STAR*D: A Tale and Trail of Bias

H. Edmund Pigott, PhD

NeuroAdvantage, LLC
Clarksville, Maryland

The 35-million-dollar Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study is the largest antidepressant effectiveness study ever conducted. STAR*D enrolled 4,041 depressed patients and provided them with exemplary free acute and continuing antidepressant care to maximize their likelihood of achieving and maintaining remission. Patients who failed to get adequate relief from their first antidepressant were provided with up to three additional trials of pharmacologically distinct treatments. This article identifies numerous instances of apparent bias in the conduct and reporting of outcomes from this study. In contrast to STAR*D's report of positive findings supporting antidepressants' effectiveness, only 108 of its 4,041 patients (2.7%) had an acute-care remission, and during the 12 months of continuing care, these patients neither relapsed nor dropped out. This article also discusses the roles of the *American Journal of Psychiatry* (AJP) and the National Institute of Mental Health (NIMH) in promoting the biased reporting of STAR*D's results.

Keywords: major depression; antidepressants; researcher bias; STAR*D

In the controlled clinical trials article describing STAR*D's methods, research design, and purpose, its authors state:

> STAR*D uses a randomized, controlled design to evaluate both the theoretical principles and clinical beliefs that currently guide the management of treatment-resistant depression in terms of symptoms, function, satisfaction, side-effect burden, and health care utilization and cost estimates. Given the dearth of controlled data, 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. (Rush, Fava, et al., 2004, p. 136)

To accomplish these objectives, STAR*D:

- Enrolled 4,041 real patients seeking care versus people responding to advertisements for depressed subjects as is common in industry-sponsored research.
- Included patients meeting a lower depression severity threshold than common in efficacy trials by requiring only a baseline Hamilton Rating Scale of Depression (HRSD) score of ≥14 versus ≥20 (e.g., see Davidson et al., 2002). This low symptom threshold, though, is similar to the many patients who are only mildly depressed when first prescribed antidepressants in routine clinical practice (Zimmerman, Mattia, & Posternak, 2002).
- Included depressed patients with comorbid medical and psychiatric conditions while only excluding those with a primary diagnosis of bipolar, psychotic, obsessive-compulsive, or eating disorders.

DOI: 10.1891/1559-4343.13.1.6

- Used "remission" versus "response" as the primary criterion of successful treatment.
- Provided 12 months of continuing care while monitoring the durability of treatment gains versus only reporting acute-care improvement.

STAR*D provided these 4,041 "real-world" depressed patients with exemplary free acute and continuing-care drug treatment while making extensive efforts to keep patients in treatment and maximize their likelihood of achieving and maintaining remission in a manner consistent with "the theoretical principles and clinical beliefs that currently guide the management of treatment-resistant depression." These efforts included:

- Providing a multistep educational program for patients and families throughout acute-care based on the neurochemical imbalance theory of depression that included "a glossy visual representation of the brain and neurotransmitters," consistently emphasizing that "depression is a disease, like diabetes or high blood pressure, and has not been caused by something the patient has or has not done. (Depression is an illness, *not* a personal weakness or character flaw.*)* The educator should emphasize that depression can be treated as effectively as other illnesses," and "explaining the basic principles of mechanism of action" for the patient's current antidepressant drug (O'Neal & Biggs, 2001, pp. 4–7).
- Providing "measurement-based care" that involved measuring symptoms and side effects at each clinic visit to guide aggressive medication dosing during each trial with a fully adequate dose for a sufficient duration to "ensure that the likelihood of achieving remission was maximized and that those who did not reach remission were truly resistant to the medication" (Trivedi, Rush, et al., 2006, p. 30). These "clinical management tools" were administered by each site's clinical research coordinator (CRC) who also provided the patient education.
- Allowing patients to select acceptable treatment options in steps 2–4 if their initial drug trial was not successful "to empower patients, strengthen the therapeutic alliance, optimize treatment adherence, and improve outcome" (Fava, Rush, et al., 2003, p. 483).
- Allowing liberal prescribing of nonstudy medications during every treatment phase (Trivedi, Rush, et al., 2006, p. 31).
- Open-label prescribing of all study medications with no placebo-control condition during any treatment phase.
- Using marketing strategies to promote patients' study affiliation via STAR*D-branded brochures, bimonthly newsletters (*The STAR*D Gram*), and an informational video emphasizing STAR*D's public health significance and the critical role played by patients (Fava, Rush, et al., 2003, p. 473).
- Using a reminder system to alert patients before appointments and calling patients on the day of any missed appointments to reschedule, and again within 24 hours, if there was no response, as well as having the patient's physician send a letter within 48 hours if contact had still not been established urging the patient to reschedule (Fava, Rush, et al., 2003, p. 474).
- Using a reminder system for all research outcome assessment phone calls and paying patients $25 for participating in said assessments (Fava, Rush, et al., 2003, p. 474).
- Permitting patients to reenter the study within 4 weeks after having dropped out (Trivedi, Stegman, Rush, Wisniewski, & Nierenberg, 2002, p. 80).
- Allowing physicians to make any therapeutic change necessary during continuing care to maximize patients' likelihood of sustaining remission, including scheduling additional visits if depressive symptoms returned and/or intolerable side effects emerged (Trivedi et al., 2002, p. 78).

In light of these "best practice" efforts, STAR*D's results should be viewed as reflecting antidepressants' optimal level of effectiveness when care is delivered in such a way to maximize patients' likelihood of achieving and maintaining remission by taking these drugs.

STAR*D's primary objective was to evaluate the relative efficacy of 11 pharmacologically distinct "next-step" treatments in up to three additional drug trials for the many patients who failed to get adequate relief from their first antidepressant. Table 1 describes each step's compared drugs. Cognitive therapy was also a treatment option in step 2, but too few patients accepted this form of treatment and it was therefore excluded from the primary step-2 switch and augmentation analyses (Rush, Trivedi, Wisniewski, Stewart, et al., 2006; Trivedi, Fava, et al., 2006).

TABLE 1. Compared Drug Treatment Strategies

Treatment Step	Compared Drug(s)
Step 1	Citalopram (Celexa) was the first-line SSRI treatment because of: • Absence of discontinuation symptoms; • Demonstrated safety in elderly and medically fragile patients; • Easy once-a-day dosing with few dose adjustment steps; and • Favorable drug–drug interaction profile (Trivedi et al., 2006b, p. 30).
Step 2 Switch study	• Sertraline (Zoloft), an SSRI with the same pharmacological profile as citalopram (Celexa); • Extended-release venlafaxine (Effexor), a "dual-action" agent that inhibits the reuptake of both serotonin and norepinephrine; and • Sustained-release bupropion (Wellbutrin SR), an "out-of-class" agent whose neurochemical action mechanisms are unknown; other than that, it does not inhibit serotonin reuptake and is believed to produce antidepressant effects by blocking the reuptake of dopamine and norepinephrine (Rush, Trivedi, Wisniewski, Stewart, et al., 2006, p. 1232).
Step 2 Citalopram (Celexa) augmentation study	• Buspirone (Buspar), a partial agonist at the postsynaptic 5-hydroxytryptamine 1A (5-HT1A) receptor that is believed to enhance the activity of SSRIs through the 5HT1A receptors; and • Sustained-release bupropion (Wellbutrin SR) whose neurochemical action mechanisms are unknown but is believed to produce antidepressant effects by blocking the reuptake of dopamine and norepinephrine (Trivedi et al., 2006a, p. 1244).
Step 3 Switch study	• Nortriptyline (Pamelor), a tricyclic antidepressant; and • Mirtazapine (Remeron), a tetracyclic antidepressant that blocks inhibitory α2-adrenoceptors on norepinephrine and serotonin neurons to enhance both norepinephrine and serotonin neurotransmission (Fava et al., 2006, p. 1169).
Step 3 Augmentation study of step 2's drug(s)	• Lithium; and • Triiodothyronine (Cytomel), a thyroid hormone.
Step 4 Switch study	• Tranylcypromine (Parnate), a monoamine oxidase inhibitor; and • Coadministered venlafaxine (Effexor) and mirtazapine (Remeron) to inhibit the reuptake of both serotonin and norepinephrine and block inhibitory α2-adrenoceptors on both norepinephrine and serotonin neurons to enhance both norepinephrine and serotonin neurotransmission.

