

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

H Edmund Pigott ,¹ Thomas Kim,² Colin Xu,² Irving Kirsch,³ Jay Amsterdam⁴

To cite: Pigott HE, Kim T, Xu C, *et al.* 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. *BMJ Open* 2023;13:e063095. doi:10.1136/bmjopen-2022-063095

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2022-063095>).

Received 31 March 2022
Accepted 07 July 2023



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr H Edmund Pigott;
HPIGOTT75@gmail.com

ABSTRACT

Objective Reanalyse the patient-level data set 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 semirandomised examining the effectiveness of up to four optimised 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 centres.

Participants 4041 adults screened positive for major depressive disorder. In contrast to most clinical trials, STAR*D enrolled patients seeking care (vs recruited) and included patients with a wide range of common comorbid medical and psychiatric conditions to enhance the generalisability 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%

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ We reanalysed the largest ever prospective antidepressant trial's patient-level data set with fidelity to the original research protocol and related publications.
- ⇒ The reanalysis was conducted under the guidelines of the *Restoring Invisible and Abandoned Trials* initiative.
- ⇒ Treatment remission, response and extent of symptom improvement rates were calculated for 14 antidepressant therapies for those patients who met Sequenced Treatment Alternatives to Relieve Depression (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 Hamilton Rating Scale for Depression (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 7030 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.

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.

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 enrol 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 generalisability 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(s). 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)' (Rush *et al*,⁸ p120). 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 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 clinicians in the implementation of the treatment protocol. Research outcomes assessments are **NOT** collected at the clinic visits. They are not collected by either clinicians or Clinical Research Coordinators (National Institute of Mental Health,¹⁰ p47,48; emphasis in the original).

In their summary article, STAR*D investigators used the QIDS-SR as the sole measure to report remission, response and extent of symptom improvement. The Abstract section of this article states that 'the overall cumulative remission rate was 67%' with no qualifiers to this claim ([Rush *et al*,⁷ p1905). Besides making this claim based on an assessment the protocol specifically excluded from use as a research measure, it is not until the article's Results section that careful readers learn this high level of treatment success did not occur. The STAR*D investigators' claim was theoretical—an estimate based on the provisos of what would have happened if there were no study dropouts, and furthermore, 'that those who exited the study would have had the same remission rates as those who stayed in the protocol' (Rush *et al*⁷ p1910). As Pigott *et al* documented though, the investigators' assumptions are not true in the real world since more patients dropped out than remitted in each STAR*D treatment level,¹¹ and furthermore, it has been found in placebo-controlled trials that patients who drop out are more likely to have had adverse treatment side effects and/or emergent suicidality.¹²

Unfortunately, the STAR*D investigators' claim of a 67% cumulative remission rate has become accepted clinical wisdom, and the provisions on which it is based are commonly not referenced when portraying STAR*D's findings. For example, in 2009, NIMH's director Dr Thomas Insel claimed that STAR*D found 'at the end of 12 months, with up to four treatment steps, roughly 70% of participants were in remission' (Insel and Wang¹³ p1466). Similarly in 2013, an editorial in the *American Journal of Psychiatry* (AJP) claimed that STAR*D found 'after four optimised, well-delivered treatments, approximately 70% of patients achieve remission' (Greden,¹⁴ p580). More recently (2022), a New York Times' article claimed that half of STAR*D's participants 'had significantly improved after using either the first or second medication, and nearly 70% of people had become symptom free by the fourth antidepressant'.¹⁵ These are not factual statements of STAR*D's findings.

The first author has made published criticisms alleging protocol violations that appear to inflate STAR*D's findings and called for the reanalysis of the data set by independent investigators.¹⁶ In 2018, the first and fourth authors collaborated with researchers from the University of Connecticut to reanalyse STAR*D's level 1 data obtained from NIMH.¹⁷ This reanalysis found substantial inflation of STAR*D's reported remission and response

rates. Furthermore, the reanalysis found that the extent of HRSD improvement in STAR*D's level 1 trial was approximately half that of open-label antidepressant comparator trials.

