapy alone or combined
20 with citalopram and chose to continue in the study, their third treatment step was designated level 2A
21 and they were randomized to one of two level 2 switch medications. For these patients, their level 2A
22 treatment was their third treatment step. For level 2A patients who did not adequately benefit from
23 this medication trial and chose to continue in the study, they entered a fourth treatment step consisting
24 of level 3 treatments.
25
26

27
28 All patients were administered the SSRI citalopram for their level 1 treatment. Each treatment level
29 consisted of 12 weeks of antidepressant therapy, with an additional 2 weeks for patients deemed close
30 to remission. Treatment was administered using a system of measurement-based care that assessed
31 symptoms and side effects at each clinic visit. STAR*D investigators state, "*To enhance the quality and*
32 *consistency of care, physicians used the clinical decision support system that relied on the measurement*
33 *of symptoms (QIDS-C and QIDS-SR), side-effects, medication adherence, and clinical judgment based on*
34 *patient progress.*"[1, p.30] This system was used to guide medication management of a fully adequate
35 dose for a sufficient time to "*ensure that the likelihood of achieving remission was maximized and that*
36 *those who did not reach remission were truly resistant to the medication.*"[1, p.30]
37
38

39 For those patients who failed to gain an adequate response from citalopram, STAR*D allowed them to
40 select acceptable treatment options for randomization in levels 2 to 4 "*to empower patients, strengthen*
41 *the therapeutic alliance, optimize treatment adherence, and improve outcome*"[20, p.483]. The
42 treatment options available for randomization involved either switching to a new treatment or
43 augmenting the patient's current treatment. Treatment levels 2 to 4 evaluated the relative
44 effectiveness of 11 pharmacologically distinct drug/drug combination treatments. Cognitive therapy
45 was also available as either a switch or citalopram augmentation option in level 2.
46
47

48 *STAR*D Follow Up Phase*

49
50 In each treatment trial for levels 1-4, patients who scored < 6 on their last QIDS-Clinician version (QIDS-C)
51 were considered clinician-rated remissions and encouraged to enter the 12-month follow-up phase.
52 During follow-up, patients continued their "previously effective acute treatment medication(s) at the
53 doses used in acute treatment but that any psychotherapy, medication, or medication dose change
54 could be used." [7, p.1908] Based on prior research, a QIDS score of < 6 was estimated by STAR*D
55 investigators to correspond to a score of < 8 on the HRSD, STAR*D's prespecified primary outcome
56
57
58
59
60

1
2
3 measure for classifying patients as remitted.[9] Clinicians strongly encouraged patients who did not
4 obtain a QIDS-defined remission to enter the next-level treatment. Patients who failed to attain a QIDS-
5 defined remission, but did have a $\geq 50\%$ reduction on the QIDS-C and did not want to be randomized to a
6 next-level treatment, were also encouraged to enter follow-up.
7

8 9 *Research Design of the STAR*D Study*

10
11 STAR*D investigators developed a new research design for the study termed “equipoise-stratified” to
12 evaluate the relative efficacy of 13 antidepressant therapies in levels 2-4 for depressed patients who
13 failed to gain adequate benefit from their level 1 medication trial.[21] In level 1, all patients received
14 citalopram as their first treatment. In level 2, patients were informed regarding seven treatment
15 options to choose from: four switch options in which citalopram was stopped and the new treatment
16 initiated and three augmentation options in which citalopram was combined with a second
17 antidepressant treatment. In level 3, patients were informed regarding four treatment options to
18 choose from: two switch options and two augmentation options. Level 4 involved randomization to one
19 of two medication/medication combination switch options.
20
21

22 *Analytic Plan of the RIAT Reanalysis*

23
24 We reanalyzed the STAR*D patient-level dataset with fidelity to the original research protocol wherever
25 possible. Where the protocol was silent, we used other STAR*D publications to guide our analysis. This
26 occurred four times. First, the protocol is silent regarding patients who entered the study without a
27 baseline ROA-administered HRSD score of ≥ 14 . In their level 1 article, STAR*D investigators deemed the
28 931 such patients who lacked this marker of depression severity ineligible for inclusion in data
29 analysis.[1] We do the same and extend this exclusion for such patients who continued on to levels 2-4
30 because their extent of depression severity at study outset is not known. Second, the protocol is silent
31 on what to do with patients who met the remission criteria on the HRSD at entry into their next-level
32 treatment. In STAR*D’s Rationale and Research Design article though, its investigators prespecify that
33 “patients who begin a level with HRSD < 8 will be excluded from analyses.”[8, p.130] We therefore
34 excluded 125 such patients from our analyses of treatment levels 2-4. Third, the protocol is silent on
35 how to analyze patients who exit a treatment without taking the HRSD. STAR*D investigators state in
36 their level 1 article, “our primary analyses classified patients with missing exit HRSD scores as
37 nonremitters a priori”[1, p.34] and repeat similar statements in their level 2-4 articles.[2-6] Therefore,
38 we do likewise.
39
40
41

42 Finally, STAR*D had many patients with missing exit HRSD scores. In their level 2-4 articles, STAR*D
43 investigators used a correspondence table to map the final QIDS-SR score to the HRSD for patients
44 missing their exit HRSD score to assess the impact of their approach to counting such patients as
45 “nonremitters a priori.”[22,23] For patients with missing exit HRSD scores, we therefore mapped their
46 last QIDS-SR score to the HRSD and used it to calculate the mean HRSD exit, mean change, and
47 combined HRSD & QIDS-SR response rates for all treatments. We also calculated STAR*D’s remission
48 rate both as prespecified based on an exit HRSD score of < 8 as well as a final QIDS-SR score of < 6 for
49 those patients missing an exit HRSD score.
50
51

52 All pre-processing and analyses were performed in R.[24] Authors 2 and 3 identified patients by their
53 subject key and used this variable to match information across datasets. Data on patients’ treatment
54 pathways, and when patients transitioned from one level to the next, were taken from the IVRA dataset
55 completed by CRCs, and verified against the data on patient level exits. Authors 2 and 3 then compared
56
57
58
59
60

1
2
3 the number of patients identified for all level 1-4 treatments to that reported in the STAR*D summary
4 article's patient flowchart, and the number of patients matched.[7]
5

6
7 Next, authors 2 and 3 applied STAR*D's level 1 inclusion for data analysis criterion to patients in
8 treatment levels 2-4 as well as excluded from analysis the 125 patients who scored <8 on the HRSD at
9 entry into their next-level treatment. We counted these 125 patients as remitted in the prior treatment
10 level but excluded them from the analyses of subsequent treatments. Supplemental Table 3 presents
11 the number of level 2-4 patients excluded from our reanalysis, and the reasons for their exclusion.
12 Supplemental Table 4 identifies the number of patients with missing entry and/or exit HRSD scores for
13 all level 1-4 treatments. As seen in Supplemental Table 4, 1,330 patients were missing their exit HRSD
14 score across all treatments.
15

16
17 We then compared STAR*D's outcomes to those found in a meta-analysis of 7,030 patients enrolled in
18 antidepressant comparator trials.[25] Similar to STAR*D, comparator trials typically are conducted
19 open-label without a control group and therefore are the appropriate comparison data for STAR*D's
20 outcomes. Continuous HRSD improvement means were provided by the first author of the meta-
21 analysis.[25]
22

23
24 Finally, we compared the STAR*D protocol's step-by-step predictions of patient drop out and the
25 number of patients who would have a satisfactory treatment response and enter follow-up care to what
26 actually occurred.[10] While the purpose of these predictions' was to estimate the number of
27 continuing patients available for randomization in treatment levels 2-4, at the meta-level these
28 predictions are an important hypothesis STAR*D tested by assessing how well its investigators could
29 predict the aggregate step-by-step successful treatment outcomes from their treat-to-remission model
30 of care.
31

32 *Patient and Public Involvement*

33
34
35 Neither patients nor the public were involved in the design, conduct, reporting, or dissemination plans
36 of our research.
37

38 **Results**

39
40
41 Figure 1 presents the overall flow of patients enrolled in the various protocol-defined treatment levels
42 and places them in groups defined by the number of treatment steps. Of the 4,041 patients enrolled
43 into STAR*D, 3,110 met the eligibility for data analysis criterion of having a ROA-administered HRSD
44 score ≥ 14 at study outset. Figure 1 also identifies the number of patients who exited the study following
45 each treatment step, the number who entered follow-up after each treatment step, and the number
46 who were randomly assigned to a next-level treatment.
47

48
49 Supplemental Table 5 describes the demographic and clinical features of the patients who entered
50 treatment in steps 1-4 based on their level 1 baseline presentation when enrolling into the study.
51 Summary statistics are presented as means and standard deviations for continuous variables and
52 percentages for discrete variables. Note that 55.7% of STAR*D patients had 2 or more comorbid axis 1
53 disorders when first enrolled based on the Psychiatric Diagnostic Screening Questionnaire and averaged
54 2.5 comorbid medical conditions based on the Cumulative Illness Rating Scale. Furthermore, the
55 average length of patients' current MDD episode was 25.9 months. In a post hoc analysis, STAR*D
56
57
58
59

investigators found that 77.8% of its enrolled patients would have been excluded from most antidepressant trials due to having two or more concurrent medical conditions, more than one comorbid psychiatric disorder, and/or a current depressive episode lasting > 2 years.[26]

Table 1 presents the mean HRSD entry, exit, and change scores for patients by the specific treatment they received in steps 1-4 as well as the HRSD remission and response rates. Table 1 also provides the HRSD cumulative remission rate after up to 4 trials on antidepressant therapies as well as the combined HRSD plus QIDS-SR remission and response rates for the 1,330 patients missing an exit HRSD score.

Table 1: Outcomes Across All Treatments

Treatment Step	HRSD Score		*Mean Change [95% confidence interval] (SD)	HRSD Remission Rate # (%)	*Combined HRSD & QIDS-SR Remission Rate # (%)	*Combined HRSD & QIDS-SR Response Rate # (%)
	Entry Mean (SD)	*Exit Mean (SD)				
Step 1 (N=3,110)	21.87 (5.21)	13.49 (8.42)	8.38 [8.10, 8.67] (8.11)	794 (25.5%)	938 (30.2%)	1261 (40.5%)
Step 2 (N=1,134)	18.76 (6.24)	13.97 (8.09)	4.79 [4.37, 5.21] (7.23)	241 (21.3%)	283 (25.0%)	329 (29.0%)
Switch strategy (N=620)	19.85 (6.08)	14.70 (8.01)	5.16 [4.59, 5.73] (7.22)	113 (18.2%)	134 (21.6%)	178 (28.7%)
Bupropion (N=190)	20.11 (6.25)	15.32 (7.85)	4.78 [3.82, 5.75] (6.78)	31 (16.3%)	37 (19.5%)	46 (24.2%)
Sertraline (N=198)	19.95 (5.98)	14.92 (8.02)	5.03 [4.04, 6.01] (7.10)	32 (16.2%)	36 (18.2%)	57 (28.8%)
Venlafaxine (N=192)	19.89 (6.19)	14.31 (8.12)	5.58 [4.53, 6.63] (7.45)	37 (19.3%)	44 (22.9%)	59 (30.7%)
Cognitive Therapy (N=40)	18.01 (4.96)	12.44 (7.93)	5.58 [2.87, 8.28] (8.73)	13 (32.5%)	17 (42.5%)	16 (40.0%)
Augmentation strategy (N=514)	17.44 (6.18)	13.10 (8.10)	4.34 [3.72, 4.97] (7.23)	128 (24.9%)	149 (29.0%)	151 (29.4%)
Bupropion (N=216)	16.88 (6.11)	12.52 (7.83)	4.36 [3.38, 5.33] (7.30)	54 (25.0%)	64 (29.6%)	66 (30.6%)

Buspirone (N=225)	17.80 (6.50)	13.36 (8.40)	4.43 [3.52, 5.35] (7.02)	58 (25.8%)	68 (30.2%)	66 (29.3%)
Cognitive Therapy (N=73)	17.99 (5.24)	13.98 (7.98)	4.01 [2.25, 5.78] (7.69)	16 (21.9%)	17 (23.3%)	19 (26.0%)
Step 3 (N=325)	19.59 (6.09)	16.38 (7.77)	3.21 [2.48, 3.94] (6.70)	43 (13.2%)	50 (15.4%)	63 (19.4%)
Level 2A (N=28)	20.89 (5.44)	16.96 (6.48)	3.93 [1.81, 6.04] (5.71)	3 (10.7%)	3 (10.7%)	5 (17.9%)
Bupropion (N=12)	19.92 (3.85)	17.58 (7.35)	2.33 [-0.81, 5.48] (5.55)	2 (16.7%)	2 (16.7%)	2 (16.7%)
Venlafaxine (N=16)	21.62 (6.41)	16.50 (5.96)	5.12 [2.33, 7.92] (5.70)	1 (6.2%)	1 (6.2%)	3 (18.8%)
Level 3 (N=297)	19.46 (6.14)	16.32 (7.88)	3.14 [2.37, 3.92] (6.79)	40 (13.5%)	47 (15.8%)	58 (19.5%)
Switch strategy (N=186)	20.01 (6.24)	17.01 (7.91)	2.99 [2.00, 3.99] (6.94)	23 (12.4%)	25 (13.4%)	31 (16.7%)
Nortriptyline (N=92)	19.67 (5.27)	16.99 (8.35)	2.67 [1.10, 4.24] (7.68)	15 (16.3%)	15 (16.3%)	16 (17.4%)
Mirtazapine (N=94)	20.34 (7.08)	17.03 (7.49)	3.30 [2.06, 4.55] (6.15)	8 (8.5%)	10 (10.6%)	15 (16.0%)
Augmentation strategy (N=111)	18.55 (5.89)	15.16 (7.74)	3.40 [2.18, 4.62] (6.56)	17 (15.3%)	22 (19.8%)	27 (24.3%)
Lithium (N=58)	18.69 (6.47)	15.91 (7.29)	2.78 [1.42, 4.15] (5.31)	7 (12.1%)	9 (15.5%)	10 (17.2%)
T3 (N=53)	18.41 (5.25)	14.34 (8.19)	4.07 [1.99, 6.14] (7.69)	10 (18.9%)	13 (24.5%)	17 (32.1%)
Step 4 (N=106)	20.65 (5.54)	16.49 (7.47)	4.16 [2.80, 5.52] (7.15)	11 (10.4%)	13 (12.3%)	22 (20.8%)
Level 3 (N=16)	20.62	17.62	3.00	2	2	3

	(4.01)	(6.87)	[-0.45, 6.45] (7.04)	(12.5%)	(12.5%)	(18.8%)
Tranlycypromine (N=43)	21.02 (6.57)	16.45 (7.89)	4.57 [2.22, 6.92] (7.87)	3 (7.0%)	5 (11.6%)	9 (20.9)
Venlafaxine XR/mirtazapine (N=47)	20.32 (5.02)	16.14 (7.38)	4.18 [2.30, 6.07] (6.59)	6 (12.8%)	6 (12.8%)	10 (21.3%)
CUMULATIVE remission rate after up to four treatment steps				1,089 (35.0%)	1,284 (41.3%)	

*For patients with missing exit HRSD scores, their last QIDS-SR score is mapped to the HRSD and used to calculate HRSD Exit Mean, Mean Change, Combined HRSD & QIDS-SR Remission Rate, and Combined HRSD & QIDS-SR Response Rate.

Table 2 presents patients' aggregate HRSD status in terms of remission, response, and extent of mean symptomatic change at entry and exit for each treatment step as well as study dropout. In step 1, 25.5% of patients remitted. Steps 2-4 show a continuous decrease in remission rates from step 2's 21.3% to step 3's 13.2% and step 4's 10.4% with increasing rates of study dropout from step 1's 34.5% to step 3's 46.2%.

Table 2: Outcomes by Treatment Step

	Step 1 (N=3,110)		Step 2 (N=1,134)		Step 3 (N=325)		Step 4 (N=106)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
HRSD score at entry into step	21.87	5.21	18.76	6.24	19.59	6.09	20.65	5.54
HRSD score at exit from step*	13.49	8.42	13.97	8.09	16.38	7.77	16.49	7.47
HRSD Mean Change*	8.38	8.11	4.79	7.23	3.21	6.7	4.16	7.15
	N	%	N	%	N	%	N	%
Remission at each step exit	794	25.5%	241	21.3%	43	13.2%	11	10.4%
Response at each step exit*	1261	40.5%	329	29.0%	63	19.4%	22	20.8%
Entered Follow-up	902	29.0%	406	35.8%	69	21.2%	38	35.9%
Study Exit/Dropout	1,074	34.5%	403	35.5%	150	46.2%		

* For patients with missing exit HRSD scores, their last QIDS-SR score is mapped to the HRSD and used to calculate HRSD Exit Mean, Mean Change, and Combined HRSD & QIDS-SR Response Rate.

Supplemental Figures 1 and 2 compare the HRSD remission, response, and extent of symptom improvement rates for STAR*D patients in steps 1-4 to that found in a meta-analysis of 7,030 patients enrolled in non-blinded antidepressant comparator trials.[25] In step 1, these measures of

1
2
3 improvement among STAR*D's patients were one-third or more less than that found in comparator
4 trials, and improvement was worse in each subsequent treatment step.
5

6
7 Figure 2 compares the STAR*D protocol's predictions of patient dropout and the number of patients
8 who would have a satisfactory treatment response and enter follow-up to what occurred. Cumulatively,
9 STAR*D's investigators predicted that 73.8% of patients would have a successful treatment response
10 and enter follow-up whereas in fact only 45.6% achieved this measure of treatment success.
11 Furthermore, whereas STAR*D investigators predicted that over the course of up to four antidepressant
12 therapies 20.7% of patients would dropout, in fact, 53.7% dropped out. On this measure of treatment
13 failure, STAR*D's dropout rate was 2.6 times greater than predicted.
14

15
16 Figure 3 presents the step-by-step cumulative remission rate in three ways. First, the 'theoretical' rate
17 propagated by STAR*D investigators based on the provisos of what would have happened if there were
18 no study dropouts and that those who did exit had the same QIDS-SR remission rates as those who
19 stayed.[7] Next, the combined HRSD plus QIDS-SR remission rate based on either an exit HRSD score of
20 <8, OR a last clinic visit QIDS-SR score of <6 for the 1,330 patients missing an exit HRSD. Finally, our RIAT
21 reanalysis rate when using the protocol-specified exit HRSD score of <8 as the sole measure of remission
22 for the 3,110 patients who met STAR*D's inclusion in data analysis criteria. The cumulative remission
23 rate after up to four antidepressant therapies using the HRSD was 35.0% versus 41.3% when combined
24 with the QIDS-SR, both of which are substantially less than the 67% cumulative remission rate claimed in
25 the summary article's Abstract.
26

27 Discussion

28 *Principal findings and comparison with original STAR*D publication*

29
30
31
32 STAR*D's results highlight the discrepancy in likely outcomes between typical antidepressant clinical
33 trials with their exclusion criteria and the real-world patients for whom these medications are
34 commonly prescribed. Our RIAT reanalysis found poorer outcomes after up to four optimized, and
35 increasingly aggressive, antidepressant therapies than reported in STAR*D's summary article published
36 in *AJP*. [7] In contrast to the 67% cumulative remission rate reported in *AJP*, the actual rate was 35.0%
37 when using the protocol-specified HRSD, and increased to 41.3% when combined with a final clinic-visit
38 QIDS-SR score of <6 for patients missing exit HRSD scores in treatment steps 1-4. The 41.3% cumulative
39 remission rate should be viewed as the "best case scenario" since it added an additional 195 QIDS-
40 defined remissions (a remission measure not specified in the protocol) from the 1,330 patients with
41 missing exit HRSD scores. As there was neither a placebo nor waitlist control group during any phase of
42 the STAR*D study, it is impossible to know to what extent the observed results were due to the
43 pharmacologic effects of the prescribed medications, placebo effects, and/or the passage of time.
44
45

46
47 Our reanalysis did not assess the durability of treatment effects during the 12-month follow up phase.
48 In their summary article though, STAR*D investigators reported an overall relapse rate of 46.1% for the
49 1,729 patients for whom they had at least one assessment (of up to 12 scheduled) during follow up
50 using a telephonic-administered version of the QIDS [7] whereas Pigott et al. found a far lower sustained
51 recovery rate when incorporating patient dropout in the analysis.[11]
52

53 *Comparison with other studies*

1
2
3 Our reanalysis found that in step 1, STAR*D's remission, response, and extent of improvement rates
4 were substantially less than those reported in other open-label antidepressant comparator trials and
5 then grew progressively worse in steps 2-4.[25] Such studies typically exclude depressed patients with
6 the range and number of comorbid medical and/or psychiatric disorders that were included in STAR*D.
7

8
9 STAR*D's step 1 remission rate was 25.5% followed by a progressive decline in remission rates for those
10 patients receiving subsequent, and increasingly aggressive treatments, such that by step 4 it was only
11 10.4%. This decline in antidepressants' effectiveness essentially mirrors the findings from randomized
12 and naturalistic, prospective studies reporting a 20-30% loss of effectiveness with each increase in the
13 number of prior antidepressant trials.[27-32] Furthermore, several recent analyses suggest that the
14 sequential application of antidepressant medications for non-remitting depression may in fact foster
15 treatment resistance for many patients.[33-36]
16

17
18 Regarding the protocol's predictions of treatment success and patient dropout, it states:
19

20 We arrived at these estimates using three experienced practitioners who
21 independently made estimates that were surprisingly close to each other. Then, via
22 teleconferencing, the final estimates were made. ***The underlying assumptions of***
23 ***these estimates come largely by inferences from results of published RCTs.***[10, p.31;
24 emphasis added]
25

26
27 STAR*D's actual measures of treatment success and failure were significantly worse than predicted. As
28 Barbui et al. noted, antidepressant study dropout rates provide a "hard measure of treatment
29 effectiveness and acceptability"[12, p.296] and STAR*D's dropout rate was 2.6 times greater than
30 predicted. This discrepancy further highlights the relative ineffectiveness of antidepressants in treating
31 real-world depressed patients, compared to those reported in conventional studies.
32

33 *Conclusion*

34
35 Bias in the clinical literature is commonly associated with industry-funded RCTs, not publicly funded
36 ones.[37] Our RIAT reanalysis though documents scientific errors in this NIMH-funded study. These
37 errors inflated STAR*D investigators' report of positive outcomes.
38
39

40 The STAR*D summary article's claim of a 67% cumulative remission rate was published in 2006. If
41 STAR*D's outcomes had been reported as prespecified, its model of care would likely have faced much
42 stronger criticism 16 years ago and fueled a more vigorous search for evidence-based treatment
43 alternatives.
44

45 **Ethics Statement:**

46
47 The study was overseen by NIMH; NCT00021528.
48
49

50 **Data Sharing Statement:**

51
52 Data is available from the NIMH-supported National Database for Clinical Trials (NDCT).
53
54

55 **Funding Statement:**

Funding for this project was provided by The RIAT Support Center.

Competing Interests:

The authors have no relevant financial disclosures or other conflicts of interest to report.

Contributorship Statement:

HEP, JDA and IK contributed to the design of the study and secured funding. TK and CX conducted all of the data analyses. HEP wrote the manuscript with input from JDA, IK, TK, and CX.

Acknowledgements:

We thank Termeh Feinberg for her early efforts on this project, particularly her correspondence with the NIMH help desk to resolve issues with the 26 data files as well as the RIAT Support Center for funding this project. Data used in the preparation of this manuscript were obtained from the controlled access datasets distributed from the NIMH-supported National Database for Clinical Trials (NDCT). NDCT is a collaborative informatics system created by the National Institute of Mental Health to provide a national resource to support and accelerate discovery related to clinical trial research in mental health. The content of this publication does not necessarily reflect the views of the RIAT Support Center nor NIMH.

References:

1. 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. 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. 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. Bupropion-SR, sertraline, or venlafaxine-XR after failure of SSRIs for depression. *N Engl J Med*. 2006;354:1231–1242.
3. 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. Medication augmentation after the failure of SSRIs for depression. *N Engl J Med*. 2006;354:1243–1252.
4. Fava M, Rush AJ, Wisniewski SR, Nierenberg AA, Alpert JE, McGrath PJ, Thase ME, Warden D, Biggs MM, Luther JF, Niederehe G, Ritz L, Trivedi MH. 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.
5. Nierenberg AA, Fava M, Trivedi MH, Wisniewski SR, Thase ME, McGrath PJ, Alpert JE, Warden D, Luther JF, Niederehe G, Lebowitz BD, Shores-Wilson K, Rush AJ. 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.
6. McGrath PJ, Stewart JW, Fava M, Trivedi MH, Wisniewski SR, Nierenberg AA, Thase ME, Davis L, Biggs MM, Shores-Wilson K, Luther JF, Niederehe G, Warden D, Rush AJ. Tranylcypromine versus venlafaxine plus mirtazapine following three failed antidepressant medication trials for depression: a STAR*D report. *Am J Psychiatry*. 2006;163:1531–1541.
7. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, Niederehe G, Thase ME, Lavori PW, Lebowitz BD, McGrath PJ, Rosenbaum JF, Sackheim HA, Kupfer DJ, Luther J, Fava

- 1
2
3 M. Acute and longer-term outcomes in depressed outpatients requiring one or several
4 treatment steps: a STAR*D report. *Am J Psychiatry*. 2006;163:1905–1917.
- 5
6 8. Rush AJ, Fava M, Wisniewski SR, Lavori PW, Trivedi MH, Sackeim HA, Thase ME, Nierenberg AA,
7 Quitkin FM, Kashner TM, Kupfer DJ, Rosenbaum JF, Alpert J, Stewart JW, McGrath PJ, Biggs MM,
8 Shores-Wilson K, Lebowitz BD, Ritz L, Niederehe G. Sequenced treatment alternatives to relieve
9 depression (STAR*D): rationale and design. *Control Clin Trials*. 2004;25:119–142.
- 10
11 9. Rush AJ, Trivedi MH, Ibrahim HM, Carmody T J, Arnow B, Klein DN, ... Manber R. The 16-item
12 Quick Inventory of Depressive Symptomatology (QIDS), Clinician Rating (QIDS-C), and Self-
13 Report (QIDS-SR): a psychometric evaluation in patients with chronic major depression.
14 *Biological Psychiatry*. 2003;54:573– 583.
- 15
16 10. National Institute of Mental Health (NIMH). Sequenced Treatment Alternatives to Relieve
17 Depression (STAR*D) Research Protocol. Washington (DC): NIMH; revised June 28, 2002.
- 18
19 11. Barbui C, Furukawa TA, Cipriani A. Effectiveness of paroxetine in the treatment of acute major
20 depression in adults: a systematic re-examination of published and unpublished data from
21 randomized trials. *CMAJ*. 2008;178:296–305.
- 22
23 12. Pigott HE, Leventhal AM, Alter GS, Boren JJ. Efficacy and effectiveness of antidepressants:
24 current status of research. *Psychother Psychosom*. 2010;79:267–279.
- 25
26 13. Insel TR, Wang PS. The STAR*D trial: revealing the need for better treatments. *Psychiatric
27 Services*. 2009;60:1466-1467.
- 28
29 14. Greden JF. Workplace depression: personalize, partner, or pay the price. *Am J Psychiatry*.
30 2013;170:578–581.
- 31
32 15. Smith DG. Antidepressants don't work the way many people think. *New York Times*. November
33 8, 2022.
- 34
35 16. Pigott HE. The STAR*D trial: it's time to reexamine the clinical beliefs which guide the treatment
36 of major depression. *Canadian J Psychiatry*. 2015;60:9–13.
- 37
38 17. Kirsch I, Huedo-Medina TB, Pigott HE, Johnson B. Do outcomes of clinical trials resemble those
39 "real world" patients? A re-analysis of STAR*D antidepressant data. *Psychology of
40 Consciousness: Theory, Research, and Practice*. 2018 <https://doi.org/10.1037/cns0000164>
- 41
42 18. Pigott HE, Dubin, W, Kirsch I, Amsterdam J. Call to action: RIAT reanalysis of the Sequenced
43 Treatment Alternatives to Relieve Depression (STAR*D) Study. *BMJ*. March 6, 2019;
44 <https://www.bmj.com/content/346/bmj.f2865/rr-10>
- 45
46 19. Doshi P, Dickersin K, Healy D, Vedula SS, Jefferson T. Restoring invisible and abandoned trials: a
47 call for people to publish the findings. *BMJ* 2013;346:f2865.
- 48
49 20. Fava M, Rush AJ, Trivedi MH, et al. Background and rationale for the Sequenced Treatment
50 Alternatives to Relieve Depression (STAR*D) study. *Psychiatr Clin North Am*. 2003;26:457–494.
- 51
52 21. Lavori PW, Rush AJ, Wisniewski SR, Alpert J, Fava M, Kupfer DJ, Nierenberg A, Quitkin FM,
53 Sackeim HA, Thase ME, Trivedi M. Strengthening clinical effectiveness trials: equipoise-stratified
54 randomization. *Biological Psychiatry*. 2001;50:792– 801.
- 55
56 22. Rush AJ, Bernstein IH, Trivedi MH, Carmody TJ, Wisniewski S, Mundt JC, Shores-Wilson K, Biggs
57 MM, Woo A, Nierenberg AA, Fava M. An evaluation of the quick inventory of depressive
58 symptomatology and the hamilton rating scale for depression: a sequenced treatment
59 alternatives to relieve depression trial report. *Biol Psychiatry*. 2006 Mar 15;59(6):493-501. doi:
60 10.1016/j.biopsych.2005.08.022. Epub 2005 Sep 30. PMID: 16199008; PMCID: PMC2929841.
- 61
62 23. Rush AJ. Inventory of Depressive Symptomatology (IDS) and Quick Inventory of Depressive
63 Symptomatology (QIDS). [IDS/QIDS \(ids-qids.org\)](https://ids-qids.org)
- 64
65 24. R Core Team. R: A language and environment for statistical computing. R Foundation for
66 Statistical Computing, Vienna, Austria.

