The New York Times Is Now Engulfed in the STAR*D Scandal

madinamerica.com/2024/04/the-new-york-times-is-now-engulfed-in-the-stard-scandal/

By Robert Whitaker April 27, 2024

After Ed Pigott and colleagues published their patient-level reanalysis of the STAR-D results this past summer in *BMJ Open,* it seemed that this scandal—which is a tale of research fraud—might finally be reported on by the mainstream media, and thus burst into the American consciousness. But eight months have now passed, and while the *Psychiatric Times,* much to its credit, did publish a cover story of the reanalysis, American newspapers have remained mute, even though Pigott and colleagues have contacted reporters at *The New York Times* and other major newspapers, urging them to set the record straight.

And now *The New York Times*, rather than report on the *BMJ Open* paper, has once again trotted out the fraudulent results as evidence of the efficacy of antidepressants.

On Thursday, April 25, *The New York Times* published a story titled "What You Really Need to Know About Antidepressants." And then it informed its readers the following:

"The largest study of multiple antidepressants—nicknamed the STAR*D trial—found that half of the participants had improved after using either the first or second medication that they tried, and nearly 70 percent of people had become symptom-free by the fourth antidepressant."

That is the bottom-line outcome that the STAR*D investigators promoted to the public in November 2006, when it published a summary of the study outcomes in the *American Journal of Psychiatry*. And here it is 18 years later, and *The New York Times* is telling of that outcome, even though it has been conclusively shown, in a paper published in a prestigious medical journal, that if the STAR*D investigators had adhered to the protocol, they would have reported a remission rate of 35%.

Although Pigott and colleagues are continuing to analyze the patient-level STAR*D data, and have already spotted evidence of a failure to report on a suicide in the study, this article in *The New York Times* is a sign that the scandal is on its way to disappearing from public sight, with most of the public—as a result of an extraordinary and inexplicable journalistic failure—never hearing of it.

As such, it is now a story of a twin institutional failure. The American Psychiatric Association, through its *American Journal of Psychiatry*, has defended the published result, and mainstream newspapers have failed to inform the public of the *BMJ Open* paper, letting the nearly 70% remission rate remain in the public mind of evidence of the "efficacy" of antidepressants.

The cost to society of this twin failure is extraordinary.

As a society, we organized our thinking around a “narrative of science” that, beginning in the 1980s, told of how researchers had discovered that chemical imbalances were the cause of major psychiatric disorders and that a second generation of psychiatric drugs fixed those chemical imbalances, like insulin for diabetes. This was a story of a great medical advance, and the announced results from the STAR*D trial, heralded by the NIMH as the “largest and longest study ever done to evaluate depression treatment,” fit into that story of medical progress, for it told of 70% of depressed patients becoming “symptom free” after repeated treatments with antidepressants.

However, it is not a story of advance that has survived the test of time. The low serotonin theory of depression fell apart long ago, and Ed Pigott and colleagues published their first report on the STAR*D transgressions in 2010, telling of protocol violations that had been employed to inflate the reported remission rate. Their paper in *BMJ Open,* if it had been reported on by the media, could have spurred a day of public reckoning: why had we been misled in this way about the results from the very NIMH trial that was expected to guide clinical care?

That possibility promised a new way forward. But now, with *The New York Times* publication on April 25, there is reason to think there will be no day of reckoning, and a research paper that should be retracted will continue to govern public understanding of the “efficacy” of antidepressants.

As such, this is a scandal that now tars American journalism, and not just psychiatry.

A Brief Recap of the Scandal

When the STAR*D study was launched, it was understood that that its results would guide future clinical care of depressed patients. While there had been numerous industry-funded trials of antidepressants, those trials utilized inclusion-exclusion criteria that prevented most “real-world” patients from entering the studies. Various studies had found that 60% to 90% of patients with depression couldn’t participate in industry-funded trials of antidepressants. STAR*D would fill this void.

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.”

The STAR*D study was not designed to measure the effectiveness of antidepressants against placebo. It was designed to assess how well real-world patients fared in clinical care that utilized antidepressants as a first-line treatment, with prescribers able to change dosages and try a second, third, or fourth antidepressant if an initial treatment didn’t work.

Moreover, it had two phases. Once patients remitted after acute treatment with an antidepressant, they were whisked into a follow-up trial to assess whether their remission could be sustained with continuing use of antidepressants. The acute phase of the trial would assess whether antidepressant use could produce a *moment* of remission (at the end of one of the stages of treatment), while the follow-up would assess whether such treatment enabled them to stay well.

