THE UNIVERSITY of ADELAIDE

8 May 2015

Elizabeth Loder, MD, MPH
BMJ Editorial Team

Dear Dr Loder

Re: BMJ.2014.022376.R2 entitled 'Restoring Study 329: A randomised, controlled trial of the efficacy and harms of paroxetine and imipramine in the treatment of adolescent major depression'

Thanks for your letter.

Although you provisionally accepted our paper on 3 March, you have raised a number of new issues and revisited some issues that we have previously addressed. It is no longer clear to us if you still wish to publish the paper. If you are going to reject the paper, we want to know that within a week.

We are not prepared to submit yet another draft for a prolonged review process with a new set of reviewers. Many of the issues that have come up in your editorial meetings are things we have discussed over and over in our encounter with this new thing – the RIAT Initiative. And your editors are encountering the same problem. Their input has been helpful, but it lacks continuity as each new crop struggles anew with what RIAT is. The purpose of RIAT is to formally correct the scientific record: correcting the invisibility of unpublished trials and 'correcting reporting biases persisting in existing trial publications' (Doshi et al. 2014).

We lay out our responses to your letter below. We still want to see our paper published in the BMJ, and we request that you work quickly with us to a point of agreement to accept or reject it. We are happy to interact with and negotiate with you in order to accommodate a final rejection or acceptance within a week of your receiving this letter. At that point, if you decide to accept, we will submit a revised version within a week.

We see the handling of adverse events as the most important sticking point.

1. **You raise concerns that our reporting of adverse events is unconvincing because we may be perceived to have a bias due to involvement in litigation.**

Your references to litigation are unfortunate. Involvement in litigation does have a potential for bias, as we have acknowledged in our paper, but it does not disqualify us from analysing data. Unlike most others publishing in BMJ or elsewhere, we are exposing our biases to scrutiny by making all data available. There will always be potential COI in RIAT. The team has assembled for a reason. In the future, in any RCTs the BMJ publishes, will you insist that the coding is not done by anyone biased by their association with the trial, as sponsoring company or CRO or expert academics drafted in as notional authors, or indeed the clinical investigators themselves? Our rater was blinded and trained. We doubt you will find better quality control over the process of adverse event analysis and reporting.

2. You suggest that our emphasis on adverse events (AEs) is like 'the tail wagging the dog'.

Far from our emphasis on AEs being like 'the tail wagging the dog', we think it is ground-breaking work that needs to be in the foreground, along with the fact that our paper is also a study in authorship and the effects on authorship of access to the data. We are therefore unwilling to weaken our analysis.

First please re-read our rationale:

> All of the initial coding from the clinical descriptions in the CSR was done blind, as was coding from the CRFs.
>
> The original protocol for Study 329 makes no mention of how AEs from this trial would be coded. The CSR specifies that the AEs noted by clinical investigators in this trial were coded using the Adverse Drug Experience Coding System (ADECS) that was being used by SKB at the time. ADECS was derived from a coding system developed by the United States Food and Drug Administration (FDA), Coding Symbols for a Thesaurus of Adverse Reaction Terms (COSTART), but is not itself a recognized system.
>
> We coded AEs using MedDRA, which has replaced COSTART for the FDA, because it is by far the most commonly used coding system today, and it is not possible to access ADECS. For coding purposes, we have taken the original terms used by the clinical investigators as transcribed from the original CRFs into the CSR, and applied MedDRA codes to these descriptions.
>
> In general, MedDRA coding stays closer to the original clinician description of the event than ADECS does. For instance, MedDRA codes 'sore throat' as 'sore throat', but SKB, using ADECS, coded it as 'pharyngitis' (inflammation of the throat). Sore throats may arise because of pharyngitis, but when someone is taking SSRIs they may indicate a dystonic reaction in the oro-pharyngeal area.[21]
>
> Classifying a problem as a 'respiratory system disorder' (inflammation) rather than as a 'dystonia' (a central nervous system disorder) can make a significant difference to the apparent AE profile of a drug. In staying closer to the original description of events, MedDRA codes suicidal events as 'suicidal ideation' or 'suicidal events' rather than the ADECS option of 'emotional lability'; similarly, aggression is more clearly flagged as 'aggressive events' rather than 'hostility'.

You put words in our mouth when saying that 'the main point of the paper ... was the reanalysis of the efficacy findings, showing that the original claim of superiority rested on post-hoc outcomes'. This statement is wrong on at least two counts. First, the point that the original claim of superiority rested on post-hoc outcomes has already been well made in a paper commissioned by your journal, rejected by you on legal advice, ultimately published elsewhere and often cited.[1]

Second, we disagree with your implication that efficacy is primary, with harms being an adjunct. Historically, regulatory bodies were tasked with ensuring safety. Efficacy was added much later. Approving an inert non-toxic drug is a far lesser sin than approving one that works but is toxic. Consequently we focused heavily on harms, which were minimized in the original paper. As mentioned above, RIAT is explicitly about correcting the scientific record, including correcting reporting biases.