25. Rutherford BR, Sneed JR, Roose SP. Does study design influence outcome? The effects of placebo control and treatment duration in antidepressant trials. *Psychotherapy and Psychosomatics*. 2009;78:172–181.
26. Wisniewski SR, Rush AJ, Nierenberg AA, Gaynes BN, Warden D, Luther JF, McGrath PJ, Lavori PW, Thase ME, Fava M, Trivedi MH: Can phase III trial results of antidepressant medications be generalized to clinical practice? A STAR * D report. *Am J Psychiatry* 2009;166:599–607.
27. Amsterdam JD, Maislin G. Fluoxetine efficacy in treatment resistant depression. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*. 1994;18:243-261.
28. Nierenberg A, Feighner JP, Rudolph R, Cole JO, Sullivan J. Venlafaxine for treatment-resistant unipolar depression. *J Clinical Psychopharmacology*. 1994;14:419-423.
29. Amsterdam JD, Shults J. MAOI safety and efficacy in advanced treatment-resistant depression: a retrospective analysis. *J Affective Disorders*. 2005; 89:183-188.
30. Amsterdam JD, Williams D, Michelson D, Adler LA, Dunner SL, Nierenberg A, Reimherr FW, Schatzberg AF. Tachyphylaxis after repeated antidepressant drug exposures in patients with recurrent major depressive disorder. *Neuropsychobiology*. 2009;59:227-233.
31. Leykin Y, Amsterdam JD, DeRubeis RJ, Shelton RC, Hollon SD. Progressive resistance to SSRI therapy but not to cognitive therapy in the treatment of major depression. *J Consulting & Clinical Psychology*. 2007;75:267-276.
32. Amsterdam JD, Lorenzo-Luaces L, DeRubeis RJ. Step-wise loss of antidepressant effectiveness after repeated antidepressant trials in bipolar II depression. *Bipolar Disorders*. 2016;18:563-570.
33. Fava GA, Offidani E. The mechanisms of tolerance in antidepressant action. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*. 2011;35:1593-602.
34. Amsterdam JD, Lorenzo-Luaces L. Increase in pharmacodynamic tolerance after repeated antidepressant trials in treatment-responsive bipolar II depressed subjects: An exploratory study. *Psychiatria Polska*. 2018;52:957–969.
35. Amsterdam JD, Kim TT. Increased risk of depressive relapse during maintenance therapy after repeated antidepressant trials in treatment-responsive subjects. *J Clinical Psychopharmacology*. 2019;39:344-350.
36. Andrews PW, Amsterdam JD. A hormetic approach to understanding antidepressant effectiveness and the development of antidepressant tolerance. *Psychiatria Polska*. 2020;54:1067–1090.
37. Amsterdam JD, McHenry LB, Jureidini JN. Industry-corrupted psychiatric trials. *Psychiatria Polska*. 2017;51:993–1008.

Figure 1 Caption: Patient Flowchart

Figure 1 Footnote:

* In level 2, 580 patients were randomized to switch medications, 441 to medication augmentation, and 113 to Cognitive Therapy as either a switch or medication augmentation treatment. In level 2A, 28 patients were randomized to one of two level 2 switch medications. For step 3/level 3 patients, 186 were randomized to medication switch and 111 to medication augmentation. For step 4/level 3 patients, 7 were randomized to medication switch and 9 to medication augmentation. For step 4/level 4 patients, 90 were randomized to one of two medication/medication combination switch options.

** Exit refers to the number of patients who exit the study and do not proceed either to the next treatment level nor enter follow-up.

1
2
3 *** Follow-up refers to the number of patients who exit a treatment and enter the 12-month follow-up
4 phase.
5

6 **Figure 2 Caption: Comparison of STAR*D Protocol Predictions to What Occurred**

7
8 **Figure 3 Caption: STAR*D's Step-by-Step Cumulative Remission Rate Presented Three Ways**

9
10
11 **Figure 3 Footnote:**

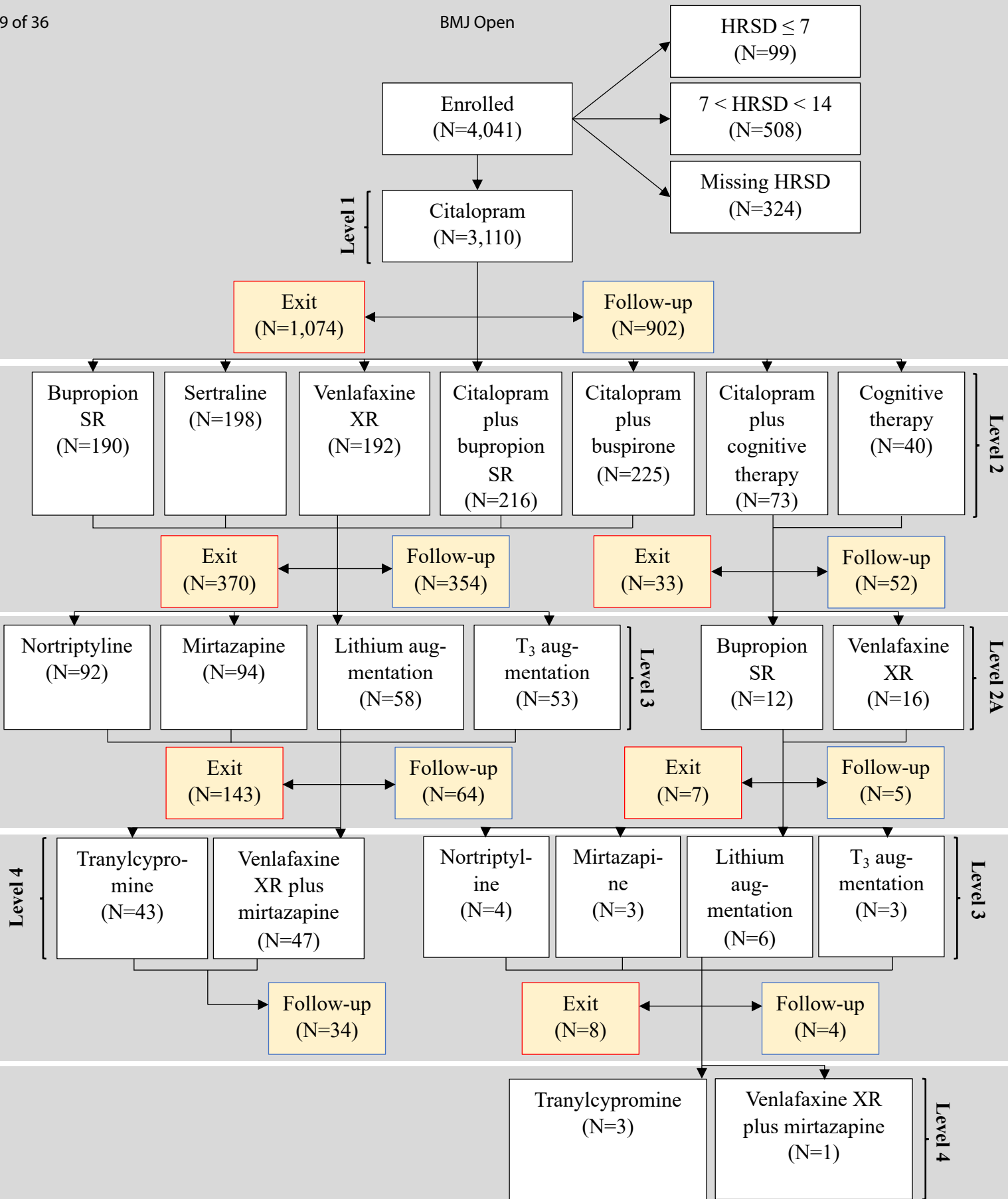
12
13 The step-by-step theoretical remission rates were obtained from the STAR*D summary article where it
14 states: "The theoretical cumulative remission rate is 67% (37+19+6+5)." [7, p.1910].
15

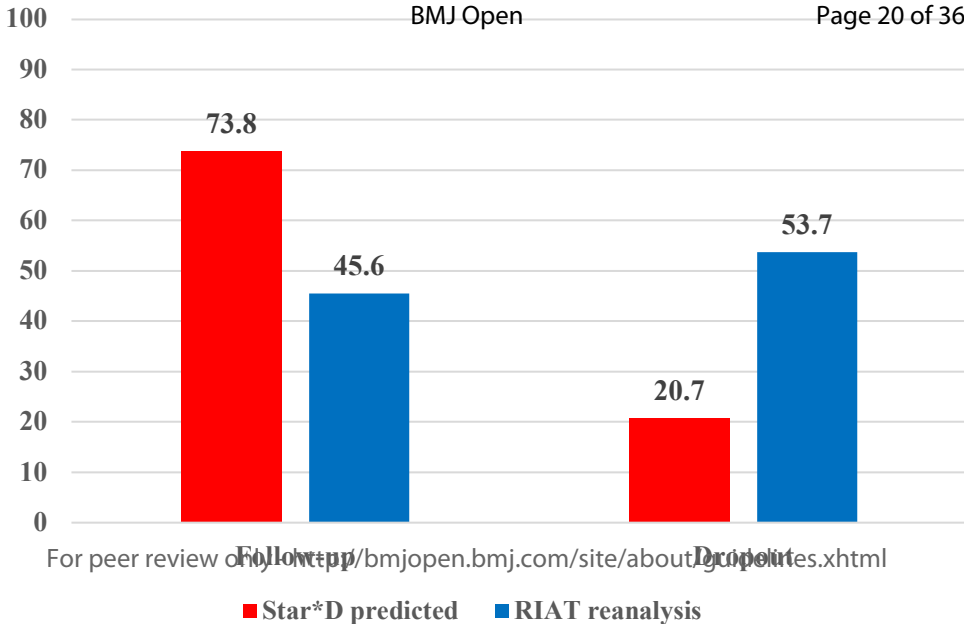
16
17 The HRSD + QIDS-SR cumulative remission rate was taken from Table 1. It combines the 1,089 patients
18 with an exit HRSD score of <8 with the 195 patients who were missing an exit HRSD score but had a final
19 clinic-visit QIDS-SR score of <6.
20

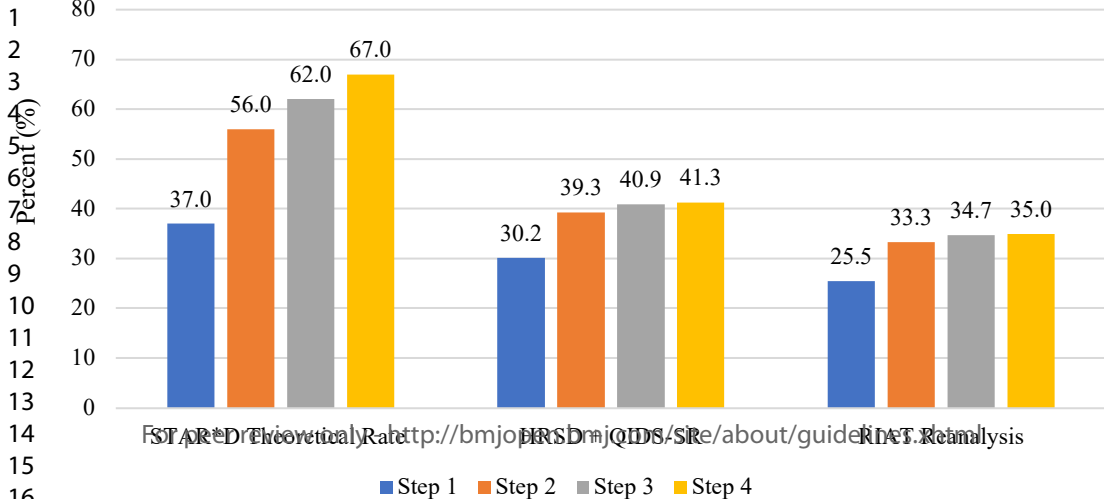
21 The RIAT Reanalysis cumulative remission rate is based on an exit HRSD score of <8 as the sole measure
22 of remission for the 3,110 patients who met STAR*D's inclusion for data analysis criteria.
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Step 1

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18



Appendix 1: Highest Quality of Acute and Continuing-Care to Maximize Remissions While Minimizing Relapse and Dropouts

Descriptor	Explanation
<p>Optimized Sustained Study Participation to Minimize Dropouts ^{20, p. 473-474}</p>	<ul style="list-style-type: none"> • Promoted patients' study affiliation via STAR*D-branded brochures, bimonthly newsletters, and an informational video emphasizing STAR*D's public health significance and the critical role played by patients; • Educated patients and families about depression and its treatment using a multi-step educational package. This included teaching the "mechanism of action" for patients' current antidepressant and educating patients that "depression is a disease, like diabetes or high blood pressure" and "can be treated as effectively as other illnesses," etc.; • Used a letter reminder system to alert patients before appointments in those clinics without such systems who had a >15% rate of missed appointments; • Ensured timely follow-up and rescheduling of missed appointments by calling patients on the day of the missed appointment, and again within 24 hours, if there was no response. Patient's physician sent letter within 48 hours if contact was not established; • Used a letter reminder system for all research outcome assessment calls during acute and continuing-care; • In every clinic visit, the Clinical Research Coordinator (CRC) discussed the research outcomes phone calls with the patient to ensure that the calls were completed on schedule and worked to resolve any problematic issues regarding said calls [Clinical Procedures Manual, page 75]; • Paid patients \$25.00 for participating in each telephonic research outcomes assessment; • Permitted patients to re-enter acute and/or continuing-care within four weeks after having dropped out [Clinical Procedures Manual, page 80]; • Recommended one-year of continuing-care for all patients who achieved a satisfactory clinical response with the essential goal of preventing relapse [Clinical Procedures Manual, page 15] and • Permitted continuing-care patients to remain in the study if they moved from the area [Clinical Procedures Manual, page 81].
<p>Acute-Care Visits</p>	<p>Physicians met with patients on entry into each new step to initiate drug treatment with follow-up visits scheduled on weeks 2, 4, 6, 9, 12, with an optional week 14 visit.</p>

Measurement-Based Care	Conducted structured evaluations of symptoms and side-effects at each visit and included a centralized treatment monitoring and physician feedback system to ensure consistent implementation of optimal care across research sites.
Aggressive Medication Dosing	Provided aggressive medication dosing with a fully adequate dose for a sufficient duration to “ <i>ensure that the likelihood of achieving remission was maximized and that those who did not reach remission were truly resistant to the medication</i> ”. ^{1, p.30}
Liberal Prescribing of Non-Study Medications	Physicians had great leeway in prescribing non-study medications to treat comorbid symptoms resulting in: <ul style="list-style-type: none"> • 17.2% taking Trazodone for sleep; • 11.9% taking an anti-anxiety medication; • 16.7% taking either a sedative or hypnotic medication; and • An undisclosed percent taking medications to address side-effects. ^{2, table 2}
Continuing-Care Visits	Patients saw their physician every 2 months and continued taking their treatment medication(s) at the same doses but their physicians were allowed to make any psychotherapy, medication, and/or medication dose changes to maximize the likelihood of maintaining patients’ remission status. ^{7, p. 1908} Additional continuing-care visits were scheduled when patients began to experience a return of depressive symptoms and/or intolerable side-effects [Clinical Procedures Manual. page 78].
Clinical Research Coordinator (CRC)	Each site had a CRC who: ^{1, p. 30} <ul style="list-style-type: none"> • Saw patients before each visit administering multiple measures to them including the QIDS-SR during each acute-care visit; • Assisted physicians in protocol implementation; and • Provided patients support and encouragement in protocol implementation.
Treatment Designed to Enhance Subject Retention	Treatment was designed to minimize drop-outs and/or non-compliance including: <ul style="list-style-type: none"> • Open label prescribing during acute and continuing-care with no placebo control condition during any study phase; • Patients chose their acceptable treatment assignments for steps two and three to eliminate any concerns they might have about receiving an unacceptable assignment. This resulted in only 21 of 1,439 (1.5%) Step-2 patients making themselves available for random assignment to all treatment options ^{2, p. 1235} while only 29 of 377 (7.7%) did so in Step-3. ^{5, p. 1521} • During each step, patients could enroll immediately into the next step if they had intolerable side-effects or had maximized their current medication(s)’ dosing without achieving a remission; and • During any step, patients could enter continuing-care directly on their current medication(s) if they were treatment responders even if they had not achieved remission. This was done to minimize responders from dropping out in order to avoid having

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

	to discontinue their current medication(s) and start a new drug regimen.
--	--

***Trivedi MH, Stegman D, Rush AJ, Wisniewski SR, Nierenberg AA: STAR * D clinical procedures manual. July 31, 2002. www.edc.pitt.edu/stard/public/study_manuals.html

For peer review only

Appendix 2: Description of Levels 1-4 Treatments

Level 1:

STAR*D investigators report that Citalopram (Celexa) was chosen as the first-line SSRI treatment because (1) absence of discontinuation symptoms; (2) demonstrated safety in elderly and medically fragile patients; (3) easy once-a-day dosing with few dose adjustments; and (4) favorable drug–drug interaction profile.¹ Citalopram was started at 20 mg/day and then raised to 40 mg/day by day 28 and up to 60 mg/day by day 43 and onward. Dose adjustments were based on how long a patient had received a particular dose, symptom changes, and side effect burden.

Level 2 switch treatments:

Citalopram was discontinued without a tapering at the initiation of each level 2 switch treatment. STAR*D investigators chose pharmacologically distinct switch medications. The level 2 treatments were:

- Sertraline (Zoloft), an SSRI with the same pharmacological profile as citalopram. Sertraline was started at a daily dose of 50 mg and increased to 100 mg at day 8, to 150 mg at day 28, and to 200 mg at day 63 and onward.
- Sustained-release bupropion (Wellbutrin SR), an “out-of-class” agent whose neurochemical action mechanisms are unknown; other than that, it does not inhibit serotonin reuptake and is believed to produce antidepressant effects by blocking the reuptake of dopamine and norepinephrine. The daily dose of sustained-release bupropion was 150 mg for week 1, 200 mg from day 8 to 27, 300 mg from day 28 to 41, and 400 mg from day 42 onward.
- Extended-release venlafaxine (Effexor), a “dual-action” agent that inhibits the reuptake of both serotonin and norepinephrine. The starting daily dose of extended-release venlafaxine was 37.5 mg for week 1 and increased to 75 mg from day 8 to 14, to 150 mg from day 15 to 27, to 225 mg from day 28 to 41, to 300 mg from day 42 to 62, and to 375 mg from day 63 onward.
- Cognitive therapy was provided by a trained psychotherapist and scheduled twice weekly for the first four weeks, then once weekly for the remaining 8 weeks (16 sessions total).

Level 2 Citalopram augmentation treatments:

During the augmentation trial, the citalopram dose was kept constant but reduced if side effects developed. The level 2 augmentation treatments were:

- Buspirone (Buspar), a partial agonist at the postsynaptic 5-hydroxytryptamine 1A (5-HT_{1A}) receptor that is believed to enhance the activity of SSRIs through the 5HT_{1A} receptors. The starting dose was 15 mg per day week 1, increasing to 30 mg per day week 2, and then to 45 mg per day for weeks 3 through 5, and a final, maximum dose of 60 mg per day week 6 and onward.
- Sustained-release bupropion (Wellbutrin SR) whose neurochemical action mechanisms are unknown but is believed to produce antidepressant effects by blocking the reuptake of dopamine and norepinephrine. The initial dose was 200 mg per day during weeks 1 and 2, increasing to 300 mg per day by week 4 and to 400 mg per day week 6 and onward.
- Cognitive therapy was provided by a trained psychotherapist and scheduled twice weekly for the first four weeks, then once weekly for the remaining 8 weeks (16 sessions total).

Level 3 switch treatments:

At entry into the Level 3 switch trial, all level 2 medications were discontinued without tapering at the initial Level 3 treatment visit. The level 3 switch treatments were:

- Nortriptyline (Pamelor), a tricyclic antidepressant. Recommended doses were 25 mg/ day for 3 days, 50 mg/day for 4 days, and then 75 mg/day by day 8, 100 mg/day by day 28, and, if necessary, 150 mg/day by day 42 and onward
- Mirtazapine (Remeron), a tetracyclic antidepressant that blocks inhibitory α -adrenoceptors on norepinephrine and serotonin neurons to enhance both norepinephrine and serotonin neurotransmission. Recommended mirtazapine doses were 15 mg/day for the first 7 days, 30 mg/day by day 8, 45 mg/day by day 28, and, if necessary, 60 mg/ day by day 42 and onward.

Level 3 augmentation treatments of level 2 medications:

The two medication augmentation options used in level 2, buspirone and sustained-release bupropion, were discontinued without tapering in the initial level 3 visit. The two medication augmentation treatments in level 3 were added to ongoing treatment with citalopram, sertraline, sustained-release bupropion, or extended-release venlafaxine. The level 3 augmentation treatments were:

- Lithium started at 450 mg/day, and at week 2 it was increased to the recommended dose of 900 mg/day. If participants could not tolerate the initial dose, it could be reduced to 225 mg/day for 1 week then increased to 450 mg/day. There was no monitoring of lithium levels.
- Triiodothyronine (T₃), a thyroid hormone, started at 25 μ g/day for 1 week and then increased to the recommended dose of 50 μ g/ day. There was no pretreatment assessment, nor ongoing monitoring, of thyroid functioning.

Level 4 switch treatments:

The level 4 switch treatments were:

- Tranylcypromine (Parnate), a monoamine oxidase inhibitor. A 2-week washout period of Level 3 medications was required for patients assigned to the tranylcypromine group. The recommended dosing for tranylcypromine was 10 mg/day for the first 2 weeks, followed by weekly increases of 10 mg/day until a maximum of 60 mg/day.
- Co-administered venlafaxine (Effexor) and mirtazapine (Remeron) to inhibit the reuptake of both serotonin and norepinephrine and block inhibitory α_2 -adrenoceptors on both norepinephrine and serotonin neurons to enhance both norepinephrine and serotonin neurotransmission. Level 3 medications were discontinued without tapering for patients assigned to this group. The dosage of extended-release venlafaxine was 37.5 mg/day for the first week, 75 mg/day for the second week, 150 mg/day for weeks 3–5, 225 mg/day for weeks 6–8, and 300 mg/day onward. Mirtazapine was started at 15 mg/day for the first 3 weeks, 30 mg/day for weeks 4 to 8, and then 45 mg/day onward.

Appendix 3: Number of Level 2-4 Participants Excluded from our RIAT Reanalysis, and the Reasons for their Exclusion, yet Included in STAR*D

Level 2 Treatments

Number of Level 2 Participants Excluded from our Reanalysis but Included in STAR*D	Bup	Sert	Ven	CT	Cit + BUP	Cit + Busp	Cit + CT	Total
Scored as Remitted at ENTRY into Level 2 yet still included in STAR*D's Level 2 analyses	22	8	14	7	30	24	4	109
Scored as only mildly depressed (HRSD >7 & <14) at entry into Level 1, and therefore excluded from STAR*D's data analysis, yet still treated in Level 1, progressed to Level 2, and then included in STAR*D's Level 2 data analyses	21	15	25	15	20	30	7	133
Scored as Remitted at entry into Level 1 (HRSD ≤ 7), and therefore excluded from STAR*D's data analysis, yet still treated in Level 1 and progressed to Level 2 and then included in STAR*D's Level 2 data analyses	6	1	4	2	2	2	2	19
Missing baseline HRSD at entry into Level 1, and therefore excluded from STAR*D's data analysis, yet still treated in Level 1, and progressed to Level 2, and then included in STAR*D's Level 2 data analyses	12	18	22	4	16	13	1	86
Number meeting 2 exclusion criterions	12	2	7	6	5	8	2	42

Bup=Sustained-release Bupropion; Sert= Sertraline; Ven= Extended-release Venlafaxine; CT=Cognitive Therapy; Cit+BUP= Citalopram + Sustained-release Bupropion; Cit+Busp=Citalopram + Buspirone; Cit+CT= Citalopram + Cognitive Therapy

Level 3 Treatments

	Nortriptyline	Mirtazapine	Lithium Augmentation	Triiodothyronine Augmentation	Total
Scored as Remitted at ENTRY into Level 3 yet still included in STAR*D's Level 3 analyses	4	0	1	5	10
Scored as only mildly depressed (HRSD >7 &	8	5	3	4	20

<14) at entry into Level 1, and therefore excluded from STAR*D's data analysis, yet still treated in Level 1, progressed to Level 2, and then 3 and included in STAR*D's Level 3 data analyses					
Scored as Remitted at entry into Level 1 (HRSD ≤ 7), and therefore excluded from STAR*D's data analysis, yet still treated in Level 1 and progressed to Level 2 and then 3 and included in STAR*D's Level 3 data analyses	2	1	0	1	4
Missing baseline HRSD at entry into Level 1, and therefore excluded from STAR*D's data analysis, yet still treated in Level 1, and progressed to Level 2, and then level 3 and included in STAR*D's Level 3 data analyses	7	8	1	7	23

Level 4 Treatments

	Tranlylcypromine	Venlafaxine + Mirtazapine	Total
Scored as Remitted at ENTRY into Level 4 yet still included in STAR*D's Level 4 analyses	5	1	6
Scored as only mildly depressed (HRSD >7 & <14) at entry into Level 1, and therefore excluded from STAR*D's data analysis, yet still treated in Level 1, progressed to Level 2, 3 and then 4 and included in STAR*D's Level 4 data analyses	3	1	4
Scored as Remitted at entry into Level 1 (HRSD ≤ 7), and therefore excluded from STAR*D's data analysis, yet still treated in Level 1 and progressed	0	0	0

to Level 2, then Level 3 and included in STAR*D's Level 4 data analyses			
Missing baseline HRSD at entry into Level 1, and therefore excluded from STAR*D's data analysis, yet still treated in Level 1, and progressed to Level 2, and then Level 3 and 4 included in STAR*D's Level 4 data analyses	5	1	6

For peer review only

Appendix 4: Number and Percent of Participants Missing Entry and/or Exit HRSD Scores

	#/(%) with Missing Entry HRSD	#/(%) with Missing Exit HRSD
Step 1 (N=3,110)	0 (0%)	926 (29.8%)
Step 2 (N=1,134)	168 (14.8%)	304 (26.8%)
Switch strategy (N=620)	90 (14.5%)	183 (29.5%)
Bupropion (N=190)	34	58
Sertraline (N=198)	27	56
Venlafaxine (N=192)	24	56
Cognitive Therapy (N=40)	5	13
Augmentation strategy (N=514)	78 (15.2%)	121 (23.5%)
Bupropion (N=216)	35	58
Bupirone (N=225)	37	52
Cognitive Therapy (N=73)	6	11
Step 3 (N=325)	42 (12.9%)	78 (24%)
Level 2A (N=28)	3	6
Bupropion (N=12)	3	2
Venlafaxine (N=16)	0	4
Level 3 (N=297)	39	72
Switch strategy (N=186)	26	49
Nortriptyline (N=92)	11	23
Mirtazapine (N=94)	15	26
Augmentation strategy (N=111)	13	23
Lithium (N=58)	9	13
Bupropion SR (N=17)	3	2
Citalopram (N=22)	5	6
Sertraline (N=11)	1	3
Venlafaxine XR (N=8)	0	2
T3 (N=53)	4	10
Bupropion SR (N=6)	1	0
Citalopram (N=26)	1	7
Sertraline (N=8)	1	1
Venlafaxine XR (N=13)	1	2
Step 4 (N=106)	15 (14.2%)	22 (20.8%)

Level 3 (N=16)	3	3
Tranlycypromine (N=43)	7	10
Venlafaxine XR/mirtazapine (N=47)	5	9
Total Across Treatment Steps	225	1,330

For peer review only

Appendix 5: Baseline Demographic and Clinical Features by Treatment Step

	Treatment Step ^a							
	Step 1 (N=3,110)		Step 2 (N=1,134)		Step 3 (N=325)		Step 4 (N=106)	
Demographic Features	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age	41.0	13.0	42.0	12.6	44.1	12.0	46.9	11.0
Education (years)	13.6	3.2	13.2	3.3	12.8	3.1	12.6	2.3
Monthly household income	2,289	2,732	1,744	1,539	1,470	1,383	1,003	887
	N	%	N	%	N	%	N	%
Female	1,469	74.6	502	73.1	113	65.3	34	65.4
Race								
White	2,328	74.9	870	76.7	259	79.7	86	81.1
Black	333	10.7	115	10.1	29	8.9	7	6.6
Other	449	14.4	149	13.1	37	11.4	13	12.3
Hispanic	402	12.9	139	12.3	45	13.8	16	15.1
Employment status								
Employed	975	58.7	314	54.2	69	46.9	19	43.2
Unemployed	612	36.9	243	42.0	72	49.0	24	54.5
Retired	73	4.4	22	3.8	6	4.1	1	2.3
Medical insurance								
Private	848	52.2	254	44.5	52	36.6	14	31.8
Public	282	17.4	109	19.2	30	21.4	10	23.3
None	534	33.2	223	39.3	60	43.2	20	46.5
Marital status								
Single	475	28.6	171	29.5	40	27.2	10	22.7
Married/cohabiting	716	43.1	238	41.0	61	41.5	18	40.9
Divorce/separated	429	25.8	155	26.7	42	28.6	14	31.8
Widowed	41	2.5	16	2.8	4	2.7	2	4.5
Clinical Features	N	%	N	%	N	%	N	%
First episode occurrence before age 18	1,200	39.0	436	38.8	120	37.0	41	38.7
Recurrent depression	1,940	66.8	718	68.3	188	63.3	59	60.8
Family history of depression	1,694	55.4	609	54.9	165	51.7	58	54.7

Duration of current episode \geq 2 years	787	25.6	311	27.7	88	27.2	34	32.1
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age at first episode (years)	24.9	14.5	24.7	14.2	25.9	14.6	25.9	14.3
Illness duration (years)	16.1	13.5	17.2	13.7	18.2	14.1	21.0	14.9
Number of episodes	4.4	9.7	4.9	11.1	4.4	10.3	5.0	12.1
Duration of current episode (months)	25.9	52.0	28.1	58.8	32.1	68.5	45.9	82.0
Median duration of current episode (months)	8.3		8.7		9.5		10.1	
Quality of Life and Enjoyment Satisfaction Questionnaire score ^b	39.1	14.3	36.5	13.6	33.7	13.5	31.6	12.8
SF-12 Mental ^c	25.6	8.1	25.0	7.7	24.4	7.7	24.0	7.5
SF-12 Physical ^c	48.6	12.1	47.0	12.4	44.5	12.1	43.8	12.3
Work and Social Adjustment Scale score ^d	25.0	8.7	26.3	8.2	28.3	7.7	29.4	7.0
HRSD ₁₇ score	21.9	5.2	22.5	5.2	23.4	5.2	23.9	5.4
IDS-C ₃₀ score ^e	39.1	9.6	40.6	9.7	42.6	9.4	43.6	9.8
QIDS-IVR score ^f	16.9	3.3	17.3	3.3	17.9	3.0	18.3	3.1
Cumulative Illness Rating Scale								
Categories endorsed	2.5	1.5	2.6	1.6	2.8	1.6	3.1	1.6
Total score	4.7	3.9	5.1	4.0	5.8	4.5	6.2	4.3
Severity score	1.8	0.8	1.8	0.8	2.0	0.9	2.0	0.7
	N	%	N	%	N	%	N	%
Psychiatric Diagnostic Screening Questionnaire								
Agoraphobia	559	18.2	240	21.4	89	27.5	32	30.2
Alcohol abuse/dependence	371	12.0	136	12.1	36	11.1	8	7.5
Bulimia	607	19.7	232	20.6	67	20.7	20	18.9

Drug abuse/dependence	234	7.6	80	7.1	21	6.5	7	6.6
Generalized anxiety disorder	736	23.9	290	25.8	94	29.0	36	34.0
Hypochondriasis	336	10.9	139	12.4	45	13.9	14	13.2
Obsessive-compulsive disorder	723	23.5	265	23.6	97	29.9	31	29.2
Panic disorder	422	13.7	183	16.3	65	20.1	21	19.8
Posttraumatic stress disorder	387	12.6	172	15.3	55	17.0	16	15.1
Social phobia	963	31.3	379	33.7	117	36.1	35	33.0
Somatoform disorder	284	9.2	105	9.3	35	10.8	9	8.5
Number of axis I comorbid psychiatric disorders								
0	606	19.7	190	16.9	48	14.8	12	11.3
1	740	24.0	257	22.9	68	21.0	23	21.7
2	577	18.7	217	19.3	62	19.1	25	23.6
3	363	11.8	139	12.4	42	13.0	14	13.2
4+	793	25.8	321	28.6	104	32.1	32	30.2

^a Sums do not always equal N due to missing values. Percentages are based on available data.