Our published criticisms of STAR*D investigators' report of outcomes are as follows¹⁸:

- ▶ While STAR*D investigators used the HRSD to report remission rates in their levels 1–4 articles,^{1–6} the QIDS-SR was used as the sole measure to report remission, response and extent of improvement rates in their summary article⁷ without disclosing that the protocol specifically excluded all non-blinded/clinic-administered assessments such as the QIDS-SR from use as outcome measures. The primary outcome measure, the HRSD, should have been used to report the summary article's outcomes.
- ▶ Using data from the 931 patients deemed ineligible for analysis in STAR*D's level 1 article because these patients lacked a baseline ROA-administered HRSD score of ≥ 14 , in STAR*D's levels 2–4 and summary articles without clear disclosure. This included 99 patients who scored < 8 on their baseline HRSD—indicating these patients met STAR*D's remission criterion at study outset and should not have been included in their report of outcomes.
- ▶ 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' (Trivedi *et al.*¹ p34). These 370 early dropout patients should have been counted as non-remitters 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' (Rush *et al.*⁸ p130).

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

METHOD

Restoring Invisible and Abandoned Trials 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.¹⁹ If investigators failed to correct a study identified as misreported, independent investigators were encouraged to correct the record by reanalysing the study's patient-level data set consistent with the research protocol and analytic plan.

On 6 March 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 reanalyse the STAR*D data set.¹⁸ We then notified STAR*D's principal investigators of our intention and requested they inform us whether they would undertake a reanalysis of the data set adhering to the research protocol. On 22 March 2019, STAR*D investigators acknowledged our email notification, indicated that the STAR*D data were in the public domain, and stated they had no interest in undertaking a reanalysis.

In July 2019, we received an 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 data set consisting of 26 text files and limited supporting study documentation. In September 2019, we obtained funding from the RIAT Support Center to reanalyse STAR*D.

Patients

STAR*D patients were 18–75 years of age, seeking care at 18 primary and 23 psychiatric care clinics. Clinical research coordinators (CRCs) screened 4790 patients for MDD. This screening included the CRCs' administering the HRSD, on which 4041 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 maximise the number of remissions while minimising dropouts (see online supplemental table 1). Online 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 randomised to one of two level 2 switch

