Pigott HE. The STAR*D trial: it is time to reexamine the clinical beliefs that guide the treatment of major depression. Can J Psychiatry. 2015;60(1):9–13.

eTable 1
Indicators of Bias Previously Documented

Indicator	Description
Changed Outcome Measures Following Data Collection	As designed, STAR*D's prespecified primary measure was the HRSD for identifying "remitted" patients (i.e., those with a <8 HSRD score) and "responder" patients (i.e., those with a 50% or greater reduction in depressive symptoms). The HRSD was obtained in phone interviews by research outcome assessors blind to treatment assignment at entry into and exit from each trial and every 3 months during the 12 months of follow-up. After completing data collection, the STAR*D investigators decided to use the clinic-administered QIDS-SR to report outcomes. The QIDS-SR was not originally intended as a research measure, but rather one of STAR*D's "clinical management tools" that was administered by non-blinded clinical research coordinators (CRSs) and used to guide care. The primary source documents excluded all clinic-administered assessments such as the QIDS-SR from use as research measures. For example the STAR*D Research Protocol states, "Recall that the research outcomes assessments are distinguished from assessments conducted at clinic visits. The latter are designed to collect information that guides clinicians in the implementation of the treatment protocol. Research outcomes assessments are not collected at the clinic visits. They are **not** collected by either clinicians or CRCs" (emphasis in the original).[9,p47-48] The STAR*D researchers used the clinic-administered QIDS-SR as the secondary measure to report patient remission rates and sole measure to report response rates in the six published steps 1-4 articles as well as the sole measure to report remission and response rates in their summary article. By using the non-blinded QIDS-SR to report outcomes, STAR*D's published remission and response rates were inflated.
Used Data from Ineligible Patients Without Disclosure	The STAR*D investigators stated in their step1 article that patients had to have a baseline HRSD score of 14 or greater to be included in data analysis (see figure 1).[1] The investigators however changed the eligibility for analysis criteria in the published results for steps 2–4 without informing readers. These changes resulted in the inclusion of 607 patients who were initially excluded in the step-1 article because their score of <14 on the baseline HRSD signified at most only mild depressive symptoms when entering the study. Similarly, an additional 324 patients who were initially reported as excluded because they lacked a baseline HRSD were also subsequently included. Thus, 931 of STAR*D's 4,041 patients (23%) did not meet STAR*D's eligibility for analysis criteria but were included in the published results of the steps 2–4 and summary articles. Including these 931 patients, 607 of whom had at most only mild depression, inflated the published results.
Failed to Disclose that All Patients Were Started on	The STAR*D investigators failed to disclose that all 4,041 patients were started on citalopram in their initial baseline visit and that they excluded from analysis the 370 patients who dropped out without any subsequent visits, although the

Citalopram in their Baseline Visit	step1 article states, "our primary analyses classified patients with missing exit HRSD scores as nonremitters a priori."[1] These early dropout patients did not take the exit HRSD and therefore should have been counted as treatment failures as prespecified. By excluding these patients, STAR*D's published remission and response rates were inflated.
Failed to Disclose How to Interpret Survival Data for Follow-up	In the summary article, the STAR*D investigators did not disclose how to interpret the quarter-by-quarter survival data for the 12 months of follow-up care. This obscured from readers the fact that only 108 of the 4,041 enrolled patients (2.7%) had a QIDS-SR determined remission after up to 4 rounds of antidepressant drug care and neither relapsed nor dropped out as evidenced by taking at least one of the months 10-to-12 QIDS-IVR assessments. It is not known how many of these 108 survivors were one of the 607 patients who, due to a change in eligibility criteria, were allowed into the study despite having a baseline HRSD score of <14 signifying at most only mild symptoms when first started on citalopram and therefore who had to score worse during follow-up than when they first entered the study to be counted as relapsed. Nor is it known how many of the 108 patients actually remained 'in remission' during follow-up care.
Made Deceptive Statements Justifying the Use of the QIDS-SR	The STAR*D investigators made deceptive statements in the summary article to justify dropping the HRSD and instead using only the QIDS-SR to report outcomes. These misrepresentations included falsely stating that "the QIDS-SR was not used to make treatment decisions"[5,p1908] despite the fact that this assertion is contradicted by the researchers themselves when they wrote in the 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" (emphasis added)[1.p30] as well as being contradicted in all primary source documents.
Inflated the Extent of Improvement From Antidepressant Drug Care	The first 'Key Point' in the CCJM article states, "Remission (ie, 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" (emphasis added).[16,p57] This statement is similar to the NIMH press release on the step2 results quoting STAR*D principal investigator Madhukar Trivedi stating, "Augmenting the first medication may be an effective way for people with depression to become symptom-free" (emphasis added).[22] STAR*D defined remission as a score of <8 on the HRSD. While this is a common criterion for classifying remission, such a score is by no means synonymous with "complete relief from a depressive episode" or indicative of the patient becoming "symptom-free" because patients could have up to 7 HRSD symptoms mildly expressed and still met this criterion. For example the HRSD suicide item, "feels like life is not worth living" is scored as only 1 and the same for "feels he/she has let people down," "feels incapable, listless, less efficient," and "has decreased sexual drive and satisfaction;" all of these symptoms when present are scored as only 1 on the HRSD.[23, p285] A patient endorsing just these 4 HRSD items would be counted as remitted with three "mild" symptoms to spare, yet no competent clinician would describe such a patient as experiencing "complete relief" from his or her depressive episode or

	becoming "symptom-free" because each of these symptoms are used in diagnosing major depression.
Failed to Report the Pre/Post, Mean Change Scores and Follow-up Results for STAR*D's 11 Prespecified Secondary Outcome Measures	STAR*D had 11 prespecified secondary outcome measures and a detailed analytic plan for evaluating the effectiveness, cost-efficiency, employment and public health impact of its steps 1-4 treatments. These measures included preassessment/postassessment of 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 as well as reassessing remitted patients every 3 months on these same measures during 12 months of free continuing care.[2 p476-479; 9 p48-51] STAR*D has only recently published a study using a small portion of this dataset. This study assessed changes in work productivity in steps 1 and 2 but not follow-up.[16] 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 AD care through reductions in healthcare utilization and costs.[2] It is troubling that STAR*D's researchers have still not published these findings despite it being 8 years since the publication of the summary article.[5]
Pattern of Rounding Up Errors in Reporting Remission and Response Rates in Steps 1-3	The STAR*D 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.[8, p16] There were no rounding errors in calculating the step4 remission and response rates. In the 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 either direction for 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 rates of intolerable side-effects. Although the inflationary effect of these rounding up errors was trivial, the pattern's consistency is troubling.