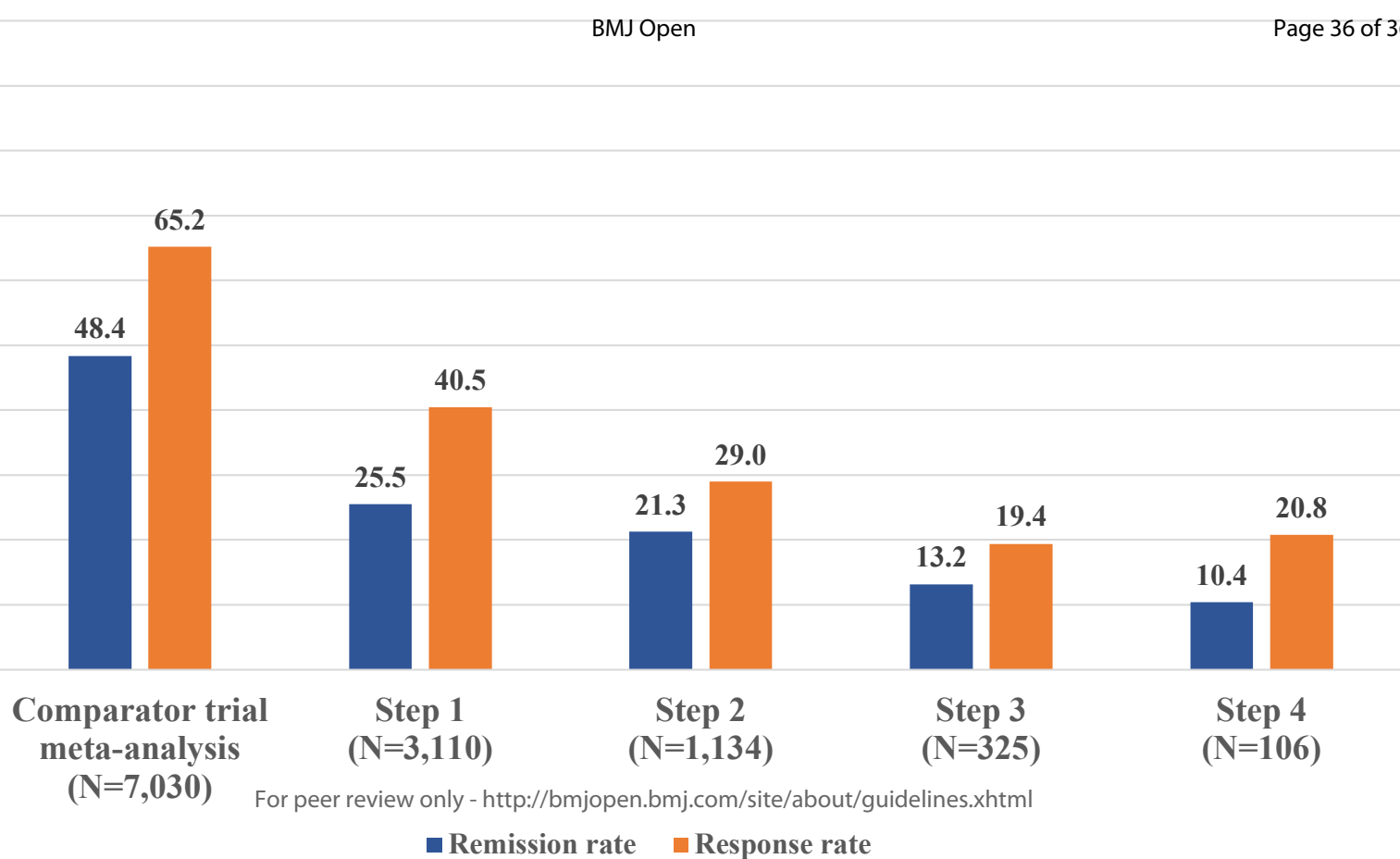
^b Integrated voice response (IVR) administered version of the Quality-of-Life Enjoyment and Satisfaction Questionnaire assessing participants' global rate of satisfaction. Higher scores (range=0–100) represent greater life enjoyment and satisfaction.

^c IVR-administered version of the SF-12 assessing perceived mental and physical health status. Two subscales (physical health factor and mental health) range from 0 to 100— higher scores indicate better functioning with a population norm for each score of 50.

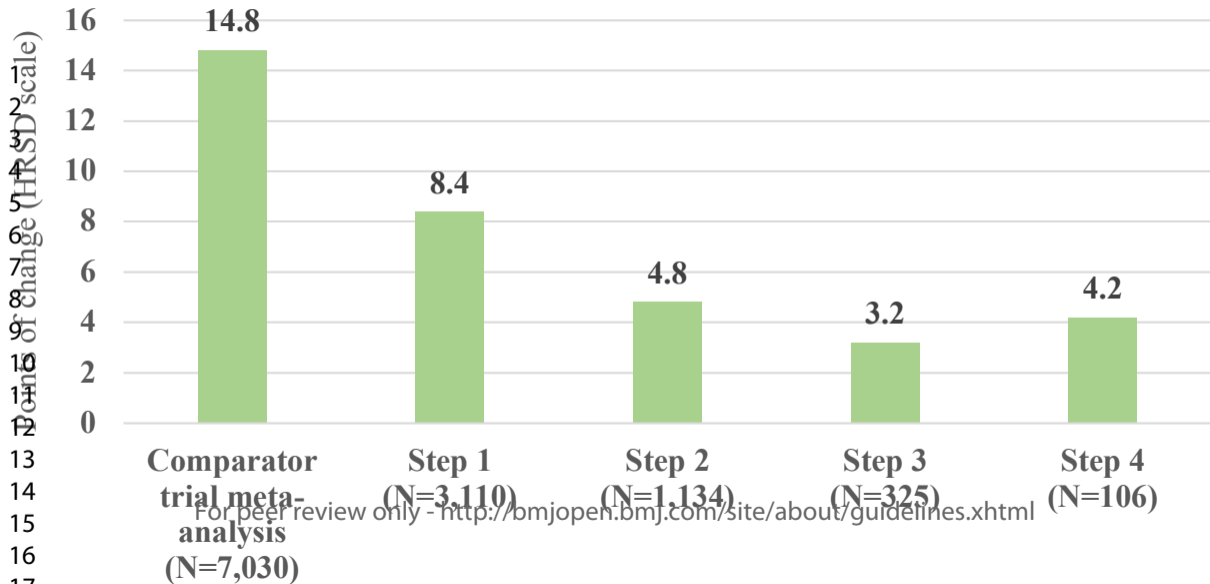
^d IVR-administered version of the Work and Social Adjustment Scale. Scores between 10 and 20 are associated with significant functional impairment while scores above 20 suggest moderate to severe functional impairment.

^e Inventory of Depressive Symptomatology administered telephonically.

^f IVR-administered version of the QIDS.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27For peer review only - <http://bmjopen.bmj.com/site/about/guidelines.xhtml>

■ Remission rate ■ Response rate



BMJ Open

What are the Treatment Remission, Response, and Extent of Improvement Rates after up to Four Trials of Antidepressant Therapies in Real-World Depressed Patients? A Reanalysis of the STAR*D Study's Patient-Level Data with Fidelity to the Original Research Protocol

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2022-063095.R3
Article Type:	Original research
Date Submitted by the Author:	10-May-2023
Complete List of Authors:	Pigott, H.; None Kim, Thomas; University of Pennsylvania, Psychology Xu, Colin; University of Pennsylvania, Psychology Kirsch, Irving; Harvard Medical School Amsterdam, Jay; University of Pennsylvania Perelman School of Medicine, Psychiatry
Primary Subject Heading:	Mental health
Secondary Subject Heading:	Pharmacology and therapeutics
Keywords:	Adult psychiatry < PSYCHIATRY, Depression & mood disorders < PSYCHIATRY, CLINICAL PHARMACOLOGY

SCHOLARONE™
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

**What are the Treatment Remission, Response, and Extent of
Improvement Rates after up to Four Trials of Antidepressant Therapies in Real-World
Depressed Patients? A Reanalysis of the STAR*D Study's Patient-Level Data
with Fidelity to the Original Research Protocol**

Authors:

H. Edmund Pigott, Thomas T. Kim, Colin Xu, Irving Kirsch, & Jay D. Amsterdam

H. Edmund Pigott, Ph.D.

Position: Clinical Psychologist

Institution: none

Email: HPIGOTT75@gmail.com

Thomas T. Kim

Position: PhD candidate in Psychology

Institution: Department of Psychology, University of Pennsylvania

Colin Xu

Position: PhD candidate in Psychology

Institution: Department of Psychology, University of Pennsylvania

Irving Kirsch, Ph.D.

Position: Associate Director, Program in Placebo Studies

Institution: Harvard Medical School

Jay D. Amsterdam, MD

Position: Professor of Psychiatry (Emeritus), Depression Research Unit

Institution: Perelman School of Medicine, University of Pennsylvania

Corresponding author: H. Edmund Pigott, Ph.D.

Abstract

Objective: Reanalyze the patient-level dataset of the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study with fidelity to the original research protocol and related publications.

Design: The study was open label and semi-randomized examining the effectiveness of up to four optimized, and increasingly aggressive, antidepressant therapies in depressed adults. Patients who failed to gain adequate relief from their level 1 trial on the SSRI citalopram could receive up to three additional treatment trials in levels 2-4.

Setting: 41 North American psychiatry and primary care treatment centers.

Participants: 4,041 adults screened positive for major depressive disorder. In contrast to most clinical trials, STAR*D enrolled patients seeking care (versus recruited) and included patients with a wide range of common co-morbid medical and psychiatric conditions to enhance the generalizability of findings to real-world clinical practice.

Interventions: STAR*D evaluated the relative effectiveness of 13 antidepressant therapies in treatment levels 2-4 for depressed patients who failed to gain adequate benefit from their level 1 medication trial.

Main Outcome Measures: According to the STAR*D protocol, the primary outcome was remission, defined as a score <8 on the blinded Hamilton Rating Scale for Depression (HRSD). Response was a secondary outcome defined as ≥50% reduction in HRSD scores. STAR*D's protocol specifically excluded all non-blinded clinic-administered assessments from use as research outcome measures.

Results: STAR*D investigators did not use the protocol-stipulated HRSD to report cumulative remission and response rates in their summary article, and instead used a non-blinded clinic-administered assessment. This inflated their report of outcomes, as did their inclusion of 99 patients who scored as remitted on the HRSD at study outset as well as 125 who scored as remitted when initiating their next-level treatment. These patients should have been excluded from data analysis. In contrast to the STAR*D-reported 67% cumulative remission rate after up to four antidepressant treatment trials, the rate was 35.0% when using the protocol-stipulated HRSD and inclusion in data analysis criteria.

Conclusion: STAR*D's cumulative remission rate was approximately half of that reported.

Strengths and limitations of this study

- We reanalyzed the largest ever prospective antidepressant trial's patient-level dataset with fidelity to the original research protocol and related publications.
- The reanalysis was conducted under the guidelines of the *Restoring Invisible and Abandoned Trials* (RIAT) initiative.

- Treatment remission, response, and extent of symptom improvement rates were calculated for 14 antidepressant therapies for those patients who met STAR*D's inclusion in data analysis criteria as well as the overall cumulative remission rate after up to four trials of antidepressant therapies.
- We calculated STAR*D's remission rate using the protocol-stipulated HRSD as well as combining the HRSD remissions with those from a non-stipulated measure of remission for patients missing an exit HRSD score. Combining STAR*D's HRSD-defined remissions with those from the non-stipulated measure increased its cumulative remission rate from 35.0% to 41.3%.
- Finally, we compared STAR*D's outcomes to those found in a meta-analysis of 7,030 patients enrolled in similar open-label antidepressant comparator trials. Whereas the treatment remission and response rates in comparator trials averaged 48.4% and 65.2% respectively, they were only 25.5% and 40.5% for STAR*D's level 1 patients and worse in treatment levels 2-4. Similarly, comparator trials patients' mean change on the HRSD was 14.8 points versus 8.4 points for STAR*D's level 1 patients and worse for patients in treatment levels 2-4.

Introduction

At a cost of 35 million US dollars, the National Institute of Mental Health (NIMH) funded Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study is the largest and most expensive prospective antidepressant trial ever conducted with over 100 journal articles published by study investigators.[1-7] In contrast to most clinical trials that enroll symptomatic volunteers (typically recruited through advertising), STAR*D enrolled 4041 patients who screened positive for major depressive disorder (MDD) while seeking routine medical or psychiatric care. STAR*D did not exclude patients with medical conditions and most comorbid psychiatric disorders, thereby increasing the generalizability of its findings to real-world clinical practice.

The STAR*D study provided up to four treatment trials per patient and was designed to give guidance in selecting the best next-level treatment option for the many patients who fail to gain sufficient relief from their first, and/or subsequent, antidepressant trial. To mimic clinical practice, STAR*D used an open-label research design with no control group during any phase of the study.

Our STAR*D reanalysis examines key methodological deviations from its research protocol and related publications, and these deviations' impact on its investigators' report of outcomes. In STAR*D's Rationale and Research Design article, and repeated in the level 1-4 published study outcomes, STAR*D investigators stated, "*the primary outcome is depressive symptom severity, measured by the 17-item Hamilton Rating Scale for Depression (HRSD)*." [8, p. 120]. STAR*D's prespecified primary outcome was remission, defined as scoring <8 on the HRSD which was administered telephonically by Research Outcome Assessors (ROAs) blind to patients' study status (treatment level entry/exit/follow-up). Response was a secondary outcome defined as a ≥50% reduction in patients' HRSD scores. Remission as defined by the HRSD (according to the protocol) was not presented in STAR*D's summary article.[7] Furthermore, despite its investigators' numerous publications, neither change in HRSD depressive symptom severity nor HRSD response rates have been reported for STAR*D's six primary studies [1-6] and summary article.[7] Instead, response rates and change in symptom severity were reported using the clinic-administered Quick Inventory of Depressive Symptomatology–Self Report (QIDS-SR), a measure developed by the STAR*D principal investigators.[9] This occurred despite the fact that STAR*D's research protocol specifically excluded all clinic-administered assessments, such as the

1
2
3 QIDS-SR, from use as research outcome measures since they were not blinded and instead, used to
4 guide patient care. The protocol states:
5

6 *Recall that the research outcomes assessments are distinguished from assessments*
7 *conducted at clinic visits. The latter are designed to collect information that guides*
8 *clinicians in the implementation of the treatment protocol. Research outcomes*
9 *assessments are **not** collected at the clinic visits. They are not collected by either*
10 *clinicians or Clinical Research Coordinators.*[10,p.47-48; emphasis in the original]
11
12

13 In their summary article, STAR*D investigators used the QIDS-SR as the sole measure to report
14 remission, response, and extent of symptom improvement. This article's Abstract states that "the
15 overall cumulative remission rate was 67%" with no qualifiers to this claim.[7, p.1905] Besides making
16 this claim based on an assessment the protocol specifically excluded from use as a research measure, it
17 is not until the article's Results section that readers learn this high level of treatment success did not
18 occur. The STAR*D investigators' claim was theoretical—an estimate based on the provisos of what
19 would have happened if there were no study dropouts, and furthermore, "that those who exited the
20 study would have had the same remission rates as those who stayed in the protocol." [7, p.1910] As
21 Pigott et al. documented though, the investigators' assumptions are not true in the real world since
22 more patients dropped out than remitted in each STAR*D treatment level,[11] and furthermore, it has
23 been found in placebo-controlled trials that patients who drop out are more likely to have had adverse
24 treatment side effects and/or emergent suicidality.[12]
25
26

27
28 Unfortunately, the STAR*D investigators' claim of a 67% cumulative remission rate has
29 become accepted clinical wisdom, and the provisions on which it is based are commonly not
30 referenced when portraying STAR*D's findings. For example, in 2009 NIMH's Director Dr.
31 Thomas Insel claimed STAR*D found "at the end of 12 months, with up to four treatment
32 steps, roughly 70% of participants were in remission." [13, p.1466] Similarly in 2013, an
33 editorial in the *American Journal of Psychiatry (AJP)* claimed STAR*D found "after four
34 optimized, well-delivered treatments, approximately 70% of patients achieve remission." [14,
35 p.580]. More recently (2022), a New York Times' article claimed that half of STAR*D's
36 participants "had significantly improved after using either the first or second medication, and
37 nearly 70 percent of people had become symptom-free by the fourth antidepressant." [15]
38 These are not factual statements of STAR*D's findings.
39
40

41 The first author has made published criticisms alleging protocol violations that appear to inflate
42 STAR*D's findings and called for the reanalysis of the dataset by independent investigators.[16] In 2018,
43 the first and fourth authors collaborated with researchers from the University of Connecticut to
44 reanalyze STAR*D's level 1 data obtained from NIMH.[17] This reanalysis found substantial inflation of
45 STAR*D's reported remission and response rates. Furthermore, the reanalysis found that the extent of
46 HRSD improvement in STAR*D's level 1 trial was approximately half that of open-label antidepressant
47 comparator trials.
48
49

50 Our published criticisms of STAR*D investigators' report of outcomes are as follows:[18]
51

- 52 • While STAR*D investigators used the HRSD to report remission rates in their levels 1-4
53 articles,[1-6] the QIDS-SR was used as the sole measure to report remission, response, and
54 extent of improvement rates in their summary article[7] without disclosing that the protocol
55 specifically excluded all non-blinded/clinic-administered assessments such as the QIDS-SR from
56
57
58
59
60

1
2
3 use as outcome measures. The primary outcome measure, the HRSD, should have been used to
4 report the summary article's outcomes.

- 5 • Using data from the 931 patients deemed ineligible for analysis in STAR*D's level 1 article
6 because these patients lacked a baseline ROA-administered HRSD score of ≥ 14 , in STAR*D's
7 levels 2-4 and summary articles without clear disclosure. This included 99 patients who scored
8 < 8 on their baseline HRSD—indicating these patients met STAR*D's remission criterion at study
9 outset and should not have been included in their report of outcomes.
- 10 • Excluding from analysis 370 patients who dropped out after starting on citalopram in their first
11 clinic visit without taking the exit HRSD despite STAR*D investigators stating, "our primary
12 analyses classified patients with missing exit HRSD scores as nonremitters a priori." [1, p.34]
13 These 370 early dropout patients should have been counted as nonremitters as prespecified.
- 14 • Including in their analyses 125 patients who scored as remitted at entry into their next-level
15 treatment. This occurred despite STAR*D investigators prespecifying that, "patients who begin
16 a level with HRSD < 8 will be excluded from analyses." [8, p.130]
17
18
19

20 This article reanalyzes STAR*D's treatment remission, response, and extent of improvement after up to
21 four trials of antidepressant therapies, using STAR*D's protocol-specified primary outcome measure, the
22 ROA-administered HRSD. This effort builds on Pigott et al's 2010 article [11] that focused on
23 deconstructing STAR*D investigators' levels 1-4 and summary articles, [1-7] by reanalyzing STAR*D's
24 patient-level dataset obtained from NIMH in 2019 with fidelity to the original research protocol and
25 inclusion in data analysis criteria. Future efforts will focus on reanalyzing STAR*D's levels 2-4 semi-
26 randomized comparator trials, including the extent of emergent suicidal ideation and 12-month follow-
27 up outcomes tied to each compared treatment.
28
29

30 Method

31 *RIAT Initiative*

32
33 The Restoring Invisible and Abandoned Trials (RIAT) initiative started in 2013 calling on funders
34 and investigators of abandoned (unpublished) or misreported studies to publish undisclosed
35 outcomes or correct misleading publications. [19] If investigators failed to correct a study
36 identified as misreported, independent investigators were encouraged to correct the record by
37 reanalyzing the study's patient-level dataset consistent with the research protocol and analytic
38 plan.
39
40
41

42 On March 6, 2019, the RIAT investigators published our response to a 'Call to Action' statement in the
43 *British Medical Journal*, in which we stated our intention to reanalyze the STAR*D dataset. [18] We then
44 notified STAR*D's principal investigators of our intention and requested they inform us whether they
45 would undertake a reanalysis of the dataset adhering to the research protocol. On March 22, 2019,
46 STAR*D investigators acknowledged our email notification, indicated the STAR*D data were in the public
47 domain, and stated they had no interest in undertaking a reanalysis.
48
49

50 In July 2019, we received a STAR*D Data Use Certificate, issued by the NIMH Data Archive Data Access
51 Committee, and gained access to the STAR*D levels 1-4 and follow-up patient-level dataset consisting of
52 26 text files, and limited supporting study documentation. In September 2019, we obtained funding
53 from the RIAT Support Center to reanalyze STAR*D.
54

55 *Patients*

1
2
3
4 STAR*D patients were 18 to 75 years of age, seeking care at 18 primary and 23 psychiatric care clinics.
5 Clinical research coordinators (CRCs) screened 4,790 patients for MDD. This screening included the
6 CRCs' administrating the HRSD, on which 4,041 patients scored ≥ 14 , met the other inclusion criteria, and
7 enrolled into the study. CRCs also gathered patients' psychiatric history, demographic information, and
8 administered both the Cumulative Illness Rating Scale and the Psychiatric Diagnostic Screening
9 Questionnaire to determine the extent of comorbid medical and psychiatric disorders.
10

11 12 *Levels/Steps of Acute Treatment*

13
14 STAR*D investigators sought to provide the highest quality of care to maximize the number of
15 remissions while minimizing dropouts (see Supplemental Table 1). Supplemental Table 2 describes the
16 antidepressant therapies available in treatment levels 1-4 while steps refer to the numeric order of
17 treatments. As seen in Figure 1, treatment steps 1 and 2 correspond to levels 1 and 2 treatments.
18 Similarly, for most patients their levels 3 and 4 treatments correspond to treatment steps 3 and 4. For
19 level/step 2 patients though who failed to respond adequately to cognitive therapy alone or combined
20 with citalopram and chose to continue in the study, their third treatment step was designated level 2A
21 and they were randomized to one of two level 2 switch medications. For these patients, their level 2A
22 treatment was their third treatment step. For level 2A patients who did not adequately benefit from
23 this medication trial and chose to continue in the study, they entered a fourth treatment step consisting
24 of level 3 treatments.
25
26

27
28 All patients were administered the SSRI citalopram for their level 1 treatment. Each treatment level
29 consisted of 12 weeks of antidepressant therapy, with an additional 2 weeks for patients deemed close
30 to remission. Treatment was administered using a system of measurement-based care that assessed
31 symptoms and side effects at each clinic visit. STAR*D investigators state, "*To enhance the quality and*
32 *consistency of care, physicians used the clinical decision support system that relied on the measurement*
33 *of symptoms (QIDS-C and QIDS-SR), side-effects, medication adherence, and clinical judgment based on*
34 *patient progress.*"[1, p.30] This system was used to guide medication management of a fully adequate
35 dose for a sufficient time to "*ensure that the likelihood of achieving remission was maximized and that*
36 *those who did not reach remission were truly resistant to the medication.*"[1, p.30]
37
38

39 For those patients who failed to gain an adequate response from citalopram, STAR*D allowed them to
40 select acceptable treatment options for randomization in levels 2 to 4 "*to empower patients, strengthen*
41 *the therapeutic alliance, optimize treatment adherence, and improve outcome*"[20, p.483]. The
42 treatment options available for randomization involved either switching to a new treatment or
43 augmenting the patient's current treatment. Treatment levels 2 to 4 evaluated the relative
44 effectiveness of 11 pharmacologically distinct drug/drug combination treatments. Cognitive therapy
45 was also available as either a switch or citalopram augmentation option in level 2.
46
47

48 *STAR*D Follow Up Phase*

49
50 In each treatment trial for levels 1-4, patients who scored < 6 on their last QIDS-Clinician version (QIDS-C)
51 were considered clinician-rated remissions and encouraged to enter the 12-month follow-up phase.
52 During follow-up, patients continued their "previously effective acute treatment medication(s) at the
53 doses used in acute treatment but that any psychotherapy, medication, or medication dose change
54 could be used." [7, p.1908] Based on prior research, a QIDS score of < 6 was estimated by STAR*D
55 investigators to correspond to a score of < 8 on the HRSD, STAR*D's prespecified primary outcome
56
57
58
59
60

1
2
3 measure for classifying patients as remitted.[9] Clinicians strongly encouraged patients who did not
4 obtain a QIDS-defined remission to enter the next-level treatment. Patients who failed to attain a QIDS-
5 defined remission, but did have a $\geq 50\%$ reduction on the QIDS-C and did not want to be randomized to a
6 next-level treatment, were also encouraged to enter follow-up.
7

8 9 *Research Design of the STAR*D Study*

10
11 STAR*D investigators developed a new research design for the study termed “equipoise-stratified” to
12 evaluate the relative efficacy of 13 antidepressant therapies in levels 2-4 for depressed patients who
13 failed to gain adequate benefit from their level 1 medication trial.[21] In level 1, all patients received
14 citalopram as their first treatment. In level 2, patients were informed regarding seven treatment
15 options to choose from: four switch options in which citalopram was stopped and the new treatment
16 initiated and three augmentation options in which citalopram was combined with a second
17 antidepressant treatment. In level 3, patients were informed regarding four treatment options to
18 choose from: two switch options and two augmentation options. Level 4 involved randomization to one
19 of two medication/medication combination switch options.
20
21