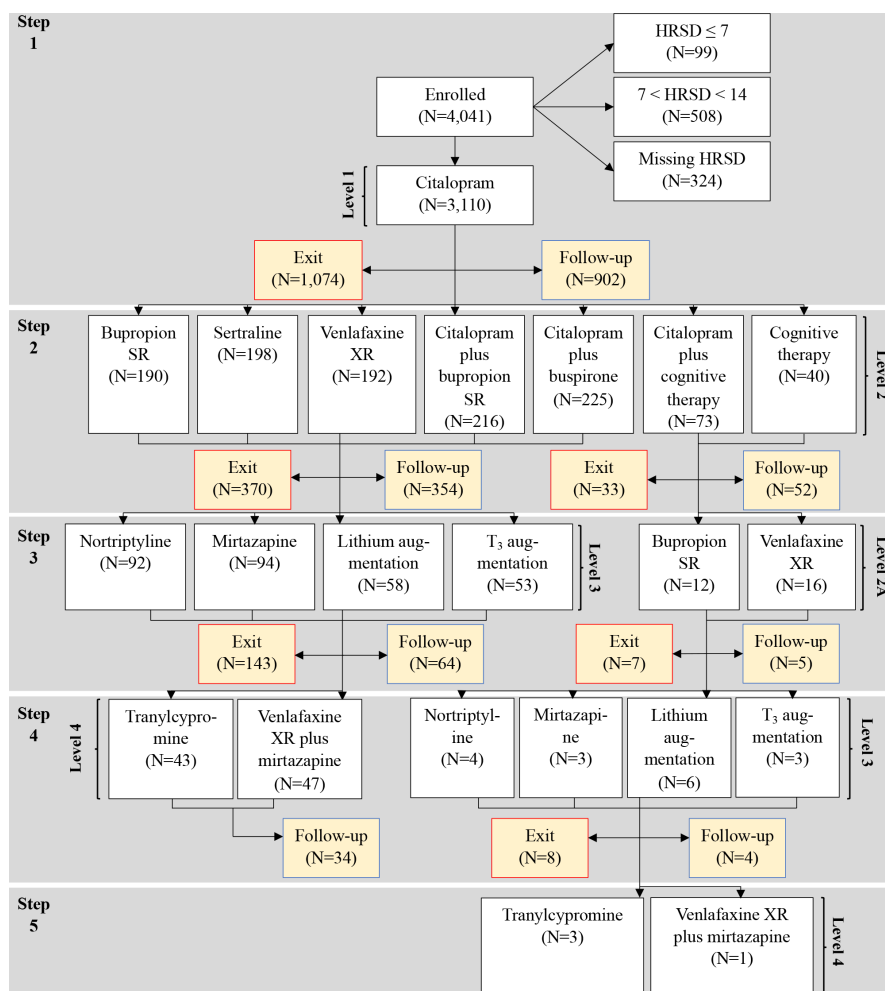


Figure 1 Patient flowchart. *In level 2, 580 patients were randomised to switch medications, 441 to medication augmentation, and 113 to cognitive therapy as either a switch or medication augmentation treatment. In level 2A, 28 patients were randomised to one of two level 2 switch medications. For step 3/level 3 patients, 186 were randomised to medication switch and 111 to medication augmentation. For step 4/level 3 patients, seven were randomised to medication switch and nine to medication augmentation. For step 4/level 4 patients, 90 were randomised 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. ***Follow-up refers to the number of patients who exit a treatment and enter the 12-month follow-up phase. HRSD, Hamilton Rating Scale for Depression.

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' (Trivedi *et al*,¹ p30). 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' (Trivedi *et al*,¹ p30).

For those patients who failed to gain an adequate response from citalopram, STAR*D allowed them to select acceptable treatment options for randomisation in levels 2 to 4 'to empower patients, strengthen the therapeutic alliance, optimize treatment adherence, and improve outcome' (Fava *et al*,²⁰ p483). The treatment options available for randomisation involved either switching to a new treatment or augmenting the patient's current treatment. Treatment levels 2–4 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’ (Rush *et al.*⁷ p1908). 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 ≥50% reduction on the QIDS-C and did not want to be randomised to a next-level treatment were also encouraged to enter follow-up.

Research design of the STAR*D study

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

Analytic plan of the RIAT reanalysis

We reanalysed the STAR*D patient-level data set with fidelity to the original research protocol wherever possible. Where the protocol was silent, we used other STAR*D publications to guide our analysis. This occurred four times. First, the protocol is silent regarding patients who entered the study without a baseline ROA-administered HRSD score of ≥14. In their level 1 article, STAR*D investigators deemed the 931 such patients who lacked this marker of depression severity ineligible for inclusion in data analysis.¹ We do the same and extend this exclusion for such patients who continued on to levels 2–4 because their extent of depression severity at study outset is not known. Second, the protocol is silent on what to do with patients who met the remission criteria on the HRSD at entry into their next-level treatment. In STAR*D’s Rationale and Research Design article though, its investigators prespecify that ‘patients who begin a level with HRSD <8 will be excluded from analyses’ (Rush *et al.*⁸ p130). We, therefore, excluded 125 such patients from

our analyses of treatment levels 2–4. Third, the protocol is silent on how to analyse patients who exit a treatment without taking the HRSD. STAR*D investigators state in their level 1 article, ‘our primary analyses classified patients with missing exit HRSD scores as nonremitters a priori’ (Trivedi *et al.*¹ p34) and repeat similar statements in their level 2–4 articles.^{2–6} Therefore, we do likewise.

Finally, STAR*D had many patients with missing exit HRSD scores. In their level 2–4 articles, STAR*D investigators used a correspondence table to map the final QIDS-SR score to the HRSD for patients missing their exit HRSD score to assess the impact of their approach to counting such patients as ‘nonremitters a priori’.^{22 23} For patients with missing exit HRSD scores, we therefore mapped their last QIDS-SR score to the HRSD and used it to calculate the mean HRSD exit, mean change and combined HRSD and QIDS-SR response rates for all treatments. We also calculated STAR*D’s remission rate both as prespecified based on an exit HRSD score of <8 as well as a final QIDS-SR score of <6 for those patients missing an exit HRSD score.

