

The STAR*D Scandal: Scientific Misconduct on a Grand Scale

 madinamerica.com/2023/09/the-stard-scandal-scientific-misconduct-on-a-grand-scale/

By Robert Whitaker

September 9, 2023

Editor's note: This article was updated on September 20.

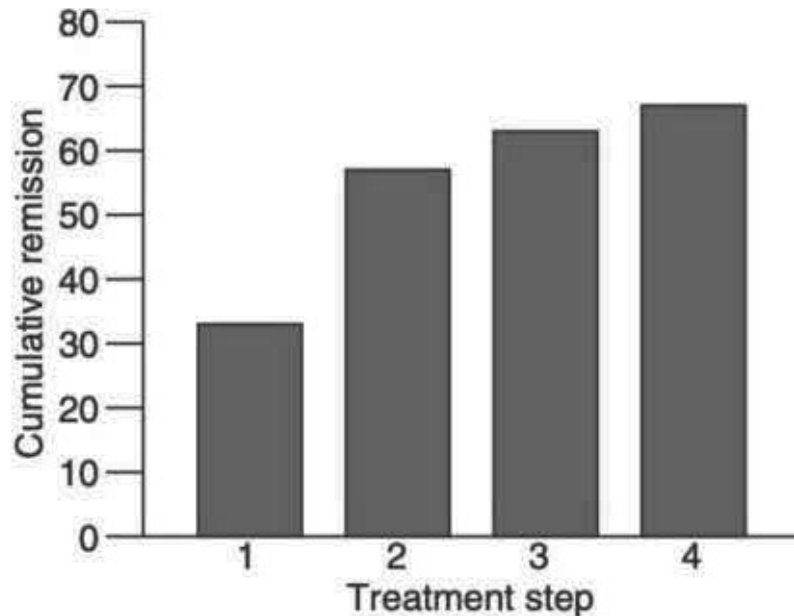
When the STAR*D study was launched more than two decades ago, the NIMH investigators promised that the results would be rapidly disseminated and used to guide clinical care. This was the “largest and longest study ever done to evaluate depression treatment,” the [NIMH noted](#), and most important, it would be conducted in “real-world” patients. Various studies had found that 60% to 90% of real-world patients couldn’t participate in industry trials of antidepressants because of exclusionary criteria.

The STAR*D investigators [wrote](#): “Given the dearth of controlled data [in real-world patient groups], results should have substantial public health and scientific significance, since they are obtained in representative participant groups/settings, using clinical management tools that can easily be applied in daily practice.”

In 2006, they published three accounts of STAR*D results, and the NIMH, in its [November press release](#), trumpeted the good news. “Over the course of all four levels, almost 70 percent of those who didn’t withdraw from the study became symptom-free,” the NIMH informed the public. Here is a graphic from a subsequent published review, titled [“What Does STAR*D Teach Us”](#), that charts that path to wellness:

Figure 2

Cumulative remission rate by
STAR*D treatment level



Source: Gaynes, et al. "What Did STAR*D Teach Us? Results from a Large-scale, Practical, Clinical Trial for Patients with Depression." *Psychiatric Services* 60 (2009):1439-1445.

This became the finding that the media highlighted. The largest and longest study of antidepressants in real-world patients had found that the drugs worked. In the STAR*D study, *The New Yorker* reported in 2010, there was a "sixty-seven-percent effectiveness rate for antidepressant medication, far better than the rate achieved by a placebo."

That happened to be the same year that psychologist Ed Pigott and colleagues published their deconstruction of the STAR*D trial. Pigott had filed a Freedom of Information Act request to obtain the STAR*D protocol and other key documents, and once he and his collaborators had the protocol, they were able to identify the various ways the NIMH investigators had deviated from the protocol to inflate the remission rate. They published patient data that showed if the protocol had been followed, the cumulative remission would have been 38%. The STAR*D investigators had also failed to report the stay-well rate at the end of one year, but Pigott and colleagues found that outcome hidden in a confusing graphic that the STAR*D investigators had published. Only 3% of the 4041 patients who entered the trial had remitted and then stayed well and in the trial to its end.

The protocol violations and publication of a fabricated “principal outcome”—the 67% cumulative remission rate—are evidence of scientific misconduct that rises to the level of fraud. Yet, as Pigott and colleagues have published their papers deconstructing the study, the NIMH investigators have never uttered a peep in protest. They have remained silent, and this was the case when Pigott and colleagues, in August of this year, published their latest paper in *BMJ Open*. In it, they analyzed patient-level data from the trial and detailed, once again, the protocol violations used to inflate the results. As *BMJ Open* wrote in the Rapid Responses section of the online article, “we invited the authors of the STAR*D study to provide a response to this article, but they declined.”

In fact, the one time a STAR*D investigator was prompted to respond, he confirmed that the 3% stay-well rate that Pigott and colleagues had published was accurate. While major newspapers have steadfastly ignored Pigott’s findings, after Pigott and colleagues published their 2010 article, *Medscape Medical News* turned to STAR*D investigator Maurizio Fava for a comment. Could this 3% figure be right? “I think their analysis is reasonable and not incompatible with what we had reported,” Fava said.

That was 13 years ago. The protocol violations, which are understood to be a form of scientific misconduct, had been revealed. The inflation of remission rates and the hiding of the astoundingly low stay-well rate had been revealed. In 2011, *Mad in America* published two blogs by Ed Pigott detailing the scientific misconduct and put documents online that provided proof of that misconduct. In 2015, Lisa Cosgrove and I—relying on Pigott’s published work and the documents he had made available—published a detailed account of the scientific misconduct in our book *Psychiatry Under the Influence*. The fraud was out there for all to see.

Pigott and colleagues subsequently obtained patient-level data through the “Restoring Invisible and Abandoned Trials” initiative (RIAT), and their analysis has confirmed the accuracy of their earlier sleuthing, when they used the protocol to deconstruct the published data. Thus, the documentation of the scientific misconduct by Pigott and colleagues has gone through two stages, the first enabled by their examination of the protocol and other trial-planning documents, and the second by their analysis of patient-level data.

Yet, there has been no public acknowledgement by the American Psychiatric Association (APA) of this scientific misconduct. There has been no call by the APA—or academic psychiatrists in the United States—to retract the studies that reported the inflated remission rates. There has been no censure of the STAR*D investigators for their scientific misconduct. Instead, they have, for the most part, retained their status as leaders in the field.

Thus, given the documented record of scientific misconduct, in the largest and most important trial of antidepressants ever conducted, there is only one conclusion to draw: In American psychiatry, scientific misconduct is an accepted practice.

This presents a challenge to the American citizenry. If psychiatry will not police its own research, then it is up to the public to make the fraud known, and to demand that the paper published in the *American Journal of Psychiatry*, which told of a 67% cumulative remission rate, be withdrawn. As STAR*D was designed to guide clinical care, it is of great public health importance that this be done.