EVIDENCE OF APPARENT BIAS IN PIGOTT ET AL. ARTICLE

Pigott, Leventhal, Alter, and Boren (2010) recently published an article on the status of antidepressants' efficacy and effectiveness research. This article's effectiveness section focused on a detailed reanalysis of STAR*D and discovered evidence suggestive of bias, which inflated STAR*D's reported acute-care remission rates while not reporting forthrightly its rate of relapse and/or dropout during the 12 months of continuing care. Specifically, Pigott et al. documented that:

- STAR*D changed its outcome measures following data collection. As designed, STAR*D's prespecified primary measure was the HRSD and the Inventory of Depressive Symptomatology—Clinician-Rated (IDS-C30), the secondary one for identifying "remitted" (i.e., those with a ≤7 HSRD score) and "responder" (i.e., those with a ≥50% reduction in depressive symptoms) patients. These measures were obtained in interviews by research outcome assessors (ROAs) blind to treatment assignment at entry into and exit from each trial and every 3 months during the 12 months of continuing care.
- STAR*D dropped the IDS-C30 and replaced it with the Quick Inventory of Depressive Symptomatology–Self-Report (QIDS-SR), a proprietary tool developed by STAR*D's principal investigators. The QIDS-SR was not a prespecified research measure, but rather one of STAR*D's "clinical management tools" that was used to guide care during every treatment visit. In the Pigott et al. paper's peer-review process, a reviewer wrote that the National Institute of Mental Health's (NIMH) Data Safety and Monitoring Board (DSMB) authorized the use of the QIDS-SR prior to "data lock and unblinding" because of STAR*D's high study dropout rate, which frequently resulted in missing exit IDS-C30 and HRSD assessments. Pigott et al. made this change in the paper, even though it could not be documented in the published literature. Subsequently, this author learned that no such DSMB authorization occurred (see the succeeding discussion).
- STAR*D changed its eligibility for analysis criteria in the steps 2–4 and summary articles without making this change explicit to readers. This change resulted in 607 patients who were initially reported as excluded because their <14 score on the ROA-administered baseline HRSD signified at most only mild depressive symptoms when starting on citalopram (Celexa) in step 1 subsequently being included. Similarly, an additional 324 patients who were initially reported as excluded because they lacked a baseline ROA-administered HRSD in step 1 were subsequently included. Thus, 931 of STAR*D's 4,041 patients (23% of all subjects) did not meet its step-1 eligibility for analysis criteria but were included in the steps 2–4 and summary articles' analyses.
- STAR*D failed to disclose that all 4,041 patients were started on citalopram (Celexa) in their initial baseline visit and that they excluded from analysis the 370 patients who dropped out without any subsequent visits, although the step-1 article states, "our primary analyses classified patients with missing exit HRSD scores as nonremitters a priori" (Trivedi, Rush, et al., 2006, p. 34). These early dropout patients did not take the exit HRSD and therefore should have been counted as treatment failures as prespecified.
- STAR*D did not disclose how to interpret the quarter-by-quarter survival data for continuing-care patients and thereby obscured from readers their startling finding that only 108 of the 1,518 remitted patients (7.1%) had not relapsed and/or dropped out by continuing-care's 12th month (see Table 2).

The cumulative effect of these decisions inflated STAR*D's reported findings, giving a false impression of antidepressants' optimal effectiveness with patients commonly prescribed these drugs. For example, STAR*D's major summary article reports a 36.8% remission rate in step 1 based on the QIDS-SR (Rush, Trivedi, Wisniewski, Nierenberg, et al., 2006, p. 1905).

TABLE 2. Survival Analysis by Treatment Step for Remitted and All Remitted/Responder Patients Who Consented to Continuing Care

Treatment Step and Status on Entry Into Continuing Care	N[a]	0–3 Months[b]	3–6 Months[b]	6–9 Months[b]	9–12 Months[b]
Step-1 remitted patients	1,085	628 (57.9%)	431 (39.7%)	290 (26.7%)	84 (7.7%)
Step-2 remitted patients	383	199 (52%)	133 (34.7%)	79 (20.6%)	20 (5.2%)
Step-3 remitted patients	35	16 (45.7%)	11 (31.4%)	8 (22.9%)	2 (5.7%)
Step-4 remitted patients	15	8 (53.3%)	5 (33.3%)	5 (33.3%)	2 (13.3%)
Survival Rate by Quarter for All Remitted Patients	1,518	851 (56.1%)	580 (38.2%)	382 (25.2%)	108 (7.1%)
Step-1 remitted/responder patients	1,475	803 (54.4%)	529 (35.7%)	347 (23.5%)	98 (6.6%)
Step-2 remitted/responder patients	622	300 (48.2%)	190 (30.5%)	115 (18.5%)	29 (4.7%)
Step-3 remitted/responder patients	102	37 (36.3%)	22 (21.6%)	15 (14.7%)	3 (2.9%)
Step-4 remitted/responder patients	49	20 (40.8%)	12 (24.5%)	9 (18.4%)	2 (4.1%)
Survival Rate by Quarter for All Patients Who Entered Continuing Care	2,248	1,160 (51.6%)	753 (33.5%)	486 (21.6%)	132 (5.9%)

[a]Number of patients entering continuing care (Rush, Trivedi, Wisniewski, Nierenberg, et al., 2006, Figures 2 and 3).

[b]Number of patients who called in at least once into the IVR system during this time period and did not score has having relapsed in this or a prior time period (Rush, Trivedi, Wisniewski, Nierenberg, et al., 2006, Figures 2 and 3).

Actually, though, only 790 of the 3,110 patients (25.4%) who had a baseline ≥14 HRSD score met the prespecified remission criterion. STAR*D's reported 36.8% step-1 remission rate for citalopram (Celexa) inflates this drug's actual rate by 44.9%.

Numerous recent meta-analyses and systematic reviews have found that selective outcome reporting bias, in which researchers fail to report the negative results for the prespecified primary measure and instead highlight positive results from a new measure as though it was their primary measure of interest, plagues industry-sponsored research and significantly inflates the reported outcomes (Mathieu, Boutron, Moher, Altman, & Ravaud, 2009; Rising, Bacchetti, & Bero, 2008; Turner, Matthews, Linardatos, Tell, & Rosenthal, 2008; Vedula, Bero, Scherer, & Dickersin, 2009). Reporting outcomes, as prespecified, is essential to ensure the integrity of clinical research because it guards against investigators selectively publishing results that merely reinforce their original expectations and beliefs by changing the prespecified measures and planned analyses

following data collection and analysis—a form of researcher bias known as *HARKing*, an acronym for "hypothesizing after the results are known" (Kerr, 1998). Although such HARKing is most commonly associated with industry-sponsored research, Chan, Krleza-Jerić, Schmid, and Altman (2004) document its occurrence in publicly funded research as well. Such was the case in STAR*D.

ADDITIONAL EVIDENCE SUGGESTIVE OF BIAS BEYOND THE SCOPE OF PIGOTT ET AL.

Deception in Justifying the Use of the QIDS-SR

As stated previously, in the Pigott et al. peer-review process, a reviewer claimed that STAR*D's DSMB had authorized using the QIDS-SR as an outcome measure prior to "data lock and unblinding." To better understand what took place in this study, the author filed a Freedom of Information Act (FOIA) request for STAR*D's contract, the contract's research protocol, its statistical analytic plan, the minutes from all DSMB meetings, and the quarterly and annual progress reports submitted by the investigators to NIMH (Pigott, 2010a). The author received from NIMH the contract and research protocol, which included the analytic plan but not the DSMB meeting minutes or any quarterly or annual progress reports. The author was informed that these later documents could not be located and may have been destroyed (NIMH, 2010b). Despite not receiving all that was requested, the contract, research protocol, and analytic plan were very helpful in providing additional information in understanding STAR*D's original purpose, measures, methods, and planned analyses (NIMH, 2002).

In its published reports, STAR*D's authors never inform readers that the QIDS-SR was explicitly NOT intended to be used as a research measure. Instead, when stating the reasons for dropping the HRSD and using only the QIDS-SR to report the step-by-step acute and continuing-care remission and relapse rates in the summary article, its authors' state:

> We used the QIDS-SR as the primary measure to define outcomes for acute and follow-up phases because 1) QIDS-SR ratings were available for all participants at each acute treatment clinic visit, 2) QIDS-SR and HRSD outcomes are highly related, 3) the QIDS-SR was not used to make treatment decisions, which minimizes the potential for clinician bias, and 4) the QIDS-SR scores obtained from the interactive voice response system, the main follow-up outcome measure, and the paper-and-pencil QIDS-SR are virtually interchangeable, which allows us to use a similar metric to summarize the acute and follow-up phase results. (Rush, Trivedi, Wisniewski, Nierenberg, et al., 2006, p. 1908)

STAR*D authors' assertion that "the QIDS-SR was not used to make treatment decisions" is a half-truth at best. Although the QIDS-SR was not the sole or final measure for making treatment decisions, it was clearly part of STAR*D's 'clinical decision support system' as the authors themselves state in their step-1 article:

> 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 (ratings of frequency, intensity, and burden), medication adherence (self-report), and clinical judgment based on patient progress. (Trivedi, Rush, et al., 2006, p. 30)

Also in the controlled clinical trials article, the QIDS-SR is identified as one of several measures used "to provide consistent information to the clinicians who use this information in the protocol" (Rush, Fava, et al., 2004, p. 128). Furthermore, on pages 32 and 47 of the research protocol, the researchers explicitly distinguish between the research outcomes assessments that were conducted by ROAs blind to what treatment(s) the patient had received and the nonblinded assessments, such as the QIDS-SR that were collected at every clinic visit and used to guide care, stating, "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 CRCs" (pp. 47–48). STAR*D authors' assertion that "the QIDS-SR was not used to make treatment decisions" is highly deceptive since by their own admission it was part of STAR*D's clinical decision support system and *not* a research measure.