22 *Analytic Plan of the RIAT Reanalysis*

23
24 We reanalyzed the STAR*D patient-level dataset with fidelity to the original research protocol wherever
25 possible. Where the protocol was silent, we used other STAR*D publications to guide our analysis. This
26 occurred four times. First, the protocol is silent regarding patients who entered the study without a
27 baseline ROA-administered HRSD score of ≥ 14 . In their level 1 article, STAR*D investigators deemed the
28 931 such patients who lacked this marker of depression severity ineligible for inclusion in data
29 analysis.[1] We do the same and extend this exclusion for such patients who continued on to levels 2-4
30 because their extent of depression severity at study outset is not known. Second, the protocol is silent
31 on what to do with patients who met the remission criteria on the HRSD at entry into their next-level
32 treatment. In STAR*D’s Rationale and Research Design article though, its investigators prespecify that
33 “patients who begin a level with HRSD < 8 will be excluded from analyses.”[8, p.130] We therefore
34 excluded 125 such patients from our analyses of treatment levels 2-4. Third, the protocol is silent on
35 how to analyze patients who exit a treatment without taking the HRSD. STAR*D investigators state in
36 their level 1 article, “our primary analyses classified patients with missing exit HRSD scores as
37 nonremitters a priori”[1, p.34] and repeat similar statements in their level 2-4 articles.[2-6] Therefore,
38 we do likewise.
39
40
41

42 Finally, STAR*D had many patients with missing exit HRSD scores. In their level 2-4 articles, STAR*D
43 investigators used a correspondence table to map the final QIDS-SR score to the HRSD for patients
44 missing their exit HRSD score to assess the impact of their approach to counting such patients as
45 “nonremitters a priori.”[22,23] For patients with missing exit HRSD scores, we therefore mapped their
46 last QIDS-SR score to the HRSD and used it to calculate the mean HRSD exit, mean change, and
47 combined HRSD & QIDS-SR response rates for all treatments. We also calculated STAR*D’s remission
48 rate both as prespecified based on an exit HRSD score of < 8 as well as a final QIDS-SR score of < 6 for
49 those patients missing an exit HRSD score.
50
51

52 All pre-processing and analyses were performed in R.[24] Authors 2 and 3 identified patients by their
53 subject key and used this variable to match information across datasets. Data on patients’ treatment
54 pathways, and when patients transitioned from one level to the next, were taken from the IVRA dataset
55 completed by CRCs, and verified against the data on patient level exits. Authors 2 and 3 then compared
56
57
58
59
60

1
2
3 the number of patients identified for all level 1-4 treatments to that reported in the STAR*D summary
4 article's patient flowchart, and the number of patients matched.[7]
5

6
7 Next, authors 2 and 3 applied STAR*D's level 1 inclusion for data analysis criterion to patients in
8 treatment levels 2-4 as well as excluded from analysis the 125 patients who scored <8 on the HRSD at
9 entry into their next-level treatment. We counted these 125 patients as remitted in the prior treatment
10 level but excluded them from the analyses of subsequent treatments. Supplemental Table 3 presents
11 the number of level 2-4 patients excluded from our reanalysis, and the reasons for their exclusion.
12 Supplemental Table 4 identifies the number of patients with missing entry and/or exit HRSD scores for
13 all level 1-4 treatments. As seen in Supplemental Table 4, 1,330 patients were missing their exit HRSD
14 score across all treatments.
15

16
17 We then compared STAR*D's outcomes to those found in a meta-analysis of 7,030 patients enrolled in
18 antidepressant comparator trials.[25] Similar to STAR*D, comparator trials typically are conducted
19 open-label without a control group and therefore are the appropriate comparison data for STAR*D's
20 outcomes. Continuous HRSD improvement means were provided by the first author of the meta-
21 analysis.[25]
22

23
24 Finally, we compared the STAR*D protocol's step-by-step predictions of patient drop out and the
25 number of patients who would have a satisfactory treatment response and enter follow-up care to what
26 actually occurred.[10] While the purpose of these predictions' was to estimate the number of
27 continuing patients available for randomization in treatment levels 2-4, at the meta-level these
28 predictions are an important hypothesis STAR*D tested by assessing how well its investigators could
29 predict the aggregate step-by-step successful treatment outcomes from their treat-to-remission model
30 of care.
31

32 *Patient and Public Involvement*

33
34
35 Neither patients nor the public were involved in the design, conduct, reporting, or dissemination plans
36 of our research.
37

38 **Results**

39
40
41 Figure 1 presents the overall flow of patients enrolled in the various protocol-defined treatment levels
42 and places them in groups defined by the number of treatment steps. Of the 4,041 patients enrolled
43 into STAR*D, 3,110 met the eligibility for data analysis criterion of having a ROA-administered HRSD
44 score ≥ 14 at study outset. Figure 1 also identifies the number of patients who exited the study following
45 each treatment step, the number who entered follow-up after each treatment step, and the number
46 who were randomly assigned to a next-level treatment.
47

48
49 Supplemental Table 5 describes the demographic and clinical features of the patients who entered
50 treatment in steps 1-4 based on their level 1 baseline presentation when enrolling into the study.
51 Summary statistics are presented as means and standard deviations for continuous variables and
52 percentages for discrete variables. Note that 55.7% of STAR*D patients had 2 or more comorbid axis 1
53 disorders when first enrolled based on the Psychiatric Diagnostic Screening Questionnaire and averaged
54 2.5 comorbid medical conditions based on the Cumulative Illness Rating Scale. Furthermore, the
55 average length of patients' current MDD episode was 25.9 months. In a post hoc analysis, STAR*D
56
57
58
59

investigators found that 77.8% of its enrolled patients would have been excluded from most antidepressant trials due to having two or more concurrent medical conditions, more than one comorbid psychiatric disorder, and/or a current depressive episode lasting > 2 years.[26]

Table 1 presents the mean HRSD entry, exit, and change scores for patients by the specific treatment they received in steps 1-4 as well as the HRSD remission and response rates. Table 1 also provides the HRSD cumulative remission rate after up to 4 trials on antidepressant therapies as well as the combined HRSD plus QIDS-SR remission and response rates for the 1,330 patients missing an exit HRSD score.

Table 1: Outcomes Across All Treatments

Treatment Step	HRSD Score		*Mean Change [95% confidence interval] (SD)	HRSD Remission Rate # (%)	*Combined HRSD & QIDS-SR Remission Rate # (%)	*Combined HRSD & QIDS-SR Response Rate # (%)
	Entry Mean (SD)	*Exit Mean (SD)				
Step 1 (N=3,110)	21.87 (5.21)	13.49 (8.42)	8.38 [8.10, 8.67] (8.11)	794 (25.5%)	938 (30.2%)	1261 (40.5%)
Step 2 (N=1,134)	18.76 (6.24)	13.97 (8.09)	4.79 [4.37, 5.21] (7.23)	241 (21.3%)	283 (25.0%)	329 (29.0%)
Switch strategy (N=620)	19.85 (6.08)	14.70 (8.01)	5.16 [4.59, 5.73] (7.22)	113 (18.2%)	134 (21.6%)	178 (28.7%)
Bupropion (N=190)	20.11 (6.25)	15.32 (7.85)	4.78 [3.82, 5.75] (6.78)	31 (16.3%)	37 (19.5%)	46 (24.2%)
Sertraline (N=198)	19.95 (5.98)	14.92 (8.02)	5.03 [4.04, 6.01] (7.10)	32 (16.2%)	36 (18.2%)	57 (28.8%)
Venlafaxine (N=192)	19.89 (6.19)	14.31 (8.12)	5.58 [4.53, 6.63] (7.45)	37 (19.3%)	44 (22.9%)	59 (30.7%)
Cognitive Therapy (N=40)	18.01 (4.96)	12.44 (7.93)	5.58 [2.87, 8.28] (8.73)	13 (32.5%)	17 (42.5%)	16 (40.0%)
Augmentation strategy (N=514)	17.44 (6.18)	13.10 (8.10)	4.34 [3.72, 4.97] (7.23)	128 (24.9%)	149 (29.0%)	151 (29.4%)
Bupropion (N=216)	16.88 (6.11)	12.52 (7.83)	4.36 [3.38, 5.33] (7.30)	54 (25.0%)	64 (29.6%)	66 (30.6%)

Buspirone (N=225)	17.80 (6.50)	13.36 (8.40)	4.43 [3.52, 5.35] (7.02)	58 (25.8%)	68 (30.2%)	66 (29.3%)
Cognitive Therapy (N=73)	17.99 (5.24)	13.98 (7.98)	4.01 [2.25, 5.78] (7.69)	16 (21.9%)	17 (23.3%)	19 (26.0%)
Step 3 (N=325)	19.59 (6.09)	16.38 (7.77)	3.21 [2.48, 3.94] (6.70)	43 (13.2%)	50 (15.4%)	63 (19.4%)
Level 2A (N=28)	20.89 (5.44)	16.96 (6.48)	3.93 [1.81, 6.04] (5.71)	3 (10.7%)	3 (10.7%)	5 (17.9%)
Bupropion (N=12)	19.92 (3.85)	17.58 (7.35)	2.33 [-0.81, 5.48] (5.55)	2 (16.7%)	2 (16.7%)	2 (16.7%)
Venlafaxine (N=16)	21.62 (6.41)	16.50 (5.96)	5.12 [2.33, 7.92] (5.70)	1 (6.2%)	1 (6.2%)	3 (18.8%)
Level 3 (N=297)	19.46 (6.14)	16.32 (7.88)	3.14 [2.37, 3.92] (6.79)	40 (13.5%)	47 (15.8%)	58 (19.5%)
Switch strategy (N=186)	20.01 (6.24)	17.01 (7.91)	2.99 [2.00, 3.99] (6.94)	23 (12.4%)	25 (13.4%)	31 (16.7%)
Nortriptyline (N=92)	19.67 (5.27)	16.99 (8.35)	2.67 [1.10, 4.24] (7.68)	15 (16.3%)	15 (16.3%)	16 (17.4%)
Mirtazapine (N=94)	20.34 (7.08)	17.03 (7.49)	3.30 [2.06, 4.55] (6.15)	8 (8.5%)	10 (10.6%)	15 (16.0%)
Augmentation strategy (N=111)	18.55 (5.89)	15.16 (7.74)	3.40 [2.18, 4.62] (6.56)	17 (15.3%)	22 (19.8%)	27 (24.3%)
Lithium (N=58)	18.69 (6.47)	15.91 (7.29)	2.78 [1.42, 4.15] (5.31)	7 (12.1%)	9 (15.5%)	10 (17.2%)
T3 (N=53)	18.41 (5.25)	14.34 (8.19)	4.07 [1.99, 6.14] (7.69)	10 (18.9%)	13 (24.5%)	17 (32.1%)
Step 4 (N=106)	20.65 (5.54)	16.49 (7.47)	4.16 [2.80, 5.52] (7.15)	11 (10.4%)	13 (12.3%)	22 (20.8%)
Level 3 (N=16)	20.62	17.62	3.00	2	2	3

	(4.01)	(6.87)	[-0.45, 6.45] (7.04)	(12.5%)	(12.5%)	(18.8%)
Tranlycypromine (N=43)	21.02 (6.57)	16.45 (7.89)	4.57 [2.22, 6.92] (7.87)	3 (7.0%)	5 (11.6%)	9 (20.9)
Venlafaxine XR/mirtazapine (N=47)	20.32 (5.02)	16.14 (7.38)	4.18 [2.30, 6.07] (6.59)	6 (12.8%)	6 (12.8%)	10 (21.3%)
CUMULATIVE remission rate after up to four treatment steps				1,089 (35.0%)	1,284 (41.3%)	

*For patients with missing exit HRSD scores, their last QIDS-SR score is mapped to the HRSD and used to calculate HRSD Exit Mean, Mean Change, Combined HRSD & QIDS-SR Remission Rate, and Combined HRSD & QIDS-SR Response Rate.

Table 2 presents patients' aggregate HRSD status in terms of remission, response, and extent of mean symptomatic change at entry and exit for each treatment step as well as study dropout. In step 1, 25.5% of patients remitted. Steps 2-4 show a continuous decrease in remission rates from step 2's 21.3% to step 3's 13.2% and step 4's 10.4% with increasing rates of study dropout from step 1's 34.5% to step 3's 46.2%.

Table 2: Outcomes by Treatment Step

	Step 1 (N=3,110)		Step 2 (N=1,134)		Step 3 (N=325)		Step 4 (N=106)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
HRSD score at entry into step	21.87	5.21	18.76	6.24	19.59	6.09	20.65	5.54
HRSD score at exit from step*	13.49	8.42	13.97	8.09	16.38	7.77	16.49	7.47
HRSD Mean Change*	8.38	8.11	4.79	7.23	3.21	6.7	4.16	7.15
	N	%	N	%	N	%	N	%
Remission at each step exit	794	25.5%	241	21.3%	43	13.2%	11	10.4%
Response at each step exit*	1261	40.5%	329	29.0%	63	19.4%	22	20.8%
Entered Follow-up	902	29.0%	406	35.8%	69	21.2%	38	35.9%
Study Exit/Dropout	1,074	34.5%	403	35.5%	150	46.2%		

* For patients with missing exit HRSD scores, their last QIDS-SR score is mapped to the HRSD and used to calculate HRSD Exit Mean, Mean Change, and Combined HRSD & QIDS-SR Response Rate.

Supplemental Figures 1 and 2 compare the HRSD remission, response, and extent of symptom improvement rates for STAR*D patients in steps 1-4 to that found in a meta-analysis of 7,030 patients enrolled in non-blinded antidepressant comparator trials.[25] In step 1, these measures of

1
2
3 improvement among STAR*D's patients were at least one-third less than that found in comparator
4 trials, and improvement was worse in each subsequent treatment step.
5

6
7 Figure 2 compares the STAR*D protocol's predictions of patient dropout and the number of patients
8 who would have a satisfactory treatment response and enter follow-up to what occurred. Cumulatively,
9 STAR*D's investigators predicted that 73.8% of patients would have a successful treatment response
10 and enter follow-up whereas in fact only 45.6% achieved this measure of treatment success.
11 Furthermore, whereas STAR*D investigators predicted that over the course of up to four antidepressant
12 therapies 20.7% of patients would dropout, in fact, 53.7% dropped out. On this measure of treatment
13 failure, STAR*D's dropout rate was 2.6 times greater than predicted.
14

15
16 Figure 3 presents the step-by-step cumulative remission rate in three ways. First, the 'theoretical' rate
17 propagated by STAR*D investigators based on the provisos of what would have happened if there were
18 no study dropouts and that those who did exit had the same QIDS-SR remission rates as those who
19 stayed.[7] Next, the combined HRSD plus QIDS-SR remission rate based on either an exit HRSD score of
20 <8, OR a last clinic visit QIDS-SR score of <6 for the 1,330 patients missing an exit HRSD. Finally, our RIAT
21 reanalysis rate when using the protocol-specified exit HRSD score of <8 as the sole measure of remission
22 for the 3,110 patients who met STAR*D's inclusion in data analysis criteria. The cumulative remission
23 rate after up to four antidepressant therapies using the HRSD was 35.0% versus 41.3% when combined
24 with the QIDS-SR, both of which are substantially less than the 67% cumulative remission rate claimed in
25 the summary article's Abstract.
26

27 Discussion

28 *Principal findings and comparison with original STAR*D publication*

29
30
31
32 STAR*D's results highlight the discrepancy in likely outcomes between typical antidepressant clinical
33 trials with their exclusion criteria and the real-world patients for whom these medications are
34 commonly prescribed. Our RIAT reanalysis found poorer outcomes after up to four optimized, and
35 increasingly aggressive, antidepressant therapies than reported in STAR*D's summary article published
36 in *AJP*. [7] In contrast to the 67% cumulative remission rate reported in *AJP*, the actual rate was 35.0%
37 when using the protocol-specified HRSD, and increased to 41.3% when combined with a final clinic-visit
38 QIDS-SR score of <6 for patients missing exit HRSD scores in treatment steps 1-4. The 41.3% cumulative
39 remission rate should be viewed as the "best case scenario" since it added an additional 195 QIDS-
40 defined remissions (a remission measure not specified in the protocol) from the 1,330 patients with
41 missing exit HRSD scores. As there was neither a placebo nor waitlist control group during any phase of
42 the STAR*D study, it is impossible to know to what extent the observed results were due to the
43 pharmacologic effects of the prescribed medications, placebo effects, and/or the passage of time.
44
45

46
47 Our reanalysis did not assess the durability of treatment effects during the 12-month follow up phase.
48 In their summary article though, STAR*D investigators reported an overall relapse rate of 46.1% for the
49 1,729 patients for whom they had at least one assessment (of up to 12 scheduled) during follow up
50 using a telephonic-administered version of the QIDS [7] whereas Pigott et al. found a far lower sustained
51 recovery rate when incorporating patient dropout in the analysis. [11]
52

53 *Comparison with other studies*

1
2
3 Our reanalysis found that in step 1, STAR*D's remission, response, and extent of improvement rates
4 were substantially less than those reported in other open-label antidepressant comparator trials and
5 then grew progressively worse in steps 2-4.[25] Such studies typically exclude depressed patients with
6 the range and number of comorbid medical and/or psychiatric disorders that were included in STAR*D.
7

8
9 STAR*D's step 1 remission rate was 25.5% followed by a progressive decline in remission rates for those
10 patients receiving subsequent, and increasingly aggressive treatments, such that by step 4 it was only
11 10.4%. This decline in antidepressants' effectiveness essentially mirrors the findings from randomized
12 and naturalistic, prospective studies reporting a 20-30% loss of effectiveness with each increase in the
13 number of prior antidepressant trials.[27-32] Furthermore, several recent analyses suggest that the
14 sequential application of antidepressant medications for non-remitting depression may in fact foster
15 treatment resistance for many patients.[33-36]
16

17
18 Regarding the protocol's predictions of treatment success and patient dropout, it states:
19

20 We arrived at these estimates using three experienced practitioners who
21 independently made estimates that were surprisingly close to each other. Then, via
22 teleconferencing, the final estimates were made. ***The underlying assumptions of***
23 ***these estimates come largely by inferences from results of published RCTs.***[10, p.31;
24 emphasis added]
25

26
27 STAR*D's actual measures of treatment success and failure were significantly worse than predicted. As
28 Barbui et al. noted, antidepressant study dropout rates provide a "hard measure of treatment
29 effectiveness and acceptability"[12, p.296] and STAR*D's dropout rate was 2.6 times greater than
30 predicted. This discrepancy further highlights the relative ineffectiveness of antidepressants in treating
31 real-world depressed patients, compared to those reported in conventional studies.
32

33 *Conclusion*

34
35 Bias in the clinical literature is commonly associated with industry-funded RCTs, not publicly funded
36 ones.[37] Our RIAT reanalysis though documents scientific errors in this NIMH-funded study. These
37 errors inflated STAR*D investigators' report of positive outcomes.
38
39

40 The STAR*D summary article's claim of a 67% cumulative remission rate was published in 2006. If
41 STAR*D's outcomes had been reported as prespecified, its model of care would likely have faced much
42 stronger criticism 16 years ago and fueled a more vigorous search for evidence-based treatment
43 alternatives.
44

45 **Ethics Statement:**

46
47
48 Ethics committee approval was not required for our reanalysis since the data was anonymized by NIMH.
49

50 **Data Sharing Statement:**

51
52 Data is available from the NIMH-supported National Database for Clinical Trials (NDCT).
53

54 **Funding Statement:**

Funding for this project was provided by The RIAT Support Center.

Competing Interests:

The authors have no relevant financial disclosures or other conflicts of interest to report.

Contributorship Statement:

HEP, JDA and IK contributed to the design of the study and secured funding. TK and CX conducted all of the data analyses. HEP wrote the manuscript with input from JDA, IK, TK, and CX.

Acknowledgements:

We thank Termeh Feinberg for her early efforts on this project, particularly her correspondence with the NIMH help desk to resolve issues with the 26 data files as well as the RIAT Support Center for funding this project. Data used in the preparation of this manuscript were obtained from the controlled access datasets distributed from the NIMH-supported National Database for Clinical Trials (NDCT). NDCT is a collaborative informatics system created by the National Institute of Mental Health to provide a national resource to support and accelerate discovery related to clinical trial research in mental health. The content of this publication does not necessarily reflect the views of the RIAT Support Center nor NIMH.

References:

1. 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. 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. 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. Bupropion-SR, sertraline, or venlafaxine-XR after failure of SSRIs for depression. *N Engl J Med*. 2006;354:1231–1242.
3. 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. Medication augmentation after the failure of SSRIs for depression. *N Engl J Med*. 2006;354:1243–1252.
4. Fava M, Rush AJ, Wisniewski SR, Nierenberg AA, Alpert JE, McGrath PJ, Thase ME, Warden D, Biggs MM, Luther JF, Niederehe G, Ritz L, Trivedi MH. 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.
5. Nierenberg AA, Fava M, Trivedi MH, Wisniewski SR, Thase ME, McGrath PJ, Alpert JE, Warden D, Luther JF, Niederehe G, Lebowitz BD, Shores-Wilson K, Rush AJ. 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.
6. McGrath PJ, Stewart JW, Fava M, Trivedi MH, Wisniewski SR, Nierenberg AA, Thase ME, Davis L, Biggs MM, Shores-Wilson K, Luther JF, Niederehe G, Warden D, Rush AJ. Tranylcypromine versus venlafaxine plus mirtazapine following three failed antidepressant medication trials for depression: a STAR*D report. *Am J Psychiatry*. 2006;163:1531–1541.
7. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, Niederehe G, Thase ME, Lavori PW, Lebowitz BD, McGrath PJ, Rosenbaum JF, Sackheim HA, Kupfer DJ, Luther J, Fava

- 1
2
3 M. Acute and longer-term outcomes in depressed outpatients requiring one or several
4 treatment steps: a STAR*D report. *Am J Psychiatry*. 2006;163:1905–1917.
- 5
6 8. Rush AJ, Fava M, Wisniewski SR, Lavori PW, Trivedi MH, Sackeim HA, Thase ME, Nierenberg AA,
7 Quitkin FM, Kashner TM, Kupfer DJ, Rosenbaum JF, Alpert J, Stewart JW, McGrath PJ, Biggs MM,
8 Shores-Wilson K, Lebowitz BD, Ritz L, Niederehe G. Sequenced treatment alternatives to relieve
9 depression (STAR*D): rationale and design. *Control Clin Trials*. 2004;25:119–142.
- 10
11 9. Rush AJ, Trivedi MH, Ibrahim HM, Carmody T J, Arnow B, Klein DN, ... Manber R. The 16-item
12 Quick Inventory of Depressive Symptomatology (QIDS), Clinician Rating (QIDS-C), and Self-
13 Report (QIDS-SR): a psychometric evaluation in patients with chronic major depression.
14 *Biological Psychiatry*. 2003;54:573– 583.
- 15
16 10. National Institute of Mental Health (NIMH). Sequenced Treatment Alternatives to Relieve
17 Depression (STAR*D) Research Protocol. Washington (DC): NIMH; revised June 28, 2002.
- 18
19 11. Barbui C, Furukawa TA, Cipriani A. Effectiveness of paroxetine in the treatment of acute major
20 depression in adults: a systematic re-examination of published and unpublished data from
21 randomized trials. *CMAJ*. 2008;178:296–305.
- 22
23 12. Pigott HE, Leventhal AM, Alter GS, Boren JJ. Efficacy and effectiveness of antidepressants:
24 current status of research. *Psychother Psychosom*. 2010;79:267–279.
- 25
26 13. Insel TR, Wang PS. The STAR*D trial: revealing the need for better treatments. *Psychiatric
27 Services*. 2009;60:1466-1467.
- 28
29 14. Greden JF. Workplace depression: personalize, partner, or pay the price. *Am J Psychiatry*.
30 2013;170:578–581.
- 31
32 15. Smith DG. Antidepressants don't work the way many people think. *New York Times*. November
33 8, 2022.
- 34
35 16. Pigott HE. The STAR*D trial: it's time to reexamine the clinical beliefs which guide the treatment
36 of major depression. *Canadian J Psychiatry*. 2015;60:9–13.
- 37
38 17. Kirsch I, Huedo-Medina TB, Pigott HE, Johnson B. Do outcomes of clinical trials resemble those
39 "real world" patients? A re-analysis of STAR*D antidepressant data. *Psychology of
40 Consciousness: Theory, Research, and Practice*. 2018 <https://doi.org/10.1037/cns0000164>
- 41
42 18. Pigott HE, Dubin, W, Kirsch I, Amsterdam J. Call to action: RIAT reanalysis of the Sequenced
43 Treatment Alternatives to Relieve Depression (STAR*D) Study. *BMJ*. March 6, 2019;
44 <https://www.bmj.com/content/346/bmj.f2865/rr-10>
- 45
46 19. Doshi P, Dickersin K, Healy D, Vedula SS, Jefferson T. Restoring invisible and abandoned trials: a
47 call for people to publish the findings. *BMJ* 2013;346:f2865.
- 48
49 20. Fava M, Rush AJ, Trivedi MH, et al. Background and rationale for the Sequenced Treatment
50 Alternatives to Relieve Depression (STAR*D) study. *Psychiatr Clin North Am*. 2003;26:457–494.
- 51
52 21. Lavori PW, Rush AJ, Wisniewski SR, Alpert J, Fava M, Kupfer DJ, Nierenberg A, Quitkin FM,
53 Sackeim HA, Thase ME, Trivedi M. Strengthening clinical effectiveness trials: equipoise-stratified
54 randomization. *Biological Psychiatry*. 2001;50:792– 801.
- 55
56 22. Rush AJ, Bernstein IH, Trivedi MH, Carmody TJ, Wisniewski S, Mundt JC, Shores-Wilson K, Biggs
57 MM, Woo A, Nierenberg AA, Fava M. An evaluation of the quick inventory of depressive
58 symptomatology and the hamilton rating scale for depression: a sequenced treatment
59 alternatives to relieve depression trial report. *Biol Psychiatry*. 2006 Mar 15;59(6):493-501. doi:
60 10.1016/j.biopsych.2005.08.022. Epub 2005 Sep 30. PMID: 16199008; PMCID: PMC2929841.
- 61
62 23. Rush AJ. Inventory of Depressive Symptomatology (IDS) and Quick Inventory of Depressive
63 Symptomatology (QIDS). [IDS/QIDS \(ids-qids.org\)](https://ids-qids.org)
- 64
65 24. R Core Team. R: A language and environment for statistical computing. R Foundation for
66 Statistical Computing, Vienna, Austria.

25. Rutherford BR, Sneed JR, Roose SP. Does study design influence outcome? The effects of placebo control and treatment duration in antidepressant trials. *Psychotherapy and Psychosomatics*. 2009;78:172–181.
26. Wisniewski SR, Rush AJ, Nierenberg AA, Gaynes BN, Warden D, Luther JF, McGrath PJ, Lavori PW, Thase ME, Fava M, Trivedi MH: Can phase III trial results of antidepressant medications be generalized to clinical practice? A STAR * D report. *Am J Psychiatry* 2009;166:599–607.
27. Amsterdam JD, Maislin G. Fluoxetine efficacy in treatment resistant depression. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*. 1994;18:243-261.
28. Nierenberg A, Feighner JP, Rudolph R, Cole JO, Sullivan J. Venlafaxine for treatment-resistant unipolar depression. *J Clinical Psychopharmacology*. 1994;14:419-423.
29. Amsterdam JD, Shults J. MAOI safety and efficacy in advanced treatment-resistant depression: a retrospective analysis. *J Affective Disorders*. 2005; 89:183-188.
30. Amsterdam JD, Williams D, Michelson D, Adler LA, Dunner SL, Nierenberg A, Reimherr FW, Schatzberg AF. Tachyphylaxis after repeated antidepressant drug exposures in patients with recurrent major depressive disorder. *Neuropsychobiology*. 2009;59:227-233.
31. Leykin Y, Amsterdam JD, DeRubeis RJ, Shelton RC, Hollon SD. Progressive resistance to SSRI therapy but not to cognitive therapy in the treatment of major depression. *J Consulting & Clinical Psychology*. 2007;75:267-276.
32. Amsterdam JD, Lorenzo-Luaces L, DeRubeis RJ. Step-wise loss of antidepressant effectiveness after repeated antidepressant trials in bipolar II depression. *Bipolar Disorders*. 2016;18:563-570.
33. Fava GA, Offidani E. The mechanisms of tolerance in antidepressant action. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*. 2011;35:1593-602.
34. Amsterdam JD, Lorenzo-Luaces L. Increase in pharmacodynamic tolerance after repeated antidepressant trials in treatment-responsive bipolar II depressed subjects: An exploratory study. *Psychiatria Polska*. 2018;52:957–969.
35. Amsterdam JD, Kim TT. Increased risk of depressive relapse during maintenance therapy after repeated antidepressant trials in treatment-responsive subjects. *J Clinical Psychopharmacology*. 2019;39:344-350.
36. Andrews PW, Amsterdam JD. A hormetic approach to understanding antidepressant effectiveness and the development of antidepressant tolerance. *Psychiatria Polska*. 2020;54:1067–1090.
37. Amsterdam JD, McHenry LB, Jureidini JN. Industry-corrupted psychiatric trials. *Psychiatria Polska*. 2017;51:993–1008.