An Intent to Deceive

The World Association of Medical Editors lists seven categories of scientific misconduct. Two in particular apply to this case:

- “Falsification of data, ranging from fabrication to deceptive selective reporting of findings and omission of conflicting data, or willful suppression and/or distortion of data.”
- “Violation of general research practices” which include “deceptive statistical or analytical manipulations, or improper reporting of results.”

The essential element in scientific misconduct is this: it does not result from honest mistakes, but rather is born from an “intent to deceive.”

In this instance, once Pigott and colleagues identified the deviations from the protocol present in the STAR*D reports, the STAR*D investigators’ “intent to deceive” was evident. By putting the protocol and other key documents up on Mad in America, Pigott made it possible for the scientific community to see for themselves the deception.

Their recent RIAT publication makes it possible to put together a precise numerical accounting of how the STAR*D investigators’ research misconduct, which unfolded step by step as they published three articles in 2006, served to inflate the reported remission rate. This MIA Report lays out that chronology of deceit. Indeed, readers might think of this MIA Report as a presentation to a jury. Does the evidence show that the STAR*D’s summary finding of a 67% cumulative remission rate was a fabrication, with this research misconduct born from a desire to preserve societal belief in the effectiveness of antidepressants?

The Study Protocol

According to the STAR*D protocol, patients enrolled into the study would need to be at least “moderately depressed,” with a score of 14 or higher on the Hamilton Depression Rating Scale (also known as HAM-D). They would be treated with citalopram (Celexa) at their baseline visit, and then, during the next 12 weeks, they would have five clinical visits. At each one, a coordinator would assess their symptoms using a tool known as the Quick Inventory of Depressive Symptomatology (QIDS-C). As this study was meant to mimic real-world care, physicians would use the QIDS data to help determine whether the citalopram dosage should be altered, and whether to prescribe other “non-study” medications, such as drugs for sleep, anxiety, or for the side effects caused by citalopram.

At each clinic visit, patients would also self-report their symptoms using this same measuring stick (QIDS-SR). The QIDS instrument had been developed by the STAR*D investigators, and they wanted to see whether the self-rated scores were consistent with the QIDS scores assessed by clinicians.

At the end of the treatment period, independent “Research Outcome Assessors” (ROAs) would assess the patients’ symptoms using both the HAM-D₁₇ and the “Inventory of Depressive Symptomatology” scale (IDS-C₃₀). The primary outcome was remission of symptoms, which was defined as a HAM-D score ≤ 7 . The protocol explicitly stated:

“The research evaluation of effectiveness will rest on the HAM-D obtained, not by the clinician or clinical research coordinator, but by telephone interviews with the ROAs.”

And:

“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 are not collected at the clinic.”

During this exit assessment, patients would also self-report their outcomes via an “interactive voice recording” system (IVR) using the QIDS questionnaire. This would be done “to determine how this method performed compared to the above two gold standards.” The protocol further stated:

“Comparing the IDS-C₃₀ collected by the ROA and the QIDS₁₆, collected by IVR, allows us to determine the degree to which a briefer symptom rating obtained by IVR can be substituted for a clinician rating. If this briefer rating can substitute for a clinician rating, the dissemination and implementation of STAR*D findings is made easier. Thus, the inclusion of QIDS₁₆ by IVR is aimed at methodological improvements.”

After the first 12-week trial with citalopram, patients who hadn’t remitted were encouraged to enter a second “treatment step,” which would involve either switching to another antidepressant or adding another antidepressant to citalopram. Patients who failed to remit during this second step of treatment could then move on to a third “treatment step” (where they would be offered a new treatment mix), and those who failed to remit in step 3 would then get one final chance to remit. In each instance, the HAM-D, administered by a Research Outcome Assessor, would be used to determine whether a patient’s depression had remitted. At the end of the four steps, the STAR*D investigators would publish the cumulative remission rate, which they predicted would be 74%.

Patients who remitted at the end of any of the four steps were urged to participate in a year-long maintenance study to assess rates of relapse and recurrence for those maintained on antidepressants. The existing literature, the protocol stated, suggested that a “worst case”

scenario was that 30% of remitted patients maintained on antidepressants “experience a depressive breakthrough within five years.” Yet, it was possible that real-world relapse rates might be higher.

“How common are relapses during continued antidepressant treatment in ‘real-world’ clinical practice?” the STAR*D investigators asked. “How long [are remitted patients] able to stay well?”

In sum, the protocol stated:

- Patients had to have a HAM-D score of 14 or higher to be eligible for the trial.
- The primary outcome would be an HAM-D assessment of symptoms administered by a Research Outcome Assessor at the end of the treatment period. Remission was defined as a HAM-D score of 7 or less.
- The secondary outcome would be an IDS-C₃₀ assessment of symptoms administered by a Research Outcomes Assessor at the end of the treatment period.
- The QIDS-C would be administered at clinic visits to guide treatment decisions, such as increasing drug dosages. Patients would also self report their symptoms on the QIDS scale (QIDS-SR) to see if their scores matched up with the clinicians’ numbers. The two QIDS evaluations during clinic visits would not be used to assess study outcomes.
- The QIDS-SR administered by IVR at the end of treatment was for the purpose of seeing whether using this automated questionnaire, which took only six minutes, could replace clinician-administered scales to guide clinical care once STAR*D findings were published. It was not to be used to assess study outcomes.
- Relapse and stay-well rates would be published at the end of the one-year follow-up.

While the protocol was silent on how drop-outs would be counted, a [2004 article](#) by the STAR*D investigators on the study’s “rationale and design” stated that patients with missing HAM-D scores at the end of each treatment step were “assumed not to have had a remission.”

Thus, the STAR*D documents were clear: those who dropped out during a treatment step without returning for an exit HAM-D assessment would be counted as non-remitters.

The Published Results

Step 1 outcomes

Trivedi, et al. “Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: Implications for clinical practice.” *Am J of Psychiatry* 163 (2006): 28-40.

In January of 2006, the STAR*D investigators reported results from the first stage of treatment. Although 4,041 patients had been enrolled in the study, there were only 2,876 “evaluable” patients. The non-evaluable group (N=1,165) was composed of 607 patients who had a baseline HAM-D score of less than 14 and thus weren’t eligible for the study; 324 patients who had never been given a baseline HAM-D score; and 234 who failed to return after their initial baseline visit. Seven hundred ninety patients remitted during stage one, with their HAM-D scores dropping to 7 or less. The STAR*D investigators reported a HAM-D remission rate of 28% (790/2,876).

At first glance, this appeared to be a careful reporting of outcomes. However, there were two elements discordant with the protocol.

As Trivedi and colleagues noted in their “statistical analysis” section, patients were designated as “not achieving remission” when their exit HAM-D score was missing. In addition, they noted that “intolerance was defined a priori as either leaving treatment before 4 weeks or leaving at or after 4 weeks with intolerance as the identified reason.”

Thus, by both of these standards, the 234 patients who had failed to return after their baseline visit, when they were first prescribed Celexa, should have been counted as treatment failures rather than as non-evaluable patients. They were “intolerant” of the drug and had left the trial without an exit HAM-D score. If the STAR*D investigators had adhered to this element of their study plan, the number of evaluable patients would have been 3,110, which would have lowered the reported remission rate to 25% (790/3,110).