Second, although STAR*D's authors report that the "QIDS-SR and HRSD outcomes are highly related," the question is: Were patients' last-administered QIDS-SR "highly related" to the HRSD for those who dropped out and did not take the HRSD? STAR*D's claim is based on their research correlating the QIDS-SR and HRSD assessments in patients continuing in active treatment at the end of step 1 (Rush, Bernstein, et al., 2006). STAR*D provides no research supporting its use of the nonblinded QIDS-SR in the last visit to determine the remission status for those patients who dropped out without taking the blinded HRSD. STAR*D states that the reason for dropping out was not obtained for the "vast majority" of such patients (Rush, Trivedi, Wisniewski, Nierenberg, et al., 2006, p. 1908) and 24% (690 of 2,876) dropped out without taking the exit HRSD in step 1 alone.

Patients dropping out are by definition different from patients continuing in treatment in their estimation of the value of the free care that they are receiving. This is particularly true in STAR*D because the QIDS-SR was overseen by the CRC who also administered in every visit a clinician-interview version of this same tool (the QIDS-C) that had the identical 16 questions and response options as the QIDS-SR (in education, this practice would be considered an extreme example of teaching to the test). Furthermore, given that the CRC also provided patient education ensuring that patients understood the basic "mechanism of action" for their current antidepressant and educating the patient that "depression is a disease, like diabetes or high blood pressure" and "can be treated as effectively as other illnesses," the demand characteristics for patients to answer the QIDS-SR in a manner more consistent with the CRC's patient education than their actual experience was likely high for many patients who dropped out without informing the CRC of their plans to do so. STAR*D's authors were well aware of these demand characteristics and their potential to bias results in this open-label study, which is why in the research protocol, they emphasized, "Research outcomes assessments are not collected at the clinic visits. They are *not* collected by either clinicians or CRCs."

Finally, STAR*D states that the "QIDS-SR scores obtained from the IVR-system, the main follow-up outcome measure, and the paper-and-pencil QIDS-SR are virtually interchangeable, which allows us to use a similar metric to summarize the acute and follow-up phase results" (Rush, Trivedi, Wisniewski, Nierenberg, et al., 2006, p. 1908). Just as the QIDS-SR was NOT a research measure, the interactive voice response (IVR) version was NOT STAR*D's "main follow-up outcome measure." Although STAR*D's objective to use "a similar metric to summarize the acute and follow-up phase results" is admirable, such outcome metrics had already been both prespecified and collected—the blindly

administered HRSD and IDS-C30. STAR*D's authors chose not to report to readers its results as prespecified nor inform readers that they were not doing so. Instead, they resorted to deception to justify their use of the QIDS-SR, a sham measure, to report outcomes.

Bias by Omission

STAR*D lacked a control group in every phase despite its authors being fully aware that it was common for antidepressants to not differentiate from placebos in controlled trials. In an editorial, STAR*D researcher Michael Thase (2007) acknowledges that drug/placebo differences are often nonexistent in antidepressant drug trials, whereas Maurizio Fava estimated that placebo's true response rate in antidepressant trials was 35%–45% since most failed trials go unpublished (Fava, Evins, Dorer, & Schoenfeld, 2003). By not including a placebo in any phase, STAR*D's researchers and NIMH avoided the difficulty of explaining the likelihood that many of the compared drugs would not differentiate from placebo.

The research protocol states in its opening abstract using capital letters that STAR*D was a comparative EFFECTIVENESS study of different treatment options for people with major depression. In this regard, STAR*D was a well-designed study that had 12 prespecified research outcome measures and a detailed analytic plan for evaluating the effectiveness and cost-efficiency of 11 pharmacologically distinct drug treatments (along with cognitive psychotherapy) for depressed patients who failed to improve from their first antidepressant trial. These measures included preassessment/postassessment of depressive symptoms, level of functioning, patient satisfaction, quality of life, side-effect burden, health care utilization and cost of care, health status, work productivity, and personal income (NIMH, revised 2002 pages 48–51; Rush, Fava, et al., 2004, p. 127–131; Fava, et al., 2003, 476–479) as well as reassessing remitted patients every 3 months on these same measures during 12 months of free continuing care.

Despite it being 5 years since the completion of STAR*D's data collection and its authors having published more than 70 peer-reviewed articles on its findings, none of these articles have reported the pre/post mean change scores for any of the 12 prespecified measures nor reported its findings in a manner consistent with the analytic plan as presented in STAR*D's research protocol (e.g., see pp. 55–62 that describe STAR*D's plan for comparing the cost-effectiveness of the antidepressant treatments, including their impact on overall health care utilization and cost of care) and background articles (Fava, et al., 2003, 476–479; Rush, Fava, et al., 2004, p. 127–131, 135–136).

STAR*D's raison d'être was to compare treatments in their ability to both relieve depressive symptoms and improve patients' health status, functioning, and quality of life while also assessing the offsets to the cost of providing such drug care through reductions in healthcare utilization and costs.

Given the "substantial public health and scientific significance" of this information, it is deeply troubling that STAR*D's researchers have still not published these findings despite it being 4 years since the publication of the summary article and having ample funding to complete this effort. Instead, out of the more than 70 articles STAR*D's authors have published (NIMH, 2009), the vast majority have little relevance from either a treatment or health care policy perspective, and such studies' trivial nature is evidenced by their not being included in the analytic plan. This triviality is exemplified in STAR*D's most recently published study, this one on insomnia, that concludes "insomnia symptoms are very common, undertreated, and indicative of a more severe depression" (Sunderajan

et al., 2010, p. 394). This article fails to report any outcomes, not even the effectiveness of adding trazodone or a sedative to insomniac patients' antidepressant, which frequently occurred in this study (e.g., Rush, Trivedi, Wisniewski, Stewart, et al., 2006, Table 2). Therefore, $35 million later, we still do not know even if the common practice of adding one drug to another drug for depressed insomniac patients improves any outcome, only that insomnia is "indicative of a more severe depression."

STAR*D's failure to publish its findings as prespecified is highly suggestive that antidepressant drug care failed to deliver the wide range of positive outcomes and offsetting costs its authors and NIMH expected so they chose not to publish this damning data.

The author therefore filed a second FOIA request for all data analytic reports provided by STAR*D's investigators to NIMH as specified in the contract as well as any correspondence regarding modifications to the contract authorizing the investigators to deviate from the analytic plan (Pigott, 2010b). This request resulted in a conference call with George Niederehe, the government program officer responsible for overseeing all aspects of STAR*D, and Stephen Wisniewski, STAR*D's chief biostatistician. When asked for these analyses, the author was told that to their knowledge, they were never performed. Niederehe also stated that the publication of STAR*D's results constituted the investigators' fulfillment to NIMH of the contract's data analytic reporting requirements, and if such analyses have not been published, they were likely never conducted (Pigott, 2010c). Both then stated that STAR*D's dataset is now available to researchers to perform such analyses as well as any others deemed warranted (NIMH, 2010a). Additional highlights from this call include:

- Both denied that STAR*D's DSMB authorized using the QIDS-SR prior to "data lock and unblinding" as Pigott et al. had been informed by the reviewer. Instead, they stated that this decision occurred in the Communications Committee meetings for which notes were not taken as they were for the DSMB meetings.
- Wisniewski acknowledged that the QIDS-SR was not a prespecified research measure but stated that the reason this was not disclosed in the steps 1–4 and summary articles was that "there was not enough journal space" and noted that they did report the HRSD remission rates as the primary outcome in the steps 1–4 articles. Wisniewski's justification for not using the HRSD in the summary article was that it was not part of the main study, only a secondary analysis; therefore, it was not necessary to use the HRSD to report outcomes in this article.
- At the call's conclusion, Niederehe explained that in government research contracts such as STAR*D, it was common for changes to be made without documentation.

After the call, the author reread the contract and found additional contractually required reports from the contractor and therefore has filed a third FOIA requesting them (Pigott, 2010d).

Biased Interpretation of Results

Besides documenting antidepressant drugs' general lack of effectiveness even when optimally administered, STAR*D's findings failed to support its neurochemical imbalance theory of depression. STAR*D's hypothesis was that patients who failed to respond adequately during the prior step did so because the drug(s) did not produce the "right" neurochemical change. Therefore, "switching" to a new drug, or combination of new drugs, with a different neurochemical "mechanism of action," or "augmenting" the current

drug(s) with a new drug that has a different neurochemical action, might trigger the "right" change resulting in remission (Boren, 2007).

To develop the science necessary to guide such "next-step" decision-making, STAR*D carefully selected each new drug and drug combinations to be evaluated based on those found most promising in prior research while ensuring that each step's compared drugs had distinct pharmacological profiles (see Table 1). Surprisingly, there were no significant differences in the five next-step comparisons of 11 pharmacologically distinct drug treatments, even though the *N* per each specific treatment ranged from 51 to 286 patients and was therefore more than sufficient to identify meaningful differences if any existed.

Despite the lack of significant differences in all five comparisons and this fact's logical meaning, STAR*D's authors assert, "These results also have provocative theoretical implications. The findings are suggestive that major depressive disorder is biologically heterogeneous such that different treatments differ in the likelihood of achieving remission in different patients" and then go on to make the obligatory observation, "However, without a placebo control at each step and without substantial differences in remission rates among treatments in the same step, such a notion remains to be fully established" (Rush, Trivedi, Wisniewski, Nierenberg, et al., 2006, p. 1913).

