Inappropriate Data and Measures Lead to Questionable Conclusions

In recent articles, Gibbons and colleagues[1,2] concluded that antidepressants lowered suicidality relative to placebo among adult patients while demonstrating no difference in suicidality among youths; they further concluded that antidepressants possessed robust efficacy vs placebo. However, there are several problems with the underlying data and their choice of suicidality measures.

Though their article purported to include data regarding patients with major depressive disorder, not all trials included patients with a primary diagnosis of depression. For example, in the LYAQ fluoxetine trial, participants either (1) met *DSM-IV* criteria for attention-deficit/hyperactivity disorder and had concurrent anxious/depressive symptoms or (2) met criteria for an anxiety or depressive disorder and had concurrent attention-deficit/hyperactivity disorder symptoms. Further, all participants taking fluoxetine were also taking atomoxetine hydrochloride, rendering the trial incapable of assessing potential fluoxetine-related suicidality; thus, this study does not provide evidence regarding either the efficacy or safety of fluoxetine relative to placebo.[3] We would have gladly combed through descriptions of the remaining studies to assess for similar problems, but most of them are difficult or impossible to obtain. Readers should not have to attempt a search of manufacturer clinical trial databases, which are far from inclusive, to find basic descriptions of included studies; at the very least, for each included trial, a clearer description of participants' diagnoses, concurrent treatment, reports of suicide-related events, and change on depression rating scales is needed to better understand the nature of the data underlying the claims of Gibbons et al.

The Gibbons et al data set included adult trials of fluoxetine and venlafaxine but data from youth taking venlafaxine were not included. This unexplained maneuver clearly biases their analyses given that clinical trial data on venlafaxine found a substantially elevated risk of suicidality relative to placebo. The original venlafaxine publication found that 11 of 182 patients receiving venlafaxine experienced suicide-related events compared with 1 patient of 179 receiving placebo.[4] After these data were carefully recoded by a second group of researchers, the numbers changed to 8 of 182 patients receiving venlafaxine compared with 0 of 179 receiving placebo.[5] Either way, venlafaxine was associated with increased suicidality in adolescents, yet this important data set was inappropriately excluded.

Gibbons et al do not provide actual data on suicide attempts and suicides broken down by drug vs placebo. Perhaps the closest tabulation of such data for youth can be found in the Bridge et al meta-analysis[5] of suicide ideation, attempts, and preparatory action, which generated the following numbers: 17 events of 287 participants receiving fluoxetine compared with 11 of 289 receiving placebo and 8 of 182 receiving venlafaxine compared with 0 of 179 receiving placebo. Thus, including the 2 drugs examined by Gibbons et al, the rate of suicidality while receiving the drug is 5.33% compared with 2.35% while receiving placebo, a relative risk of 2.27. This comes as no surprise given that multiple meta-analyses have found increased rates of suicidal events while receiving antidepressants relative to placebo among youth.[5,6]

The choice of primary outcome for the Gibbons et al suicidality study, scores on the Children's Depression Rating Scale (CDRS) and Hamilton Depression Rating Scale (HAM-D) suicidality items, is problematic. The CDRS and HAM-D are primarily used to detect a signal of treatment efficacy; neither Wyeth nor Lilly designed their trials with the idea of systematically assessing potential treatment-emergent suicidality. These items are low-threshold measures, potentially generating background noise, leading to a diminished ability to detect the true signal of less common but more troubling suicidal events. Changes in rating scale scores regarding suicidality are surely less meaningful than actual reports of suicide-related adverse events, particularly those that have been independently coded by experts in the field of suicide research, as was the case in the aforementioned meta-analyses.[5,6] While statistical power may be increased by the use of a rating scale item completed at multiple points by hundreds of participants, such power is meaningless and potentially misleading if it obscures the detection of true suicidal events. The director of the Food and Drug Administration Psychiatry Products Division, Thomas Laughren, MD, expressed that relying on rating scales to detect suicidality "turned out not to be very helpful."[7] Robert Temple, MD, the director of the Office of Medical Policy at the Center for Drug Evaluation and Research at the Food and Drug Administration, also noted that a depression rating scale is "not where it shows up," referring to the signal of suicidality among youth taking antidepressants.[8]

For venlafaxine in youth depression, the effect size on the CDRS in favor of drug is a meager $d=0.14$, falling short of the conventional definition of even a small treatment effect.[5] In addition, Gibbons et al excluded fluoxetine study HCCJ, apparently because it did not use the CDRS. However, on the HAM-D, the change shown by youth taking fluoxetine was only 11% greater than the change shown by youth taking placebo.[9] We are aware of only 1 Wyeth-sponsored venlafaxine trial with depressed geriatric patients; it found that treatment with either venlafaxine or fluoxetine was no more effective than pla-

cebo.[10] If these data are added to the mix, the combined efficacy of venlafaxine and fluoxetine for youth becomes much less impressive and the already middling efficacy data regarding the geriatric population likely become yet less favorable.