Figure 1 Caption: Patient Flowchart

Figure 1 Footnote:

* In level 2, 580 patients were randomized to switch medications, 441 to medication augmentation, and 113 to Cognitive Therapy as either a switch or medication augmentation treatment. In level 2A, 28 patients were randomized to one of two level 2 switch medications. For step 3/level 3 patients, 186 were randomized to medication switch and 111 to medication augmentation. For step 4/level 3 patients, 7 were randomized to medication switch and 9 to medication augmentation. For step 4/level 4 patients, 90 were randomized to one of two medication/medication combination switch options.

** Exit refers to the number of patients who exit the study and do not proceed either to the next treatment level nor enter follow-up.

1
2
3 *** Follow-up refers to the number of patients who exit a treatment and enter the 12-month follow-up
4 phase.
5

6 **Figure 2 Caption: Comparison of STAR*D Protocol Predictions to What Occurred**

7
8 **Figure 3 Caption: STAR*D's Step-by-Step Cumulative Remission Rate Presented Three Ways**

9
10
11 **Figure 3 Footnote:**

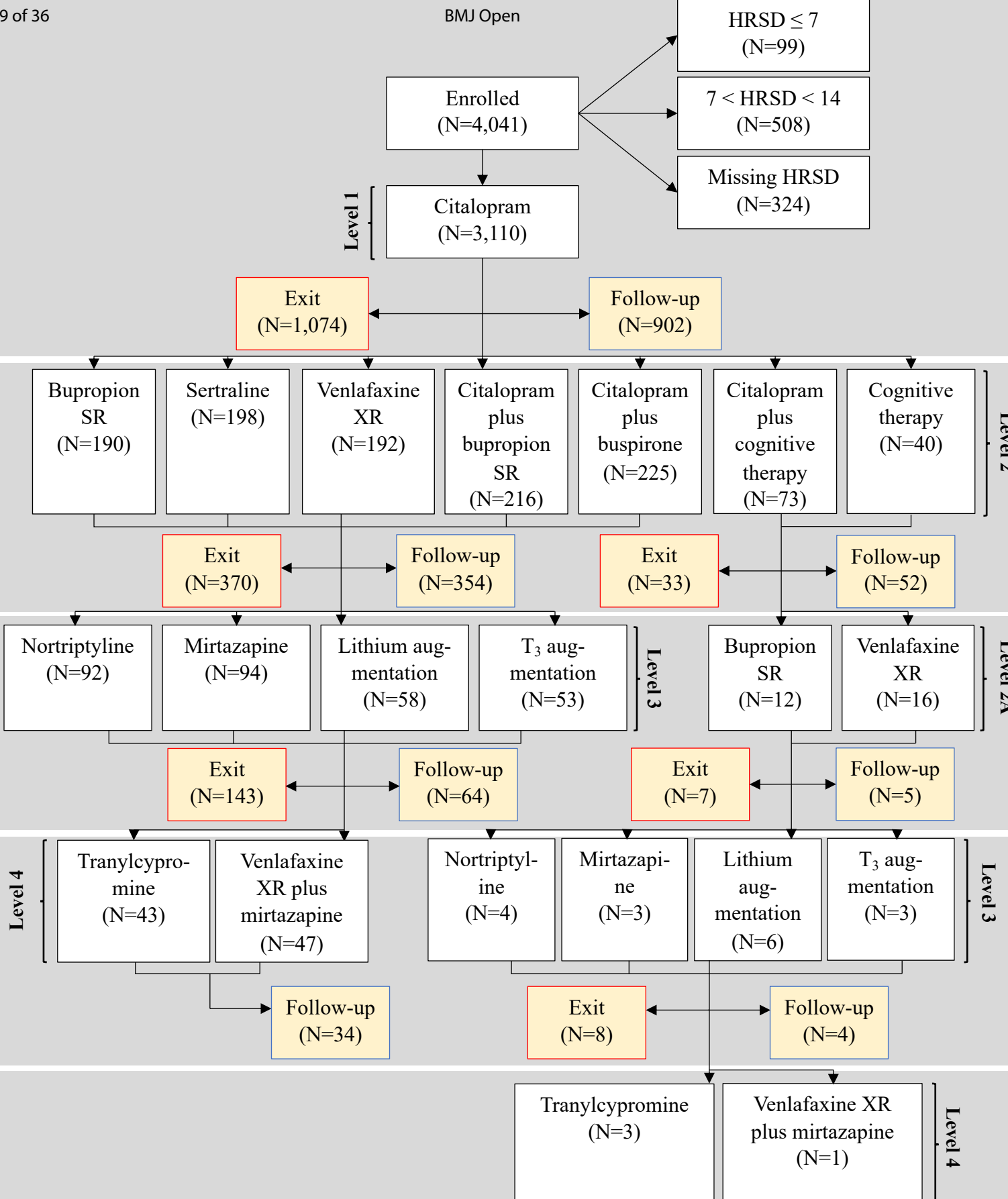
12
13 The step-by-step theoretical remission rates were obtained from the STAR*D summary article where it
14 states: "The theoretical cumulative remission rate is 67% (37+19+6+5)." [7, p.1910].
15

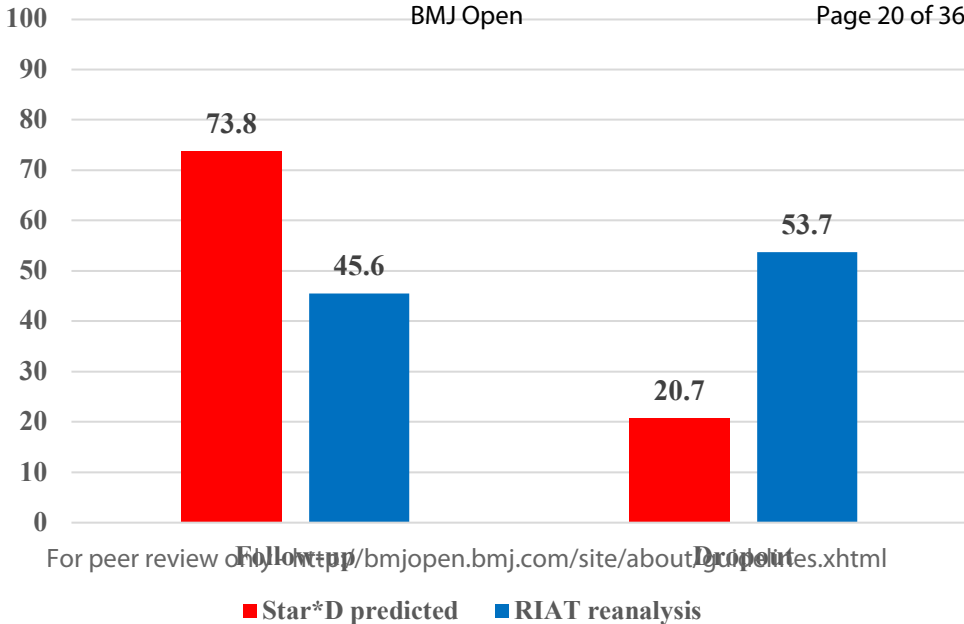
16
17 The HRSD + QIDS-SR cumulative remission rate was taken from Table 1. It combines the 1,089 patients
18 with an exit HRSD score of <8 with the 195 patients who were missing an exit HRSD score but had a final
19 clinic-visit QIDS-SR score of <6.
20

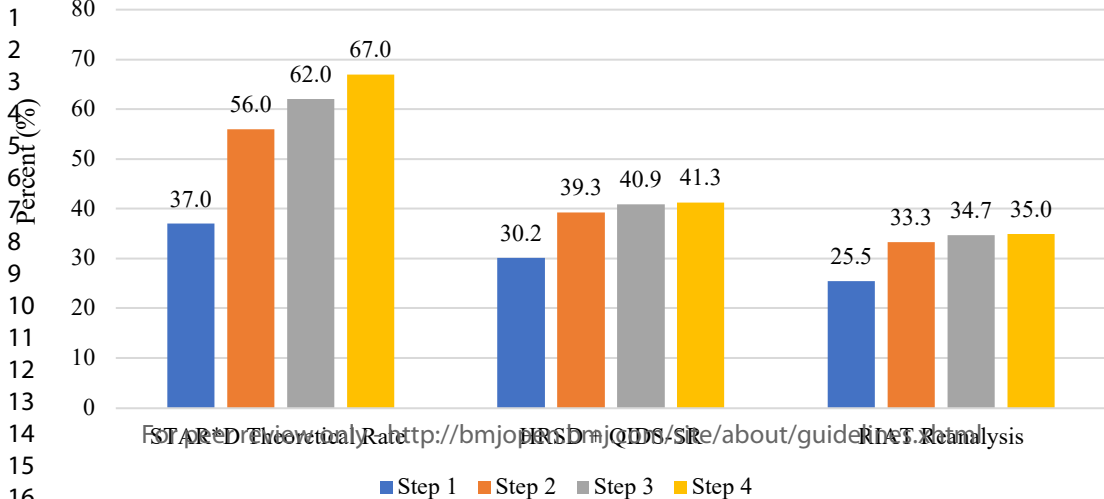
21 The RIAT Reanalysis cumulative remission rate is based on an exit HRSD score of <8 as the sole measure
22 of remission for the 3,110 patients who met STAR*D's inclusion for data analysis criteria.
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Step 1

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18



**Supplementary Table 1:
Highest Quality of Acute and Continuing-Care to
Maximize Remissions While Minimizing Relapse and Dropouts**

Descriptor	Explanation
<p>Optimized Sustained Study Participation to Minimize Dropouts ²⁰, p. 473-474</p>	<ul style="list-style-type: none"> • Promoted patients' study affiliation via STAR*D-branded brochures, bimonthly newsletters, and an informational video emphasizing STAR*D's public health significance and the critical role played by patients; • Educated patients and families about depression and its treatment using a multi-step educational package. This included teaching the "mechanism of action" for patients' current antidepressant and educating patients that "depression is a disease, like diabetes or high blood pressure" and "can be treated as effectively as other illnesses," etc.; • Used a letter reminder system to alert patients before appointments in those clinics without such systems who had a >15% rate of missed appointments; • Ensured timely follow-up and rescheduling of missed appointments by calling patients on the day of the missed appointment, and again within 24 hours, if there was no response. Patient's physician sent letter within 48 hours if contact was not established; • Used a letter reminder system for all research outcome assessment calls during acute and continuing-care; • In every clinic visit, the Clinical Research Coordinator (CRC) discussed the research outcomes phone calls with the patient to ensure that the calls were completed on schedule and worked to resolve any problematic issues regarding said calls [Clinical Procedures Manual, page 75]; • Paid patients \$25.00 for participating in each telephonic research outcomes assessment; • Permitted patients to re-enter acute and/or continuing-care within four weeks after having dropped out [Clinical Procedures Manual, page 80]; • Recommended one-year of continuing-care for all patients who achieved a satisfactory clinical response with the essential goal of preventing relapse [Clinical Procedures Manual, page 15] and • Permitted continuing-care patients to remain in the study if they moved from the area [Clinical Procedures Manual, page 81].
<p>Acute-Care Visits</p>	<p>Physicians met with patients on entry into each new step to initiate drug treatment with follow-up visits scheduled on weeks 2, 4, 6, 9, 12, with an optional week 14 visit.</p>

Measurement-Based Care	Conducted structured evaluations of symptoms and side-effects at each visit and included a centralized treatment monitoring and physician feedback system to ensure consistent implementation of optimal care across research sites.
Aggressive Medication Dosing	Provided aggressive medication dosing with a fully adequate dose for a sufficient duration 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}
Liberal Prescribing of Non-Study Medications	Physicians had great leeway in prescribing non-study medications to treat comorbid symptoms resulting in: <ul style="list-style-type: none"> • 17.2% taking Trazodone for sleep; • 11.9% taking an anti-anxiety medication; • 16.7% taking either a sedative or hypnotic medication; and • An undisclosed percent taking medications to address side-effects. ^{2, table 2}
Continuing-Care Visits	Patients saw their physician every 2 months and continued taking their treatment medication(s) at the same doses but their physicians were allowed to make any psychotherapy, medication, and/or medication dose changes to maximize the likelihood of maintaining patients’ remission status. ^{7, p. 1908} Additional continuing-care visits were scheduled when patients began to experience a return of depressive symptoms and/or intolerable side-effects [Clinical Procedures Manual. page 78].
Clinical Research Coordinator (CRC)	Each site had a CRC who: ^{1, p. 30} <ul style="list-style-type: none"> • Saw patients before each visit administering multiple measures to them including the QIDS-SR during each acute-care visit; • Assisted physicians in protocol implementation; and • Provided patients support and encouragement in protocol implementation.
Treatment Designed to Enhance Subject Retention	Treatment was designed to minimize drop-outs and/or non-compliance including: <ul style="list-style-type: none"> • Open label prescribing during acute and continuing-care with no placebo control condition during any study phase; • Patients chose their acceptable treatment assignments for steps two and three to eliminate any concerns they might have about receiving an unacceptable assignment. This resulted in only 21 of 1,439 (1.5%) Step-2 patients making themselves available for random assignment to all treatment options ^{2, p. 1235} while only 29 of 377 (7.7%) did so in Step-3. ^{5, p. 1521} • During each step, patients could enroll immediately into the next step if they had intolerable side-effects or had maximized their current medication(s)’ dosing without achieving a remission; and • During any step, patients could enter continuing-care directly on their current medication(s) if they were treatment responders even if they had not achieved remission. This was done to minimize responders from dropping out in order to avoid having

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

	to discontinue their current medication(s) and start a new drug regimen.
--	--

***Trivedi MH, Stegman D, Rush AJ, Wisniewski SR, Nierenberg AA: STAR * D clinical procedures manual. July 31, 2002. www.edc.pitt.edu/stard/public/study_manuals.html

For peer review only

Supplementary Table 2: Description of Levels 1-4 Treatments

Level 1:

STAR*D investigators report that Citalopram (Celexa) was chosen as the first-line SSRI treatment because (1) absence of discontinuation symptoms; (2) demonstrated safety in elderly and medically fragile patients; (3) easy once-a-day dosing with few dose adjustments; and (4) favorable drug–drug interaction profile.¹ Citalopram was started at 20 mg/day and then raised to 40 mg/day by day 28 and up to 60 mg/day by day 43 and onward. Dose adjustments were based on how long a patient had received a particular dose, symptom changes, and side effect burden.

Level 2 switch treatments:

Citalopram was discontinued without a tapering at the initiation of each level 2 switch treatment. STAR*D investigators chose pharmacologically distinct switch medications. The level 2 treatments were:

- Sertraline (Zoloft), an SSRI with the same pharmacological profile as citalopram. Sertraline was started at a daily dose of 50 mg and increased to 100 mg at day 8, to 150 mg at day 28, and to 200 mg at day 63 and onward.
- Sustained-release bupropion (Wellbutrin SR), an “out-of-class” agent whose neurochemical action mechanisms are unknown; other than that, it does not inhibit serotonin reuptake and is believed to produce antidepressant effects by blocking the reuptake of dopamine and norepinephrine. The daily dose of sustained-release bupropion was 150 mg for week 1, 200 mg from day 8 to 27, 300 mg from day 28 to 41, and 400 mg from day 42 onward.
- Extended-release venlafaxine (Effexor), a “dual-action” agent that inhibits the reuptake of both serotonin and norepinephrine. The starting daily dose of extended-release venlafaxine was 37.5 mg for week 1 and increased to 75 mg from day 8 to 14, to 150 mg from day 15 to 27, to 225 mg from day 28 to 41, to 300 mg from day 42 to 62, and to 375 mg from day 63 onward.
- Cognitive therapy was provided by a trained psychotherapist and scheduled twice weekly for the first four weeks, then once weekly for the remaining 8 weeks (16 sessions total).

Level 2 Citalopram augmentation treatments:

During the augmentation trial, the citalopram dose was kept constant but reduced if side effects developed. The level 2 augmentation treatments were:

- Buspirone (Buspar), a partial agonist at the postsynaptic 5-hydroxytryptamine 1A (5-HT_{1A}) receptor that is believed to enhance the activity of SSRIs through the 5HT_{1A} receptors. The starting dose was 15 mg per day week 1, increasing to 30 mg per day week 2, and then to 45 mg per day for weeks 3 through 5, and a final, maximum dose of 60 mg per day week 6 and onward.
- Sustained-release bupropion (Wellbutrin SR) whose neurochemical action mechanisms are unknown but is believed to produce antidepressant effects by blocking the reuptake of dopamine and norepinephrine. The initial dose was 200 mg per day during weeks 1 and 2, increasing to 300 mg per day by week 4 and to 400 mg per day week 6 and onward.
- Cognitive therapy was provided by a trained psychotherapist and scheduled twice weekly for the first four weeks, then once weekly for the remaining 8 weeks (16 sessions total).

Level 3 switch treatments:

At entry into the Level 3 switch trial, all level 2 medications were discontinued without tapering at the initial Level 3 treatment visit. The level 3 switch treatments were:

- Nortriptyline (Pamelor), a tricyclic antidepressant. Recommended doses were 25 mg/ day for 3 days, 50 mg/day for 4 days, and then 75 mg/day by day 8, 100 mg/day by day 28, and, if necessary, 150 mg/day by day 42 and onward
- Mirtazapine (Remeron), a tetracyclic antidepressant that blocks inhibitory α -adrenoceptors on norepinephrine and serotonin neurons to enhance both norepinephrine and serotonin neurotransmission. Recommended mirtazapine doses were 15 mg/day for the first 7 days, 30 mg/day by day 8, 45 mg/day by day 28, and, if necessary, 60 mg/ day by day 42 and onward.

Level 3 augmentation treatments of level 2 medications:

The two medication augmentation options used in level 2, buspirone and sustained-release bupropion, were discontinued without tapering in the initial level 3 visit. The two medication augmentation treatments in level 3 were added to ongoing treatment with citalopram, sertraline, sustained-release bupropion, or extended-release venlafaxine. The level 3 augmentation treatments were:

- Lithium started at 450 mg/day, and at week 2 it was increased to the recommended dose of 900 mg/day. If participants could not tolerate the initial dose, it could be reduced to 225 mg/day for 1 week then increased to 450 mg/day. There was no monitoring of lithium levels.
- Triiodothyronine (T₃), a thyroid hormone, started at 25 μ g/day for 1 week and then increased to the recommended dose of 50 μ g/ day. There was no pretreatment assessment, nor ongoing monitoring, of thyroid functioning.

Level 4 switch treatments:

The level 4 switch treatments were:

- Tranylcypromine (Parnate), a monoamine oxidase inhibitor. A 2-week washout period of Level 3 medications was required for patients assigned to the tranylcypromine group. The recommended dosing for tranylcypromine was 10 mg/day for the first 2 weeks, followed by weekly increases of 10 mg/day until a maximum of 60 mg/day.
- Co-administered venlafaxine (Effexor) and mirtazapine (Remeron) to inhibit the reuptake of both serotonin and norepinephrine and block inhibitory α -2-adrenoceptors on both norepinephrine and serotonin neurons to enhance both norepinephrine and serotonin neurotransmission. Level 3 medications were discontinued without tapering for patients assigned to this group. The dosage of extended-release venlafaxine was 37.5 mg/day for the first week, 75 mg/day for the second week, 150 mg/day for weeks 3–5, 225 mg/day for weeks 6–8, and 300 mg/day onward. Mirtazapine was started at 15 mg/day for the first 3 weeks, 30 mg/day for weeks 4 to 8, and then 45 mg/day onward.

**Supplementary Table 3:
Number of Level 2-4 Participants Excluded from our RIAT
Reanalysis, and the Reasons for their Exclusion, yet Included in STAR*D**

Level 2 Treatments

Number of Level 2 Participants Excluded from our Reanalysis but Included in STAR*D	Bup	Sert	Ven	CT	Cit + BUP	Cit + Busp	Cit + CT	Total
Scored as Remitted at ENTRY into Level 2 yet still included in STAR*D's Level 2 analyses	22	8	14	7	30	24	4	109
Scored as only mildly depressed (HRSD >7 & <14) at entry into Level 1, and therefore excluded from STAR*D's data analysis, yet still treated in Level 1, progressed to Level 2, and then included in STAR*D's Level 2 data analyses	21	15	25	15	20	30	7	133
Scored as Remitted at entry into Level 1 (HRSD ≤ 7), and therefore excluded from STAR*D's data analysis, yet still treated in Level 1 and progressed to Level 2 and then included in STAR*D's Level 2 data analyses	6	1	4	2	2	2	2	19
Missing baseline HRSD at entry into Level 1, and therefore excluded from STAR*D's data analysis, yet still treated in Level 1, and progressed to Level 2, and then included in STAR*D's Level 2 data analyses	12	18	22	4	16	13	1	86
Number meeting 2 exclusion criterions	12	2	7	6	5	8	2	42

Bup=Sustained-release Bupropion; Sert= Sertraline; Ven= Extended-release Venlafaxine;
CT=Cognitive Therapy; Cit+BUP= Citalopram + Sustained-release Bupropion;
Cit+Busp=Citalopram + Buspirone; Cit+CT= Citalopram + Cognitive Therapy

Level 3 Treatments

	Nortriptyline	Mirtazapine	Lithium Augmentation	Triiodothyronine Augmentation	Total
Scored as Remitted at ENTRY into Level 3 yet still included in STAR*D's Level 3 analyses	4	0	1	5	10
Scored as only mildly depressed (HRSD >7 &	8	5	3	4	20

1 2 3 4 5 6 7 8 9 10 11 12 13	<14) at entry into Level 1, and therefore excluded from STAR*D's data analysis, yet still treated in Level 1, progressed to Level 2, and then 3 and included in STAR*D's Level 3 data analyses					
14 15 16 17 18 19 20 21 22 23 24 25 26	Scored as Remitted at entry into Level 1 (HRSD ≤ 7), and therefore excluded from STAR*D's data analysis, yet still treated in Level 1 and progressed to Level 2 and then 3 and included in STAR*D's Level 3 data analyses	2	1	0	1	4
27 28 29 30 31 32 33 34 35 36 37 38 39	Missing baseline HRSD at entry into Level 1, and therefore excluded from STAR*D's data analysis, yet still treated in Level 1, and progressed to Level 2, and then level 3 and included in STAR*D's Level 3 data analyses	7	8	1	7	23

Level 4 Treatments

	Tranlycypromine	Venlafaxine + Mirtazapine	Total	
42 43 44 45 46	Scored as Remitted at ENTRY into Level 4 yet still included in STAR*D's Level 4 analyses	5	1	6
47 48 49 50 51 52	Scored as only mildly depressed (HRSD >7 & <14) at entry into Level 1, and therefore excluded from STAR*D's data analysis, yet still treated in Level 1, progressed to Level 2, 3 and then 4 and included in STAR*D's Level 4 data analyses	3	1	4
53 54 55 56 57 58 59 60	Scored as Remitted at entry into Level 1 (HRSD ≤ 7), and therefore excluded from STAR*D's data analysis, yet still treated in Level 1 and progressed	0	0	0

to Level 2, then Level 3 and included in STAR*D's Level 4 data analyses			
Missing baseline HRSD at entry into Level 1, and therefore excluded from STAR*D's data analysis, yet still treated in Level 1, and progressed to Level 2, and then Level 3 and 4 included in STAR*D's Level 4 data analyses	5	1	6

For peer review only

**Supplementary Table 4:
Number and Percent of Participants Missing Entry and/or
Exit HRSD Used for Last Observation Carried Forward Analyses**

	#/(%) with Missing Entry HRSD	#/(%) with Missing Exit HRSD
Step 1 (N=3,110)	0 (0%)	926 (29.8%)
Step 2 (N=1,134)	168 (14.8%)	304 (26.8%)
Switch strategy (N=620)	90 (14.5%)	183 (29.5%)
Bupropion (N=190)	34	58
Sertraline (N=198)	27	56
Venlafaxine (N=192)	24	56
Cognitive Therapy (N=40)	5	13
Augmentation strategy (N=514)	78 (15.2%)	121 (23.5%)
Bupropion (N=216)	35	58
Bupirone (N=225)	37	52
Cognitive Therapy (N=73)	6	11
Step 3 (N=325)	42 (12.9%)	78 (24%)
Level 2A (N=28)	3	6
Bupropion (N=12)	3	2
Venlafaxine (N=16)	0	4
Level 3 (N=297)	39	72
Switch strategy (N=186)	26	49
Nortriptyline (N=92)	11	23
Mirtazapine (N=94)	15	26
Augmentation strategy (N=111)	13	23
Lithium (N=58)	9	13
Bupropion SR (N=17)	3	2
Citalopram (N=22)	5	6
Sertraline (N=11)	1	3
Venlafaxine XR (N=8)	0	2
T3 (N=53)	4	10
Bupropion SR (N=6)	1	0
Citalopram (N=26)	1	7
Sertraline (N=8)	1	1
Venlafaxine XR (N=13)	1	2

Step 4 (N=106)	15 (14.2%)	22 (20.8%)
Level 3 (N=16)	3	3
Tranlycypromine (N=43)	7	10
Venlafaxine XR/mirtazapine (N=47)	5	9

For peer review only

Supplementary Table 5: Baseline Demographic and Clinical Features by Treatment Step

	Treatment Step ^a							
	Step 1 (N=3,110)		Step 2 (N=1,134)		Step 3 (N=325)		Step 4 (N=106)	
Demographic Features	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age	41.0	13.0	42.0	12.6	44.1	12.0	46.9	11.0
Education (years)	13.6	3.2	13.2	3.3	12.8	3.1	12.6	2.3
Monthly household income	2,289	2,732	1,744	1,539	1,470	1,383	1,003	887
	N	%	N	%	N	%	N	%
Female	1,469	74.6	502	73.1	113	65.3	34	65.4
Race								
White	2,328	74.9	870	76.7	259	79.7	86	81.1
Black	333	10.7	115	10.1	29	8.9	7	6.6
Other	449	14.4	149	13.1	37	11.4	13	12.3
Hispanic	402	12.9	139	12.3	45	13.8	16	15.1
Employment status								
Employed	975	58.7	314	54.2	69	46.9	19	43.2
Unemployed	612	36.9	243	42.0	72	49.0	24	54.5
Retired	73	4.4	22	3.8	6	4.1	1	2.3
Medical insurance								
Private	848	52.2	254	44.5	52	36.6	14	31.8
Public	282	17.4	109	19.2	30	21.4	10	23.3
None	534	33.2	223	39.3	60	43.2	20	46.5
Marital status								
Single	475	28.6	171	29.5	40	27.2	10	22.7
Married/cohabiting	716	43.1	238	41.0	61	41.5	18	40.9
Divorce/separated	429	25.8	155	26.7	42	28.6	14	31.8
Widowed	41	2.5	16	2.8	4	2.7	2	4.5
Clinical Features	N	%	N	%	N	%	N	%
First episode occurrence before age 18	1,200	39.0	436	38.8	120	37.0	41	38.7
Recurrent depression	1,940	66.8	718	68.3	188	63.3	59	60.8
Family history of depression	1,694	55.4	609	54.9	165	51.7	58	54.7

Duration of current episode \geq 2 years	787	25.6	311	27.7	88	27.2	34	32.1
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age at first episode (years)	24.9	14.5	24.7	14.2	25.9	14.6	25.9	14.3
Illness duration (years)	16.1	13.5	17.2	13.7	18.2	14.1	21.0	14.9
Number of episodes	4.4	9.7	4.9	11.1	4.4	10.3	5.0	12.1
Duration of current episode (months)	25.9	52.0	28.1	58.8	32.1	68.5	45.9	82.0
Median duration of current episode (months)	8.3		8.7		9.5		10.1	
Quality of Life and Enjoyment Satisfaction Questionnaire score ^b	39.1	14.3	36.5	13.6	33.7	13.5	31.6	12.8
SF-12 Mental ^c	25.6	8.1	25.0	7.7	24.4	7.7	24.0	7.5
SF-12 Physical ^c	48.6	12.1	47.0	12.4	44.5	12.1	43.8	12.3
Work and Social Adjustment Scale score ^d	25.0	8.7	26.3	8.2	28.3	7.7	29.4	7.0
HRSD ₁₇ score	21.9	5.2	22.5	5.2	23.4	5.2	23.9	5.4
IDS-C ₃₀ score ^e	39.1	9.6	40.6	9.7	42.6	9.4	43.6	9.8
QIDS-IVR score ^f	16.9	3.3	17.3	3.3	17.9	3.0	18.3	3.1
Cumulative Illness Rating Scale								
Categories endorsed	2.5	1.5	2.6	1.6	2.8	1.6	3.1	1.6
Total score	4.7	3.9	5.1	4.0	5.8	4.5	6.2	4.3
Severity score	1.8	0.8	1.8	0.8	2.0	0.9	2.0	0.7
	N	%	N	%	N	%	N	%
Psychiatric Diagnostic Screening Questionnaire								
Agoraphobia	559	18.2	240	21.4	89	27.5	32	30.2
Alcohol abuse/dependence	371	12.0	136	12.1	36	11.1	8	7.5
Bulimia	607	19.7	232	20.6	67	20.7	20	18.9

Drug abuse/dependence	234	7.6	80	7.1	21	6.5	7	6.6
Generalized anxiety disorder	736	23.9	290	25.8	94	29.0	36	34.0
Hypochondriasis	336	10.9	139	12.4	45	13.9	14	13.2
Obsessive-compulsive disorder	723	23.5	265	23.6	97	29.9	31	29.2
Panic disorder	422	13.7	183	16.3	65	20.1	21	19.8
Posttraumatic stress disorder	387	12.6	172	15.3	55	17.0	16	15.1
Social phobia	963	31.3	379	33.7	117	36.1	35	33.0
Somatoform disorder	284	9.2	105	9.3	35	10.8	9	8.5
Number of axis I comorbid psychiatric disorders								
0	606	19.7	190	16.9	48	14.8	12	11.3
1	740	24.0	257	22.9	68	21.0	23	21.7
2	577	18.7	217	19.3	62	19.1	25	23.6
3	363	11.8	139	12.4	42	13.0	14	13.2
4+	793	25.8	321	28.6	104	32.1	32	30.2

^a Sums do not always equal N due to missing values. Percentages are based on available data.