Fundamentally, STAR*D's findings suggest no such "biologically heterogeneous" theory in which "different treatments differ in the likelihood of achieving remission in different patients." What is "provocative" is STAR*D's authors ignoring the obvious. Because of the similar remission rates in all five comparisons, the most parsimonious explanation is that it did not matter what drugs were prescribed because every compared drug or drug combination yielded about the same effect as every other drug or drug combination. To reconcile STAR*D's theory with their findings requires that in every comparison:

- Each class of biologically needed change was equally dispersed across treatments (a reasonable assumption given random assignment); and
- Each class's size was always essentially equivalent, allowing no statistically significant separation in outcomes between them (a highly dubious assumption).

For instance, it requires in step 2's switch comparison where the three drugs triggered an indistinguishable remission rate averaging 21.2% that:

- Those patients needing a selective serotonin reuptake inhibitor (SSRI) versus "dual-action" versus "out-of-class" agent were equally dispersed into the three compared strategies; and
- Each of the three compared groups' *N* was always approximately 21% of the total *N*, with an additional 37% needing an as-of-yet untested neurochemically derived change.

For STAR*D's theory to remain plausible, this scenario would have to then be repeated in all four other comparisons.

Second, step 2's switch comparison also contradicts this theory. In this comparison, sertraline (Zoloft), an SSRI neurochemically similar to step 1's SSRI citalopram (Celexa), was no different in effectiveness and tolerability as bupropion (Wellbutrin) and venlafaxine (Effexor), even though 56% of step 2's "switch" patients were found intolerant to citalopram (Celexa) in step 1 (Trivedi, Fava, et al., 2006, p. 1240).

Since patients who failed to gain a step-1 remission were by STAR*D's definition "SSRI medication resistant" (Trivedi, Rush, et al., 2006, p. 30), and 56% of step 2's switch patients were "citalopram intolerant," why was a similar SSRI as effective and tolerable in

step 2 for such highly "SSRI resistant and intolerant" patients as were bupropion (Wellbutrin) and venlafaxine (Effexor) with their different pharmacological profiles? This finding directly contradicts STAR*D's theory, yet was not even discussed in step 2's switch article nor the summary article.

Third, the fact that STAR*D allowed liberal prescribing of nonstudy drugs during every step precludes its authors' ability to assert anything regarding different drug treatments "achieving remission in different patients" because it is impossible to disentangle what were the unique effects of the compared drugs versus their coadministration with other nonstudy drugs. For example, in step 2's switch comparison, 17.1% of patients were also prescribed trazodone, 16.5% a sedative or hypnotic drug, and 11.8% an anxiolytic (Rush, Trivedi, Wisniewski, Stewart, et al., 2006, Table 2). Given this, the only "provocative" observation STAR*D's authors should have made is that even with such polypharmacy efforts, their outcomes were far less than expected.

A final fact contradicting STAR*D's theory is its staggering relapse rate such that by follow-up's 12-month conclusion, 94.2% of all "remitted" patients had relapsed and/or dropped out. If "different treatments differ in the likelihood of achieving remission in different patients," why the loss of efficacy for so many patients so quickly while continuing to take the same drug(s) that allegedly were the bases for their remissions? Again, STAR*D's authors never address this, or any of the other glaring inconsistencies to their "biologically heterogeneous" theory, yet instead claim support for that which their results disprove.

Small Biases Reflecting BIG BIAS

STAR*D's authors demonstrate a pattern of rounding up their findings, thereby inflating the reported remission and response rates. In the step-2 switch study in which the combined HRSD remission rate for the three switch medications was 21.2%, the authors state in the abstract's conclusion section that "approximately one in four patients had a remission of symptoms after switching to another antidepressant" (Rush, Trivedi, Wisniewski, Stewart, et al., 2006, p. 1231). Evidently, STAR*D's authors failed to realize that 21.2% is closer to "one in five" than "one in four" patients.

STAR*D's summary article's "Acute Treatment Outcomes by Treatment Step" Table 3 has rounding "errors" in steps 1–3 each time inflating STAR*D's reported remission and response rates by 0.1 to 0.2 points (Rush, Trivedi, Wisniewski, Nierenberg, et al., 2006, p. 1910). In step 1, STAR*D reports that 1,346 of 3,671 patients had a QIDS-defined remission, and STAR*D then reports this as a 36.8% remission rate. Actually, 1,346 divided by 3,671 is equal to 0.36665; therefore, STAR*D should have rounded it up to only 36.7% versus 36.8%. Similar rounding-up "errors" occurred in calculating steps 2 and 3's remission rates.

This same pattern occurs in STAR*D's reported response rates. In step 1, 1,776 of 3,671 patients had a QIDS-defined response, and STAR*D reports this as a 48.6% response rate when actually 1,776 divided by 3,671 is equal to 0.48379; therefore, it should have been rounded up to only 48.4% versus 48.6%. Again, similar rounding-up "errors" occurred in steps 2 and 3's reported response rates.

In this same table, STAR*D also calculated the step-by-step rates of intolerable side effects in the row immediately below those for remission and response. There were no rounding errors in these four calculations. Therefore, out of 12 calculations—8 of whom

reported antidepressants' remission and response rates—STAR*D had rounding-up errors in 75% (6 of 8) of those reporting its already inflated QIDS-SR rates and none in those reporting the intolerable side-effect rates. Although the inflationary effect of these rounding UP "errors" was trivial, the pattern's consistency speaks volumes.

In the summary article's result section, STAR*D's authors calculated a theoretical cumulative remission rate of 67% based on the inflated QIDS-SR whose findings they inflated further. In this article's discussion section, they assert that "the overall cumulative remission rate would approach 70% after four steps (if needed)" (Rush, Trivedi, Wisniewski, Nierenberg, et al., 2006, p. 1912). In an article published in the *Cleveland Clinic Journal of Medicine* targeting primary care physicians (PCPs), STAR*D's authors' final "Key Point" states, "With persistent and vigorous treatment, most patients will enter remission: about 33% after one step, 50% after two steps, 60% after three steps, and 70% after four steps (assuming patients stay in treatment)" (Gaynes et al., 2008, p. 57).

When it comes to reporting remission rates, STAR*D's progression is always UP. The highly inflated 67% becomes "approach 70%" and finally "70% after four steps," with no acknowledgment in the Cleveland Clinic article that this theoretical rate is based on the assumption that dropouts would have had the same remission rate as those who did not dropout nor acknowledge the often short-lived duration of said "remissions" nor acknowledge that this alleged "cumulative" rate was based on a sham measure.

Inflating the Extent of Improvement From Antidepressant Drug Treatment

STAR*D's authors inflated the extent of improvement that patients obtained from achieving remission. In the Cleveland Clinic article's "Key Points" section, the first point states, "Remission (i.e., complete relief from a depressive episode) rather than response (merely substantial improvement) should be the goal of treatment, as it is associated with a better prognosis and better function" (Gaynes et al., 2008, p. 57). This description of remission is similar to STAR*D's statement in its controlled clinical trials article that describes remission as "the complete absence of depressive symptoms" and then defined remission as a score of ≤7 on the HRSD (Rush, Fava, et al., 2004, p. 121).

Although an HRSD ≤7 score is a common criterion for classifying remission in depression research, such a score is by no means synonymous with "complete relief from a depressive episode" or "the complete absence of depressive symptoms" because patients could have up to seven HRSD symptoms mildly expressed and still met this criterion. For example, on the HRSD suicide question, "feels like life is not worth living" is scored as 1; "recurrent thoughts or wishes about death of self" is scored as 2; "active suicidal thoughts, threats, gestures" is scored as 3; and a recent "serious suicide attempt" is scored as 4. Similar progressions in severity are found on all of the questions such that for the guilt and delusions question, "feels he/she has let people down" is scored as 1; for the work and interest question, "feels incapable, listless, less efficient" is scored as 1; and for the libido question, "has decreased sexual drive and satisfaction" is scored as 1. A patient scoring 1 on just these four HRSD questions would be counted as remitted with three "mild" symptoms to spare, yet few would describe such a patient as experiencing "complete relief" from his or her depressive episode because each of these symptoms is used in diagnosing major depression.

As should be obvious, in the absence of STAR*D reporting on the quality of life, level of functioning, and side-effect burden research measures that they collected, no reader can

judge the extent of improvement that the drug treatments actually bought about for the "remitted" patients. We do know though that it was certainly not "complete relief" and a return to euthymia because patients could have up to seven depressive symptoms and still be counted as having obtained remission.