All preprocessing and analyses were performed in R.²⁴ Authors 2 and 3 identified patients by their subject key and used this variable to match information across data sets. Data on patients’ treatment pathways, and when patients transitioned from one level to the next, were taken from the Integrated Voice Response Alert data set completed by CRCs, and verified against the data on patient-level exits. Authors 2 and 3 then compared the number of patients identified for all level 1–4 treatments to that reported in the STAR*D summary article’s patient flowchart and the number of patients matched.⁷

Next, authors 2 and 3 applied STAR*D’s level 1 inclusion for data analysis criterion to patients in treatment levels 2–4 as well as excluded from analysis the 125 patients who scored <8 on the HRSD at entry into their next-level treatment. We counted these 125 patients as remitted in the prior treatment level but excluded them from the analyses of subsequent treatments. Online supplemental table 3 presents the number of level 2–4 patients excluded from our reanalysis, and the reasons for their exclusion. Online supplemental table 4 identifies the number of patients with missing entry and/or exit HRSD scores for all level 1–4 treatments. As seen in online supplemental table 4, 1330 patients were missing their exit HRSD score across all treatments.

We then compared STAR*D’s outcomes to those found in a meta-analysis of 7030 patients enrolled in antidepressant comparator trials.²⁵ 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.²⁵

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 the purpose of these predictions' was to estimate the number of continuing patients available for randomisation 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 4041 patients enrolled into STAR*D, 3110 met the eligibility for data analysis criterion of having an 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.

Online supplemental table 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 SDs for continuous variables and percentages for discrete variables. Note that 55.7% of STAR*D patients had two 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 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 four trials on antidepressant therapies as well as the combined HRSD plus QIDS-SR remission and response rates for the 1330 patients missing an exit HRSD score.

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%.

Online supplemental figure 1 and online supplemental figure 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 7030 patients enrolled in non-blinded antidepressant comparator trials.²⁵ In step 1, these measures of improvement among STAR*D's patients were at least one-third less than that found in comparator trials, and improvement was worse in each subsequent treatment step.

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 STAR*D 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 in 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 1330 patients missing an exit HRSD. Finally, the RIAT reanalysis rate when using the protocol-specified exit HRSD score of <8 as the sole measure of remission for the 3110 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 35.0% vs 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

STAR*D's results highlight the discrepancy in likely outcomes between typical antidepressant clinical trials with their exclusion criteria and the real-world patients for whom these medications are commonly prescribed. Our RIAT reanalysis found poorer outcomes after up to four optimised, and increasingly aggressive, antidepressant therapies than reported in STAR*D's summary article published in *AJP*.⁷ In contrast to the 67% cumulative remission rate reported in *AJP*, the actual rate was 35.0% when using the protocol-specified HRSD and increased to 41.3% when combined with a final clinic-visit QIDS-SR score of <6 for patients' missing exit HRSD scores in treatment steps 1–4. The 41.3% cumulative remission rate should be viewed as

Table 1 Outcomes across all treatments

Treatment step	HRSD Score		*Mean change (95% CI) (SD)	HRSD remission rate # (%)	*Combined HRSD and QIDS-SR remission rate # (%)	*Combined HRSD and QIDS-SR response rate # (%)
	Entry mean (SD)	*Exit mean (SD)				
Step 1 (N=3110)	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=1134)	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%)

Continued

Table 1 Continued

Treatment step	HRSD Score		*Mean change (95% CI) (SD)	HRSD remission rate # (%)	*Combined HRSD and QIDS-SR remission rate # (%)	*Combined HRSD and QIDS-SR response rate # (%)
	Entry mean (SD)	*Exit mean (SD)				
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 (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				1089 (35.0%)	1284 (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 and QIDS-SR remission rate and combined HRSD and QIDS-SR response rate.
HRSD, Hamilton Rating Scale for Depression; QIDS-SR, Quick Inventory of Depressive Symptomatology–Self Report.

the ‘best case scenario’ since it added an additional 195 QIDS-defined remissions (a remission measure not specified in the protocol) from the 1330 patients with missing exit HRSD scores. As there was neither a placebo nor waitlist control group during any phase of the STAR*D study, it is impossible to know to what extent the observed results were due to the pharmacologic effects of the prescribed medications, placebo effects and/or the passage of time.