The second discordant element in this first publication tells more clearly of an “intent to deceive.” In their summary of results, they wrote:

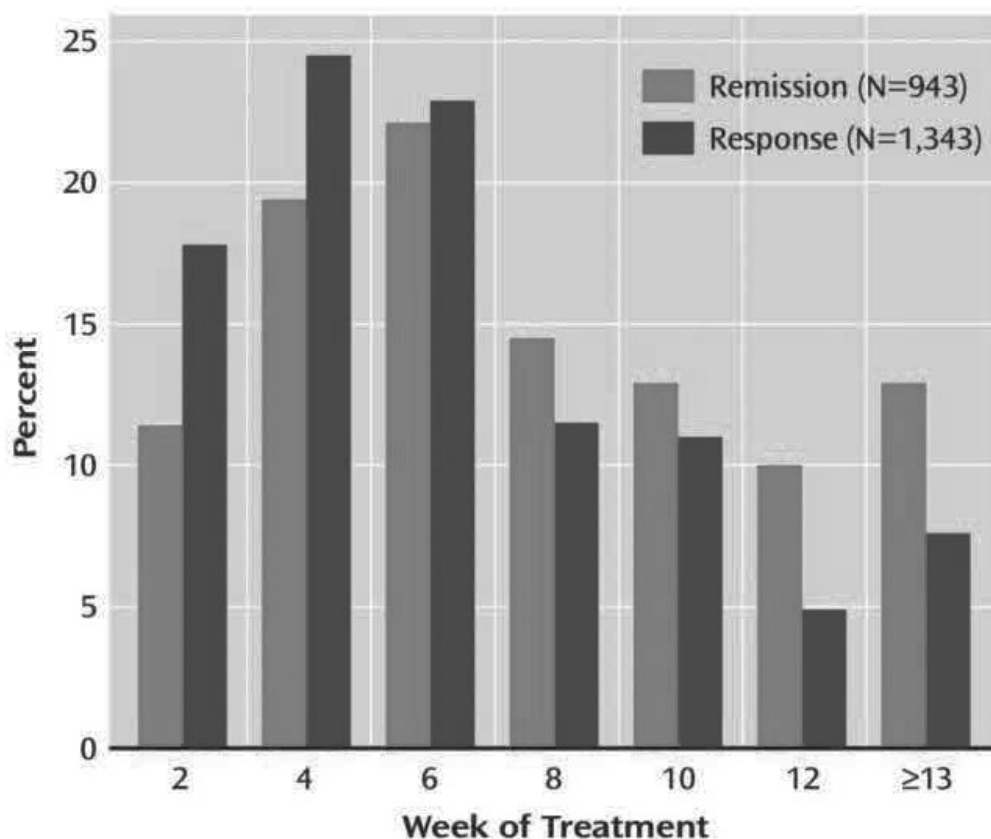
“Remission was defined as an exit score of ≤ 7 on the 17-item Hamilton Depression Rating Scale (HAM-D) (primary outcome) or a score of ≤ 5 on the 16-item Quick Inventory of Depressive Symptomatology, Self-Report (QIDS-SR) (secondary outcome).”

They were now presenting QIDS-SR as a secondary outcome measure, even though the protocol explicitly stated that the secondary outcome measure would be an IDS-C₃₀ score administered by a Research Outcome Assessor. Moreover, they were now reporting remission using the QIDS-SR score at the patient’s “last treatment visit,” even though the protocol explicitly stated that “research outcomes are not collected at the clinic.”

This switch to a QIDS-SR score from the clinic made it possible to count those who had no exit HAM-D score as remitters if their last in-clinic QIDS-SR score was five or less. This deviation from the protocol added 153 to their remitted count, such that on the QIDS-SR scale, 33% were said to achieve remission (943/2,876).

The STAR*D investigators even published a graphic of remission rates with the QIDS-SR, setting the stage for it to be presented, when the cumulative remission rate was announced, as the *primary method* for assessing effectiveness outcomes.

FIGURE 3. Percent of 2,876 Outpatients With Nonpsychotic Major Depressive Disorder Who Achieved Response or Remission Defined by 16-Item Quick Inventory of Depressive Symptomatology, Self-Report (QIDS-SR), Scores by Week of Treatment^a



^a Response was defined as improvement of $\geq 50\%$ in QIDS-SR score from baseline. Remission was defined as a QIDS-SR score of ≤ 5 at endpoint.

Step 2 outcomes

Rush, et al. "Bupropion-SR, sertraline, or venlafaxine-XR after failure of SSRIs for depression." NEJM 354 (2006): 1231-42. Also: Trivedi, et al. "Medication augmentation after the failure of SSRIs for depression." NEJM 354 (2006): 1243-1252.

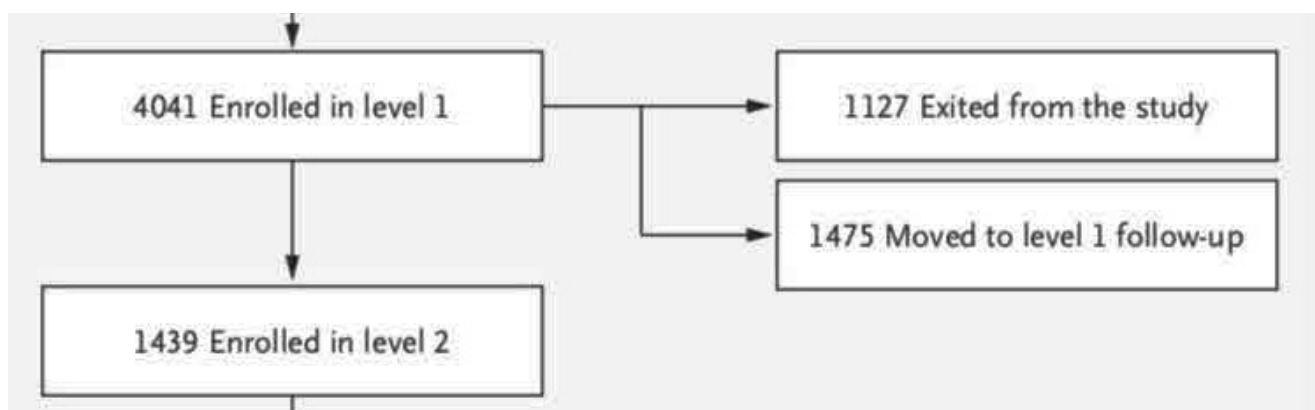
Two months later, the STAR*D investigators published two articles detailing remission rates for those who had failed to remit on citalopram and had entered the second treatment step (N=1,439).

One publication told of patients who had been withdrawn from citalopram and then randomized to either bupropion, sertraline, or venlafaxine. There were 729 patients so treated in step 2, and the remission rate was 21% on the HAM-D scale and 26% on the QIDS-SR scale. The investigators concluded that “after unsuccessful treatment, approximately one in four patients had a remission of symptoms after switching to another antidepressant.” That conclusion presented the QIDS-SR as the preferred scale for assessing remission.