Apparent Bias by the *American Journal of Psychiatry*

The *American Journal of Psychiatry* (AJP) published five of STAR*D's seven outcome articles. Many of the apparent biases indentified in this paper should have been corrected through competent peer review. Examples include:

- Insisting that STAR*D's authors disclose the fact that all 4,041 patients were started on citalopram (Celexa) in their baseline visit. Even with careful reading, this fact is not disclosed in either the step-1 or the summary articles. Instead, the summary article's patient flowchart misleads readers into believing that only 3,671 patients were in the step-1 citalopram (Celexa) trial. This is because in the patient flowchart's top level is a box stating "Enrolled (N=4,041)," followed by a box to the side with "No postbaseline visit (N=370)," and the figure's "Level 1" treatment box stating "Citalopram (N=3,671)" (Rush, Trivedi, Wisniewski, Nierenberg, et al., 2006, Figure 1), leading to the false assumption by many that the citalopram (Celexa) trial consisted of only 3,671 patients. In evaluating antidepressants use with real-world patients under optimal conditions, it is critical for readers to know that 9.2% of depressed patients who consented to antidepressant care, had started STAR*D's multistep "depression-as-disease" model educational program, and begun free SSRI treatment dropped out without a follow-up visit.
- Insisting that STAR*D report its acute and continuing-care remission and relapse rates using the HRSD. The HRSD was identified as STAR*D's primary outcome measure in the steps 1–4 articles. By dropping the HRSD in its summary article, STAR*D significantly inflated its acute-care remission rates with unknown effects on its purported relapse rate. AJP reviewers should not have allowed this switch because simply reading the steps 1–4 abstracts stating the HRSD and QIDS-SR remission rates would have alerted them to its inflationary effects.
- Not allowing STAR*D's authors to falsely assert that "the QIDS-SR was not used to make treatment decisions" as justification for using it as the sole measure to report its summary findings. Again, simply reading AJP's step-1 article makes explicit that the QIDS-SR was part of STAR*D's "clinical decision support system."
- Insisting that STAR*D's authors exclude the 931 patients who were enrolled into the study without a ROA-administered baseline ≥14 HRSD or, at a minimum, disclosing these patients' enrollment in a straightforward manner. Instead, this disclosure was dispersed over three pages in the summary article. Again, simply comparing the step-1 AJP article's QIDS-SR remission rate of 32.8% to the step-1 QIDS-SR remission rate of 36.8% as reported in the summary article demonstrates the inflationary effects of this decision.
- Insisting that STAR*D's authors report in the summary article how to interpret the three survival analysis tables and discuss the survival analyses' astonishing findings that out of the 1,518 remitted patients, only 108 (7.1%) survived continuing care without relapsing and/or dropping out.

In that AJP published most of the STAR*D outcome studies, Pigott et al. submitted their paper to AJP and only after rejection to Psychotherapy and Psychosomatics. This submission included a detailed letter to Robert Freedman, the journal's editor, documenting multiple instances of bias in the AJP-published STAR*D articles that warranted correction to better inform readers. The letter highlighted how an unbiased presentation of STAR*D's findings discredits the American Psychiatric Association's (APA) continuation

phase guideline that "following remission, patients who have been treated with antidepressant medications in the acute phase should be maintained on these agents to prevent relapse" despite this recommendation having received its expert panel's highest "clinical confidence" rating (APA, 2000, p. 15).

The letter also offered to send Freedman the author's July 2008 e-mail exchanges with Wisniewski confirming the accuracy of the paper's analysis so that this information could be provided to AJP peer reviewers. Freedman never requested the Wisniewski e-mail exchanges. Pigott et al.'s AJP submission resulted in a form-letter rejection with no comment on the paper's substance or indication that it was sent out for peer review. AJP's refusal to allow an examination of apparent bias in its STAR*D publications combined with incompetent peer review suggests a disregard for open and honest science on its leadership's part.

Clear Bias by NIMH

STAR*D was an NIMH-initiated research contract starting in September 1999 and continuing for 7 years with additional follow-on studies that are still in progress today—not a mere research grant award in response to a request for proposals. NIMH was therefore intimately involved in STAR*D's oversight because of its expense and public health significance. This involvement included three NIMH employees being coauthors on four or more of the steps 1–4 and summary articles; two being research branch chiefs, Barry Lebowitz and George Niederehe who was also STAR*D's government program officer with ultimate oversight responsibility for the study. Louise Ritz, the third NIMH employee/coauthor, was also STAR*D's program officer with day-to-day oversight responsibilities.

It is troubling, with glaring conflicts of interest, when those charged with oversight in taxpayer-funded research are also included in the spoils of publication. These spoils in STAR*D were not inconsequential. Through their oversight roles, both Niederehe and Ritz became coauthors of articles in the *New England Journal of Medicine* (NEJM) and AJP; Niederehe coauthored one article in NEJM and six in AJP and Ritz coauthored two articles for each journal. Who was watching out for the public interest by ensuring the scientific integrity and proper reporting of findings in this study? Was it John Rush, STAR*D's principal investigator, who disclosed 19 conflicts of interests, the vast majority of which were with pharmaceutical companies, OR was it the federal employees included in prestigious publications? Who? Evidently not the other coauthors who averaged 9.8 disclosed conflicts each; ten of whom report receiving money from Forest Pharmaceuticals, the maker of citalopram/Celexa based on their conflict of interest disclosures in the steps 2-4 and summary articles. As noted previously, the summary article inflated this drug's remission rate by 44.9%.

NIMH was fully complicit in the biased reporting of STAR*D's results as evidenced by the contract provision that none of the study findings could be "released, presented at meetings, or published" without the prior "review and approval" of the Government Program Officer (NIMH STAR*D Contract, 1999, p. 21). This provision is similar to those commonly found in pharmaceutical-industry sponsored studies and renders meaningless the statement that "The content of this article does not necessarily reflect the views or policies of the Department of Health and Human Services" which appears at the end of each STAR*D article.

Rather than insisting on forthright and honest reporting in peer-reviewed articles, NIMH issued STAR*D press releases that are highly misleading to both professionals and

people with depression on the likely benefits from antidepressant drug care. Examples of this bias include:

- The title of NIMH's press release and web page summarizing the step-2 results published in NEJM is: "New Strategies Help Depressed Patients Become Symptom-Free" and includes a quote from NIMH's Director Thomas Insel stating:

  > If the first treatment attempt fails, patients should not give up. By remaining in treatment, and working closely with clinicians to tailor the most appropriate next steps, many patients may find the best single or combination *[drug]* treatment that will enable them to become symptom-free. (NIMH, 2006a, para. 5)

- The step-2 NIMH press release refers eight times to remitted patients becoming "symptom-free" and includes a quote from STAR*D author Madhukar Trivedi stating, "Augmenting the first medication may be an effective way for people with depression to become symptom-free" (NIMH, 2006a, para. 11).
- NIMH's press releases summarizing the step-3 and step-4 results continue to state that remitted patients became "symptom-free," repeating this false claim four times in the step-3 press release (NIMH, 2006d) and five more times in the step-4 press release (NIMH, 2006b). In the step-4 press release, NIMH adds to the misrepresentation by stating, "Over the course of all four levels, about 70 percent of those who did not withdraw from the study became symptom-free." (para. 4)
- NIMH's (2006c) summary results' press release states in the opening paragraph, "In STAR*D, the outcome measure was a 'remission' of depressive symptoms—becoming symptom-free" ("5. What were the results?" para. 1) and then repeats this false "symptom-free" claim 17 more times.

Both NIMH and STAR*D's authors were well aware that remitted patients' HRSD ≤7 score was not synonymous with them becoming "symptom-free," given the severity of symptoms such as "feels like life is not worth living" and "feels incapable, listless, less efficient" that are scored as only 1 on this measure. NIMH leadership's decision to repeatedly state in its press releases and Internet website that "almost 70%" of those patients who persisted through STAR*D's various drug trials became "symptom-free" makes medical claims for antidepressant drugs' level of effectiveness that are simply not true and suggests a profound pro-drug bias within this taxpayer-funded agency.

Additional evidence of NIMH's pro-drug bias is that instead of insisting that STAR*D's results be published as prespecified, NIMH rewarded STAR*D's lead investigators with new taxpayer funding to conduct the *Combining Medication to Enhance Outcomes of Depression* (CO-MED) study, evaluating polypharmacy drug treatments of depression (ScienceDaily, 2008b). Furthermore, these investigators secured an additional 1.2 million taxpayer dollars from the Agency for Healthcare Research and Quality to develop and evaluate a computerized version of their proprietary "measurement-based system" for ensuring "high-quality" antidepressant drug care (ScienceDaily, 2008a). Both of these taxpayer funded initiatives appear designed to foster further a drug-centric approach to treating depression.

DISCUSSION

This article identifies numerous instances of apparent bias by STAR*D's authors as well as AJP's and NIMH's leadership. This bias significantly inflated the alleged benefits of

antidepressant drug treatment while not disclosing the staggering relapse and/or dropout rate that occurred in those patients who initially responded favorably.

In contrast to the STAR*D authors' and NIMH's false representations Pigott et al.'s analysis found that of the 4,041 patients initially started on citalopram (Celexa), only 1,518 patients (37.6%) obtained remission after up to four medication trials and entered STAR*D's free continuing care. In every drug trial, more patients dropped out than were remitted and this dropout rate increased throughout the study. Of these 1,518 remitted patients, only 108 (7.1%) survived continuing care without relapsing and/or dropping out. Moreover, it is not known how many of these few patients were one of the 607 patients whose baseline <14 HRSD signified at most only mild symptoms when first started on citalopram (Celexa) and therefore had to score worse during continuing care than when they first entered the study to be counted as relapsed, nor how many actually remained "in remission" during continuing care. This reality directly counters NIMH leaderhip's false claim that "about 70 percent of those who did not withdraw from the study became symptom-free."

