Point-Counterpoint with Gibbons et al.

Drafted 01.10.2013 and revised 01.23.2013
Glen Spielmans, Ph.D.

The numbered points describe the points made in our letter to the editor in the January 2013 issue of JAMA Psychiatry. The Counterpoints are the responses of Gibbons et al. The Rejoinders represent our responses to the counterpoints raised by Gibbons et al. Each point is then concluded with a summary.

1. LYAQ participants did not all have major depressive disorder diagnoses

Counterpoint: Furthermore, although patients in this study had comorbidities including attention-deficit/hyperactivity disorder, the majority (81%) of patients had a depressive disorder.

Rejoinder: “The majority (81%) of patients had a depressive disorder.” Kratochvil CJ et al., J Am Acad Child Adolesc Psychiatry 2005;44:915-924 [a paper that published findings of LYAQ] indicate that 45.7% of LYAQ participants had a diagnosis of major depressive disorder. Gibbons et al must be adding in participants with dysthymia and mood disorder NOS to reach 81%. Gibbons et al’s papers claim to focus on major depressive disorder (MDD); if less than half of participants had MDD, then LYAQ is clearly not appropriate for inclusion.

Further, 99% of LYAQ participants had an ADHD diagnosis, also making it inappropriate for inclusion among studies that focused on participants with purer depression diagnoses.

We did not raise the point regarding study duration in our letter, but Gibbons et al raise it in their response. Their statement that LYAQ participants were assessed at “days 1-42 postbaseline depending on the subject” is misleading. According to the above-cited paper which published results of LYAQ - a paper not cited by Gibbons et al (Kratochvil et al., 2005): “At study entry, patients were randomly assigned to receive either fluoxetine (20 mg daily) or placebo under double-blind conditions. After approximately 3 weeks (two evenly spaced visits) to allow for attainment of steady-state fluoxetine concentrations, atomoxetine was added to each patient’s regimen for the final 5 weeks of treatment... (pg. 917).” The average participant thus received double-blind treatment for approximately 3 weeks. Yet Gibbons et al report in their analysis that “6 weeks... was the minimum trial duration (Gibbons et al. Efficacy of Antidepressants E2).”

Summary: LYAQ is inappropriate for inclusion due to its short duration, over half of participants not meeting criteria for a MDD diagnosis, and all participants meeting criteria for an ADHD diagnosis.

2. LYAQ participants were all taking concomitant atomoxetine.

Counterpoints: In fact, the first 2 active treatment visits did not include atomoxetine and our analysis of the data from this study was restricted to this period (days 1-42 postbaseline depending on the subject).
Rejoinder: Gibbons is correct that LYAQ participants did not immediately take atomoxetine. In conducting research for our letter, we gathered information from Lilly’s online registry report for LYAQ. Lilly’s registry report does not clearly indicate that the first 2 visits did not include atomoxetine treatment. We subsequently discovered a publication (Kratochvil et al., 2005) which described LYAQ more clearly. After our letter was finally accepted for publication, Spielmans asked the editor if we could update our description of LYAQ and was told this was not possible.

**Summary:** Gibbons et al were correct on this point. Our point would not have been raised were data reported transparently in Gibbons et al’s papers.

3. For each trial, a description of patient diagnoses, concurrent treatment, number of suicide-related events, and change on depression rating scales should be provided

Counterpoint: None.

**Summary:** Our point was not addressed.

4. Youth data on venlafaxine indicates increased suicidality but data on venlafaxine in youth were not included.

Counterpoint: We did not have access to pediatric data for venlafaxine. The primary focus of our article was on fluoxetine. The adult venlafaxine data were used to determine the extent to which our overall findings for adults only generalized to other antidepressants. Our article is clear that for elderly individuals and youth, our results are for fluoxetine only and that we had not determined whether these results generalize to other antidepressants for the same age groups.

Rejoinder: In the abstract of both papers, Gibbons et al wrote that “All sponsor-conducted randomized controlled trials of fluoxetine and venlafaxine” were their source of data. This is inaccurate. Anyone who read the abstract without reading the paper would come to an inaccurate conclusion based on reading the abstracts of these papers.