In their protocol, the STAR*D investigators told of the critical importance of this follow-up phase. “How common are relapses during continued antidepressant treatment in ‘real-world’ clinical practice?” the STAR*D investigators wrote. “How long [are remitted patients] able to stay well?”

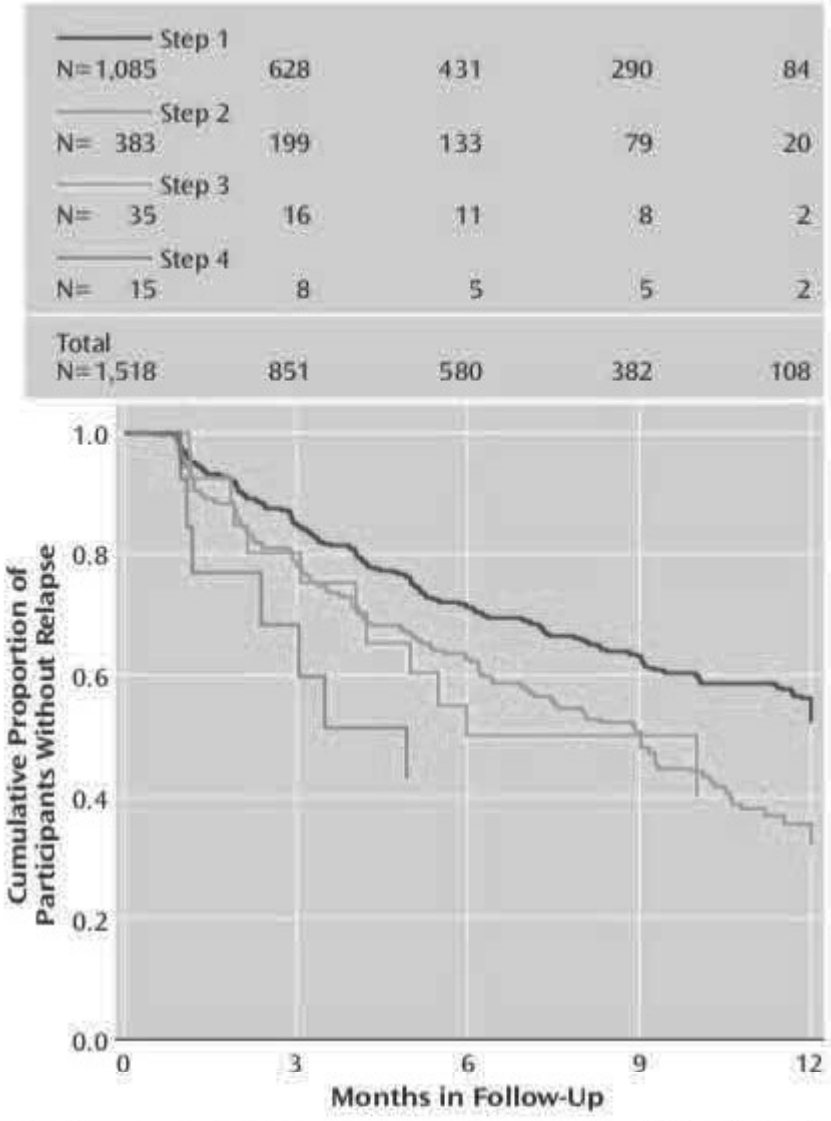
In the first months of 2006, the STAR*D investigators published three reports in the *American Journal of Psychiatry* detailing remission rates following the first two stages of acute treatment. Then, in November 2006, they published a summary report of outcomes. Their paper was titled: “Acute and long-term outcomes in depressed outpatients requiring one or several treatment steps: A STAR*D report.”

Anyone who had read the first three articles could only be confused by this November report. The number of “evaluable patients” had changed in each of the reports; there was no report on remission rates according to a HAM-D assessment of symptoms, which in the earlier reports had been presented as a primary instrument for doing so; and although the abstract didn’t mention this, the authors confessed in their discussion that the reported 67% remission rate was a *theoretical* calculation, based on the notion that if those who had dropped out during the acute phase of the study had instead stayed in through all four stages of treatment they would have remitted at the same rate as those who did stay in through all four stages.

In other words, close readers of the paper would understand that the reported 67% remission rate in the abstract was something of a made-up number.