In a 2009 article, Insel acknowledges the severe inadequacy of the neurochemical imbalance theory of mental disorders to advance treatment outcomes and later states that there is "no evidence that the morbidity or mortality of mental disorders has dropped substantially in the past decades" despite the increased use of second-generation antidepressant and antipsychotic drugs to treat them with these drugs having "a combined market of $25 billion" in 2007 in the United States alone (Insel, 2009, pp. 701, 703). Rather than not "substantially" dropping though, the morbidity and chronicity of mental disorders appears to be increasing with a twofold to threefold increase between 1987 and 2007 in the number of Americans receiving disability payments for such disorders (Whitaker, 2005, 2010). Although there are certainly multiple factors affecting this increased disability rate, the fact that this dramatic increase has occurred during the same time as the dramatic increased use of second-generation psychotropic drugs to treat these disorders—combined with the emergence of APA's continuation phase treatment guidelines calling for essentially the open-ended use of same—demands serious investigation and not simply dismissing Whitaker's investigative journalism out of hand because it is counter to accepted wisdom.

In the same article, Insel observes that in every large comparative effectiveness study of second-generation drugs for depression, schizophrenia, and bipolar disorder, these drugs have repeatedly been found to be no better than their first-generation cousins from the 1950s, 1960s, and 1970s despite their added costs (and this reality is despite the tens of billions spent over decades in public and private research efforts to make improvements in said drugs). In this same section, Insel also acknowledges that even "after 14 weeks of optimal treatment with the second-generation medication citalopram," STAR*D's step-1 success rate was no different from that commonly found by placebos in controlled trials. Insel then goes on to state, "The unfortunate reality is that current medications help too few people to get better and very few people to get well" (Insel, 2009, pp. 703–704).

Insel's belated acknowledgment of the dismal outcomes from psychotropic drug care is a far cry from his 2006 claims regarding STAR*D's results. This author searched in vain on NIMH's website to find anything indicative of this more sober assessment of psychotropic drugs effectiveness, yet instead found information more in common with the pharmaceutical industry's marketing efforts than an unbiased taxpayer-funded research institute committed to fostering sound science to advance the treatment of mental disorders.

By failing to insist on an accurate reporting in 2006 of STAR*D's comparative effectiveness findings as specified in this 35-million-dollar contract—and instead repeatedly making false claims of the alleged "symptom-free" benefits achieved through try-try-try-and-try-again drug care—NIMH's leadership subverted and delayed an honest reappraisal of antidepressant drugs' appropriate role in the treatment of depression.

NIMH's long-standing pro-drug bias as documented by Whitaker has had profound public health consequences (Whitaker, 2010). Each year, major depression disorder affects approximately 6.7% of American adults (Kessler, Chiu, Demler, & Walters, 2005), with costs exceeding $80 billion per year, two thirds of which are caused by the disability and workplace-related costs that are associated with it (Greenberg et al., 2003; Kessler, et al., 2006). This reality is not caused by a lack of pharmaceutical efforts. Between 1996 and 2005, Americans' prescribed antidepressant drugs almost doubled, increasing from 5.8% to 10.1% of all those 6 years of age or older, with a concurrent significant increase in the average number of antidepressant drug prescriptions filled per patient/per year from 5.6 to 6.93 (Olfson & Marcus, 2009). During this same period, psychiatrists significantly increased their use of polypharmacy such that outpatient visits resulting in two or more prescribed psychotropic drugs increased from 42.6% in 1996 to 59.8% in 2005 and psychiatry visits resulting in three or more such drugs being prescribed doubled, increasing from 16.9% to 33.2% (Mojtabai & Olfson, 2010). This increased polypharmacy has occurred despite the fact that meta-analyses of antidepressant drug augmentation strategies have found significant increased risk of adverse events with only exceptionally modest (if any) added benefit from such polypharmacy efforts whether the augmentation be with another antidepressant or an atypical antipsychotic drug (Nelson & Papakostas, 2009; Yury, Fisher, Antonuccio, Valenstein, & Matuszak, 2009).

Antidepressant drugs' failure to promote sustained recovery is not something newly discovered by Pigott et al.'s STAR*D reanalysis. As Pigott et al. noted, although APA's continuation phase guideline recommending open-ended use of antidepressant drugs is consistent with meta-analyses reporting large effect sizes for them in preventing relapse (e.g., Geddes et al., 2003), these analyses do not control for publication bias nor selective outcome reporting, both of whom significantly inflate the report of positive findings.

Instead, despite the APA's highest "clinical confidence" rating for this guideline, prospective studies have documented excessively high relapse and/or dropout rates resulting from the continued use of these drugs. For example, Rush, Trivedi, et al. (2004) found a sustained remission rate of only 5.1% over 12 months in 118 depressed patients treated by following a decision "algorithm" to guide drug treatment and providing patient/family education teaching the importance of these drugs to treat depression. Even this 5.1% rate though is an overestimation because the study used last observation carried forward (LOCF) analysis such that patients who dropped out before study completion, but had not scored as relapsed in their last assessment, were counted as having achieved a sustained remission; evidently, LOCF enthusiasts are unaware that depressed patients commonly discontinue drug treatment when their once helpful drug stops working and their depression returns. In 2008, Bockting et al. reported the results of 172 patients with recurrent depression and found that only 42% used antidepressants continuously during 2 years of continuation phase treatment of which 60.4% relapsed while taking these drugs, whereas patients who stopped using them experienced significantly less relapse, with only 8% of those who received preventive cognitive therapy relapsing. In STAR*D, even for step 1's 1,085 remitted patients,

only 84 (7.7%) did not relapse and/or dropout during continuing care (see Table 2), and these few patients had the greatest likelihood of achieving a sustained remission.

The concept that continuation phase antidepressant drug treatment may increase versus decrease depression's chronicity and likelihood of relapse is not new. While endorsing the use of antidepressants during acute-care treatment for major depressive disorder, in 1994, Giovanni Fava proposed that long-term use of antidepressant and antianxiety drugs increases the neurobiological vulnerability of some patients to affective disorders while decreasing the likelihood of positive response to new drug treatments when symptoms reemerge (Fava, 1994). In a 2003 systematic review, Fava found evidence suggesting very poor long-term outcomes from continuation phase antidepressant drug treatment and that in some patients, such continued drug use appears to foster changes that counter the drug's initial beneficial effects resulting in (1) loss of efficacy with symptom worsening, (2) increased risk of relapse during drug withdrawal, (3) drug-induced switching to mania and rapid mood cycling in bipolar patients, and (4) diminished responsiveness to subsequent drug treatment (Fava, 2003). Fava termed this hypothesized phenomena *oppositional tolerance*. In an updated 2010 systematic review, Fava and Offidani found additional support for this hypothesis and conclude by stating

> When we prolong treatment over 6–9 months we may recruit processes that oppose the initial acute effects of antidepressant drugs (loss of clinical effects). We may also propel the illness to a malignant and treatment-unresponsive course that may take the form of resistance or episode acceleration. When drug treatment ends, these processes may be unopposed and yield withdrawal symptoms and increased vulnerability to relapse. Such processes are not necessarily reversible. The more we switch or potentiate antidepressant drugs the more likely is oppositional tolerance to take place. (Fava & Offidani, 2010)

Fava's oppositional tolerance hypothesis is supported by both the dismal continuing-care findings from antidepressant drug–effectiveness studies and the apparent 1987-to-2007 increase in disability caused by mental disorders that is correlated with the increased usage of these drugs. Viewed from the perspective of oppositional tolerance, there is little wonder why STAR*D's "try-try-try-and-try-again" approach to care yielded such poor "real-world" results and highlights the folly of relying on drug efficacy trials to guide practice because of the selective publication and outcome reporting biases that plague this body of research.

In addition to the risk of oppositional tolerance from continued care on these drugs, STAR*D found that 8.6% of step-1 patients reported increased suicidal ideation while taking citalopram (Celexa) during acute phase treatment (Perlis et al., 2007). Furthermore, in another article STAR*D reported that 71.3% of those who had a remission during acute-care treatment reported increased weight gain while taking citalopram (Celexa) and 71.7% reported residual symptoms of sleep disturbance despite having achieved "remission" status (Nierenberg et al., 2010). The increased weight gain during acute-care SSRI drug treatment is of particular significance because a recent study of 165,958 depressed patients found that long-term treatment with SSRIs doubled their risk of developing diabetes (Andersohn, Schade, Suissa, & Garbe, 2009).

In light of the modest to no drug/placebo advantage for antidepressants in the now five meta-analyses free from publication bias, combined with the substantial adverse risks from using these drugs, it is hard to find any reason for their first-line "come-one-come-all" use to treat depression other than convention, ease to prescribe for physicians, and the success of pharmaceutical

companies' relentless marketing efforts (Barbui, Furukawa, & Cipriani, 2008; Eyding et al., 2010; Kirsch, Moore, Scoboria, & Nicholls, 2002; Kirsch et al., 2008; Turner et al., 2008).

As emphasized in Pigott et al., APA's continuation phase guideline is profoundly misguided because there is no apparent benefit for most patients from continued antidepressant drug treatment, yet this "evidence-based" practice unnecessarily exposes such patients to significant adverse risks.