The second publication told of remission rates for 565 patients treated with citalopram augmented by either bupropion or buspirone. The remission rate was 30% using HAM-D, and 36% using QIDS-SR. The investigators concluded that these two remission rates “were not significantly different,” yet another comment designed to legitimize reporting remission rates with QIDS.

There were two other deviations from the protocol in these reports on step 2 outcomes, although neither was easily discovered by reading the articles. The first was that the 931 patients cited as being “unevaluable” in the step 1 report, either because they had a baseline HRSD score less than 14 (607 patients) or no baseline score at all (324), were now being included in calculations of remitted patients. This could be seen in a graphic in Rush’s article, which stated that of the 4,041 enrolled into the trial, by the start of second step 1, 1,127 had dropped out, 1,475 had moved into the one-year follow-up, and 1,439 had entered step 2. The 931 patients were now simply flowing into one of those three categories.

Here is the flow chart from the Rush paper that shows this fact:



The re-characterization of the 931 patients as evaluable patients could be expected, of course, to markedly inflate cumulative remission rates. Not only were 607 not depressed enough to enter the study, but Pigott and colleagues, with their access to patient-level data, determined that 99 in this group had baseline HAM-D scores below 8. They met criteria for remission before they had been given their first dose of citalopram.

The second deviation was that patients who, at a clinic visit, scored as remitted on the QIDS-SR and sustained that remission for “at least 2 weeks,” could now be counted as having remitted and “move to follow-up.” Depressive symptoms are known to wax and wane, and with this new laxer standard, patients were being given multiple chances to be counted as “remitted” during any treatment step, and doing so using a self-report scale that they had filled out many times.

Final report on outcomes

*Rush, et al. “Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: A STAR*D report.” Am J Psychiatry 163 (2006): 1905-1917.*

In November 2006, the STAR*D investigators provided a comprehensive report on outcomes from both the acute and maintenance phases of the study. The protocol deviations, and thus the intent to deceive, are on full display in this paper.

Acute outcomes

Their reported remission rate of 67% relied on three deviations from the protocol, and a fourth “theoretical” calculation that transformed 606 drop-outs into imaginary remitters.

However, the patient numbers involved in the protocol deviations can seem confusing for this reason: in the summary paper, the count of evaluable patients has once again changed. The step 1 report told of 2,876 evaluable patients. The step 2 report added the 931 patients without a qualifying baseline HAM-D score back into the mix, which seemingly produced a total of 3,807 evaluable patients. But the final summary paper tells of 3,671 evaluable patients.

So where did this drop of 136 in the number of evaluable patients come from? In the step 1 report, the Star*D investigator stated there were 234 patients, out of the entering group of 4,041 patients, who didn’t return for a second visit and thus weren’t included in the evaluable group. In this summary paper, the STAR*D authors state there are 370 in this group. They provide no explanation for why this “did not return” number increased by 136 patients. (See footnote at end of this report for two possibilities.)

As for the 3,671 evaluable patients, the paper states that this group is composed of the 2,876 evaluable patients listed in the step 1 report, plus participants whose baseline HAM-D score was less than 14. The STAR*D authors do not explain why they are including patients who were not depressed enough to meet inclusion criteria in their count of evaluable patients. Nor do they state, in this summary report, how many are in this group. They also don’t mention their inclusion of patients who lacked a baseline HAM-D score.

As such, what is evident in this paper is that numbers are once again being juggled. However, if the reader does the arithmetic, it becomes apparent that the count of 3,671 evaluable patients consists of the 2,876 patients deemed evaluable in the step 1 report, plus 795

patients who lacked a qualifying HAM-D score (out of 931 initially stated to be in this group). What the STAR*D authors did in this final report—for reasons unknown—is remove 136 from the group of 931 who lacked a qualifying HAM-D score and added them to the “didn’t show up for a second clinic visit” group.

While the patient count numbers have changed, it is still possible to provide a precise count, based on the new numbers in the final summary report, of how all three protocol deviations served the purpose of inflating the remission rate, and did so in one of two ways: either increasing the number of remitted patients, or decreasing the number of evaluable patients.

1. Categorizing early dropouts as non-evaluable patients

The step 1 report listed 234 participants who had baseline scores of 14 or higher who didn’t return for a “post baseline” visit. As noted above, the protocol called for these patients to be chalked up as treatment failures. In their summary report, the STAR*D investigators added 136 to this count of non-evaluable patients, a change that further lowers the denominator in their calculation of a cumulative remission rate (remitters/evaluable patients).

2. Including ineligible patients in their count of remitted patients

The step 1 report excluded 931 patients whose baseline HAM-D scores were either less than 14 (607 patients) or were missing (324 patients). The final summary report includes 795 participants who lacked a qualifying HAM-D score, and as will be seen below, this group of 795 patients, who didn’t meet inclusion criteria, added 570 to the tally of remitted patients.

3. Switching Outcome Measures

The STAR*D investigators did not report HAM-D remission rates. Instead, they only reported remission rates based on QIDS-SR scores obtained during clinic visits. They justified doing so by declaring that “QIDS-SR and HRSD₁₇ outcomes are highly related,” and that QIDS-SR “paper and pencil scores” collected at clinic visits were “virtually interchangeable” with scores “obtained from the interactive voice response system.” The protocol, of course, had stated:

- that the HAM-D was to be the primary measure of remission outcomes
- that QIDS was not to be used for this purpose
- that symptom assessments made during clinic visits were not to be used for research purposes

The justification that the STAR*D investigators gave for reporting only QIDS-SR scores suggested there was an equivalency between HAM-D and QIDS, when, in fact, the use of QIDS-SR regularly produced higher remission rates. The statement presented a false equivalency to readers.

With these protocol deviations fueling their calculations, the STAR*D investigators reported the following remission rates for each of the four treatment steps.

STAR*D Reported Remission Rates

	Percentage Remitted	Number Remitted
Step 1 (N=3,671)	36.8%	1,346
Step 2 (N=1,439)	30.6%	439
Step 3 (N=390)	13.7%	53
Step 4 (N=123)	13.0%	16

Source: A. Rush, et al. "Acute and longer-term outcomes in depressed outpatients requiring or several treatment steps: A STAR*D report." *Am J Psychiatry* 163 (2006): 1905-1917.

Thus, the cumulative remission rate at the end of four steps was said to be 51% (1,854/3,671).

In their recent reanalysis, Pigott and colleagues reported what the remission rates would have been if the protocol had been followed. First, the evaluable group should have been 3,110 patients (4,041 minus the 931 patients who didn't have a baseline HAM-D score, or didn't have a HAM-D score of 14 or higher). Second, HAM-D scores should have been used to define remission. Here is the data:

Remission Rates If Protocol Had Been Followed

	Percentage Remitted	Number Remitted
Step 1 (N=3,110)	25.5%	794
Step 2 (N=1,134)	21.3%	241
Step 3 (N=325)	13.2%	43
Step 4 (N=106)	10.4%	11

Source: H. Pigott, et al. What are the treatment, remission, response and extent of improvement rates after up to four trials of antidepressants 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

Thus, if the protocol had been followed, the cumulative remission rate at the end of four steps would have been 35% (1,089/3,110). The protocol deviations added 765 remitters to the "got well" camp.