In this paper, the STAR*D investigators also reported on the one-year outcomes. However, that section of the summary paper was very short, and the investigators didn’t tell of the percentage of remitted patients who stayed well during the one-year follow-up, even though the protocol had stated that assessing the sustained stay-well rate was the primary purpose of this second phase of the study. All they stated in their section on “longer-term outcomes” was that relapse rates were higher for those who had taken several treatment steps in the acute phase of the study to remit than for those who remitted after the first treatment stage. There was a table that, in its title, told of relapse rates for patients who entered the follow-up study in remission, but it was nearly impossible to decipher, and so the bottom-line result—how many patients remitted and stayed well and in clinical care for one year—was missing from this report. Here is the table that was published:

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 (χ^2=23, df=3, p<0.0001). Significant post-hoc comparisons with Bonferroni corrections revealed significant differences between steps 1 and 2.

This was the very confusion that prompted psychologist Ed Pigott and colleagues to obtain the protocol through a Freedom of Information Act request and see if they could make sense of why this stream of four papers didn't tell a coherent story. In 2010, they published an article that told of how the STAR*D investigators had violated their protocol to inflate the reported remission rate.

At that time, Pigott and colleagues reported that only 38% of the patients who met inclusion criteria had remitted after four stages of treatment. Furthermore, and most important, they made sense of the confusing graphic that the STAR*D investigators had published in their summary report that told of the outcomes in the one-year follow-up. Their finding was

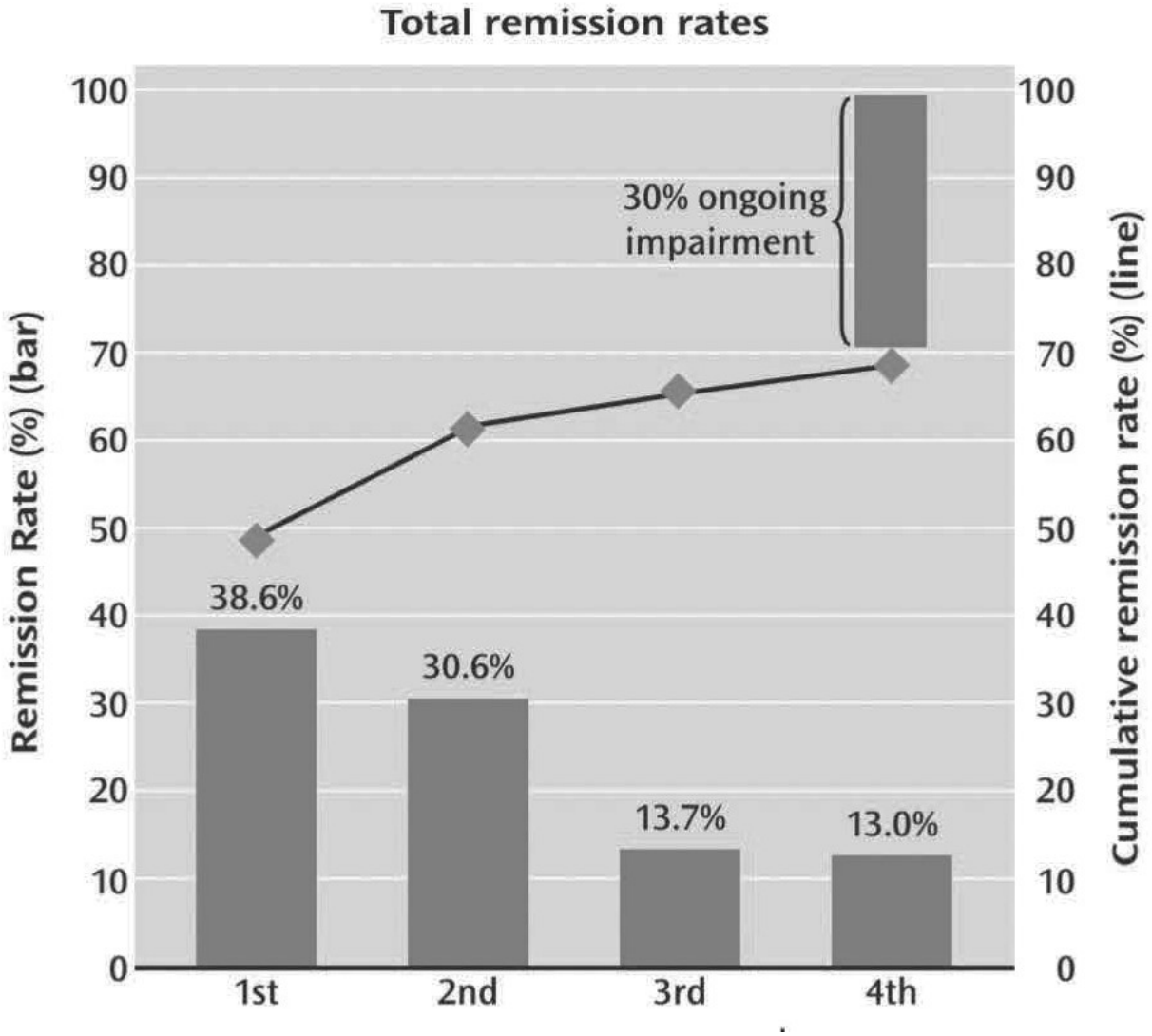
stunning: only 108 of the 4,041 patients (3%) who had entered the trial had remitted and stayed well and in the trial to the end of the one-year follow-up. All of the others had either never remitted, remitted and then relapsed, or dropped out at some point during the study.