Unfortunately, APA's depression guidelines have been widely adopted by American health plans, and the National Committee for Quality Assurance (NCQA) has incorporated APA's acute and continuation phase guidelines into its measurement system to use in determining each plan's accreditation by tracking the percentage of newly diagnosed depressed adults who are treated with an antidepressant drug and remain on an antidepressant drug for at least 6 months (NCQA, 2009). To gain (and maintain) NCQA accreditation, health insurance plans are evaluated on their success in keeping depressed patients taking their antidepressant drugs, with such plans, in turn, increasingly using this same metric to evaluate and grade physicians.

Furthermore, under the Affordable Care Act all health insurance policies must begin providing 100% reimbursement for depression screening by PCPs with no co-pay required by patients as part of this act's mandated preventative services (Affordable Care Act, 2010). Although, in theory, more widespread identification of depressed patients may be beneficial, in practice, PCPs already prescribe most antidepressant drugs and this will only increase further, thereby channeling even more patients into open-ended drug treatment because of NCQA's accreditation requirements. The potential adverse public health consequences from the act's mandatory depression screening coverage are significant because of the very low depression threshold that is commonly applied when prescribing antidepressants (Zimmerman et al., 2002) particularly in the prescribing of these drugs by PCPs (Mojtabai & Olfson, 2008).

"First, do no harm" has become an abandoned concept when treating depression because of the APA and NIMH leaderships' long-standing pro-drug biases. These biases have resulted in misguided public health policy as evidenced by NCQA's Orwellian-like metric and now the U.S. government's depression screening coverage mandate.

Until such misguided efforts and pro-drug biases are corrected, America will likely see continuing increases in the chronicity and disability caused by depression and other mental disorders. It is far past time for an honest reappraisal of antidepressant drugs' role in the treatment of depression. Critical to this reappraisal is an analysis of STAR*D's comparative effectiveness dataset by independent researchers according to STAR*D's prespecified outcome measures and analytic plan. People with depression and their families, the public interest, and honest science deserve no less because of STAR*D's "substantial public health and scientific significance" once its findings are accurately analyzed and reported.

REFERENCES

Affordable Care Act. (2010). *Preventive services covered under the Affordable Care Act*. Retrieved November 12, 2010, from http://www.healthcare.gov/law/about/provisions/services/lists.html

American Psychiatric Association. (2000). *Practice guideline for the treatment of patients with major depressive disorder* (2nd ed.). Washington, DC: Author.

Andersohn, F., Schade, R., Suissa, S., & Garbe, E. (2009). Long-term use of antidepressants for depressive disorders and the risk of diabetes mellitus. *The American Journal of Psychiatry*, *166*(5), 591–598.

Barbui, C., Furukawa, T. A., & Cipriani, A. (2008). Effectiveness of paroxetine in the treatment of acute major depression in adults: A systematic re-examination of published and unpublished data from randomized trials. *Canada Medical Association Journal*, *178*(3), 296–305.

Bockting, C. L. H., ten Doesschate, M. C., Spijker, J., Spinhoven, P., Koeter, M. W., & Schene, A. H. (2008). Continuation and maintenance use of antidepressants in recurrent depression. *Psychotherapy and Psychosomatics*, *77*(1), 17–26.

Boren, J. (2007). The effectiveness of antidepressant medications: Results from a major new study. *The Behavior Therapist*, *30*, 96–98.

Chan, A. W., Krleza-Jerić, K., Schmid, I., & Altman, D. G. (2004). Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *Canadian Medical Association Journal*, *171*(7), 735–740.

Davidson, J. R. T., Gadde, K. M., Fairbank, J. A., Krishnan, R. R., Califf, R. M., & Binanay, C., et al. (2002). Effect of Hypericum perforatum (St John's Wort) in major depressive disorder: A randomized controlled trial. *The Journal of the American Medical Association*, *287*(14), 1807–1814.

Eyding, D., Lelgemann, M., Grouven, U., Harter, M., Kromp, M., Kaiser, T., et al. (2010). Reboxetine for acute treatment of major depression: Systematic review and meta-analysis of published and unpublished placebo and selective serotonin reuptake inhibitor controlled trials. *British Medical Journal*, *341*, c4737. doi:10.1136/bmj.c4737.

Fava, G. A. (1994). Do antidepressant and antianxiety drugs increase chronicity in affective disorders? *Psychotherapy and Psychosomatics*, *61*(3–4), 125–131.

Fava, G. A. (2003). Can long-term treatment with antidepressant drugs worsen the course of depression? *The Journal of Clinical Psychiatry*, *64*(2), 123–133.

Fava, G. A., & Offidani, E. (2010). The mechanisms of tolerance in antidepressant action. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*. Prepublication accessed online at doi:10.1016/j.pnpbp.2010.07.026

Fava, M., Evins, A. E., Dorer, D. J., & Schoenfeld, D. A. (2003). The problem of the placebo response in clinical trials for psychiatric disorders: Culprits, possible remedies, and a novel study design approach. *Psychotherapy and Psychosomatics*, *72*(3), 115–127.

Fava, M., Rush, A. J., Trivedi, M. H., Nierenberg, A. A., Thase, M. E., Sackeim, H. A., et al. (2003). Background and rationale for the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study. *Psychiatric Clinics of North America*, *26*(2), 457–494.

Fava, M., Rush, A. J., Wisniewski, S. R., Nierenber, A. A., Alpert, J. E., McGrath, P. J., et al. (2006). A comparison of mirtazapine and nortriptyline following two consecutive failed medication treatments for depressed outpatients: A STAR*D report. *American Journal of Psychiatry*, *163*, 1161–1172.

Gaynes, B. N., Rush, A. J., Trivedi, M. H., Wisniewski, S. R., Spencer, D., & Fava, M. (2008). The STAR*D study: Treating depression in the real world. *Cleveland Clinic Journal of Medicine*, *75*(1), 57–66.

Geddes, J. R., Carney, S. M., Davies, C., Furukawa, T. A., Kupfer, D. J., Frank, E., et al. (2003). Relapse prevention with antidepressant drug treatment in depressive disorders: A systematic review. *Lancet*, *361*(9358), 653–661.

Greenberg, P. E., Kessler, R. C., Birnbaum, H. G., Leong, S. A., Lowe, S. W., Beglund, P. A., et al. (2003). The economic burden of depression in the United States: How did it change between 1990 and 2000? *The Journal of Clinical Psychiatry*, *64*(12), 1465–1475.

Insel, T. R. (2009). Disruptive insights in psychiatry: Transforming a clinical discipline. *Journal of Clinical Investigation*, *119*(4), 700–705.

Kerr, N. L. (1998). HARKing: hypothesizing after the results are known. *Personality and Social Psychology Review*, *2*(3), 196–217.

Kessler, R. C., Akiskal, H. S., Ames, M., Birnbaum, H., Greenberg, P. A., Hirschfeld, R. M., et al. (2006). Prevalence and effects of mood disorders on work performance in a nationally representative sample of U.S. workers. *The American Journal of Psychiatry*, *163*(9), 1561–1568.

Kessler, R. C., Chiu, W. T., Demler, O., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication (NCS-R). *Archives of General Psychiatry*, *62*(6), 617–627.

Kirsch, I., Deacon, B. J., Huedo-Medina, T. B., Scoboria, A., Moore, T. J., & Johnson, B. T. (2008). Initial severity and antidepressant benefits: A meta-analysis of data submitted to the Food and Drug Administration. *PLoS Medicine*, *5*(2), e45.

Kirsch, I., Moore, T. J., Scoboria, A., & Nicholls, S. S. (2002). The emperor's new drugs: An analysis of antidepressant medication data submitted to the U. S. Food and Drug Administration. *Prevention & Treatment*, *5*, Article 23.

Mathieu, S., Boutron, I., Moher, D., Altman, D. G., & Ravaud, P. (2009). Comparison of registered and published primary outcomes in randomized controlled trials. *The Journal of the American Medical Association*, *302*(9), 977–984.

Mojtabai, R., & Olfson, M. (2008). National patterns in antidepressant treatment by psychiatrists and general medical providers: Results from the national comorbidity survey replication. *Journal of Clinical Psychiatry*, 69(7), 1064–1074.

Mojtabai, R., & Olfson, M. (2010). National trends in psychotropic medication polypharmacy in office-based psychiatry. *Archives of General Psychiatry*, *67*(1), 26–36.

National Committee for Quality Assurance. (2009). *Antidepressant medication management (effective continuation phase treatment): Percentage of members who were diagnosed with a new episode of major depression, treated with antidepressant medication, and who remained on an antidepressant medication for at least 180 days (6 months)*. Retrieved November 12, 2010, from http://www.qualitymeasures.ahrq.gov/content.aspx?id=14961

National Institute of Mental Health. (2002). *Sequenced Treatment Alternatives to Relieve Depression (STAR*D) contract protocol* (Rev. ed.). Retrieved October 1, 2010, from Freedom of Information Act.