Pigott's 2023 report also makes it possible to identify the precise number of added remissions that came from including the 931 ineligible patients in their reports, and from switching to QIDs as the primary outcome measure.

Added Remissions From Two Protocol Violations

	Inclusion of 931 Ineligible Patients	Switch from HAM-D to QIDS	Total
Step 1	408	144	552
Step 2	156	42	198
Step 3	3	7	10
Step 4	3	2	5
Total	570	195	765

Source: Calculations based on H. Pigott, et al. What are the treatment, remission, response and extent of improvement rates after up to four trials of antidepressants 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

Even after these machinations, the STAR*D investigators still needed a boost if they were going to get close to their predicted get-well rate of 74%. To do so, they imagined that if the drop-outs had stayed in the study through all four steps of the study and remitted at the same rate as those who did stay to the end, then another 606 patients would have remitted. And voila, this produced a remission rate of 67% (2,460/3,671).

This theoretical calculation, as absurd as it was from a research standpoint, also violated the protocol. Those who dropped out without an exit HAM-D score less than 8 were deemed to be non-remitters. This theoretical calculation transformed 606 treatment failures into treatment successes.

Here is the final tally of how the STAR*D investigators' research misconduct transformed a 35% remission rate into one nearly double that:

The Numerical Fruits of Research Misconduct

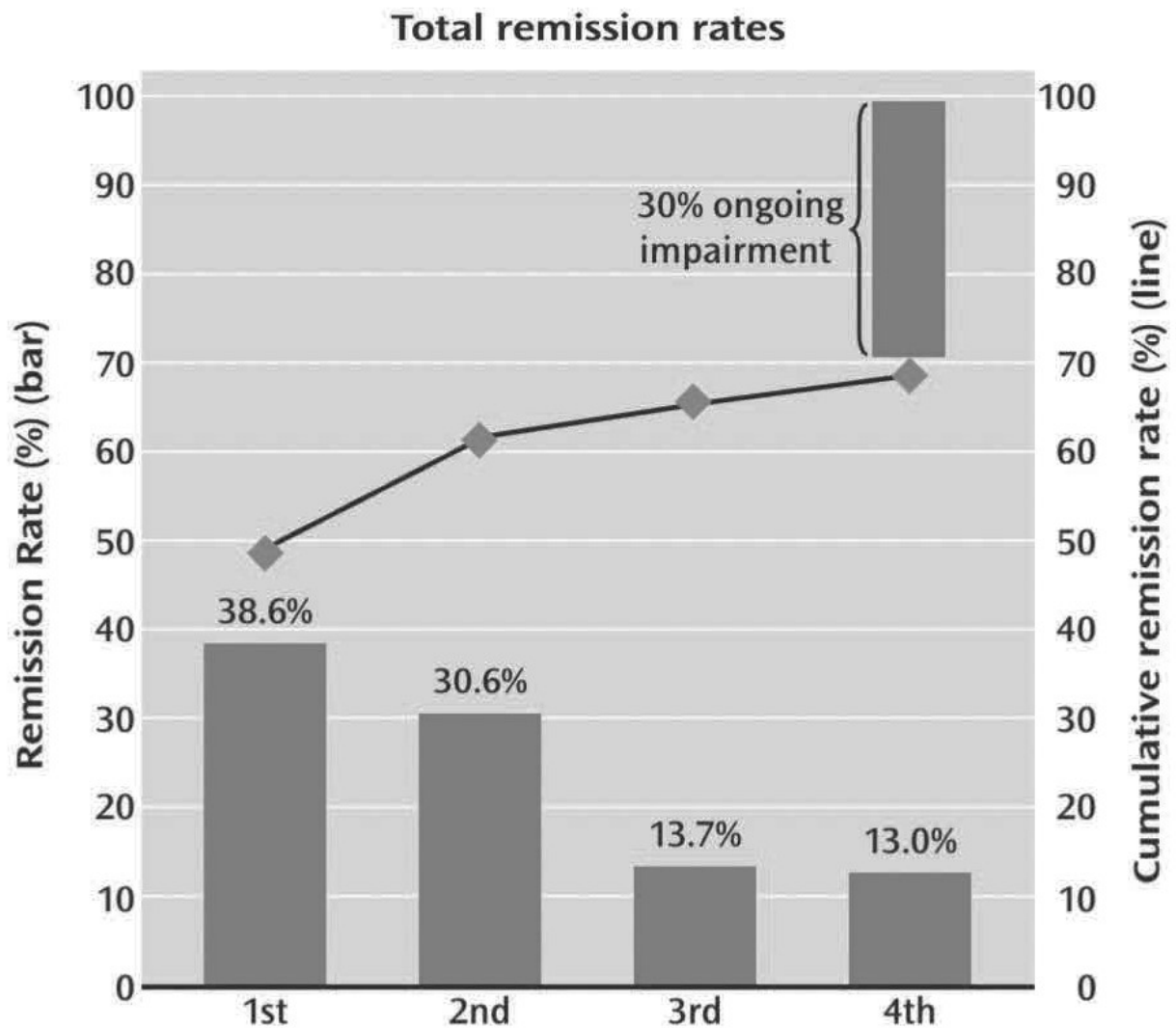
Deviation from Protocol	Number of Evaluable Patients	Number of Remitted Patients	Remission Rate
No deviation from protocol	3,110	1,089	35%
Exclusion of 234 patients who had been prescribed citalopram and never returned for a post-baseline visit. These patients should have been counted as treatment failures.	2,876	1,089	38%
Inclusion of 931 patients with missing baseline HAM-D score or baseline HAM-D less than 14	3,807	1,659	44%
Exclusion of additional 136 patients said to have never returned for a post-baseline visit.	3,671	1,659	45%
Switch from HAM-D to QIDS-SR	3,671	1,854	51%
Calculated theoretical remission rate that transformed study dropouts into remitted patients	3,671	2,460	67%

That is the account of research misconduct that took place in the acute phase of the STAR*D study. The abstract of the summary report told of an “overall cumulative remission rate,” without mentioning the theoretical element. As can be seen in this screenshot, the fabrication was presented as a bottom-line result:

Results: The QIDS-SR₁₆ remission rates were 36.8%, 30.6%, 13.7%, and 13.0% for the first, second, third, and fourth acute treatment steps, respectively. The overall cumulative remission rate was 67%. Over-

This, in turn, became the fake number peddled to the public. For instance:

- The NIMH touted this number in a [press release](#).
- *The New Yorker*, famed for its fact-checking, pointed to the [67% remission rate](#) as evidence of the real-world effectiveness of antidepressants.
- Many subsequent [articles](#) in the research literature told of this outcome.
- A [2013 editorial](#) in the *American Journal of Psychiatry* stated that in the STAR*D trial, “after four optimised, well-delivered treatments, approximately 70% of patients achieve remission.” A graphic depicted this stay-well rate:



Source: J. Greden. Workplace depression: personalize, partner, or pay the price. *Am J Psychiatry* 2013;170:578–81.