In the body of both papers, they did not claim to have youth data for venlafaxine - but they chose to ignore published clinical trials of venlafaxine which showed elevated risk of suicidality in youth.

Further, their claim that “we had not determined whether these results generalize to other antidepressants for the same age groups” is false. If they are not generalizing across all antidepressants, then how can they make the following claim? “These findings should also favor reconsideration of the risk-benefit equation that led to the black box warning for suicidal thinking and antidepressants in children (Gibbons et al efficacy paper E6).” They cannot claim that they are not generalizing across drugs while simultaneously calling for the risk-benefit equation across antidepressants to be reconsidered.

**Summary:** Gibbons et al. called for a reconsideration of the black box warning in their efficacy paper - this is generalizing across antidepressants. Thus, the considerable volume of data regarding other antidepressants (such as venlafaxine) is relevant and should have been addressed
by Gibbons et al - but was not.

5. Gibbons et al do not provide the total number of suicide attempts and suicides broken down by drug vs. placebo. If youth data from venlafaxine and fluoxetine are included from Bridge et al (2007), then the relative risk of suicidal ideation, attempts, and preparatory action for drug vs. placebo is 2.27.

Counterpoint: Furthermore, our analysis preserved the ordinal nature of the suicide risk scale, correctly placing suicidal behavior at a higher level in the hierarchy than suicidal thoughts. Spielmans and colleagues’ crude odds ratios treat suicidal thoughts and behavior equally, which is a less informative analysis. Their analysis and related meta-analyses also ignore the timing of these events, ignoring the fact that as a whole suicide risk decreases substantially over time regardless of intervention status. Current studies have more complete suicide rating scale data that permit improved prospective quantification of suicide risk.

Rejoinder: Gibbons et al do not provide a total number of suicide attempts and suicides broken down by drug vs. placebo. We agree that suicide attempts are more important than suicidal thoughts - why are these data not clearly presented in Gibbons et al’s papers?

“Suicide risk decreases substantially over time regardless of intervention status.” This is not relevant to anything we raised in our letter. Their point regarding studies with “more complete suicide rating scale data” is also not relevant to any of our points.

Summary: Our point was not addressed. The number of suicide attempts and suicides for drug vs. placebo was not provided.

6. Gibbons et al used the CDRS and HAM-D suicidality items as measures of “suicide risk.” Two leading FDA officials are quoted in our letter as saying these are not appropriate measures of suicidality.

Counterpoint: With respect to the use of rating scales to measure suicide-related events, there are both strengths and weaknesses. The strength is that these are prospective ratings by trained clinicians and not spontaneous reports that are subject to a variety of ascertainment biases.

Rejoinder: Ascertainment bias has not been well-supported empirically (Stone M et al. BMJ 2009;339:b2880). A strong explanation for why rating scale items should be valued more highly than actual suicide-related events is not provided.

Summary: We stand by the comments of FDA officials Laughren and Temple that scales such as the HAM-D and CDRS suicidality items are not particularly adept at detecting suicidality.

7. The included trials were not designed to assess potential treatment-emergent suicidality, nor were the suicidality rating scale items valid measures of assessing suicidality. These are low-threshold measures that may obscure the ability to detect true suicide-related events (much of what counts as suicidality on these rating scales is not serious suicidality).
Counterpoint: With respect to the use of rating scales to measure suicide-related events, there are both strengths and weaknesses. The strength is that these are prospective ratings by trained clinicians and not spontaneous reports that are subject to a variety of ascertainment biases.

**Summary:** Same as point 6. In addition, there was no evidence provided to thwart our claim that these rating scales are low-threshold measures that do not accurately detect suicidality.

8. Venlafaxine has no efficacy for youth depression - this has been demonstrated in prior research.

Counterpoint: Spielmans and colleagues cite an effect size of 0.14 for youth venlafaxine studies, based on a meta-analysis of endpoints without the benefit of complete longitudinal data as we used.

Rejoinder: There is no reason to suspect that any statistical method would somehow find that venlafaxine is effective for youth depression. The current data show no signal of efficacy; this point was not challenged.