National Institute of Mental Health. (2006a). *New strategies help depressed patients become symptom-free*. Retrieved July 5, 2010, from http://www.nimh.nih.gov/science-news/2006/new-strategies-help-depressed-patients-become-symptom-free.shtml

National Institute of Mental Health. (2006b). *Odds of beating depression diminish as additional treatment strategies are needed*. Retrieved July 5, 2010, from http://www.nimh.nih.gov/science-news/2006/odds-of-beating-depression-diminish-as-additional-treatment-strategies-are-needed.shtml

National Institute of Mental Health. (2006c). *Questions and answers about the NIMH Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study—All medication levels*. Retrieved July 5, 2010, from http://www.nimh.nih.gov/trials/practical/stard/allmedicationlevels.shtml

National Institute of Mental Health. (2006d). *Switching to a third antidepressant medication may prove helpful to some with treatment-resistant depression*. Retrieved July 5, 2010, from http://www.nimh.nih.gov/science-news/2006/switching-to-a-third-antidepressant-medication-may-prove-helpful-to-some-with-treatment-resistant-depression.shtml

National Institute of Mental Health. (2009). *Sequenced Treatment Alternatives to Relieve Depression (STAR*D)*. Retrieved July 5, 2010, from http://www.clinicaltrials.gov/ct/show/NCT00021528?order=1

National Institute of Mental Health. (2010a, September 29). [Letter in response to FOIA] *case # 37933*.

National Institute of Mental Health. (2010b). *Available limited access datasets from NIMH clinical trials; Sequenced Treatment Alternatives to Relieve Depression (STAR*D)*. Retrieved December 3, 2010, from http://www.nimh.nih.gov/trials/datasets/nimh-procedures-for-requesting-data-sets.shtml

*National Institute of Mental Health STAR*D Contract # N01MH90003*. (1999, September 29). Obtained through the Freedom of Information Act, October 1, 2010.

Nelson, J. C., & Papakostas, G. I. (2009). Atypical antipsychotic augmentation in major depressive disorder: A meta-analysis of placebo-controlled randomized trials. *The American Journal of Psychiatry*, *166*(9), 980–991.

Nierenberg, A. A., Husain, M. M., Trivedi, M. H., Fava, M., Warden, D., Wisniewski S. R., et al. (2010). Residual symptoms after remission of major depressive disorder with citalopram and risk of relapse: A STAR*D report. *Psychological Medicine*, *40*(1), 41–50.

Olfson, M., & Marcus, S. C. (2009). National patterns in antidepressant medication treatment. *Archives of General Psychiatry*, *66*(8), 848–856.

O'Neal, B., & Biggs, M. (2001). *STAR*D patient education manual*. Retrieved July 5, 2010, from http://www.edc.pitt.edu/stard/public/study_manuals.html

Perlis, R. H., Purcell, S., Fava, M., Fagerness, J., Rush, A. J., Trivedi, M. H., et al. (2007). Association between treatment-emergent suicidal ideation with citalopram and polymorphisms near cyclic adenosine monophosphate response element binding protein in the STAR*D study. *Archives of General Psychiatry*, *64*(6), 689–697.

Pigott, H. E. (2010a, August 17). [FOIA letter requesting STAR*D's Contract Protocol, Statistical Analytic Plan, minutes from Data Safety and Monitoring Board meetings, and the quarterly and annual reports filed by the investigators as part of contract] *N01 MH-90003*.

Pigott, H. E. (2010b, October 14). [FOIA letter requesting all data analytic reports provided by STAR*D's investigators to NIMH as part of contract] *N01 MH-90003*.

Pigott, H. E. (2010c, December 3). *Notes from December 3rd 2010 conference call with George Niederehe, government program officer responsible for STAR*D, Stephen Wisniewski, STAR*D chief biostatistician, and Lisa Alberts, NIMH Freedom of Information coordinator*.

Pigott, H. E. (2011d, January 12). *FOIA letter requesting for contract N01 MH-90003 the contractually mandated (pages 19–20): (1) quarterly reports to the Data Safety and Monitoring Board; (2) the annual reports; and (3) the Study Final Report including all "statistical analyses performed in text, tabular, and graphical form."*

Pigott, H. E., Leventhal, A. M., Alter, G. S., & Boren, J. J. (2010). Efficacy and effectiveness of antidepressants: Current status of research. *Psychotherapy and Psychosomatics*, *79*(5), 267–279.

Rising, K., Bacchetti, P., & Bero, L. (2008). Reporting bias in drug trials submitted to the Food and Drug Administration: A review of publication and presentation. *PLoS Medicine*, *5*(11), e217.

Rush, A. J., Bernstein, I. H., Trivedi, M. H., Carmody, T. J., Wisniewski, S. R., Mundt, J. C., et al. (2006). 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. *Biological Psychiatry*, *59*(6), 493–501.

Rush, A. J, Fava, M., Wisniewski, S. R., Lavori, P. W., Trivedi, M. H., Sackeim, H. A., et al. (2004). Sequenced treatment alternatives to relieve depression (STAR*D): Rationale and design. *Controlled Clinical Trials*, *25*(1), 119–142.

Rush, A. J., Trivedi, M. H., Carmody, T. J., Biggs, M. M., Shores-Wilson, K., Ibrahim, H., et al. (2004). One-year clinical outcomes of depressed public sector outpatients: A benchmark for subsequent studies. *Biological Psychiatry*, *56*(1), 46–53.

Rush A. J., Trivedi, M. H., Wisniewski S. R., Nierenberg A. A., Stewart J. W., Warden D., et al. (2006). Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: A STAR*D report. *The American Journal of Psychiatry*, *163*(11), 1905–1917.

Rush, A. J., Trivedi, M. H., Wisniewski, S. R., Stewart, J. W., Nierenberg, A. A., Thase M., et al. (2006). Bupropion-SR, sertraline, or venlafaxine-XR after failure of SSRIs for depression. *The New England Journal of Medicine*, *354*(12), 1231–1242.

ScienceDaily. (2008a, June 16). *Groundbreaking depression research being tested in real-world setting*. Retrieved November 13, 2010, from http://www.sciencedaily.com/releases/2008/06/080612070402.htm

ScienceDaily. (2008b, July 22). *Testing multiple medication treatment of depression*. Retrieved November 13, 2010, from http://www.sciencedaily.com/releases/2008/07/080722072021.htm

Sunderajan, P., Gaynes, B. N., Wisniewski, S. R., Miyahara, S., Fava, M., Akingbala, F., et al. (2010). Insomnia in patients with depression: A STAR*D report. *CNS Spectrums*, *15*(6), 394–404.

Thase, M. E. (2007). Two new sections offer further checks and balances: Negative and failed clinical trial reports and invited commentary. *Psychopharmacology Bulletin*, *40*, 5.

Trivedi, M. H., Fava, M., Wisniewski, S., Thase, M. E., Quitkin, F., Warden, D., et al. (2006). Medication augmentation after the failure of SSRIs for depression. *The New England Journal of Medicine*, *354*, 1243–1252.

Trivedi, M. H., Rush, A. J., Wisniewski, S. R., Nierenberg, A. A., Warden, D., Ritz, L., et al. (2006). Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: Implications for clinical practice. *The American Journal of Psychiatry*, *163*(1), 28–40.

Trivedi, M. H., Stegman, D., Rush, A. J., Wisniewski, S. R., & Nierenberg, A. A. (2002). *STAR*D clinical procedures manual*. Retrieved August 1, 2010, from http://www.edc.pitt.edu/stard/public/study_manuals.html

Turner, E. H., Matthews, A. M., Linardatos, E., Tell, R. A., & Rosenthal, R. (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. *The New England Journal of Medicine*, *358*(3), 252–260.

Vedula, S. S., Bero, L., Scherer, R. W., & Dickersin, K. (2009). Outcome reporting in industry-sponsored trials of gabapentin for off-label use. *The New England Journal of Medicine*, *361*(20), 1963–1971.

Whitaker, R. (2005). Anatomy of an epidemic: Psychiatric drugs and the astonishing rise of mental illness in America. *Ethical Human Psychology and Psychiatry*, *7*(1), 23–35.

Whitaker, R. (2010). *Anatomy of an epidemic: Magic bullets, psychiatric drugs, and the astonishing rise of mental illness in America*. New York: Crown Publishers.

Yury, C. A., Fisher, J. E., Antonuccio, D. O., Valenstein, M., & Matuszak, J. (2009). Meta-analysis of antidepressant augmentation: Piling on in the absence of evidence. *Ethical Human Psychology and Psychiatry*, *11*(3), 171–182.

Zimmerman, M., Mattia, J. I., & Posternak, M. A. (2002). Are subjects in pharmacological treatment trials of depression representative of patients in routine clinical practice? *The American Journal of Psychiatry*, *159*(3), 469–473.

Acknowledgments. The author would like to thank Allan Leventhal for his helpful comments on an early draft of this article, and Robert Whitaker and the peer reviewers whose suggestions were invaluable in improving the final article.

Conflicts of Interest. H. Edmund Pigott is a partner in NeuroAdvantage, LLC, a neurotherapy company. In 2010, Ed Pigott has consulted for Amen Clinics, Brain Resources, CNS Response, EEG Spectrum International, International Society of Neurofeedback and Research, and Neuronetics.

Correspondence regarding this article should be directed to H. Edmund Pigott, PhD, 7111 Moorland Drive, Clarksville, MD 21029. E-mail: pathware@erols.com