More recently, after an [article by Moncrieff and colleagues](#) debunked, yet again, the chemical imbalance theory of depression, several major newspapers, including *The New York Times*, [trotted out the 67% figure](#) to reassure the public that they needn't worry, antidepressants worked, and worked well.

One-year outcomes

There were 1,518 who entered the follow-up trial in remission. The protocol called for regular clinical visits during the year, during which their symptoms would be evaluated using QIDS-SR. Clinicians would use these self-report scores to guide their clinical care: they could change medication dosages, prescribe other medications, and recommend psychotherapy to help the patients stay well. Every three months their symptoms would be evaluated using the HAM-D. Relapse was defined as a HAM-D score of 14 or higher.

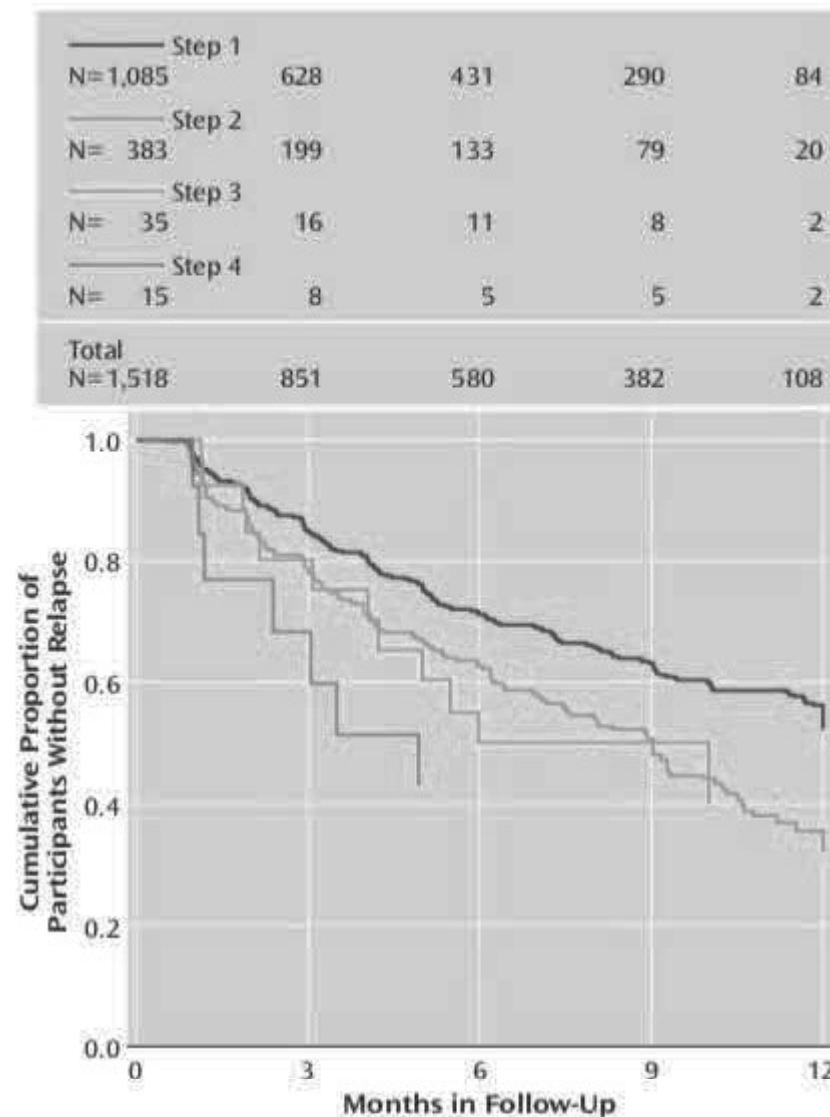
This was the larger question posed by STAR*D: What percentage of depressed patients treated with antidepressants remitted and stayed well? Yet, in the discussion section of their final report, the STAR*D investigators devoted only two short paragraphs to the one-year results. They did not report relapse rates, but rather simply wrote that “relapse rates were higher for those who entered follow-up after more treatment steps.”

Table five in the report provided the relapse rate statistics: 33.5% for the step 1 remitters, 47.4% for step 2, 42.9% for step 3, and 50% for step 4. At least at first glance, this suggested that perhaps 60% of the 1,518 patients had stayed well during the one-year maintenance study.

However, missing from the discussion and the relapse table was any mention of dropouts. How many had stayed in the trial to the one-year end?

There was a second graphic that appeared to provide information regarding “relapse rates” over the 12-month period. But without an explanation for the data in the graphic, it was impossible to decipher its meaning. Here it is:

FIGURE 3. Relapse During Follow-Up Phase by Number of Acute Treatment Steps for STAR*D Participants Who Entered Follow-Up Phase in Remission^a



^a Significant overall difference among steps ($\chi^2=23$, $df=3$, $p<0.0001$). Significant post-hoc comparisons with Bonferroni corrections revealed significant differences between steps 1 and 2.

Once Pigott launched his sleuthing efforts, he was able to figure it out. The numbers in the top part of the graphic told of how many remitted patients remained well *and in the trial* at three months, six months, nine months and one year. In other words, the top part of this graphic provided a running account of relapses plus dropouts. This is where the drop-outs lay hidden.

Before Pigott published his finding, he checked with the STAR*D biostatistician, Stephen Wisniewski, to make sure he was reading the graphic right. Wisniewski replied:

“Two things can happen during the course of follow-up that can impact on the size of the sample being analyzed. One is the event, in this case, relapse, occurring. The other is drop out. So the N’s over time represent that size of the population that is remaining in the sample (that is, has not dropped out or relapsed at an earlier time).”

Here, then, was the one-year result that the STAR*D investigators declined to make clear. Of the 1,518 remitted patients who entered the follow-up, only 108 patients remained well and in the trial at the end of 12 months. The other 1,410 patients either relapsed (439) or dropped out (971).

Pigott and colleagues, when they published their 2010 deconstruction of the STAR*D study, summed up the one-year results in this way: Of 4,041 patients who entered the study, only 108 remitted and then stayed well and in the study to its one-year end. That was a documented get-well and stay-well rate of 3%.

Improper Reporting of One-Year Results

The World Association of Medical Editors lists “improper reporting of results” as research misconduct. The hiding of the dismal long-term results fits into that definition of misconduct.

In the protocol, the STAR*D researchers stated they would determine the stay-well rate at the end of one year. However, they didn’t discuss this figure in their published report of the one-year outcomes, and to MIA’s knowledge, none of the STAR*D investigators has subsequently written about it. The 3% number isn’t to be found in psychiatric textbooks, and again, to the best of MIA’s knowledge, no major U.S. newspaper has ever published this result. The only acknowledgement by a STAR*D investigator of this dismal outcome came when *Medscape News* asked Maurizio Fava about Pigott’s finding, and he acknowledged that it wasn’t “incompatible” with what they had reported.