This was, in fact, the most important result from the STAR*D Study. In a study of real-world depressed patients, there was no evidence that regular clinical care with antidepressants helped patients get well and *stay well.* That 3% figure—if it had been promoted to the public—would have prompted a rethinking of the use of these drugs.

However, few in the public were hearing of Pigott's work. Instead, they were hearing of how the STAR*D study had proven that antidepressants were quite effective for most people.

"Over the course of all four levels, almost 70 percent of those who didn't withdraw from the study became symptom free," the NIMH had informed the public when the November 2006 paper was published in the *American Journal of Psychiatry*.

The NIMH press release made no mention of the 3% sustained remission outcome, and soon that "almost 70 percent" remission rate became the result known by the public. Moreover, it was presented as the cumulative remission rate at the end of one year, which indicated that those who had remitted had remained well. Here is a graphic, published in 2013 in the *American Journal of Psychiatry,* that told of how there was "ongoing impairment" in only 30% of the patients.



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

Given that public soundbite and promotion of the 70% remission figure in the psychiatric literature, the 2010 paper by Pigott and colleagues told of a scandal. The STAR*D investigators had grossly inflated the reported remission rate and done so though various data manipulations that violated their own protocol, and equally important, they had hidden the fact that almost no patients remitted and then stayed well and in the trial to its one-year end.

Over the next few years, Pigott and colleagues published several more articles deconstructing the STAR*D results. Mad in America published two of his blogs and reported on his journal publications, and he provided Mad in America with STAR*D documents, including the protocol, that we posted on our website. Readers could see for themselves how the STAR*D investigators had violated their own protocol. Yet, no mainstream media reported on this scandal.

However, when Pigott and colleagues published their patient-level analysis of the STAR*D outcomes in July 2023, it seemed that surely this was now a story that the media could no longer ignore.

Their "reanalysis" was published in a prestigious medical journal (*BMJ Open*), and it detailed a scientific scandal of epic proportions. In September, MIA published its report on the scandal, and in it, we provided for the lay public a precise numerical accounting of how the STAR*D protocol violations inflated the remission rate. Here were the four principal ways they had inflated their results:

Not counting early dropouts as treatment failures

The protocol called for patients who failed to return after their baseline visit, when they were first prescribed an antidepressant, to be deemed "intolerant" to the treatment and thus counted as treatment failures. However, in their final report, the STAR*D investigators removed the 234 patients who had failed to return after their baseline visit from their tally of evaluable patients (thus decreasing the number of treatment failures).

Including ineligible patients in their count of remitted patients

The protocol required patients to have a baseline HAM-D score of 14 or greater (the scale used to measure the severity of depression symptoms.) However, in their final 2006 report, the STAR*D investigators included 931 patients who either lacked a baseline HAM-D score (324 patients) or had a baseline HAM-D score less than 14, and thus weren't depressed enough to be eligible for the study (607 patients). By including this group of 931 patients in their final report, they added 570 to their tally of remitted patients.

Switching outcome measures

The protocol stated that the primary outcome would be an HAM-D assessment of symptoms administered in blinded fashion by a Research Outcome Assessor at the end of each of the four stages of treatment. Remission was defined as a HAM-D score of 7 or less, and anyone who scored in remission at the end of a treatment step was whisked into a year-long follow-up to see if they would stay well for this longer period of time.