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. 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. We are also puzzled by the reported 5.7% response rate while receiving placebo in youth fluoxetine trials. 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).[9] 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%.[11] Response rates were not reported in HCCJ, but the average participant receiving placebo improved by greater than 50% on the HAM-D-21.[9] 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.[12] 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.[5]

While we applaud the use of patient-level data to examine the potential relationship between initial severity and treatment outcome, the exclusion of data from relevant trials and inclusion of data from inappropriate trials along with the use of a poor measure to detect suicidality renders the conclusions of Gibbons et al questionable at best. We hope that a more comprehensive, appropriate, clearly described data set may help to address important questions of antidepressant safety and efficacy.

Glen I. Spielmans, PhD
Jon Jureidini, PhD, MBBS, FRANZCP
David Healy, MD
Robert Purssey, MBBS, FRANZCP

Author Affiliations: Department of Psychology, Metropolitan State University, St Paul, Minnesota (Dr Spielmans); Discipline of Psychiatry, Adelaide University, Adelaide, South Australia (Dr Jureidini), and Discipline of Psychiatry, University of Queensland, St Lucia, Queensland (Dr Purssey), Australia; and Department of Psychological Medicine, Cardiff University, Hergest Unit, Bangor, Wales (Dr Healy).

Correspondence: Dr Spielmans, Department of Psychology, Metropolitan State University, 1450 Energy Park Dr, St Paul, MN 55108 (glen.spielmans@metrostate.edu).
Conflict of Interest Disclosures: Dr Spielmans is a member of Healthy Skepticism and holds shares of less than $10 000 in Vanguard Health Care, a mutual fund that invests heavily in pharmaceutical companies. Dr Jureidini is a member of Healthy Skepticism. Dr Healy has had consultancies with, been a clinical trialist for, been a speaker at symposia for, or received support to attend meetings from AstraZeneca, Boots-Knoll, Eli Lilly, Janssen-Cilag, Lorex Synthelabo, Lundbeck, Organon, Pharmacia and Upjohn, Pierre-Fabre, Pfizer, Rhone-Poulenc Rorer, Roche, SmithKline Beecham, and Solvay-Duphar. He has provided expert witness testimony for the plantiff in cases involving alleged links between antidepressants manufactured by Lilly, Pfizer, GlaxoSmithKline, and Forest and suicidality. Dr Healy was also an expert witness in a patent case involving an antipsychotic drug. He is also a shareholder in Data Based Medicine, a company aimed at detecting adverse events on treatment. Dr Purssey is a member of Healthy Skepticism.

1. Gibbons RD, Brown CH, Hur K, Davis JM, Mann JJ. Suicidal thoughts and behavior with antidepressant treatment: reanalysis of the randomized placebo-controlled studies of fluoxetine and venlafaxine. *Arch Gen Psychiatry*. 2012; 69(6):580-587.
2. Gibbons RD, Hur K, Brown CH, Davis JM, Mann JJ. Benefits from antidepressants: synthesis of 6-week patient-level outcomes from double-blind placebo-controlled randomized trials of fluoxetine and venlafaxine. *Arch Gen Psychiatry*. 2012;69(6):572-579.
3. Eli Lilly. Strattera clinical trials summary. http://www.lillytrials.com/results/Strattera.pdf. Accessed March 6, 2012.
4. Emslie GJ, Findling RL, Yeung PP, Kunz NR, Li Y. Venlafaxine ER for the treatment of pediatric subjects with depression: results of two placebo-controlled trials. *J Am Acad Child Adolesc Psychiatry*. 2007;46(4):479-488.
5. Bridge JA, Iyengar S, Salary CB, Barbe RP, Birmaher B, Pincus HA, Ren L, Brent DA. Clinical response and risk for reported suicidal ideation and suicide attempts in pediatric antidepressant treatment: a meta-analysis of randomized controlled trials. *JAMA*. 2007;297(15):1683-1696.
6. Hammad TA, Laughren T, Racoosin J. Suicidality in pediatric patients treated with antidepressant drugs. *Arch Gen Psychiatry*. 2006;63(3):332-339.
7. Department of Health and Human Services, Food and Drug Administration Center for Drug Evaluation and Research. Psychopharmacologic Drugs Advisory Committee with the Pediatric Committee of the Anti-Infective Drugs Advisory Committee. http://www.fda.gov/ohrms/dockets/ac/04/transcripts/4006T1.htm. Published February 2, 2004. Accessed February 23, 2012.
8. FDA's role in protecting the public health: examining the FDA's review of safety and efficacy concerns in antidepressant use by children. Hearing before the Subcommittee of Oversight and Investigations of the Committee on Energy and Commerce, House of Representatives. http://www.gpo.gov/fdsys/pkg/CHRG-108hhrg96099/html/CHRG-108hhrg96099.htm. Published September 23, 2004. Accessed February 23, 2012.
9. Eli Lilly. Prozac clinical trials summary. http://www.lillytrials.com/results/Prozac.pdf. Accessed March 7, 2012.
10. Schatzberg A, Roose S. A double-blind, placebo-controlled study of venlafaxine and fluoxetine in geriatric outpatients with major depression. *Am J Geriatr Psychiatry*. 2006;14(4):361-370.
11. March J, Silva S, Petrycki S, Curry J, Wells K, Fairbank J, Burns B, Domino M, McNulty S, Vitiello B, Severe J; Treatment for Adolescents With Depression Study (TADS) Team. Fluoxetine, cognitive-behavioral therapy, and their combination for adolescents with depression: Treatment for Adolescents With Depression Study (TADS) randomized controlled trial. *JAMA*. 2004;292(7):807-820.
12. Bridge JA, Birmaher B, Iyengar S, Barbe RP, Brent DA. Placebo response in randomized controlled trials of antidepressants for pediatric major depressive disorder. *Am J Psychiatry*. 2009;166(1):42-49.