As such, the STAR*D investigators have mostly kept it hidden from the public and their own profession, and it likely would never have surfaced had it not been for Ed Pigott’s obsession with fleshing out the true results from the “largest and longest trial of antidepressants ever conducted.”

Indeed, in 2009, NIMH director Thomas Insel stated that “at the end of 12 months, with up to four treatment steps, roughly 70% of participants were in remission.” He was now informing the public that 70% of the 4,041 patients who entered the study *got well and stayed well*, a statement that exemplifies the grand scale of the STAR*D fraud. Seventy percent versus a reality of 3%—those are the bottom-line numbers for the public to remember when it judges whether, in the reporting of outcomes in the STAR*D study, there is evidence of an “intent to deceive.”

Institutional Corruption

In *Psychiatry Under the Influence*, Lisa Cosgrove and I wrote about the STAR*D trial as a notable example of “institutional corruption.” There were two “economies of influence” driving this corruption: psychiatry’s guild interests, and the extensive financial ties that the STAR*D investigators had to pharmaceutical companies.

The American Psychiatric Association, which is best understood as a trade association that promotes the financial and professional interests of its members, has long touted antidepressants as an effective and safe treatment. After Prozac was brought to market in 1988, the APA, together with the makers of antidepressants, informed the public that major depression was a brain disease, and that the drugs fixed a chemical imbalance in the brain. The prescribing of these drugs took off in the 1990s, and has continued to climb ever since, such that today more than one in eight American adults takes an antidepressant every day.

The STAR*D results, if they had been accurately reported, would have derailed that societal belief. If the public had been told that in this NIMH study, which had been conducted in real-world patients, only 35% remitted, even after four treatment steps, and that only 3% remitted and were still well at the end of one year, then prescribing of these drugs—and societal demand for these drugs—surely would have plummeted. The STAR*D investigators, through their protocol deviations and their imagined remissions in patients that had dropped out, plus their hiding of the one-year results, turned the study into a story of the efficacy of these drugs. They were, in a business sense, protecting one of their primary “products.”

In addition, through their research misconduct, they were protecting the public image of their profession. The 67% remission rate told of skillful psychiatrists who, by trying various combinations of antidepressants and other drugs, eventually helped two-thirds of all patients become “symptom free.” The remitted patients were apparently completely well.

Even though STAR*D was funded by the NIMH, the corrupting influence of pharmaceutical money was still present in this study. The STAR*D investigators had numerous financial ties to the manufacturers of antidepressants. Here is a graphic that Lisa Cosgrove and I published in *Psychiatry Under the Influence*, which counted the number of such ties the various investigators had to pharmaceutical companies.

Financial Disclosures of STAR*D Investigators

(Number of ties to pharmaceutical companies)

Investigator	Academic Affiliation	Has Served as Advisor, Consultant or Speaker, or Received Research Funds	Tie to Forest, Maker of Study Drug
John Rush, M.D.	Southwestern Medical Center (U Texas)	10	Yes
Madhukar Trivedi, M.D.	Southwestern Medical Center (U Texas)	21	Yes
Stephen Wisniewski, Ph.D.	Univ. of Pittsburgh	1	
Andrew Nierenberg, M.D.	Massachusetts General Hospital	16	Yes
Jonathan Stewart, M.D.	Columbia Univ.	5	
Michael Thase, M.D.	Univ. of Pennsylvania	16	Yes
Philip Lavori	Stanford Univ.	6	Yes
Patrick McGrath, M.D.	Columbia Univ.	5	
Jerrold Rosenbaum, M.D.	Massachusetts General Hospital	26	Yes
Harold Sackeim, Ph.D.	Columbia Univ.	6	
David Kupfer, M.D.	Univ. of Pittsburgh	6	Yes
Maurizio Fava, M.D.	Massachusetts General Hospital	33	Yes

Source: Disclosure statement in A. John Rush, "Acute and longer-term outcomes in depressed outpatients requiring one or more treatment steps." *Am J Psychiatry* 163 (2006):1905-17.

In total, the 12 STAR*D investigators had 151 ties to pharmaceutical companies. Eight of the 12 had ties to Forest, the manufacturer of citalopram.

The drug companies that sold antidepressants, of course, would not have been pleased if their key opinion leaders published results from a NIMH trial that told of real-world outcomes so much worse than outcomes from industry-funded trials of their drugs. The real-world efficacy that emerged in the STAR*D trial belied the advertisements that told of highly effective drugs that could make depression miraculously lift.

Thus, the poisoned fruit of institutional corruption: newspapers still today point to the 67% remission rate as evidence of the efficacy of antidepressants, while most of the public—and prescribers of these drugs—remain unaware of the true results.

The Harm Done

The articles published by Pigott and colleagues since 2010 have provided a record of the scientific misconduct of the STAR*D investigators. This MIA Report simply presents a chronology of the fraud, and, relying on their work, a numerical accounting of how each element of research misconduct boosted remission rates. The purpose of this MIA Report is to make clear the “intent to deceive” that was present in the STAR*D investigators’ deviations from the protocol, their publication of a fraudulent cumulative remission rate, and their hiding of the one-year outcome that told of a failure of this paradigm of care.

This research misconduct has done extraordinary harm to the American public, and, it can be argued, to the global public. As this was *the study* designed to assess outcomes in real-world patients and guide future clinical care, if the outcomes had been honestly reported, consistent with accepted scientific standards, the public would have had reason to question the effectiveness of antidepressants and thus, at the very least, been cautious about their use. But the fraud created a soundbite—a 67% remission rate in real-world patients—that provided reason for the public to believe in their effectiveness, and a soundbite for media to trot out when new questions were raised about this class of drugs.

This, of course, is fraud that violates informed consent principles in medicine. The NIMH and the STAR*D investigators, with their promotion of a false remission rate, were committing an act that, if a doctor knowingly misled his or her patient in this way, would constitute medical battery.

This cataloging of harm done extends to those who prescribe antidepressants. Primary care physicians, psychiatrists, and others in the mental health field who want to do right by their patients have been misled about their effectiveness in real-world patients by this fraud.

The harm also extends to psychiatry’s reputation with the public. The STAR*D scandal, as it becomes known, fuels the public criticism of psychiatry that the field so resents.

Yet, and this may seem counterintuitive, there is now an opportunity for psychiatry to grasp. The American Psychiatric Association, and the international community of psychiatrists, could take a great step forward in regaining public trust if they spoke out about the STAR*D fraud and requested a retraction of the published articles. Doing so would be an action that told of a profession’s commitment, as it moves forward, to uphold research standards, and to provide the public with an honest accounting of the “evidence base” for psychiatric drugs.

However, failing to do so will only deepen justified criticism of the field. It will be a continuance of the past 15 years, when psychiatry has shown, through its inaction, that research misconduct in this domain of medicine—misconduct that rises to the level of scientific fraud—is acceptable practice, even though it may do great harm.