In addition, the protocol stated that at each clinic visit patients would self-assess their symptoms, using a tool known as the Quick Inventory of Depressive Symptomatology-Self-Report (QIDS-SR). This would be done to see how well the QIDS-SR scores correlated with HAM-D scores, with the thought that in the future QIDS could become a useful tool for quickly assessing symptoms during a clinic visit. The protocol specifically stated that QIDS scores collected during clinics would not be used to report outcomes.

However, in their summary report of outcomes, the STAR*D investigators did not report remission rates based on HAM-D scores. They only reported remission based on QIDS scores. This switch of outcome measures added 195 patients to their tally of remitted patients.

Calculating a "theoretical" remission rate

The protocol called for those who dropped out of the study without having remitted to be counted as treatment failures. However, in their summary report, the STAR*D investigators theorized that if the drop-outs had stayed in the study through all four steps of the study they would have remitted at the same rate as those who did stay to the end, and they added these imagined remitters to their final tally of remitted patients. This theoretical calculation transformed 606 treatment failures into treatment successes.

In their *BMJ Open* paper, Pigott and colleagues reported that if the STAR*D investigators had adhered to their protocol, they would have announced that 1,089 of 3,110 eligible patients had remitted (35%.) Their four protocol violations removed 234 failures from the final tally of evaluable patients and added 1,371 to the tally of remitters, and in that manner, they produced a much different bottom line calculation: 2,460 remitters in a population of 3,671 evaluable patients (67% remission rate.)

In short, the protocol violations and imagined remissions accounted for 56% of the remissions that led to the announced cumulative remission rate of 67%.

After we published our September MIA Report, we put up a petition on change.org calling for the *American Journal of Psychiatry* to retract the study. More than 2,700 people signed the petition, which we sent to Ned Kalin, editor in chief of the *American Journal of Psychiatry*, on October 10.

We saw this as a defining moment for American psychiatry. Would the *American Journal of Psychiatry* do the right thing and retract the study? We also saw it as a defining moment for mainstream media in the United States: would they finally report on this extraordinary scandal in America medicine?

A Scandal Handed on a Platter to Newspapers

Pigott's report, once it was given this public push by MIA, did trigger a brief flare-up of attention to the scandal. In December, the *American Journal of Psychiatry* published a response by the STAR*D investigators to Pigott's work, and more importantly, *Psychiatric Times* made Pigott's re-analysis its cover story that month. In his essay, John Miller, editor-in-chief of *Psychiatric Times*, wrote that since 2006, the STAR*D study had stood "out as a beacon guiding treatment decisions," and he succinctly identified what was now at issue:

"For us in psychiatry, if the BMJ authors are correct, this is a huge setback, as all of the publications and policy decisions based on the STAR*D findings that became clinical dogma since 2006 will need to be reviewed, revisited, and possibly retracted."

At that point, Ed Pigott and his colleagues contacted *The New York Times* and other mainstream publications. The essay by *Psychiatric Times* had laid out what was at stake for the public, and it was nothing less than guidelines for treating depression and our societal

understanding of the efficacy of antidepressants. *BMJ* had validated Pigott's reanalysis, and now a psychiatric journal had validated its importance.

When the *Psychiatric Times* cover story was published, I thought that the media's silence on this subject would surely now be broken. This was a story being handed to reporters on a ready-made platter. You had documents that told of the protocol violations. You had a patient-level reanalysis published in a prestigious medical journal that detailed how the STAR*D investigators had employed protocol violations to grossly inflate the reported remission rate. You had a psychiatric journal telling of the importance of this study, and how, if the reanalysis was correct, the profession had been led astray by the NIMH-funded investigators and had adopted prescribing protocols that were not warranted.

This was a big story! A story, in fact, that would stun readers in countries around the world, given the importance of the STAR*D study to the global psychiatric community. And when *Psychiatric Times* subsequently published a letter from two psychiatrists titled "STAR*D: It's Time to Atone and Retract" it seemed that surely newspapers would jump on the story. The two authors of the letter, Nicolas Badre and Jason Compton, told of the extraordinary influence of the STAR*D study, describing it as probably the "most important study" in the "recent history of psychiatry."

They then provided these examples:

"The results of STAR*D are familiar to all psychiatrists as commonly tested on boards and residency training exams. According to its abstract, the main conclusion of STAR*D was that 'the overall cumulative remission rate was 67%.'"