**Summary:** This point was not challenged.

9. Lilly study HCCJ on fluoxetine for youth showed a very small degree of efficacy for fluoxetine over placebo but was not included in Gibbons et al’s analysis. It was likely excluded because it did not use the CDRS as an outcome measure. This biases their findings.

Counterpoint: Our exclusion of study HCCJ is unfortunate but at the time necessary because the statistical method used in our research synthesis uses the same measure in all studies. Additional work is under way to harmonize these measures.

Rejoinder: The exclusion of the study is understandable given their methods. But the study is nonetheless relevant and appears to find minimal advantage for fluoxetine over placebo. The exclusion of the study biases their findings.

**Summary:** The point was not challenged.

10. A trial of venlafaxine for a geriatric sample found that neither venlafaxine nor fluoxetine were better than placebo. This trial was not included by Gibbons et al., leading to an overestimate of antidepressant efficacy for older adults.

Counterpoint: Our article is clear that for elderly individuals and youth, our results are for fluoxetine only and that we had not determined whether these results generalize to other antidepressants for the same age groups.

Rejoinder: As noted earlier, their abstracts incorrect convey the message that all manufacturer-sponsored studies of venlafaxine and fluoxetine were included. They do not address our point that excluding a study of venlafaxine and fluoxetine for older adults biases their analyses.
11. The reporting of the advantage for fluoxetine over placebo exaggerates its benefits. A drop of 15.96 for placebo and 20.58 for fluoxetine was reported as a 29% greater improvement for fluoxetine; this could equally well have been reported as placebo achieving 78% of the change that occurred while receiving fluoxetine.

Counterpoint: None.

12. The categories of response (50% reduction in CDRS-R score) and remission (CDRS-R score ≤ 28) were arbitrarily created out of a continuous measure, as demonstrated by the implausible finding that remission was more frequent than response. A sensitivity analysis should have been carried out to see if the differences survived different definitions of response and remission.

Counterpoint: None

13. The reported placebo response rate of 5.7% in fluoxetine youth trials is puzzling. Using the Gibbons et al criterion for response, the reported placebo response rate was 16.8% in the HCJE study and approximately 18% in the X065 study (extrapolated from the Figure in the clinical trial report). Data from the LYAQ trial are irrelevant. Study reports from the Treatment for Adolescents With Depression Study did not use the Gibbons et al response criterion, but the response rate on the Clinical Global Impression scale for placebo was 34.8%. Response rates were not reported in HCCJ, but the average participant receiving placebo improved by greater than 50% on the HAM-D-21.

Counterpoint: Spielmans and colleagues cite a variety of response rates from various sources from various selected trials or meta-analyses, none of which used the more rigorous research synthesis provided in our article. Extrapolating values from figures that ignore missing data is no substitute for the more complete longitudinal analysis provided in our article.

Rejoinder: None of our specific numbers are addressed. They claim their research was “more rigorous” without actually addressing our points. Simply providing their calculated response rates for each trial would have been a simple solution to this problem, but this was not done. We would not need to extrapolate response rates from a figure if Gibbons et al had reported their data transparently.

14. Though not mentioned in their article, Gibbons et al selected the youth antidepressant trials with the lowest placebo response rate, helping to maximize apparent drug-placebo differences.

Counterpoint: None
Summary: Our point was not challenged.

15. The verdict on the efficacy of antidepressants in youth remains unchanged; the overall effect size of $d=0.20$ on clinician-rated depressive symptoms—and likely less on self-report measures—is unimpressive.

Counterpoint: The observed effect of fluoxetine on depression in youth is impressive and important to public health. Whether this benefit generalizes to other antidepressants remains an open scientific question.

Rejoinder: The benefits that Gibbons et al attribute to fluoxetine are questionable, as noted in several points above. Their claim of remaining agnostic regarding other antidepressants is disingenuous, as noted by their statement: “These findings should also favor reconsideration of the risk-benefit equation that led to the black box warning for suicidal thinking and antidepressants in children (Gibbons et al efficacy paper E6).”

Summary: Our point regarding the general literature on antidepressants was not challenged.