Our reanalysis did not assess the durability of treatment effects during the 12-month follow-up phase. In their summary article though, STAR*D investigators reported an overall relapse rate of 46.1% for the 1729 patients who had at least one assessment (of up to 12 scheduled) during follow-up using a telephonic-administered version of the QIDS,⁷ whereas Pigott *et al* found a far lower sustained

Table 2 Outcomes by treatment step

	Step 1 (N=3110)		Step 2 (N=1134)		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	1074	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 and QIDS-SR response rate.
HRSD, Hamilton Rating Scale for Depression; QIDS-SR, Quick Inventory of Depressive Symptomatology–Self Report.

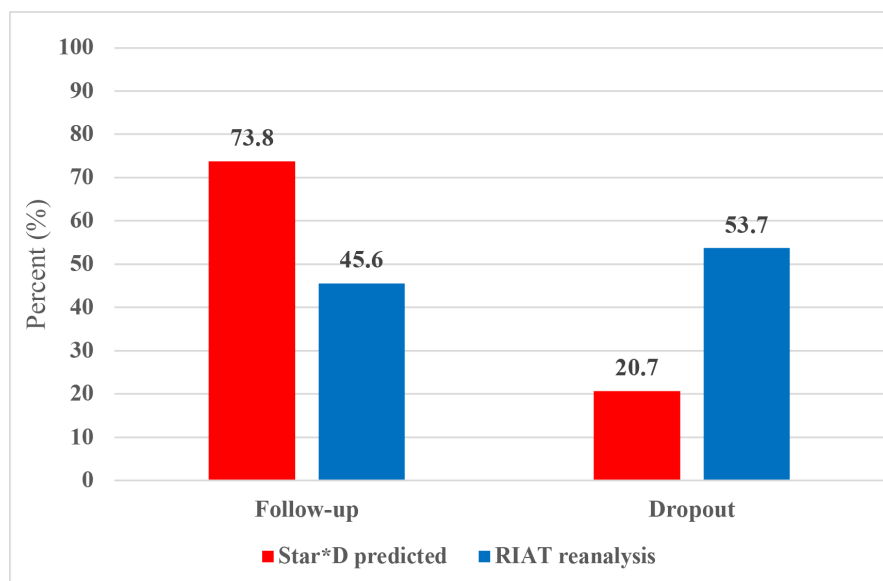


Figure 2 Comparison of STAR*D protocol predictions to what occurred. RIAT, Restoring Invisible and Abandoned Trial; STAR*D, Sequenced Treatment Alternatives to Relieve Depression.

recovery rate when incorporating patient dropout in the analysis.¹²

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 randomised 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