"The impact of STAR*D was outstanding; it is highly cited in our textbooks. The latest edition of Tasman's textbook of psychiatry makes 54 references to the STAR*D trial and includes a full chart of the study."

"Similarly, the latest edition of the widely read Maudsley's prescribing guide makes 9 references to STAR*D . . . More than 100 peer-reviewed articles have been written about STAR*D, and the original paper was cited 207 times on PubMed.org in 2023—the most of any year."

"Newspapers have regularly promoted its findings and continue to do so. As recently as last year, *The New York Times* was citing STAR*D as 'the largest study of antidepressants to date,' and touting its results that more 'than 60 percent of those patients actually had a very good response.' "

"Some may think finding that antidepressants are effective in 67% of patients is trivial; however, the efficacy of antidepressants was not as widely accepted prior to STAR*D. The saturation of psychiatric textbook with STAR*D, more than ever before, solidified that teaching."

For a journalist, that letter tied a red ribbon around the entire story. There were documents to base the story on, there was a paper published in a prestigious journal that detailed the protocol violations and the gross inflation of the reported remission rate, and there were psychiatrists to interview who could tell of how the inflated STAR*D results had misled their entire profession and the public.

And yet, so far, the mainstream media has remained mute. And now, on April 25, *The New York Times,* rather than turn its attention to the scandal, repeated the very falsehood that has poisoned the psychiatric literature and its textbooks since 2006. It did so in an article that promised to tell readers the "facts" about antidepressants.

That story told of a remarkable journalistic failure. I submitted the following comment to the article:

"The New York Times needs to inform its readers that the STAR*D investigators violated their own protocol, in four ways, to dramatically inflate the remission rate of nearly 70% that they reported in their 2006 paper. A re-analysis of patient level data by Ed Pigott and colleagues, which was published in BMJ Open last July, found that if the STAR*D investigators had adhered to their study protocol, the true remission rate at the end of four stages was 35%, not 70%. Please do a story on this re-analysis; the public needs to be informed of the true results from this NIMH-funded trial, and not be misled, yet again, by this report of a 70% remission rate in this NIMH study. The story of the STAR*D trial is a story of research fraud, and the paper published in the *American Journal of Psychiatry* that told of that fraudulent result needs to be retracted."

The *New York Times* moderator did not allow that comment to be published.

An Intent to Deceive

In their letter published in the *American Journal of Psychiatry*, John Rush and four of his STAR*D investigators defended their work, stating that in a study like STAR*D, it was important to provide results for all of the patients in the study, and chided Pigott for removing 931 patients from their analysis. They wrote:

"The analytic approach taken by Pigott et al. has significant methodological flaws . . . in total, 941 patients included in our original analyses were eliminated from Pigott et al.'s reanalyses based on their post-hoc criteria. The rationale for removing these participants from the longitudinal analysis appears to reflect a studious misunderstanding of the aims Rush et. al. paper, with the resulting large difference in remission rates most likely the result of exclusion by Pigott et al. of hundreds of patients with low symptoms scores at the time of study exit."

And:

“Effectiveness trials by design aim to be more inclusive and more representative of the real world than efficacy trials. By removing the data of over 900 study participants from their reanalyses, Pigott et al. failed to recognize the purpose of inclusiveness. It appears that the authors created rules to define post hoc which subjects to include, which eliminated many subjects who experienced large improvements during one or another of the study’s levels. By doing so, the sample is biased to underestimate the actual remission rate.”

This letter, by itself, tells of an intent to deceive. It is a stunningly dishonest letter, and the editors of the *American Journal of Psychiatry*, having published the STAR*D reports in 2006, were surely aware that it was so. There is no acknowledgement in Rush’s letter that the 931 patients excluded by Pigott et al. either weren’t depressed enough to qualify for the study or lacked a baseline HAM-D score. They hid this essential fact, and instead claimed that Pigott and colleagues had created “post-hoc criteria” to wrongly remove them from their re-analysis.

While this is an extraordinarily brazen lie, it is easy to understand their motivation for coming up with this “defense” of their work. If they had told of how these 931 patients were excluded because they didn’t have baseline scores that made them eligible for inclusion, *American Journal of Psychiatry* readers would immediately understand that the STAR*D investigators, by including them in their final tally of “evaluable patients,” were guilty of research misconduct.