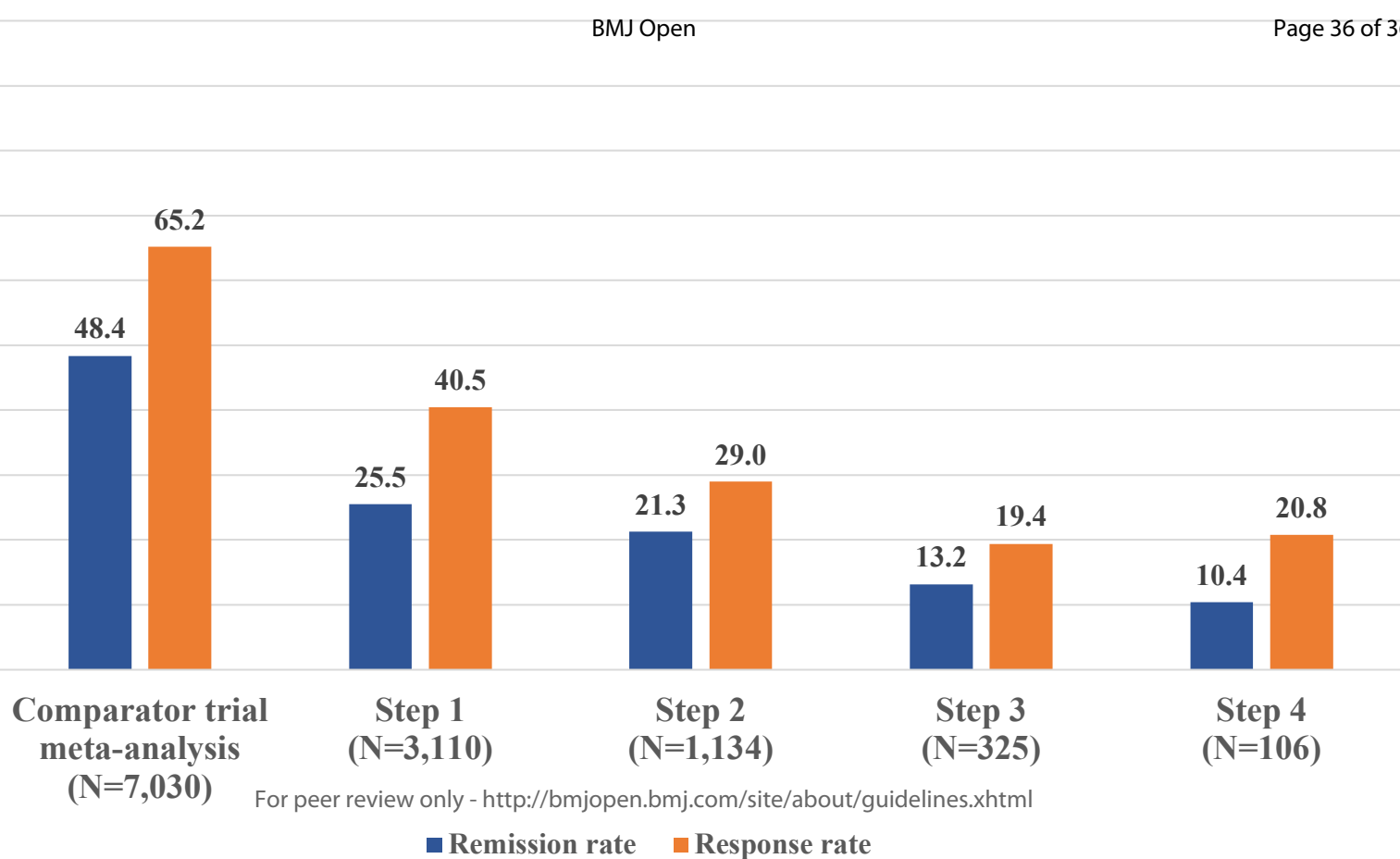
^b Integrated voice response (IVR) administered version of the Quality-of-Life Enjoyment and Satisfaction Questionnaire assessing participants' global rate of satisfaction. Higher scores (range=0–100) represent greater life enjoyment and satisfaction.

^c IVR-administered version of the SF-12 assessing perceived mental and physical health status. Two subscales (physical health factor and mental health) range from 0 to 100— higher scores indicate better functioning with a population norm for each score of 50.

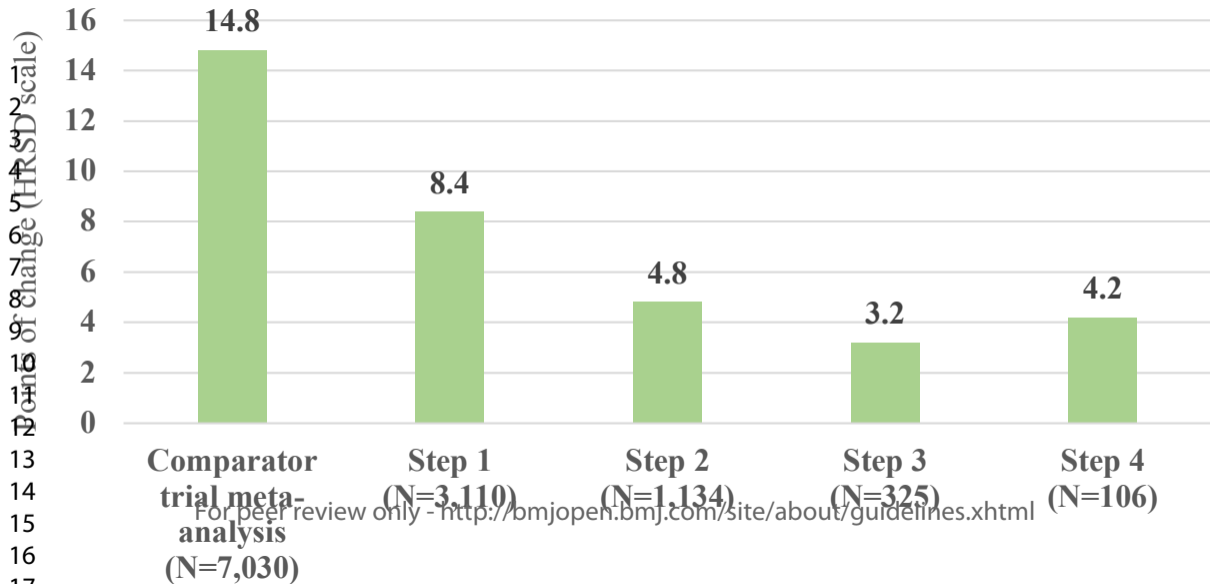
^d IVR-administered version of the Work and Social Adjustment Scale. Scores between 10 and 20 are associated with significant functional impairment while scores above 20 suggest moderate to severe functional impairment.

^e Inventory of Depressive Symptomatology administered telephonically.

^f IVR-administered version of the QIDS.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27For peer review only - <http://bmjopen.bmj.com/site/about/guidelines.xhtml>

■ Remission rate ■ Response rate



BMJ Open

What are the Treatment Remission, Response, and Extent of Improvement Rates after up to Four Trials of Antidepressant Therapies in Real-World Depressed Patients? A Reanalysis of the STAR*D Study's Patient-Level Data with Fidelity to the Original Research Protocol

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2022-063095.R4
Article Type:	Original research
Date Submitted by the Author:	22-May-2023
Complete List of Authors:	Pigott, H.; None Kim, Thomas; University of Pennsylvania, Psychology Xu, Colin; University of Pennsylvania, Psychology Kirsch, Irving; Harvard Medical School Amsterdam, Jay; University of Pennsylvania Perelman School of Medicine, Psychiatry
Primary Subject Heading:	Mental health
Secondary Subject Heading:	Pharmacology and therapeutics
Keywords:	Adult psychiatry < PSYCHIATRY, Depression & mood disorders < PSYCHIATRY, CLINICAL PHARMACOLOGY

SCHOLARONE™
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1
2
3 **What are the Treatment Remission, Response, and Extent of**
4 **Improvement Rates after up to Four Trials of Antidepressant Therapies in Real-World**
5 **Depressed Patients? A Reanalysis of the STAR*D Study's Patient-Level Data**
6 **with Fidelity to the Original Research Protocol**
7

8 Authors:

9 H. Edmund Pigott, Thomas T. Kim, Colin Xu, Irving Kirsch, & Jay D. Amsterdam
10
11
12

13 H. Edmund Pigott, Ph.D.

14 Position: Clinical Psychologist

15 Institution: none

16 Email: HPIGOTT75@gmail.com
17
18

19 Thomas T. Kim

20 Position: PhD candidate in Psychology

21 Institution: Department of Psychology, University of Pennsylvania
22
23

24 Colin Xu

25 Position: PhD candidate in Psychology

26 Institution: Department of Psychology, University of Pennsylvania
27
28

29 Irving Kirsch, Ph.D.

30 Position: Associate Director, Program in Placebo Studies

31 Institution: Harvard Medical School
32
33

34 Jay D. Amsterdam, MD

35 Position: Professor of Psychiatry (Emeritus), Depression Research Unit

36 Institution: Perelman School of Medicine, University of Pennsylvania
37
38

39 Corresponding author: H. Edmund Pigott, Ph.D.
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Abstract

Objective: Reanalyze the patient-level dataset of the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study with fidelity to the original research protocol and related publications.

Design: The study was open label and semi-randomized examining the effectiveness of up to four optimized, and increasingly aggressive, antidepressant therapies in depressed adults. Patients who failed to gain adequate relief from their level 1 trial on the SSRI citalopram could receive up to three additional treatment trials in levels 2-4.

Setting: 41 North American psychiatry and primary care treatment centers.

Participants: 4,041 adults screened positive for major depressive disorder. In contrast to most clinical trials, STAR*D enrolled patients seeking care (versus recruited) and included patients with a wide range of common co-morbid medical and psychiatric conditions to enhance the generalizability of findings to real-world clinical practice.

Interventions: STAR*D evaluated the relative effectiveness of 13 antidepressant therapies in treatment levels 2-4 for depressed patients who failed to gain adequate benefit from their level 1 medication trial.

Main Outcome Measures: According to the STAR*D protocol, the primary outcome was remission, defined as a score <8 on the blinded Hamilton Rating Scale for Depression (HRSD). Response was a secondary outcome defined as ≥50% reduction in HRSD scores. STAR*D's protocol specifically excluded all non-blinded clinic-administered assessments from use as research outcome measures.

Results: STAR*D investigators did not use the protocol-stipulated HRSD to report cumulative remission and response rates in their summary article, and instead used a non-blinded clinic-administered assessment. This inflated their report of outcomes, as did their inclusion of 99 patients who scored as remitted on the HRSD at study outset as well as 125 who scored as remitted when initiating their next-level treatment. These patients should have been excluded from data analysis. In contrast to the STAR*D-reported 67% cumulative remission rate after up to four antidepressant treatment trials, the rate was 35.0% when using the protocol-stipulated HRSD and inclusion in data analysis criteria.

Conclusion: STAR*D's cumulative remission rate was approximately half of that reported.

Strengths and limitations of this study

- We reanalyzed the largest ever prospective antidepressant trial's patient-level dataset with fidelity to the original research protocol and related publications.
- The reanalysis was conducted under the guidelines of the *Restoring Invisible and Abandoned Trials* (RIAT) initiative.

- Treatment remission, response, and extent of symptom improvement rates were calculated for 14 antidepressant therapies for those patients who met STAR*D's inclusion in data analysis criteria as well as the overall cumulative remission rate after up to four trials of antidepressant therapies.
- We calculated STAR*D's remission rate using the protocol-stipulated HRSD as well as combining the HRSD remissions with those from a non-stipulated measure of remission for patients missing an exit HRSD score. Combining STAR*D's HRSD-defined remissions with those from the non-stipulated measure increased its cumulative remission rate from 35.0% to 41.3%.
- Finally, we compared STAR*D's outcomes to those found in a meta-analysis of 7,030 patients enrolled in similar open-label antidepressant comparator trials. Whereas the treatment remission and response rates in comparator trials averaged 48.4% and 65.2% respectively, they were only 25.5% and 40.5% for STAR*D's level 1 patients and worse in treatment levels 2-4. Similarly, comparator trials patients' mean change on the HRSD was 14.8 points versus 8.4 points for STAR*D's level 1 patients and worse for patients in treatment levels 2-4.

Introduction

At a cost of 35 million US dollars, the National Institute of Mental Health (NIMH) funded Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study is the largest and most expensive prospective antidepressant trial ever conducted with over 100 journal articles published by study investigators.[1-7] In contrast to most clinical trials that enroll symptomatic volunteers (typically recruited through advertising), STAR*D enrolled 4041 patients who screened positive for major depressive disorder (MDD) while seeking routine medical or psychiatric care. STAR*D did not exclude patients with medical conditions and most comorbid psychiatric disorders, thereby increasing the generalizability of its findings to real-world clinical practice.

The STAR*D study provided up to four treatment trials per patient and was designed to give guidance in selecting the best next-level treatment option for the many patients who fail to gain sufficient relief from their first, and/or subsequent, antidepressant trial. To mimic clinical practice, STAR*D used an open-label research design with no control group during any phase of the study.

Our STAR*D reanalysis examines key methodological deviations from its research protocol and related publications, and these deviations' impact on its investigators' report of outcomes. In STAR*D's Rationale and Research Design article, and repeated in the level 1-4 published study outcomes, STAR*D investigators stated, "*the primary outcome is depressive symptom severity, measured by the 17-item Hamilton Rating Scale for Depression (HRSD)*." [8, p. 120]. STAR*D's prespecified primary outcome was remission, defined as scoring <8 on the HRSD which was administered telephonically by Research Outcome Assessors (ROAs) blind to patients' study status (treatment level entry/exit/follow-up). Response was a secondary outcome defined as a ≥50% reduction in patients' HRSD scores. Remission as defined by the HRSD (according to the protocol) was not presented in STAR*D's summary article.[7] Furthermore, despite its investigators' numerous publications, neither change in HRSD depressive symptom severity nor HRSD response rates have been reported for STAR*D's six primary studies [1-6] and summary article.[7] Instead, response rates and change in symptom severity were reported using the clinic-administered Quick Inventory of Depressive Symptomatology–Self Report (QIDS-SR), a measure developed by the STAR*D principal investigators.[9] This occurred despite the fact that STAR*D's research protocol specifically excluded all clinic-administered assessments, such as the

1
2
3 QIDS-SR, from use as research outcome measures since they were not blinded and instead, used to
4 guide patient care. The protocol states:
5

6 *Recall that the research outcomes assessments are distinguished from assessments*
7 *conducted at clinic visits. The latter are designed to collect information that guides*
8 *clinicians in the implementation of the treatment protocol. Research outcomes*
9 *assessments are **not** collected at the clinic visits. They are not collected by either*
10 *clinicians or Clinical Research Coordinators.*[10,p.47-48; emphasis in the original]
11
12

13 In their summary article, STAR*D investigators used the QIDS-SR as the sole measure to report
14 remission, response, and extent of symptom improvement. This article's Abstract states that "the
15 overall cumulative remission rate was 67%" with no qualifiers to this claim.[7, p.1905] Besides making
16 this claim based on an assessment the protocol specifically excluded from use as a research measure, it
17 is not until the article's Results section that readers learn this high level of treatment success did not
18 occur. The STAR*D investigators' claim was theoretical—an estimate based on the provisos of what
19 would have happened if there were no study dropouts, and furthermore, "that those who exited the
20 study would have had the same remission rates as those who stayed in the protocol." [7, p.1910] As
21 Pigott et al. documented though, the investigators' assumptions are not true in the real world since
22 more patients dropped out than remitted in each STAR*D treatment level,[11] and furthermore, it has
23 been found in placebo-controlled trials that patients who drop out are more likely to have had adverse
24 treatment side effects and/or emergent suicidality.[12]
25
26

27
28 Unfortunately, the STAR*D investigators' claim of a 67% cumulative remission rate has
29 become accepted clinical wisdom, and the provisions on which it is based are commonly not
30 referenced when portraying STAR*D's findings. For example, in 2009 NIMH's Director Dr.
31 Thomas Insel claimed STAR*D found "at the end of 12 months, with up to four treatment
32 steps, roughly 70% of participants were in remission." [13, p.1466] Similarly in 2013, an
33 editorial in the *American Journal of Psychiatry (AJP)* claimed STAR*D found "after four
34 optimized, well-delivered treatments, approximately 70% of patients achieve remission." [14,
35 p.580]. More recently (2022), a New York Times' article claimed that half of STAR*D's
36 participants "had significantly improved after using either the first or second medication, and
37 nearly 70 percent of people had become symptom-free by the fourth antidepressant." [15]
38 These are not factual statements of STAR*D's findings.
39
40

41 The first author has made published criticisms alleging protocol violations that appear to inflate
42 STAR*D's findings and called for the reanalysis of the dataset by independent investigators.[16] In 2018,
43 the first and fourth authors collaborated with researchers from the University of Connecticut to
44 reanalyze STAR*D's level 1 data obtained from NIMH.[17] This reanalysis found substantial inflation of
45 STAR*D's reported remission and response rates. Furthermore, the reanalysis found that the extent of
46 HRSD improvement in STAR*D's level 1 trial was approximately half that of open-label antidepressant
47 comparator trials.
48
49

50 Our published criticisms of STAR*D investigators' report of outcomes are as follows:[18]
51

- 52 • While STAR*D investigators used the HRSD to report remission rates in their levels 1-4
53 articles,[1-6] the QIDS-SR was used as the sole measure to report remission, response, and
54 extent of improvement rates in their summary article[7] without disclosing that the protocol
55 specifically excluded all non-blinded/clinic-administered assessments such as the QIDS-SR from
56
57
58
59
60

1
2
3 use as outcome measures. The primary outcome measure, the HRSD, should have been used to
4 report the summary article's outcomes.

- 5 • Using data from the 931 patients deemed ineligible for analysis in STAR*D's level 1 article
6 because these patients lacked a baseline ROA-administered HRSD score of ≥ 14 , in STAR*D's
7 levels 2-4 and summary articles without clear disclosure. This included 99 patients who scored
8 < 8 on their baseline HRSD—indicating these patients met STAR*D's remission criterion at study
9 outset and should not have been included in their report of outcomes.
- 10 • Excluding from analysis 370 patients who dropped out after starting on citalopram in their first
11 clinic visit without taking the exit HRSD despite STAR*D investigators stating, "our primary
12 analyses classified patients with missing exit HRSD scores as nonremitters a priori." [1, p.34]
13 These 370 early dropout patients should have been counted as nonremitters as prespecified.
- 14 • Including in their analyses 125 patients who scored as remitted at entry into their next-level
15 treatment. This occurred despite STAR*D investigators prespecifying that, "patients who begin
16 a level with HRSD < 8 will be excluded from analyses." [8, p.130]
17
18
19

20 This reanalysis article uses the patient-level dataset obtained from NIMH to replicate the STAR*D
21 summary article which used descriptive statistics to present the remission, response, and extent of
22 symptomatic improvement for 14 antidepressant therapies based on the QIDS-SR.[7] We perform the
23 same descriptive analyses with the key differences compared to those presented in STAR*D's summary
24 article being: 1) ours is based on the protocol-specified HRSD and only uses the QIDS-SR for those
25 patients missing their exit HRSD and 2) we only included patients who met the inclusion for data analysis
26 criteria stipulated in the research protocol and related publications. Future efforts will use inferential
27 statistics to reanalyze STAR*D's levels 2-4 semi-randomized comparator trials, including the extent of
28 emergent suicidal ideation and 12-month follow-up outcomes tied to each compared treatment.
29
30

31 **Method**

32 *RIAT Initiative*

33
34 The Restoring Invisible and Abandoned Trials (RIAT) initiative started in 2013 calling on funders
35 and investigators of abandoned (unpublished) or misreported studies to publish undisclosed
36 outcomes or correct misleading publications.[19] If investigators failed to correct a study
37 identified as misreported, independent investigators were encouraged to correct the record by
38 reanalyzing the study's patient-level dataset consistent with the research protocol and analytic
39 plan.
40
41
42

43
44 On March 6, 2019, the RIAT investigators published our response to a 'Call to Action' statement in the
45 *British Medical Journal*, in which we stated our intention to reanalyze the STAR*D dataset.[18] We then
46 notified STAR*D's principal investigators of our intention and requested they inform us whether they
47 would undertake a reanalysis of the dataset adhering to the research protocol. On March 22, 2019,
48 STAR*D investigators acknowledged our email notification, indicated the STAR*D data were in the public
49 domain, and stated they had no interest in undertaking a reanalysis.
50

51
52 In July 2019, we received a STAR*D Data Use Certificate, issued by the NIMH Data Archive Data Access
53 Committee, and gained access to the STAR*D levels 1-4 and follow-up patient-level dataset consisting of
54 26 text files, and limited supporting study documentation. In September 2019, we obtained funding
55 from the RIAT Support Center to reanalyze STAR*D.
56
57
58
59
60

Patients

STAR*D patients were 18 to 75 years of age, seeking care at 18 primary and 23 psychiatric care clinics. Clinical research coordinators (CRCs) screened 4,790 patients for MDD. This screening included the CRCs' administering the HRSD, on which 4,041 patients scored ≥ 14 , met the other inclusion criteria, and enrolled into the study. CRCs also gathered patients' psychiatric history, demographic information, and administered both the Cumulative Illness Rating Scale and the Psychiatric Diagnostic Screening Questionnaire to determine the extent of comorbid medical and psychiatric disorders.

Levels/Steps of Acute Treatment

STAR*D investigators sought to provide the highest quality of care to maximize the number of remissions while minimizing dropouts (see Supplemental Table 1). Supplemental Table 2 describes the antidepressant therapies available in treatment levels 1-4 while steps refer to the numeric order of treatments. As seen in Figure 1, treatment steps 1 and 2 correspond to levels 1 and 2 treatments. Similarly, for most patients their levels 3 and 4 treatments correspond to treatment steps 3 and 4. For level/step 2 patients though who failed to respond adequately to cognitive therapy alone or combined with citalopram and chose to continue in the study, their third treatment step was designated level 2A and they were randomized to one of two level 2 switch medications. For these patients, their level 2A treatment was their third treatment step. For level 2A patients who did not adequately benefit from this medication trial and chose to continue in the study, they entered a fourth treatment step consisting of level 3 treatments.

All patients were administered the SSRI citalopram for their level 1 treatment. Each treatment level consisted of 12 weeks of antidepressant therapy, with an additional 2 weeks for patients deemed close to remission. Treatment was administered using a system of measurement-based care that assessed symptoms and side effects at each clinic visit. STAR*D investigators state, *"To enhance the quality and consistency of care, physicians used the clinical decision support system that relied on the measurement of symptoms (QIDS-C and QIDS-SR), side-effects, medication adherence, and clinical judgment based on patient progress."*[1, p.30] This system was used to guide medication management of a fully adequate dose for a sufficient time 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]

For those patients who failed to gain an adequate response from citalopram, STAR*D allowed them to select acceptable treatment options for randomization in levels 2 to 4 *"to empower patients, strengthen the therapeutic alliance, optimize treatment adherence, and improve outcome"*[20, p.483]. The treatment options available for randomization involved either switching to a new treatment or augmenting the patient's current treatment. Treatment levels 2 to 4 evaluated the relative effectiveness of 11 pharmacologically distinct drug/drug combination treatments. Cognitive therapy was also available as either a switch or citalopram augmentation option in level 2.

STAR*D Follow Up Phase

In each treatment trial for levels 1-4, patients who scored < 6 on their last QIDS-Clinician version (QIDS-C) were considered clinician-rated remissions and encouraged to enter the 12-month follow-up phase. During follow-up, patients continued their "previously effective acute treatment medication(s) at the doses used in acute treatment but that any psychotherapy, medication, or medication dose change

1
2
3 could be used.”[7, p.1908] Based on prior research, a QIDS score of <6 was estimated by STAR*D
4 investigators to correspond to a score of <8 on the HRSD, STAR*D’s prespecified primary outcome
5 measure for classifying patients as remitted.[9] Clinicians strongly encouraged patients who did not
6 obtain a QIDS-defined remission to enter the next-level treatment. Patients who failed to attain a QIDS-
7 defined remission, but did have a $\geq 50\%$ reduction on the QIDS-C and did not want to be randomized to a
8 next-level treatment, were also encouraged to enter follow-up.
9

10 11 *Research Design of the STAR*D Study*

12
13 STAR*D investigators developed a new research design for the study termed “equipoise-stratified” to
14 evaluate the relative efficacy of 13 antidepressant therapies in levels 2-4 for depressed patients who
15 failed to gain adequate benefit from their level 1 medication trial.[21] In level 1, all patients received
16 citalopram as their first treatment. In level 2, patients were informed regarding seven treatment
17 options to choose from: four switch options in which citalopram was stopped and the new treatment
18 initiated and three augmentation options in which citalopram was combined with a second
19 antidepressant treatment. In level 3, patients were informed regarding four treatment options to
20 choose from: two switch options and two augmentation options. Level 4 involved randomization to one
21 of two medication/medication combination switch options.
22
23

24 *Analytic Plan of the RIAT Reanalysis*

25
26 We reanalyzed the STAR*D patient-level dataset with fidelity to the original research protocol wherever
27 possible. Where the protocol was silent, we used other STAR*D publications to guide our analysis. This
28 occurred four times. First, the protocol is silent regarding patients who entered the study without a
29 baseline ROA-administered HRSD score of ≥ 14 . In their level 1 article, STAR*D investigators deemed the
30 931 such patients who lacked this marker of depression severity ineligible for inclusion in data
31 analysis.[1] We do the same and extend this exclusion for such patients who continued on to levels 2-4
32 because their extent of depression severity at study outset is not known. Second, the protocol is silent
33 on what to do with patients who met the remission criteria on the HRSD at entry into their next-level
34 treatment. In STAR*D’s Rationale and Research Design article though, its investigators prespecify that
35 “patients who begin a level with HRSD <8 will be excluded from analyses.”[8, p.130] We therefore
36 excluded 125 such patients from our analyses of treatment levels 2-4. Third, the protocol is silent on
37 how to analyze patients who exit a treatment without taking the HRSD. STAR*D investigators state in
38 their level 1 article, “our primary analyses classified patients with missing exit HRSD scores as
39 nonremitters a priori”[1, p.34] and repeat similar statements in their level 2-4 articles.[2-6] Therefore,
40 we do likewise.
41
42
43

44 Finally, STAR*D had many patients with missing exit HRSD scores. In their level 2-4 articles, STAR*D
45 investigators used a correspondence table to map the final QIDS-SR score to the HRSD for patients
46 missing their exit HRSD score to assess the impact of their approach to counting such patients as
47 “nonremitters a priori.”[22,23] For patients with missing exit HRSD scores, we therefore mapped their
48 last QIDS-SR score to the HRSD and used it to calculate the mean HRSD exit, mean change, and
49 combined HRSD & QIDS-SR response rates for all treatments. We also calculated STAR*D’s remission
50 rate both as prespecified based on an exit HRSD score of <8 as well as a final QIDS-SR score of <6 for
51 those patients missing an exit HRSD score.
52
53

54 All pre-processing and analyses were performed in R.[24] Authors 2 and 3 identified patients by their
55 subject key and used this variable to match information across datasets. Data on patients’ treatment
56
57
58
59

1
2
3 pathways, and when patients transitioned from one level to the next, were taken from the IVRA dataset
4 completed by CRCs, and verified against the data on patient level exits. Authors 2 and 3 then compared
5 the number of patients identified for all level 1-4 treatments to that reported in the STAR*D summary
6 article's patient flowchart, and the number of patients matched.[7]
7

8
9 Next, authors 2 and 3 applied STAR*D's level 1 inclusion for data analysis criterion to patients in
10 treatment levels 2-4 as well as excluded from analysis the 125 patients who scored <8 on the HRSD at
11 entry into their next-level treatment. We counted these 125 patients as remitted in the prior treatment
12 level but excluded them from the analyses of subsequent treatments. Supplemental Table 3 presents
13 the number of level 2-4 patients excluded from our reanalysis, and the reasons for their exclusion.
14 Supplemental Table 4 identifies the number of patients with missing entry and/or exit HRSD scores for
15 all level 1-4 treatments. As seen in Supplemental Table 4, 1,330 patients were missing their exit HRSD
16 score across all treatments.
17

18
19 We then compared STAR*D's outcomes to those found in a meta-analysis of 7,030 patients enrolled in
20 antidepressant comparator trials.[25] Similar to STAR*D, comparator trials typically are conducted
21 open-label without a control group and therefore are the appropriate comparison data for STAR*D's
22 outcomes. Continuous HRSD improvement means were provided by the first author of the meta-
23 analysis.[25]
24

25
26 Finally, we compared the STAR*D protocol's step-by-step predictions of patient drop out and the
27 number of patients who would have a satisfactory treatment response and enter follow-up care to what
28 actually occurred.[10] While the purpose of these predictions' was to estimate the number of
29 continuing patients available for randomization in treatment levels 2-4, at the meta-level these
30 predictions are an important hypothesis STAR*D tested by assessing how well its investigators could
31 predict the aggregate step-by-step successful treatment outcomes from their treat-to-remission model
32 of care.
33

34 *Patient and Public Involvement*

35
36
37 Neither patients nor the public were involved in the design, conduct, reporting, or dissemination plans
38 of our research.
39

40 **Results**

41
42
43 Figure 1 presents the overall flow of patients enrolled in the various protocol-defined treatment levels
44 and places them in groups defined by the number of treatment steps. Of the 4,041 patients enrolled
45 into STAR*D, 3,110 met the eligibility for data analysis criterion of having a ROA-administered HRSD
46 score ≥ 14 at study outset. Figure 1 also identifies the number of patients who exited the study following
47 each treatment step, the number who entered follow-up after each treatment step, and the number
48 who were randomly assigned to a next-level treatment.
49

50
51 Supplemental Table 5 describes the demographic and clinical features of the patients who entered
52 treatment in steps 1-4 based on their level 1 baseline presentation when enrolling into the study.
53 Summary statistics are presented as means and standard deviations for continuous variables and
54 percentages for discrete variables. Note that 55.7% of STAR*D patients had 2 or more comorbid axis 1
55 disorders when first enrolled based on the Psychiatric Diagnostic Screening Questionnaire and averaged
56
57
58
59

2.5 comorbid medical conditions based on the Cumulative Illness Rating Scale. Furthermore, the average length of patients' current MDD episode was 25.9 months. In a post hoc analysis, STAR*D investigators found that 77.8% of its enrolled patients would have been excluded from most antidepressant trials due to having two or more concurrent medical conditions, more than one comorbid psychiatric disorder, and/or a current depressive episode lasting > 2 years.[26]

Table 1 presents the mean HRSD entry, exit, and change scores for patients by the specific treatment they received in steps 1-4 as well as the HRSD remission and response rates. Table 1 also provides the HRSD cumulative remission rate after up to 4 trials on antidepressant therapies as well as the combined HRSD plus QIDS-SR remission and response rates for the 1,330 patients missing an exit HRSD score.