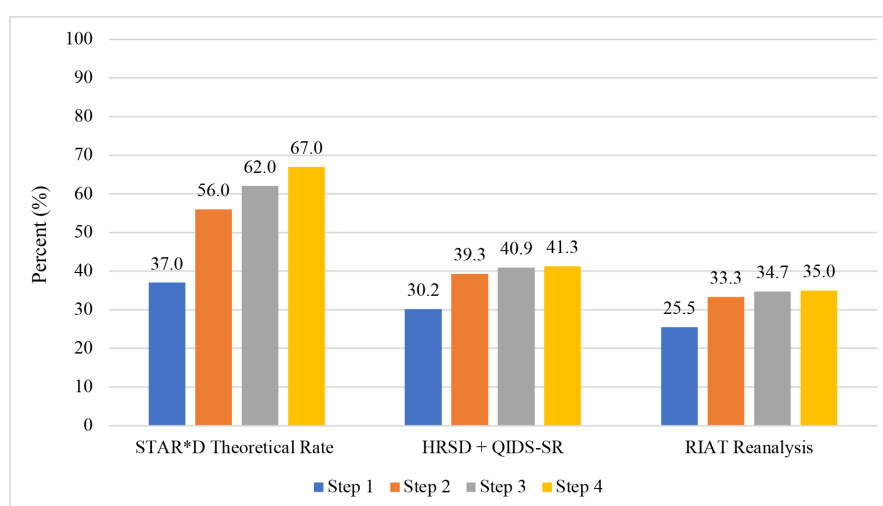


Figure 3 STAR*D's step-by-step cumulative remission rate presented three ways. The step-by-step theoretical remission rates were obtained from the STAR*D summary article where it states: 'The theoretical cumulative remission rate is 67% (37+19+6+5).' [Rush AJ *et al.*,⁷ p1910]. The HRSD+QIDS SR cumulative remission rate was taken from table 1. It combines the 1089 patients with an exit HRSD score of <8 with the 195 patients who were missing an exit HRSD score but had a final clinic-visit QIDS-SR score of <6. The RIAT Reanalysis cumulative remission rate is based on an exit HRSD score of <8 as the sole measure of remission for the 3110 patients who met STAR*D's inclusion for data analysis criteria. HRSD, Hamilton Rating Scale for Depression; QIDS-SR, Quick Inventory of Depressive Symptomatology–Self Report; RIAT, Restoring Invisible and Abandoned Trial; STAR*D, Sequenced Treatment Alternatives to Relieve Depression.

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* (National Institute of Mental Health,¹⁰ p31; 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' (Barbui *et al*,¹² p296) 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 with 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 17 years ago and fuelled a more vigorous search for evidence-based treatment alternatives.

Author affiliations

¹None, Wakefield, Rhode Island, USA

²Psychology, University of Pennsylvania, Philadelphia, Pennsylvania, USA

³Harvard Medical School, Arlington, Massachusetts, USA

⁴Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA

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.

Contributors 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. HEP is responsible for the overall content as the guarantor.

Funding Funding for this project was provided by The RIAT support center.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

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

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. Data is available from the NIMH-supported National Database for Clinical Trials (NDCT).

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

H Edmund Pigott <http://orcid.org/0000-0001-5805-0797>

REFERENCES

- Trivedi MH, Rush AJ, Wisniewski SR, *et al*. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry* 2006;163:28–40.
- Rush AJ, Trivedi MH, Wisniewski SR, *et al*. Bupropion-SR, sertraline, or venlafaxine-XR after failure of SSRIs for depression. *N Engl J Med* 2006;354:1231–42.
- Trivedi MH, Fava M, Wisniewski SR, *et al*. Medication augmentation after the failure of SSRIs for depression. *N Engl J Med* 2006;354:1243–52.
- Fava M, Rush AJ, Wisniewski SR, *et al*. A comparison of Mirtazapine and nortriptyline following two consecutive failed medication treatments for depressed outpatients: a STAR*D report. *Am J Psychiatry* 2006;163:1161–72.
- Nierenberg AA, Fava M, Trivedi MH, *et al*. A comparison of lithium and T3 augmentation following two failed medication treatments for depression: a STAR*D report. *Am J Psychiatry* 2006;163:1519–30.
- McGrath PJ, Stewart JW, Fava M, *et al*. Tranylcypromine versus venlafaxine plus Mirtazapine following three failed antidepressant medications trials for depression: a STAR*D report. *Am J Psychiatry* 2006;163:1531–41.
- Rush AJ, Trivedi MH, Wisniewski SR, *et al*. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry* 2006;163:1905–17.
- Rush AJ, Fava M, Wisniewski SR, *et al*. Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design. *Control Clin Trials* 2004;25:119–42.
- Rush AJ, Trivedi MH, Ibrahim HM, *et al*. The 16-item quick inventory of depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biol Psychiatry* 2003;54:573–83.
- National Institute of Mental Health (NIMH). Sequenced treatment alternatives to relieve depression (STAR*D) research protocol. Washington (DC): NIMH, 2002.
- Pigott HE, Leventhal AM, Alter GS, *et al*. Efficacy and effectiveness of antidepressants: Current status of research. *Psychother Psychosom* 2010;79:267–79.
- 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.
- Insel TR, Wang PS. The STAR*D trial: revealing the need for better treatments. *Psychiatr Serv* 2009;60:1466–7.
- Greden JF. Workplace depression: personalize, partner, or pay the price. *Am J Psychiatry* 2013;170:578–81.