The STAR*D investigators were of course duty bound to present results based on their protocol for assessing outcomes. However, if they had also wanted to report remission rates for the 931 patients who didn’t meet eligibility criteria for the study, they could have done so in a transparent manner and that would have been fine. Their summary paper should have told of the inclusion criteria, and in addition to reporting remission rates based on their protocol, the STAR*D investigators could have stated “and here are the remission rates for the 931 patients who hadn’t met eligibility criteria for the study.”

But there is no such transparency in their 2006 report. Instead, there is evidence, at every step, of an intent to deceive.

First, here is how they presented their results in the abstract:

“Results: The QIDS-SR_{16} 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%.”

That paragraph informs readers that the QIDS was the primary outcome measure, and that 67% of the patients entered into the study remitted. The abstract is omitting these facts: that 931 of the patients in in this summary report didn’t meet eligibility criteria, that the HAM-D was designated as the primary outcome measure, and that this 67% is a “theoretical” number that relied on converting study dropouts into imagined remitters.

Second, in their “methods” section, they did not mention that patients needed to have a baseline HAM-D score of 14 or higher to be eligible for the trial. Instead, they wrote this:

“Participants met DSM-IV criteria for nonpsychotic major depressive disorder at study entry as determined by clinical diagnosis and confirmed with a DSM-IV checklist by the clinical research coordinator.”

Now, in their report on remission rates after the first stage of treatment they noted that patients needed to have a HAM-D score of 14 or higher to be eligible in the trial, and thus they concluded that 931 patients who lacked a baseline HAM-D score or had a HAM-D score less than 14 were not “evaluable patients.” But in this summary report, they snuck them back into the roster of evaluable patients, and they did so by implying that the DSM-IV checklist for major depressive disorder was the instrument used to assess eligibility.

Third, they did not disclose that the protocol had declared that the HAM-D would be used to assess whether a patient had “remitted.” Instead, the STAR*D investigators wrote:

“We used the Quick Inventory of Depressive Symptomatology—Self-Report (QIDS-SR_{16}) as the primary measure to define outcomes for acute and follow-up phases.”

That declaration made it seem that they had planned to use this tool for reporting outcomes all along. They did not note that the protocol called for the use of HAM-D to assess remission rates, and they did not confess that the protocol explicitly stated that the QIDS-SR, which was administered at clinic visits, was not to be used to assess research results. Furthermore, they did not report HAM-D remission rates in this paper.

Fourth, as noted above, in their presentation of the one-year outcomes, they did not tell of how many patients stayed well and in the trial to its end. Instead, they wrote: “Relapse rates were higher for those who entered follow-up after more treatment steps.”

Those are the footprints of deception that can be found in their summary paper of outcomes, and it was a deception that accomplished the following:

- It turned a 35% remission rate after four stages of acute treatment into a “bottom-line” finding that with multiple trials of antidepressants eventually 67% of depressed patients became symptom free.
- It hid the extraordinarily poor results from the one-year follow-up trial, which told of how very few patients had remitted, stayed well and in the trial to its one-year end.

When the STAR*D investigators defended their 2006 by report stating that they wanted to be “inclusive” of all patients, there was a way they easily could have done that. They simply would have needed to first report outcomes based on the protocol for patients who were eligible for the trial, and then they could have added a section on remission rates for those who didn’t meet eligibility criteria. Similarly, they could have reported remission rates based

on the QIDS-SR, but they would have first needed to report remission rates based on the HAM-D. They would also have needed to confess that the protocol stated that the QIDS-SR would not be used to report research results.

Such transparency would tell of an effort to honestly report results. Instead, the “footprints of deception” visible in the published article tell of an intent to deceive, and thus provide evidence of scientific misconduct.

Will the American Media Step Up?

At this point, American psychiatry, as an institution, seems content to let this story die down. The *American Journal of Psychiatry* is not going to retract the paper, and that really isn’t a surprise: American psychiatry has shown, again and again, that it is devoted to telling the public a story that maintains societal belief in the merits of its treatment, and there is a long record of negative results, such as the results from this STAR*D trial, being kept from the public.

But now it is *The New York Times* and the mainstream media that are facing a defining moment in their coverage of American medicine: will the media do its duty and inform the public of this scandal? Or will it, for reasons that are difficult to understand, remain mute? At this point, it appears that the media will take the latter course, and by doing so, fail the American public in a profound way.

***

42 COMMENTS

LEAVE A REPLY