Table 1: Outcomes Across All Treatments

Treatment Step	HRSD Score		*Mean Change [95% confidence interval] (SD)	HRSD Remission Rate # (%)	*Combined HRSD & QIDS-SR Remission Rate # (%)	*Combined HRSD & QIDS-SR Response Rate # (%)
	Entry Mean (SD)	*Exit Mean (SD)				
Step 1 (N=3,110)	21.87 (5.21)	13.49 (8.42)	8.38 [8.10, 8.67] (8.11)	794 (25.5%)	938 (30.2%)	1261 (40.5%)
Step 2 (N=1,134)	18.76 (6.24)	13.97 (8.09)	4.79 [4.37, 5.21] (7.23)	241 (21.3%)	283 (25.0%)	329 (29.0%)
Switch strategy (N=620)	19.85 (6.08)	14.70 (8.01)	5.16 [4.59, 5.73] (7.22)	113 (18.2%)	134 (21.6%)	178 (28.7%)
Bupropion (N=190)	20.11 (6.25)	15.32 (7.85)	4.78 [3.82, 5.75] (6.78)	31 (16.3%)	37 (19.5%)	46 (24.2%)
Sertraline (N=198)	19.95 (5.98)	14.92 (8.02)	5.03 [4.04, 6.01] (7.10)	32 (16.2%)	36 (18.2%)	57 (28.8%)
Venlafaxine (N=192)	19.89 (6.19)	14.31 (8.12)	5.58 [4.53, 6.63] (7.45)	37 (19.3%)	44 (22.9%)	59 (30.7%)
Cognitive Therapy (N=40)	18.01 (4.96)	12.44 (7.93)	5.58 [2.87, 8.28] (8.73)	13 (32.5%)	17 (42.5%)	16 (40.0%)
Augmentation strategy (N=514)	17.44 (6.18)	13.10 (8.10)	4.34 [3.72, 4.97] (7.23)	128 (24.9%)	149 (29.0%)	151 (29.4%)

Bupropion (N=216)	16.88 (6.11)	12.52 (7.83)	4.36 [3.38, 5.33] (7.30)	54 (25.0%)	64 (29.6%)	66 (30.6%)
Buspirone (N=225)	17.80 (6.50)	13.36 (8.40)	4.43 [3.52, 5.35] (7.02)	58 (25.8%)	68 (30.2%)	66 (29.3%)
Cognitive Therapy (N=73)	17.99 (5.24)	13.98 (7.98)	4.01 [2.25, 5.78] (7.69)	16 (21.9%)	17 (23.3%)	19 (26.0%)
Step 3 (N=325)	19.59 (6.09)	16.38 (7.77)	3.21 [2.48, 3.94] (6.70)	43 (13.2%)	50 (15.4%)	63 (19.4%)
Level 2A (N=28)	20.89 (5.44)	16.96 (6.48)	3.93 [1.81, 6.04] (5.71)	3 (10.7%)	3 (10.7%)	5 (17.9%)
Bupropion (N=12)	19.92 (3.85)	17.58 (7.35)	2.33 [-0.81, 5.48] (5.55)	2 (16.7%)	2 (16.7%)	2 (16.7%)
Venlafaxine (N=16)	21.62 (6.41)	16.50 (5.96)	5.12 [2.33, 7.92] (5.70)	1 (6.2%)	1 (6.2%)	3 (18.8%)
Level 3 (N=297)	19.46 (6.14)	16.32 (7.88)	3.14 [2.37, 3.92] (6.79)	40 (13.5%)	47 (15.8%)	58 (19.5%)
Switch strategy (N=186)	20.01 (6.24)	17.01 (7.91)	2.99 [2.00, 3.99] (6.94)	23 (12.4%)	25 (13.4%)	31 (16.7%)
Nortriptyline (N=92)	19.67 (5.27)	16.99 (8.35)	2.67 [1.10, 4.24] (7.68)	15 (16.3%)	15 (16.3%)	16 (17.4%)
Mirtazapine (N=94)	20.34 (7.08)	17.03 (7.49)	3.30 [2.06, 4.55] (6.15)	8 (8.5%)	10 (10.6%)	15 (16.0%)
Augmentation strategy (N=111)	18.55 (5.89)	15.16 (7.74)	3.40 [2.18, 4.62] (6.56)	17 (15.3%)	22 (19.8%)	27 (24.3%)
Lithium (N=58)	18.69 (6.47)	15.91 (7.29)	2.78 [1.42, 4.15] (5.31)	7 (12.1%)	9 (15.5%)	10 (17.2%)
T3 (N=53)	18.41 (5.25)	14.34 (8.19)	4.07 [1.99, 6.14] (7.69)	10 (18.9%)	13 (24.5%)	17 (32.1%)
Step 4 (N=106)	20.65	16.49	4.16	11	13	22

	(5.54)	(7.47)	[2.80, 5.52] (7.15)	(10.4%)	(12.3%)	(20.8%)
Level 3 (N=16)	20.62 (4.01)	17.62 (6.87)	3.00 [-0.45, 6.45] (7.04)	2 (12.5%)	2 (12.5%)	3 (18.8%)
Tranlycypromine (N=43)	21.02 (6.57)	16.45 (7.89)	4.57 [2.22, 6.92] (7.87)	3 (7.0%)	5 (11.6%)	9 (20.9)
Venlafaxine XR/mirtazapine (N=47)	20.32 (5.02)	16.14 (7.38)	4.18 [2.30, 6.07] (6.59)	6 (12.8%)	6 (12.8%)	10 (21.3%)
CUMULATIVE remission rate after up to four treatment steps				1,089 (35.0%)	1,284 (41.3%)	

*For patients with missing exit HRSD scores, their last QIDS-SR score is mapped to the HRSD and used to calculate HRSD Exit Mean, Mean Change, Combined HRSD & QIDS-SR Remission Rate, and Combined HRSD & QIDS-SR Response Rate.

Table 2 presents patients' aggregate HRSD status in terms of remission, response, and extent of mean symptomatic change at entry and exit for each treatment step as well as study dropout. In step 1, 25.5% of patients remitted. Steps 2-4 show a continuous decrease in remission rates from step 2's 21.3% to step 3's 13.2% and step 4's 10.4% with increasing rates of study dropout from step 1's 34.5% to step 3's 46.2%.

Table 2: Outcomes by Treatment Step

	Step 1 (N=3,110)		Step 2 (N=1,134)		Step 3 (N=325)		Step 4 (N=106)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
HRSD score at entry into step	21.87	5.21	18.76	6.24	19.59	6.09	20.65	5.54
HRSD score at exit from step*	13.49	8.42	13.97	8.09	16.38	7.77	16.49	7.47
HRSD Mean Change*	8.38	8.11	4.79	7.23	3.21	6.7	4.16	7.15
	N	%	N	%	N	%	N	%
Remission at each step exit	794	25.5%	241	21.3%	43	13.2%	11	10.4%
Response at each step exit*	1261	40.5%	329	29.0%	63	19.4%	22	20.8%
Entered Follow-up	902	29.0%	406	35.8%	69	21.2%	38	35.9%
Study Exit/Dropout	1,074	34.5%	403	35.5%	150	46.2%		

* For patients with missing exit HRSD scores, their last QIDS-SR score is mapped to the HRSD and used to calculate HRSD Exit Mean, Mean Change, and Combined HRSD & QIDS-SR Response Rate.

1
2
3 Supplemental Figures 1 and 2 compare the HRSD remission, response, and extent of symptom
4 improvement rates for STAR*D patients in steps 1-4 to that found in a meta-analysis of 7,030 patients
5 enrolled in non-blinded antidepressant comparator trials.[25] In step 1, these measures of
6 improvement among STAR*D's patients were at least one-third less than that found in comparator
7 trials, and improvement was worse in each subsequent treatment step.
8
9

10 Figure 2 compares the STAR*D protocol's predictions of patient dropout and the number of patients
11 who would have a satisfactory treatment response and enter follow-up to what occurred. Cumulatively,
12 STAR*D's investigators predicted that 73.8% of patients would have a successful treatment response
13 and enter follow-up whereas in fact only 45.6% achieved this measure of treatment success.
14 Furthermore, whereas STAR*D investigators predicted that over the course of up to four antidepressant
15 therapies 20.7% of patients would dropout, in fact, 53.7% dropped out. On this measure of treatment
16 failure, STAR*D's dropout rate was 2.6 times greater than predicted.
17
18

19 Figure 3 presents the step-by-step cumulative remission rate in three ways. First, the 'theoretical' rate
20 propagated by STAR*D investigators based on the provisos of what would have happened if there were
21 no study dropouts and that those who did exit had the same QIDS-SR remission rates as those who
22 stayed.[7] Next, the combined HRSD plus QIDS-SR remission rate based on either an exit HRSD score of
23 <8, OR a last clinic visit QIDS-SR score of <6 for the 1,330 patients missing an exit HRSD. Finally, our RIAT
24 reanalysis rate when using the protocol-specified exit HRSD score of <8 as the sole measure of remission
25 for the 3,110 patients who met STAR*D's inclusion in data analysis criteria. The cumulative remission
26 rate after up to four antidepressant therapies using the HRSD was 35.0% versus 41.3% when combined
27 with the QIDS-SR, both of which are substantially less than the 67% cumulative remission rate claimed in
28 the summary article's Abstract.
29
30

31 Discussion

32 *Principal findings and comparison with original STAR*D publication*

33
34
35
36 STAR*D's results highlight the discrepancy in likely outcomes between typical antidepressant clinical
37 trials with their exclusion criteria and the real-world patients for whom these medications are
38 commonly prescribed. Our RIAT reanalysis found poorer outcomes after up to four optimized, and
39 increasingly aggressive, antidepressant therapies than reported in STAR*D's summary article published
40 in *AJP*. [7] In contrast to the 67% cumulative remission rate reported in *AJP*, the actual rate was 35.0%
41 when using the protocol-specified HRSD, and increased to 41.3% when combined with a final clinic-visit
42 QIDS-SR score of <6 for patients missing exit HRSD scores in treatment steps 1-4. The 41.3% cumulative
43 remission rate should be viewed as the "best case scenario" since it added an additional 195 QIDS-
44 defined remissions (a remission measure not specified in the protocol) from the 1,330 patients with
45 missing exit HRSD scores. As there was neither a placebo nor waitlist control group during any phase of
46 the STAR*D study, it is impossible to know to what extent the observed results were due to the
47 pharmacologic effects of the prescribed medications, placebo effects, and/or the passage of time.
48
49

50 Our reanalysis did not assess the durability of treatment effects during the 12-month follow up phase.
51 In their summary article though, STAR*D investigators reported an overall relapse rate of 46.1% for the
52 1,729 patients for whom they had at least one assessment (of up to 12 scheduled) during follow up
53 using a telephonic-administered version of the QIDS [7] whereas Pigott et al. found a far lower sustained
54 recovery rate when incorporating patient dropout in the analysis.[11]
55
56
57
58
59
60

Comparison with other studies

Our reanalysis found that in step 1, STAR*D's remission, response, and extent of improvement rates were substantially less than those reported in other open-label antidepressant comparator trials and then grew progressively worse in steps 2-4.[25] Such studies typically exclude depressed patients with the range and number of comorbid medical and/or psychiatric disorders that were included in STAR*D.

STAR*D's step 1 remission rate was 25.5% followed by a progressive decline in remission rates for those patients receiving subsequent, and increasingly aggressive treatments, such that by step 4 it was only 10.4%. This decline in antidepressants' effectiveness essentially mirrors the findings from randomized and naturalistic, prospective studies reporting a 20-30% loss of effectiveness with each increase in the number of prior antidepressant trials.[27-32] Furthermore, several recent analyses suggest that the sequential application of antidepressant medications for non-remitting depression may in fact foster treatment resistance for many patients.[33-36]

Regarding the protocol's predictions of treatment success and patient dropout, it states:

We arrived at these estimates using three experienced practitioners who independently made estimates that were surprisingly close to each other. Then, via teleconferencing, the final estimates were made. ***The underlying assumptions of these estimates come largely by inferences from results of published RCTs.***[10, p.31; emphasis added]

STAR*D's actual measures of treatment success and failure were significantly worse than predicted. As Barbui et al. noted, antidepressant study dropout rates provide a "hard measure of treatment effectiveness and acceptability"[12, p.296] and STAR*D's dropout rate was 2.6 times greater than predicted. This discrepancy further highlights the relative ineffectiveness of antidepressants in treating real-world depressed patients, compared to those reported in conventional studies.

Conclusion

Bias in the clinical literature is commonly associated with industry-funded RCTs, not publicly funded ones.[37] Our RIAT reanalysis though documents scientific errors in this NIMH-funded study. These errors inflated STAR*D investigators' report of positive outcomes.

The STAR*D summary article's claim of a 67% cumulative remission rate was published in 2006. If STAR*D's outcomes had been reported as prespecified, its model of care would likely have faced much stronger criticism 16 years ago and fueled a more vigorous search for evidence-based treatment alternatives.

Ethics Statement:

Ethics committee approval was not required for our reanalysis since the data was anonymized by NIMH.

Data Sharing Statement:

Data is available from the NIMH-supported National Database for Clinical Trials (NDCT).

Funding Statement:

Funding for this project was provided by The RIAT Support Center.

Competing Interests:

The authors have no relevant financial disclosures or other conflicts of interest to report.

Contributorship Statement:

HEP, JDA and IK contributed to the design of the study and secured funding. TK and CX conducted all of the data analyses. HEP wrote the manuscript with input from JDA, IK, TK, and CX.

Acknowledgements:

We thank Termeh Feinberg for her early efforts on this project, particularly her correspondence with the NIMH help desk to resolve issues with the 26 data files as well as the RIAT Support Center for funding this project. Data used in the preparation of this manuscript were obtained from the controlled access datasets distributed from the NIMH-supported National Database for Clinical Trials (NDCT). NDCT is a collaborative informatics system created by the National Institute of Mental Health to provide a national resource to support and accelerate discovery related to clinical trial research in mental health. The content of this publication does not necessarily reflect the views of the RIAT Support Center nor NIMH.

References:

1. 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. 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. 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. Bupropion-SR, sertraline, or venlafaxine-XR after failure of SSRIs for depression. *N Engl J Med*. 2006;354:1231–1242.
3. 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. Medication augmentation after the failure of SSRIs for depression. *N Engl J Med*. 2006;354:1243–1252.
4. Fava M, Rush AJ, Wisniewski SR, Nierenberg AA, Alpert JE, McGrath PJ, Thase ME, Warden D, Biggs MM, Luther JF, Niederehe G, Ritz L, Trivedi MH. 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.
5. Nierenberg AA, Fava M, Trivedi MH, Wisniewski SR, Thase ME, McGrath PJ, Alpert JE, Warden D, Luther JF, Niederehe G, Lebowitz BD, Shores-Wilson K, Rush AJ. 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.
6. McGrath PJ, Stewart JW, Fava M, Trivedi MH, Wisniewski SR, Nierenberg AA, Thase ME, Davis L, Biggs MM, Shores-Wilson K, Luther JF, Niederehe G, Warden D, Rush AJ. Tranylcypromine versus venlafaxine plus mirtazapine following three failed antidepressant medication trials for depression: a STAR*D report. *Am J Psychiatry*. 2006;163:1531–1541.

7. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, Niederehe G, Thase ME, Lavori PW, Lebowitz BD, McGrath PJ, Rosenbaum JF, Sackheim HA, Kupfer DJ, Luther J, Fava M. 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.
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. Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design. *Control Clin Trials*. 2004;25:119–142.
9. Rush AJ, Trivedi MH, Ibrahim HM, Carmody T J, Arnow B, Klein DN, ... Manber R. The 16-item Quick Inventory of Depressive Symptomatology (QIDS), Clinician Rating (QIDS-C), and Self-Report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biological Psychiatry*. 2003;54:573–583.
10. National Institute of Mental Health (NIMH). Sequenced Treatment Alternatives to Relieve Depression (STAR*D) Research Protocol. Washington (DC): NIMH; revised June 28, 2002.
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. Pigott HE, Leventhal AM, Alter GS, Boren JJ. Efficacy and effectiveness of antidepressants: current status of research. *Psychother Psychosom*. 2010;79:267–279.
13. Insel TR, Wang PS. The STAR*D trial: revealing the need for better treatments. *Psychiatric Services*. 2009;60:1466–1467.
14. Greden JF. Workplace depression: personalize, partner, or pay the price. *Am J Psychiatry*. 2013;170:578–581.
15. Smith DG. Antidepressants don't work the way many people think. *New York Times*. November 8, 2022.
16. Pigott HE. The STAR*D trial: it's time to reexamine the clinical beliefs which guide the treatment of major depression. *Canadian J Psychiatry*. 2015;60:9–13.
17. Kirsch I, Huedo-Medina TB, Pigott HE, Johnson B. Do outcomes of clinical trials resemble those "real world" patients? A re-analysis of STAR*D antidepressant data. *Psychology of Consciousness: Theory, Research, and Practice*. 2018 <https://doi.org/10.1037/cns0000164>
18. Pigott HE, Dubin, W, Kirsch I, Amsterdam J. Call to action: RIAT reanalysis of the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) Study. *BMJ*. March 6, 2019; <https://www.bmj.com/content/346/bmj.f2865/rr-10>
19. Doshi P, Dickersin K, Healy D, Vedula SS, Jefferson T. Restoring invisible and abandoned trials: a call for people to publish the findings. *BMJ* 2013;346:f2865.
20. 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.
21. Lavori PW, Rush AJ, Wisniewski SR, Alpert J, Fava M, Kupfer DJ, Nierenberg A, Quitkin FM, Sackeim HA, Thase ME, Trivedi M. Strengthening clinical effectiveness trials: equipoise-stratified randomization. *Biological Psychiatry*. 2001;50:792–801.
22. Rush AJ, Bernstein IH, Trivedi MH, Carmody TJ, Wisniewski S, Mundt JC, Shores-Wilson K, Biggs MM, Woo A, Nierenberg AA, Fava M. An evaluation of the quick inventory of depressive symptomatology and the hamilton rating scale for depression: a sequenced treatment alternatives to relieve depression trial report. *Biol Psychiatry*. 2006 Mar 15;59(6):493–501. doi: 10.1016/j.biopsych.2005.08.022. Epub 2005 Sep 30. PMID: 16199008; PMCID: PMC2929841.
23. Rush AJ. Inventory of Depressive Symptomatology (IDS) and Quick Inventory of Depressive Symptomatology (QIDS). [IDS/QIDS \(ids-qids.org\)](https://ids-qids.org)

24. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
25. Rutherford BR, Sneed JR, Roose SP. Does study design influence outcome? The effects of placebo control and treatment duration in antidepressant trials. *Psychotherapy and Psychosomatics*. 2009;78:172–181.
26. Wisniewski SR, Rush AJ, Nierenberg AA, Gaynes BN, Warden D, Luther JF, McGrath PJ, Lavori PW, Thase ME, Fava M, Trivedi MH: Can phase III trial results of antidepressant medications be generalized to clinical practice? A STAR * D report. *Am J Psychiatry* 2009;166:599–607.
27. Amsterdam JD, Maislin G. Fluoxetine efficacy in treatment resistant depression. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*. 1994;18:243-261.
28. Nierenberg A, Feighner JP, Rudolph R, Cole JO, Sullivan J. Venlafaxine for treatment-resistant unipolar depression. *J Clinical Psychopharmacology*. 1994;14:419-423.
29. Amsterdam JD, Shults J. MAOI safety and efficacy in advanced treatment-resistant depression: a retrospective analysis. *J Affective Disorders*. 2005; 89:183-188.
30. Amsterdam JD, Williams D, Michelson D, Adler LA, Dunner SL, Nierenberg A, Reimherr FW, Schatzberg AF. Tachyphylaxis after repeated antidepressant drug exposures in patients with recurrent major depressive disorder. *Neuropsychobiology*. 2009;59:227-233.
31. Leykin Y, Amsterdam JD, DeRubeis RJ, Shelton RC, Hollon SD. Progressive resistance to SSRI therapy but not to cognitive therapy in the treatment of major depression. *J Consulting & Clinical Psychology*. 2007;75:267-276.
32. Amsterdam JD, Lorenzo-Luaces L, DeRubeis RJ. Step-wise loss of antidepressant effectiveness after repeated antidepressant trials in bipolar II depression. *Bipolar Disorders*. 2016;18:563-570.
33. Fava GA, Offidani E. The mechanisms of tolerance in antidepressant action. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*. 2011;35:1593-602.
34. Amsterdam JD, Lorenzo-Luaces L. Increase in pharmacodynamic tolerance after repeated antidepressant trials in treatment-responsive bipolar II depressed subjects: An exploratory study. *Psychiatria Polska*. 2018;52:957–969.
35. Amsterdam JD, Kim TT. Increased risk of depressive relapse during maintenance therapy after repeated antidepressant trials in treatment-responsive subjects. *J Clinical Psychopharmacology*. 2019;39:344-350.
36. Andrews PW, Amsterdam JD. A hormetic approach to understanding antidepressant effectiveness and the development of antidepressant tolerance. *Psychiatria Polska*. 2020;54:1067–1090.
37. Amsterdam JD, McHenry LB, Jureidini JN. Industry-corrupted psychiatric trials. *Psychiatria Polska*. 2017;51:993–1008.

Figure 1 Caption: Patient Flowchart

Figure 1 Footnote:

* In level 2, 580 patients were randomized to switch medications, 441 to medication augmentation, and 113 to Cognitive Therapy as either a switch or medication augmentation treatment. In level 2A, 28 patients were randomized to one of two level 2 switch medications. For step 3/level 3 patients, 186 were randomized to medication switch and 111 to medication augmentation. For step 4/level 3 patients, 7 were randomized to medication switch and 9 to medication augmentation. For step 4/level 4 patients, 90 were randomized to one of two medication/medication combination switch options.

1
2
3 ** Exit refers to the number of patients who exit the study and do not proceed either to the next
4 treatment level nor enter follow-up.

5 *** Follow-up refers to the number of patients who exit a treatment and enter the 12-month follow-up
6 phase.
7

8
9 **Figure 2 Caption: Comparison of STAR*D Protocol Predictions to What Occurred**

10
11 **Figure 3 Caption: STAR*D's Step-by-Step Cumulative Remission Rate Presented Three Ways**

12
13 **Figure 3 Footnote:**

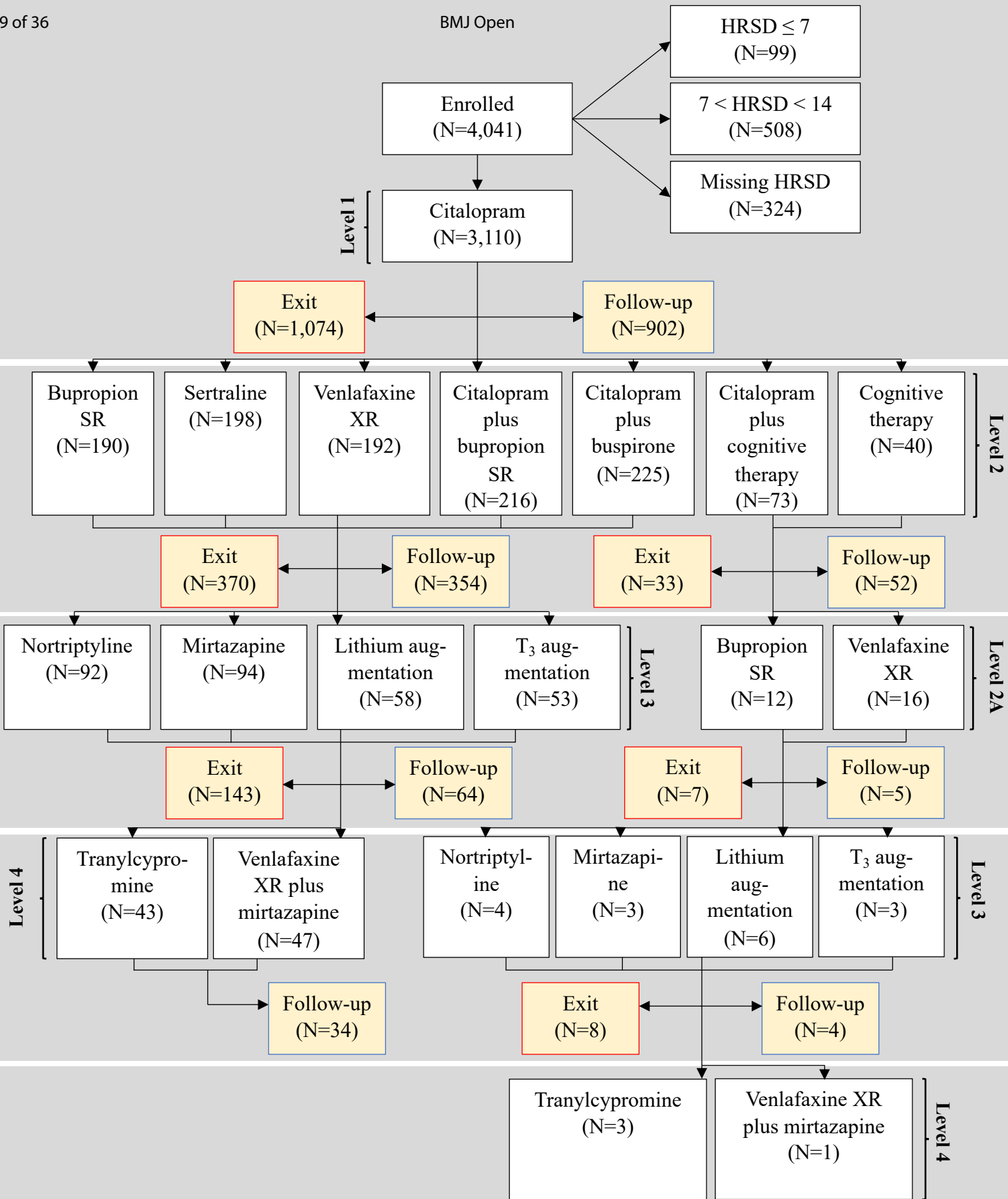
14
15 The step-by-step theoretical remission rates were obtained from the STAR*D summary article where it
16 states: "The theoretical cumulative remission rate is 67% (37+19+6+5)."[7, p.1910].
17

