

Dear

Nine of us whose names are attached to this email (we did not have time to create electronic signatures) were authors on the study originally published in 2001 in the *Journal of the American Academy of Child and Adolescent Psychiatry* entitled, "Efficacy of paroxetine in the treatment of adolescent major depression: a randomized controlled trial," and have read the reanalysis of our article, which is entitled, "Restoring Study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence", currently embargoed for publication in the *British Medical Journal* (BMJ) early this week. We are providing you with a brief summary response to several of the points in that article that which with we have strong disagreement. Given the length and detail of the BMJ publication and the multitude of specific concerns we have with its approach and conclusions, we will be writing and submitting to the BMJ's editor an in-depth letter rebutting the claims and accusations made in the article. It will take a significant amount of work to make this scholarly and thorough and do not have a time table; but that level of analysis by us far exceeds the time frame needed to give you that more comprehensive response by today.

The study was planned and designed between 1991-1992. Subject enrollment began in 1994, and was completed in 1997, at which time analysis of the data commenced. The study authors comprised virtually all of the academic researchers studying the treatment of child depression in North America at the time. The study was designed by academic psychiatrists and adopted with very little change by GSK, who funded the study in an academic / industry partnership. The two statisticians who helped design the study are among the most esteemed in psychiatry. The goal of the study designers was to do the best study possible to advance the treatment of depression in youth, not primarily as a drug registration trial. Some design issues would be made differently today -- best practices methodology have changed over the ensuing 24-year interval since inception of our study.

In the interval from when we sat down to plan the study to when we approached the data analysis phase, *but prior to the blind being broken*, the academic authors, not the sponsor, added several additional measures of depression as secondary outcomes. We did so because the field of pediatric-age depression had reached a consensus that the Hamilton Depression Rating Scale (our primary outcome measure) had significant limitations in assessing mood disturbance in younger patients. Accordingly, taking this into consideration, and in advance of breaking the blind, we added secondary outcome measures agreed upon by all authors of the paper. We found statistically significant indications of efficacy in these measures. This was clearly reported in our article, as were the negative findings.

In the "BMJ-Restoring Study 329 ..." reanalysis, the following statement is used to justify non-examination of a range of secondary outcome measures:

*Both before and after breaking the blind, however, the sponsors made changes to the secondary outcomes as previously detailed. We could not find any document that provided any scientific rationale for these post hoc changes and the outcomes are therefore not reported in this paper.*

This is not correct. The secondary outcomes were decided *by the authors prior to the blind being broken*. We believe now, as we did then, that the inclusion of these measures in the study and in our analysis was entirely appropriate and was clearly and fully reported in our paper. While secondary outcome measures may be irrelevant for purposes of governmental approval of a pharmaceutical indication, they were and to this day are frequently and appropriately included in study reports even in those cases when the primary measures do not

reach statistical significance. The authors of “Restoring Study 329” state “there were no discrepancies between any of our analyses and those contained in the CSR [clinical study report]”. In other words, the disagreement on treatment outcomes rests entirely on the arbitrary dismissal of our secondary outcome measures.

We also have areas of significant disagreement on the “Restoring Study 329” analysis of side effects (which the author’s label “harms”). Their reanalysis uses the FDA MedDRA approach to side effect data, *which was not available when our study was done*. We agree that this instrument is a meaningful advance over the approach we used at the time, which was based on the FDA’s then current COSTART approach. That one can do better reanalyzing adverse event data using refinements in approach that have accrued in the 15 years since a study’s publication is unsurprising and not a valid critique of our study as performed and presented.

A second area of disagreement (concerning the side effect data) is with their statement, “We have not undertaken statistical tests for harms.” The authors of “Restoring Study 329” with this decision are saying that we need very high and rigorous statistical standards for declaring a treatment to be beneficial but for declaring a treatment to be harmful then statistics can’t help us and whatever an individual reader thinks based on raw tabulation that looks like a harm is a harm. Statistics of course does offer several approaches to the question of when is there a *meaningful* difference in the side effect rates between different groups. There are pros and cons to the use of P values, but alternatives like confidence intervals are available.

“Restoring Study 329” asserts that this paper was ghostwritten, citing an early publication by one of the coauthors of that article. There was absolutely nothing about the process involved in the drafting, revision, or completion of our paper that constitutes “ghostwriting”. This study was initiated by academic investigators, undertaken as an academic / industry partnership, and the resulting report was authored mainly by the academic investigators with industry collaboration.

Finally the “Restoring Study 329” authors discuss an initiative to correct publications called “restoring invisible and abandoned trials (RIAT)” (BMJ, 2013; 346-f4223). “Restoring Study 329” states “We reanalyzed the data from Study 329 according to the RIAT recommendations” but gives no reference for a specific methodology for RIAT reanalysis. The RIAT approach may have general “recommendations” but we find no evidence that there is a consensus on precisely how such a RIAT analysis makes the myriad decisions inherent in any reanalysis nor do we think there is any consensus in the field that would allow the authors of this reanalysis or any other potential reanalysis to definitively say they got it right.

In summary, to describe our trial as “misreported” is pejorative and wrong, both from consideration of best research practices at the time, and in terms of a retrospective from the standpoint of current best practices.

Martin B. Keller, M.D.

Boris Birmacher, M.D.

Gregory N. Clarke, Ph.D.

Graham J. Emslie, M.D.

Harold Koplewicz, M.D.

Stan Kutcher, M.D.

Neal Ryan, M.D.

William H. Sack, M.D.

Michael Strober, Ph.D.