A Public Petition to Retract the STAR*D Summary Article

As we believe this is a matter of great importance to public health, Mad in America has put up a [petition](#) on change.org urging the *American Journal of Psychiatry* to retract the November 2006 summary article of the STAR*D results. A [2011 article](#) on the subject of retraction in a medical journal noted the following:

Articles may be retracted when their findings are no longer considered trustworthy due to scientific misconduct or error, they plagiarize previously published work, or they are found to violate ethical guidelines. . . . Although retractions are relatively rare, the retraction process is essential for correcting the literature and maintaining trust in the scientific process.

In this case, the facts are clear: the 67% remission rate published in the *American Journal of Psychiatry* in November 2006 can no longer be “considered trustworthy due to scientific misconduct,” and that retraction of the article “is essential for correcting the literature and maintaining trust in the scientific process.” The article also noted that “there is no statute of limitation on retractions.”

Moreover, the World Association of Medical Editors, in its “[Professional Code of Conduct](#),” specifically states that “editors should correct or retract publications as needed to ensure the integrity of the scientific record and pursue any allegations of misconduct relating to the research, the reviewer, or editor until the matter is resolved.”

And here is one final fact that makes the case for retraction. The [NIMH’s November 2006 press release](#), which announced that “almost 70% of those who did not withdraw from the study became symptom free,” contains evidence that the NIMH itself, or at least its press office, was duped by its own investigators. Either that, or the NIMH silently countenanced the fraud.

First, it notes that of the 4,041 who entered the trial, “1,165 were excluded because they either did not meet the study requirements of having “at least moderate” depression (based on a rating scale used in the study) or they chose not to participate.” Thus, it stated, there were 2,871 “evaluable patients.” The NIMH press office either didn’t know that 931 patients who lacked a HAM-D baseline score that met eligibility criteria had been added back into the mix of remitted patients, or else it deliberately hid this fact from the public.

Second, the press release stated that the purpose of the QIDS-SR assessments during clinic visits was to inform ongoing care: “Patients were asked to self-rate their symptoms. The study demonstrated that most depressed patients can quickly and easily self-rate their symptoms and estimate their side effect burden in a very short time. Their doctors can rely on these self-rated tools for accurate and useful information to make informed judgments about treatment.” This, of course, was consistent with the protocol, that QIDs would be used for this purpose, but instead was the instrument that the STAR*D investigators used to report remission rates in their summary paper.

Here is Ed Pigott's opinion on the call for retraction: "I started investigating STAR*D in 2006 and with colleagues have published six articles documenting significant scientific errors in the conduct and reporting of outcomes in the STAR*D trial. STAR*D's summary article should clearly be retracted. This is perhaps best seen by the fact that its own authors lack the courage to defend it. By rights and the norms of ethical research practice, STAR*D authors should either defend their work and point out the errors in our reanalysis or issue corrections in the *American Journal of Psychiatry* and *New England Journal of Medicine* where they published their 7 main articles. What they can't defend must be retracted."

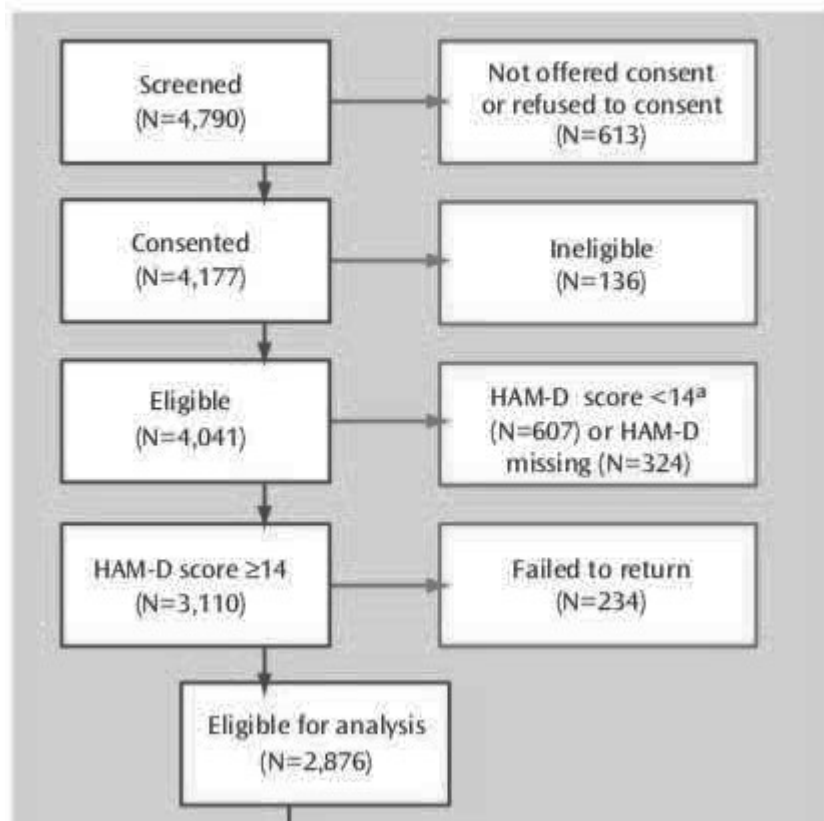
Our hope is that information about this Mad in America petition will circulate widely on social media, producing a public call for retraction that will grow too loud for the *American Journal of Psychiatry* to ignore. Indeed, the publication of the RIAT re-analysis of the STAR*D results in a prestigious medical journal presents a Rubicon moment for American psychiatry: either it retracts the paper that told of a fabricated outcome, or it admits to itself, and to the public, that scientific misconduct and misleading the public about research findings is accepted behavior in this field of medicine.

The petition can be signed [here](#).

Footnote: There are two possible explanations for the increase in the number of the 4,041 participants who were said to have failed to return for a second visit (from 234 in the step 1 report to 370 in the final summary). One possibility is that 136 of the 931 patients said to lack a qualifying HAM-D score never returned for a second visit, and thus in the summary report, they were added to this "didn't return" group and removed from the 931 group, leaving 795 participants who lacked a qualifying HAM-D score included in the count of evaluable patients.

A second possibility can be found in the patient flow chart published in the step 1 report. Here it is:

FIGURE 1. Participant Flow (CONSORT Chart) for the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) Study



Source:

You can see here that there were 4,177 who consented to being in the study. Then 136 were deemed to be ineligible for some reason, and thus weren't included in the count of 4,041 patients who entered the study. However, the 136 ineligible patients in the 4,177 count would have had a first screening visit, and once they were told they were ineligible, would not have returned for a second clinic visit. So the possibility here is that the STAR*D authors took this group of 136 ineligible patients, who never entered the study, and added them into the "did not return" group in order to further decrease the denominator in their final remission count.

Thus, there are two possibilities. The first tells of an extraordinary numerical coincidence. There were 136 patients who were declared ineligible before the study began, and a second, different group of 136 patients who lacked a qualifying HAM-D score, yet were allowed to enter the study but failed to return for a second clinic visit. The second possibility tells of falsification of data, a particularly egregious form of research misconduct.