In reply

We appreciate Spielmans and colleagues' interest in our article; however, their concerns about the data in our article are unfounded.

Spielmans and colleagues state that study LYAQ included atomoxetine as a treatment and should have been excluded. 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). Furthermore, although patients in this study had comorbidities including attention-deficit/hyperactivity disorder, the majority (81%) of patients had a depressive disorder. As such, our inclusion of this study was scientifically appropriate.

Spielmans and colleagues suggest that our failure to include youth studies of venlafaxine in our research synthesis is an "unexplained maneuver" that "clearly biases their analyses." 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.

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.[1] *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.*[2]

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

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. 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.

Robert D. Gibbons, PhD
C. Hendricks Brown, PhD
Kwan Hur, PhD
John M. Davis, MD
J. John Mann, MD

Author Affiliations: Departments of Medicine, Health Studies, and Psychiatry (Dr Gibbons) and Center for Health Statistics (Drs Gibbons, Brown, and Hur), University of Chicago, and Department of Psychiatry, University of Illinois at Chicago (Dr Davis), Chicago, and Center for Medication Safety, Pharmacy Benefit Management Services, Hines Veterans Affairs Hospital, Hines (Dr Hur); Prevention Science and Methodology Group, Center for Family Studies, Department of Epidemiology and Public Health, University of Miami, Miami, Florida (Dr Brown); and Department of Molecular Imaging and Neuropathology, New York State Psychiatric Institute and Department of Psychiatry, Columbia University College of Physicians and Surgeons, New York (Dr Mann).
Correspondence: Dr Gibbons, University of Chicago, 5841 S Maryland Ave, MC 2007, Office W260, Chicago, IL 60637 (rdg@uchicago.edu).
Conflict of Interest Disclosures: Dr Gibbons has served as an expert witness for the US Department of Justice, Wyeth, and Pfizer Pharmaceuticals in cases related to antidepressants and anticonvulsants and suicide. Dr Brown directed a suicide prevention program at the University of South Florida that received funding from JDS Pharmaceuticals. Dr Mann has received research support from GlaxoSmithKline and Novartis.
Funding/Support: This work was supported by grants MH062185 (Dr Mann), R56 MH078580 (Drs Gibbons and Brown), MH8012201 (Drs Gibbons and Brown), and MH040859 (Dr Brown) from the National Institute of Mental Health and grant 1U18HS016973 from the Agency for Healthcare Research and Quality (Dr Gibbons).

1. Posner K, Oquendo MA, Gould M, Stanley B, Davies M. Columbia Classification Algorithm of Suicide Assessment (C-CASA): classification of suicidal events in the FDA's pediatric suicidal risk analysis of antidepressants. *Am J Psychiatry*. 2007;164(7):1035-1043.
2. Posner K, Brown GK, Stanley B, Brent DA, Yershova KV, Oquendo MA, Currier GW, Melvin GA, Greenhill L, Shen S, Mann JJ. The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. *Am J Psychiatry*. 2011; 168(12):1266-1277.

Suicide Risk and Efficacy of Antidepressant Drugs

In their reports[1,2] addressing the black box warning about antidepressant drugs and suicide risk, Gibbons et al conducted risk/efficacy reanalyses of selected data sets. They concluded that

> no significant effects of treatment on suicidal thoughts and behavior were found. . . . No evidence of increased suicide risk was observed in youths receiving active medication.[1(p580)]

These conclusions are unsound. The data are contaminated by trial LYAQ,[2-4] which studied atomoxetine hydrochloride with or without fluoxetine in patients with attention-deficit/hyperactivity disorder with depressive or anxious symptoms. Less than half had major depression diagnoses[4]; pretreatment depression scores matched posttreatment scores in the other youth studies.[2] These problematic details were not addressed. By lack of conforming diagnoses and design, LYAQ outright invalidates the aggregate analyses.

The key adverse outcome was termed *suicide risk*, but Gibbons et al did not measure suicide risk. They im-