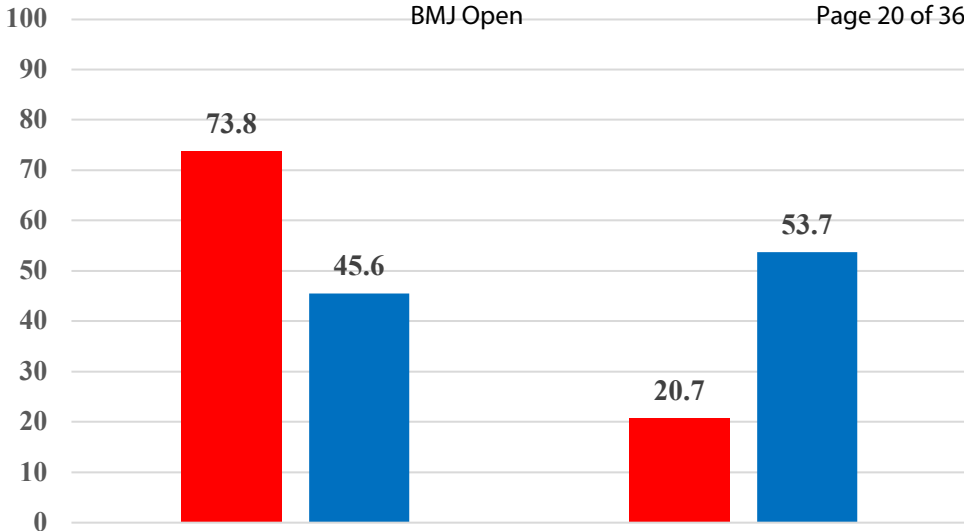
18
19 The HRSD + QIDS-SR cumulative remission rate was taken from Table 1. It combines the 1,089 patients
20 with an exit HRSD score of <8 with the 195 patients who were missing an exit HRSD score but had a final
21 clinic-visit QIDS-SR score of <6.
22

23
24 The RIAT Reanalysis cumulative remission rate is based on an exit HRSD score of <8 as the sole measure
25 of remission for the 3,110 patients who met STAR*D's inclusion for data analysis criteria.
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Step 1

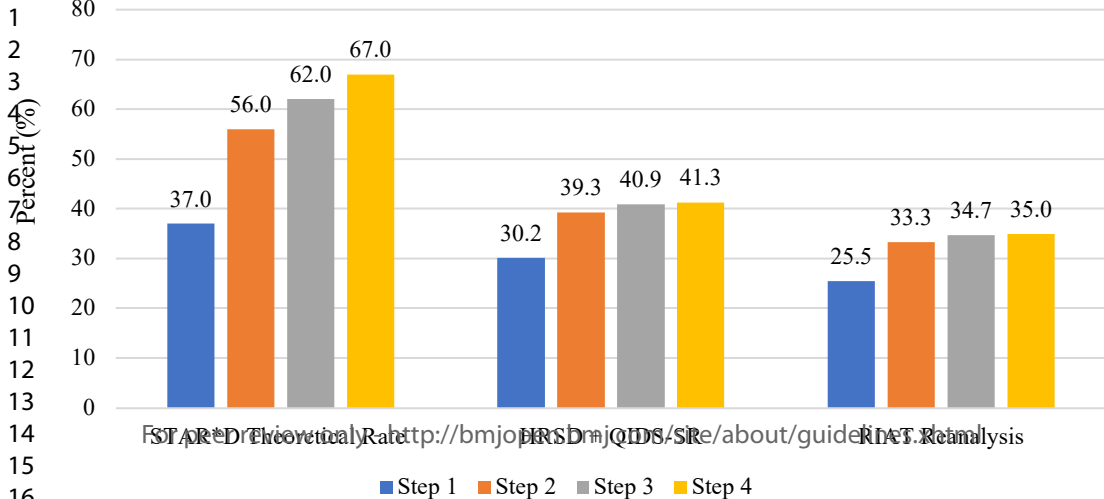
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56



1
2
3
4
Percent (%)

For peer review only: <http://bmjopen.bmj.com/site/about/governing/policies.xhtml>

■ Star*D predicted ■ RIAT reanalysis



**Supplementary Table 1:
Highest Quality of Acute and Continuing-Care to
Maximize Remissions While Minimizing Relapse and Dropouts**

Descriptor	Explanation
Optimized Sustained Study Participation to Minimize Dropouts ²⁰ , p. 473-474	<ul style="list-style-type: none"> • Promoted patients' study affiliation via STAR*D-branded brochures, bimonthly newsletters, and an informational video emphasizing STAR*D's public health significance and the critical role played by patients; • Educated patients and families about depression and its treatment using a multi-step educational package. This included teaching the "mechanism of action" for patients' current antidepressant and educating patients that "depression is a disease, like diabetes or high blood pressure" and "can be treated as effectively as other illnesses," etc.; • Used a letter reminder system to alert patients before appointments in those clinics without such systems who had a >15% rate of missed appointments; • Ensured timely follow-up and rescheduling of missed appointments by calling patients on the day of the missed appointment, and again within 24 hours, if there was no response. Patient's physician sent letter within 48 hours if contact was not established; • Used a letter reminder system for all research outcome assessment calls during acute and continuing-care; • In every clinic visit, the Clinical Research Coordinator (CRC) discussed the research outcomes phone calls with the patient to ensure that the calls were completed on schedule and worked to resolve any problematic issues regarding said calls [Clinical Procedures Manual, page 75]; • Paid patients \$25.00 for participating in each telephonic research outcomes assessment; • Permitted patients to re-enter acute and/or continuing-care within four weeks after having dropped out [Clinical Procedures Manual, page 80]; • Recommended one-year of continuing-care for all patients who achieved a satisfactory clinical response with the essential goal of preventing relapse [Clinical Procedures Manual, page 15] and • Permitted continuing-care patients to remain in the study if they moved from the area [Clinical Procedures Manual, page 81].
Acute-Care Visits	Physicians met with patients on entry into each new step to initiate drug treatment with follow-up visits scheduled on weeks 2, 4, 6, 9, 12, with an optional week 14 visit.

Measurement-Based Care	Conducted structured evaluations of symptoms and side-effects at each visit and included a centralized treatment monitoring and physician feedback system to ensure consistent implementation of optimal care across research sites.
Aggressive Medication Dosing	Provided aggressive medication dosing with a fully adequate dose for a sufficient duration to “ <i>ensure that the likelihood of achieving remission was maximized and that those who did not reach remission were truly resistant to the medication</i> ”. ^{1, p.30}
Liberal Prescribing of Non-Study Medications	Physicians had great leeway in prescribing non-study medications to treat comorbid symptoms resulting in: <ul style="list-style-type: none"> • 17.2% taking Trazodone for sleep; • 11.9% taking an anti-anxiety medication; • 16.7% taking either a sedative or hypnotic medication; and • An undisclosed percent taking medications to address side-effects. ^{2, table 2}
Continuing-Care Visits	Patients saw their physician every 2 months and continued taking their treatment medication(s) at the same doses but their physicians were allowed to make any psychotherapy, medication, and/or medication dose changes to maximize the likelihood of maintaining patients’ remission status. ^{7, p. 1908} Additional continuing-care visits were scheduled when patients began to experience a return of depressive symptoms and/or intolerable side-effects [Clinical Procedures Manual. page 78].
Clinical Research Coordinator (CRC)	Each site had a CRC who: ^{1, p. 30} <ul style="list-style-type: none"> • Saw patients before each visit administering multiple measures to them including the QIDS-SR during each acute-care visit; • Assisted physicians in protocol implementation; and • Provided patients support and encouragement in protocol implementation.
Treatment Designed to Enhance Subject Retention	Treatment was designed to minimize drop-outs and/or non-compliance including: <ul style="list-style-type: none"> • Open label prescribing during acute and continuing-care with no placebo control condition during any study phase; • Patients chose their acceptable treatment assignments for steps two and three to eliminate any concerns they might have about receiving an unacceptable assignment. This resulted in only 21 of 1,439 (1.5%) Step-2 patients making themselves available for random assignment to all treatment options ^{2, p. 1235} while only 29 of 377 (7.7%) did so in Step-3. ^{5, p. 1521} • During each step, patients could enroll immediately into the next step if they had intolerable side-effects or had maximized their current medication(s)’ dosing without achieving a remission; and • During any step, patients could enter continuing-care directly on their current medication(s) if they were treatment responders even if they had not achieved remission. This was done to minimize responders from dropping out in order to avoid having

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

	to discontinue their current medication(s) and start a new drug regimen.
--	--

***Trivedi MH, Stegman D, Rush AJ, Wisniewski SR, Nierenberg AA: STAR * D clinical procedures manual. July 31, 2002. www.edc.pitt.edu/stard/public/study_manuals.html

For peer review only

Supplementary Table 2: Description of Levels 1-4 Treatments

Level 1:

STAR*D investigators report that Citalopram (Celexa) was chosen as the first-line SSRI treatment because (1) absence of discontinuation symptoms; (2) demonstrated safety in elderly and medically fragile patients; (3) easy once-a-day dosing with few dose adjustments; and (4) favorable drug–drug interaction profile.¹ Citalopram was started at 20 mg/day and then raised to 40 mg/day by day 28 and up to 60 mg/day by day 43 and onward. Dose adjustments were based on how long a patient had received a particular dose, symptom changes, and side effect burden.

Level 2 switch treatments:

Citalopram was discontinued without a tapering at the initiation of each level 2 switch treatment. STAR*D investigators chose pharmacologically distinct switch medications. The level 2 treatments were:

- Sertraline (Zoloft), an SSRI with the same pharmacological profile as citalopram. Sertraline was started at a daily dose of 50 mg and increased to 100 mg at day 8, to 150 mg at day 28, and to 200 mg at day 63 and onward.
- Sustained-release bupropion (Wellbutrin SR), an “out-of-class” agent whose neurochemical action mechanisms are unknown; other than that, it does not inhibit serotonin reuptake and is believed to produce antidepressant effects by blocking the reuptake of dopamine and norepinephrine. The daily dose of sustained-release bupropion was 150 mg for week 1, 200 mg from day 8 to 27, 300 mg from day 28 to 41, and 400 mg from day 42 onward.
- Extended-release venlafaxine (Effexor), a “dual-action” agent that inhibits the reuptake of both serotonin and norepinephrine. The starting daily dose of extended-release venlafaxine was 37.5 mg for week 1 and increased to 75 mg from day 8 to 14, to 150 mg from day 15 to 27, to 225 mg from day 28 to 41, to 300 mg from day 42 to 62, and to 375 mg from day 63 onward.
- Cognitive therapy was provided by a trained psychotherapist and scheduled twice weekly for the first four weeks, then once weekly for the remaining 8 weeks (16 sessions total).

Level 2 Citalopram augmentation treatments:

During the augmentation trial, the citalopram dose was kept constant but reduced if side effects developed. The level 2 augmentation treatments were:

- Buspirone (Buspar), a partial agonist at the postsynaptic 5-hydroxytryptamine 1A (5-HT_{1A}) receptor that is believed to enhance the activity of SSRIs through the 5HT_{1A} receptors. The starting dose was 15 mg per day week 1, increasing to 30 mg per day week 2, and then to 45 mg per day for weeks 3 through 5, and a final, maximum dose of 60 mg per day week 6 and onward.
- Sustained-release bupropion (Wellbutrin SR) whose neurochemical action mechanisms are unknown but is believed to produce antidepressant effects by blocking the reuptake of dopamine and norepinephrine. The initial dose was 200 mg per day during weeks 1 and 2, increasing to 300 mg per day by week 4 and to 400 mg per day week 6 and onward.
- Cognitive therapy was provided by a trained psychotherapist and scheduled twice weekly for the first four weeks, then once weekly for the remaining 8 weeks (16 sessions total).

Level 3 switch treatments:

At entry into the Level 3 switch trial, all level 2 medications were discontinued without tapering at the initial Level 3 treatment visit. The level 3 switch treatments were:

- Nortriptyline (Pamelor), a tricyclic antidepressant. Recommended doses were 25 mg/day for 3 days, 50 mg/day for 4 days, and then 75 mg/day by day 8, 100 mg/day by day 28, and, if necessary, 150 mg/day by day 42 and onward
- Mirtazapine (Remeron), a tetracyclic antidepressant that blocks inhibitory α -adrenoceptors on norepinephrine and serotonin neurons to enhance both norepinephrine and serotonin neurotransmission. Recommended mirtazapine doses were 15 mg/day for the first 7 days, 30 mg/day by day 8, 45 mg/day by day 28, and, if necessary, 60 mg/day by day 42 and onward.

Level 3 augmentation treatments of level 2 medications:

The two medication augmentation options used in level 2, buspirone and sustained-release bupropion, were discontinued without tapering in the initial level 3 visit. The two medication augmentation treatments in level 3 were added to ongoing treatment with citalopram, sertraline, sustained-release bupropion, or extended-release venlafaxine. The level 3 augmentation treatments were:

- Lithium started at 450 mg/day, and at week 2 it was increased to the recommended dose of 900 mg/day. If participants could not tolerate the initial dose, it could be reduced to 225 mg/day for 1 week then increased to 450 mg/day. There was no monitoring of lithium levels.
- Triiodothyronine (T₃), a thyroid hormone, started at 25 µg/day for 1 week and then increased to the recommended dose of 50 µg/day. There was no pretreatment assessment, nor ongoing monitoring, of thyroid functioning.

Level 4 switch treatments:

The level 4 switch treatments were:

- Tranylcypromine (Parnate), a monoamine oxidase inhibitor. A 2-week washout period of Level 3 medications was required for patients assigned to the tranylcypromine group. The recommended dosing for tranylcypromine was 10 mg/day for the first 2 weeks, followed by weekly increases of 10 mg/day until a maximum of 60 mg/day.
- Co-administered venlafaxine (Effexor) and mirtazapine (Remeron) to inhibit the reuptake of both serotonin and norepinephrine and block inhibitory α -2-adrenoceptors on both norepinephrine and serotonin neurons to enhance both norepinephrine and serotonin neurotransmission. Level 3 medications were discontinued without tapering for patients assigned to this group. The dosage of extended-release venlafaxine was 37.5 mg/day for the first week, 75 mg/day for the second week, 150 mg/day for weeks 3–5, 225 mg/day for weeks 6–8, and 300 mg/day onward. Mirtazapine was started at 15 mg/day for the first 3 weeks, 30 mg/day for weeks 4 to 8, and then 45 mg/day onward.

**Supplementary Table 3:
Number of Level 2-4 Participants Excluded from our RIAT
Reanalysis, and the Reasons for their Exclusion, yet Included in STAR*D**

Level 2 Treatments

Number of Level 2 Participants Excluded from our Reanalysis but Included in STAR*D	Bup	Sert	Ven	CT	Cit + BUP	Cit + Busp	Cit + CT	Total
Scored as Remitted at ENTRY into Level 2 yet still included in STAR*D's Level 2 analyses	22	8	14	7	30	24	4	109
Scored as only mildly depressed (HRSD >7 & <14) at entry into Level 1, and therefore excluded from STAR*D's data analysis, yet still treated in Level 1, progressed to Level 2, and then included in STAR*D's Level 2 data analyses	21	15	25	15	20	30	7	133
Scored as Remitted at entry into Level 1 (HRSD ≤ 7), and therefore excluded from STAR*D's data analysis, yet still treated in Level 1 and progressed to Level 2 and then included in STAR*D's Level 2 data analyses	6	1	4	2	2	2	2	19
Missing baseline HRSD at entry into Level 1, and therefore excluded from STAR*D's data analysis, yet still treated in Level 1, and progressed to Level 2, and then included in STAR*D's Level 2 data analyses	12	18	22	4	16	13	1	86
Number meeting 2 exclusion criterions	12	2	7	6	5	8	2	42

Bup=Sustained-release Bupropion; Sert= Sertraline; Ven= Extended-release Venlafaxine;
CT=Cognitive Therapy; Cit+BUP= Citalopram + Sustained-release Bupropion;
Cit+Busp=Citalopram + Buspirone; Cit+CT= Citalopram + Cognitive Therapy

Level 3 Treatments

	Nortriptyline	Mirtazapine	Lithium Augmentation	Triiodothyronine Augmentation	Total
Scored as Remitted at ENTRY into Level 3 yet still included in STAR*D's Level 3 analyses	4	0	1	5	10
Scored as only mildly depressed (HRSD >7 &	8	5	3	4	20

1 2 3 4 5 6 7 8 9 10 11 12 13	<14) at entry into Level 1, and therefore excluded from STAR*D's data analysis, yet still treated in Level 1, progressed to Level 2, and then 3 and included in STAR*D's Level 3 data analyses					
14 15 16 17 18 19 20 21 22 23 24 25 26	Scored as Remitted at entry into Level 1 (HRSD ≤ 7), and therefore excluded from STAR*D's data analysis, yet still treated in Level 1 and progressed to Level 2 and then 3 and included in STAR*D's Level 3 data analyses	2	1	0	1	4
27 28 29 30 31 32 33 34 35 36 37 38 39	Missing baseline HRSD at entry into Level 1, and therefore excluded from STAR*D's data analysis, yet still treated in Level 1, and progressed to Level 2, and then level 3 and included in STAR*D's Level 3 data analyses	7	8	1	7	23

Level 4 Treatments

	Tranlycypromine	Venlafaxine + Mirtazapine	Total	
42 43 44 45 46	Scored as Remitted at ENTRY into Level 4 yet still included in STAR*D's Level 4 analyses	5	1	6
47 48 49 50 51 52	Scored as only mildly depressed (HRSD >7 & <14) at entry into Level 1, and therefore excluded from STAR*D's data analysis, yet still treated in Level 1, progressed to Level 2, 3 and then 4 and included in STAR*D's Level 4 data analyses	3	1	4
53 54 55 56 57 58 59 60	Scored as Remitted at entry into Level 1 (HRSD ≤ 7), and therefore excluded from STAR*D's data analysis, yet still treated in Level 1 and progressed	0	0	0

to Level 2, then Level 3 and included in STAR*D's Level 4 data analyses			
Missing baseline HRSD at entry into Level 1, and therefore excluded from STAR*D's data analysis, yet still treated in Level 1, and progressed to Level 2, and then Level 3 and 4 included in STAR*D's Level 4 data analyses	5	1	6

For peer review only

Supplementary Table 4:
Number and Percent of Participants Missing Entry and/or
Exit HRSD Used for Last Observation Carried Forward Analyses

	#/(%) with Missing Entry HRSD	#/(%) with Missing Exit HRSD
Step 1 (N=3,110)	0 (0%)	926 (29.8%)
Step 2 (N=1,134)	168 (14.8%)	304 (26.8%)
Switch strategy (N=620)	90 (14.5%)	183 (29.5%)
Bupropion (N=190)	34	58
Sertraline (N=198)	27	56
Venlafaxine (N=192)	24	56
Cognitive Therapy (N=40)	5	13
Augmentation strategy (N=514)	78 (15.2%)	121 (23.5%)
Bupropion (N=216)	35	58
Bupirone (N=225)	37	52
Cognitive Therapy (N=73)	6	11
Step 3 (N=325)	42 (12.9%)	78 (24%)
Level 2A (N=28)	3	6
Bupropion (N=12)	3	2
Venlafaxine (N=16)	0	4
Level 3 (N=297)	39	72
Switch strategy (N=186)	26	49
Nortriptyline (N=92)	11	23
Mirtazapine (N=94)	15	26
Augmentation strategy (N=111)	13	23
Lithium (N=58)	9	13
Bupropion SR (N=17)	3	2
Citalopram (N=22)	5	6
Sertraline (N=11)	1	3
Venlafaxine XR (N=8)	0	2
T3 (N=53)	4	10
Bupropion SR (N=6)	1	0
Citalopram (N=26)	1	7
Sertraline (N=8)	1	1
Venlafaxine XR (N=13)	1	2

Step 4 (N=106)	15 (14.2%)	22 (20.8%)
Level 3 (N=16)	3	3
Tranlycypromine (N=43)	7	10
Venlafaxine XR/mirtazapine (N=47)	5	9

For peer review only

Supplementary Table 5: Baseline Demographic and Clinical Features by Treatment Step

	Treatment Step ^a							
	Step 1 (N=3,110)		Step 2 (N=1,134)		Step 3 (N=325)		Step 4 (N=106)	
Demographic Features	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age	41.0	13.0	42.0	12.6	44.1	12.0	46.9	11.0
Education (years)	13.6	3.2	13.2	3.3	12.8	3.1	12.6	2.3
Monthly household income	2,289	2,732	1,744	1,539	1,470	1,383	1,003	887
	N	%	N	%	N	%	N	%
Female	1,469	74.6	502	73.1	113	65.3	34	65.4
Race								
White	2,328	74.9	870	76.7	259	79.7	86	81.1
Black	333	10.7	115	10.1	29	8.9	7	6.6
Other	449	14.4	149	13.1	37	11.4	13	12.3
Hispanic	402	12.9	139	12.3	45	13.8	16	15.1
Employment status								
Employed	975	58.7	314	54.2	69	46.9	19	43.2
Unemployed	612	36.9	243	42.0	72	49.0	24	54.5
Retired	73	4.4	22	3.8	6	4.1	1	2.3
Medical insurance								
Private	848	52.2	254	44.5	52	36.6	14	31.8
Public	282	17.4	109	19.2	30	21.4	10	23.3
None	534	33.2	223	39.3	60	43.2	20	46.5
Marital status								
Single	475	28.6	171	29.5	40	27.2	10	22.7
Married/cohabiting	716	43.1	238	41.0	61	41.5	18	40.9
Divorce/separated	429	25.8	155	26.7	42	28.6	14	31.8
Widowed	41	2.5	16	2.8	4	2.7	2	4.5
Clinical Features	N	%	N	%	N	%	N	%
First episode occurrence before age 18	1,200	39.0	436	38.8	120	37.0	41	38.7
Recurrent depression	1,940	66.8	718	68.3	188	63.3	59	60.8
Family history of depression	1,694	55.4	609	54.9	165	51.7	58	54.7

Duration of current episode \geq 2 years	787	25.6	311	27.7	88	27.2	34	32.1
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age at first episode (years)	24.9	14.5	24.7	14.2	25.9	14.6	25.9	14.3
Illness duration (years)	16.1	13.5	17.2	13.7	18.2	14.1	21.0	14.9
Number of episodes	4.4	9.7	4.9	11.1	4.4	10.3	5.0	12.1
Duration of current episode (months)	25.9	52.0	28.1	58.8	32.1	68.5	45.9	82.0
Median duration of current episode (months)	8.3		8.7		9.5		10.1	
Quality of Life and Enjoyment Satisfaction Questionnaire score ^b	39.1	14.3	36.5	13.6	33.7	13.5	31.6	12.8
SF-12 Mental ^c	25.6	8.1	25.0	7.7	24.4	7.7	24.0	7.5
SF-12 Physical ^c	48.6	12.1	47.0	12.4	44.5	12.1	43.8	12.3
Work and Social Adjustment Scale score ^d	25.0	8.7	26.3	8.2	28.3	7.7	29.4	7.0
HRSD ₁₇ score	21.9	5.2	22.5	5.2	23.4	5.2	23.9	5.4
IDS-C ₃₀ score ^e	39.1	9.6	40.6	9.7	42.6	9.4	43.6	9.8
QIDS-IVR score ^f	16.9	3.3	17.3	3.3	17.9	3.0	18.3	3.1
Cumulative Illness Rating Scale								
Categories endorsed	2.5	1.5	2.6	1.6	2.8	1.6	3.1	1.6
Total score	4.7	3.9	5.1	4.0	5.8	4.5	6.2	4.3
Severity score	1.8	0.8	1.8	0.8	2.0	0.9	2.0	0.7
	N	%	N	%	N	%	N	%
Psychiatric Diagnostic Screening Questionnaire								
Agoraphobia	559	18.2	240	21.4	89	27.5	32	30.2
Alcohol abuse/dependence	371	12.0	136	12.1	36	11.1	8	7.5
Bulimia	607	19.7	232	20.6	67	20.7	20	18.9

Drug abuse/dependence	234	7.6	80	7.1	21	6.5	7	6.6
Generalized anxiety disorder	736	23.9	290	25.8	94	29.0	36	34.0
Hypochondriasis	336	10.9	139	12.4	45	13.9	14	13.2
Obsessive-compulsive disorder	723	23.5	265	23.6	97	29.9	31	29.2
Panic disorder	422	13.7	183	16.3	65	20.1	21	19.8
Posttraumatic stress disorder	387	12.6	172	15.3	55	17.0	16	15.1
Social phobia	963	31.3	379	33.7	117	36.1	35	33.0
Somatoform disorder	284	9.2	105	9.3	35	10.8	9	8.5
Number of axis I comorbid psychiatric disorders								
0	606	19.7	190	16.9	48	14.8	12	11.3
1	740	24.0	257	22.9	68	21.0	23	21.7
2	577	18.7	217	19.3	62	19.1	25	23.6
3	363	11.8	139	12.4	42	13.0	14	13.2
4+	793	25.8	321	28.6	104	32.1	32	30.2

^a Sums do not always equal N due to missing values. Percentages are based on available data.

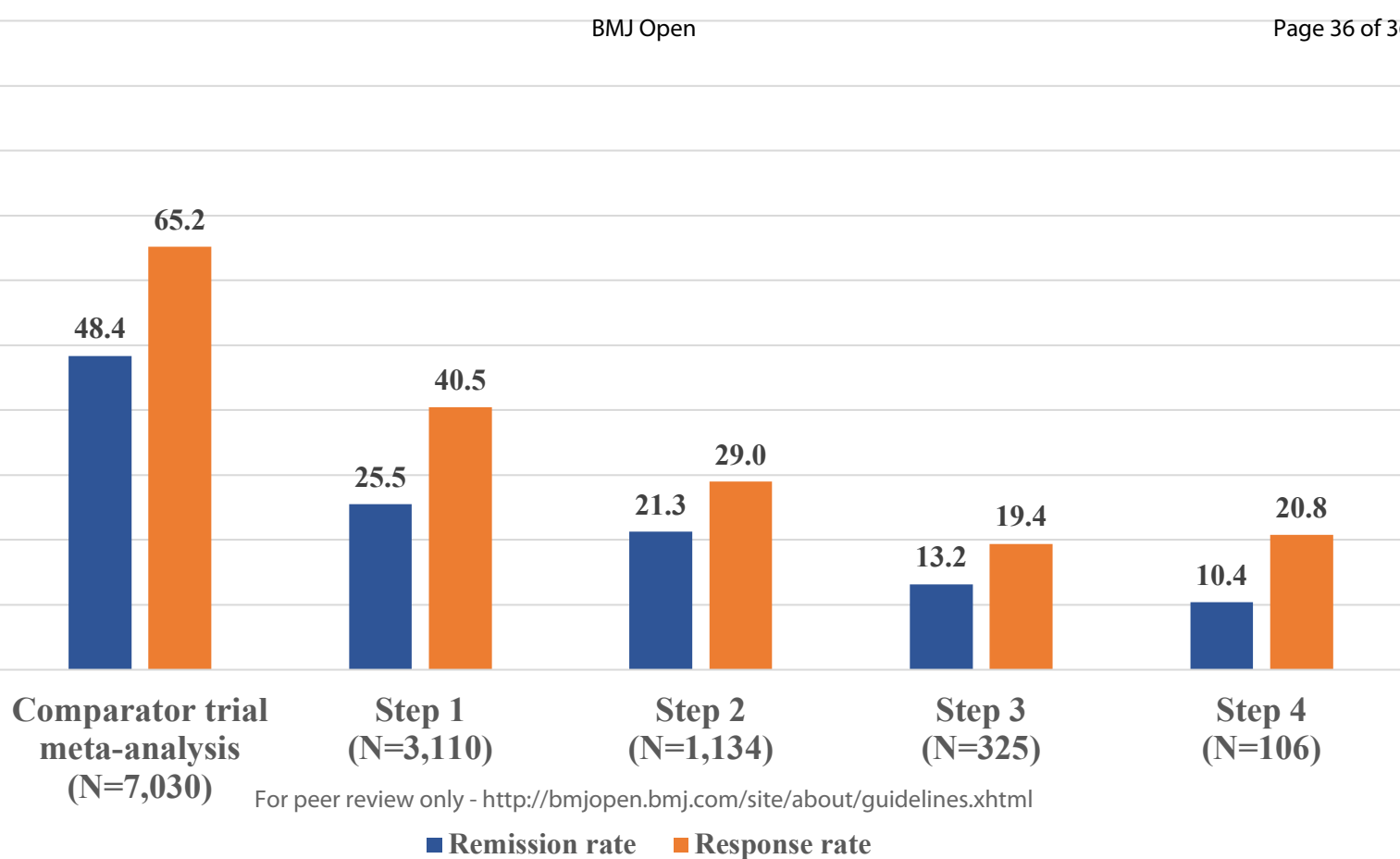
^b Integrated voice response (IVR) administered version of the Quality-of-Life Enjoyment and Satisfaction Questionnaire assessing participants' global rate of satisfaction. Higher scores (range=0–100) represent greater life enjoyment and satisfaction.

^c IVR-administered version of the SF-12 assessing perceived mental and physical health status. Two subscales (physical health factor and mental health) range from 0 to 100— higher scores indicate better functioning with a population norm for each score of 50.

^d IVR-administered version of the Work and Social Adjustment Scale. Scores between 10 and 20 are associated with significant functional impairment while scores above 20 suggest moderate to severe functional impairment.

^e Inventory of Depressive Symptomatology administered telephonically.

^f IVR-administered version of the QIDS.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27For peer review only - <http://bmjopen.bmj.com/site/about/guidelines.xhtml>

■ Remission rate ■ Response rate