- 15 Smith DG. Antidepressants don't work the way many people think. *New York Times* 8 November 2022.
- 16 Pigott HE. The STAR*D trial: it's time to reexamine the clinical beliefs which guide the treatment of major depression. *Can J Psychiatry* 2015;60:9–13.
- 17 Kirsch I, Huedo-Medina TB, Pigott HE, *et al.* "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;5:339–45.
- 18 Pigott HE, Dubin W, Kirsch I, *et al.* Call to action: RIAT Reanalysis of the sequenced treatment alternatives to relieve depression (STAR*D) study. *BMJ* March 6, 2019. Available: <https://www.bmj.com/content/346/bmj.f2865/rr-10>
- 19 Doshi P, Dickersin K, Healy D, *et al.* 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;26:457–94.
- 21 Lavori PW, Rush AJ, Wisniewski SR, *et al.* Strengthening clinical effectiveness trials: Equipoise-stratified randomization. *Biol Psychiatry* 2001;50:792–801.
- 22 Rush AJ, Bernstein IH, Trivedi MH, *et al.* 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;59:493–501.
- 23 Rush AJ. Inventory of depressive Symptomatology (IDS) and quick inventory of depressive Symptomatology (QIDS). Available: [IDS/QIDS \(ids-qids.org\)](http://ids-qids.org)
- 24 R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2013.
- 25 Rutherford BR, Sneed JR, Roose SP. Does study design influence outcome? the effects of placebo control and treatment duration in antidepressant trials. *Psychother Psychosom* 2009;78:172–81.
- 26 Wisniewski SR, Rush AJ, Nierenberg AA, *et al.* 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, Potter L. Fluoxetine efficacy in treatment resistant depression. *Prog Neuropsychopharmacol Biol Psychiatry* 1994;18:243–61.
- 28 Nierenberg AA, Feighner JP, Rudolph R, *et al.* Venlafaxine for treatment-resistant Unipolar depression. *J Clin Psychopharmacol* 1994;14:419–23.
- 29 Amsterdam JD, Shults J. MAOI safety and efficacy in advanced treatment-resistant depression: a retrospective analysis. *J Affect Disord* 2005;89:183–8.
- 30 Amsterdam JD, Williams D, Michelson D, *et al.* Tachyphylaxis after repeated antidepressant drug exposures in patients with recurrent major depressive disorder. *Neuropsychobiology* 2009;59:227–33.
- 31 Leykin Y, Amsterdam JD, DeRubeis RJ, *et al.* Progressive resistance to SSRI therapy but not to cognitive therapy in the treatment of major depression. *J Consulting & Clinical Psychology* 2007;75:267–76.
- 32 Amsterdam JD, Lorenzo-Luaces L, DeRubeis RJ. Step-wise loss of antidepressant effectiveness after repeated antidepressant trials in bipolar II depression. *Bipolar Disord* 2016;18:563–70.
- 33 Fava GA, Offidani E. The mechanisms of tolerance in antidepressant action. *Prog Neuropsychopharmacol Biol Psychiatry* 2011;35:1593–602.
- 34 Amsterdam JD, Lorenzo-Luaces L. Increase in pharmacodynamic tolerance after repeated antidepressant trials in treatment-responsive bipolar I depressed subjects: an exploratory study. *Psychiatr Pol* 2018;52:957–69.
- 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;39:344–50.
- 36 Andrews P, Amsterdam J. A Hormetic approach to understanding antidepressant effectiveness and the development of antidepressant tolerance - A conceptual view. *Psychiatr Pol* 2020;54:1067–90.
- 37 Amsterdam JD, McHenry LB, Jureidini JN. Industry-corrupted psychiatric trials. *Psychiatr Pol* 2017;51:993–1008.