3. You criticise us for going 'beyond what would have been done at the time of the trial' by recoding some of the AEs, and you request that we present them as they were originally coded.

We could not use ADECS (Adverse Drug Events Coding System), as it is unavailable (as stated in our Coding of Adverse Events section).

---

[1] Jureidini J, McHenry L, Mansfield P. Clinical trials and drug promotion: Selective reporting of study 329. Int J Risk Saf Med. 2008;20:73–81

You are correct that the sort of analysis we used was not specified in the original study. However, the coding dictionary was not specified in the protocol, nor was any AE analysis, so it is not true that we did not follow the protocol.

Perhaps our approach will be made clearer if you see a random sample of the material available in the appendix of the CSR:

Appendix D.1

Listing of Adverse Experiences by Treatment Group, ADECS Body System and Preferred Term
Intent-to-Treat Population

Treatment Group=PAROXETINE Body System=Body as a Whole

Preferred Term	Verbatim Term	Patient ID	AE Onset Date	Relative Days *	Duration	Onset Dose (mg)	No. Epi	Inv Int	Act-ion	Inv Rel	Corr Ther	SAE
Asthenia	FATIGUE WITH DIZZINESS AND DROWSINESS [FATIGUE WITH DROWSINESS]	329.005.00011	28JAN95	47, -10,-109	Not Stated	40	CON	MIL	NO	PSR	No	No
	TIRED	329.003.00075	14JUN95	141, 85, -99	2 Days	20		MIL	NO	UNR	No	No
		329.003.00089	10MAR95	4, ., -56	Not Stated	20	CON	MIL	NO	PSR	No	No
	TIREDNESS INCREASED FATIGUE	329.005.00011	19DEC94	7, -50,-119	10 Days	20	4	MIL	NO	PSR	No	No
	VERY TIRED	329.006.00039	22MAR95	43, -19, -49	7 Days	20	CON	SEV	NO	PSR	No	No
Back Pain	BACK PAIN	329.005.00151	06SEP95	1, -57,-238	1 Days	20	1	MOD	NO	UNR	Yes	No
			19OCT95	44, -14,-195	02:00 Hrs	20	1	MIL	NO	UNR	Yes	No
	BACKACHE	329.005.00336	28MAR97	25, ., -33	2 Days	20	CON	MIL	NO	UNR	No	No
			20APR97	48, ., -10	4 Days	20	CON	MIL	NO	UNR	Yes	No
		329.007.00294	21MAR97	22, ., -20	3 Days	20	CON	MOD	NO	UNR	Yes	No
	NECK PAIN	329.007.00268	26MAY96	56, -4,-206	5 Days	30	CON	MOD	NO	UNR	Yes	No
Chest Pain	CHEST PAIN	329.002.00319	09JUL96	71, 10,-174	16.00 Hrs	30	1	MIL	NO	UNR	No	No
		329.005.00002	19JUN94	26, -31,-121	06:30 Hrs	20	CON	MOD	NO	PSR	No	No
	CHEST PAIN-APPROXIMATELY 3 OCCASIONS-INTENSE JAB THEN REMITS	329.006.00038	27MAR95	41, ., -16	3 Days	20	3	SEV	NO	PSR	No	No
Chills	SHIVERING AND SHAKING	329.004.00017	28MAR95	7, -50,-231	12 Days	20	CON	MIL	NO	PSR	No	No

* days relative to start of acute phase, days relative to start of continuation phase, days relative to stop of study medication
Number of Episodes [No. Epi]: CON = Continuous
Investigator Intensity [Inv Int] : MIL = Mild, MOD = Moderate, SEV = Severe
Action Taken on Study Medication [Action] : DCR = Dose Decreased, INC = Dose Increased, NO = None, STP = Drug Stopped
Investigator Relationship [Inv Rel]: PBU = Probably Unrelated, PSR = Possibly Related, REL = Related, UNR = Not Related
Corrective Therapy [Corr Ther]
Serious AE as Judged according to SB Criteria by Investigator [SAE]

The 'preferred term' is the coding of the 'verbatim term' as carried out by SKB; from the number of uncoded or opaque coding decisions, it was clear to us that either ADECS was an idiosyncratic system or it was being improperly used. It seems bizarre that you have asked us to adhere to a coding dictionary that hides suicidality under 'emotional lability'.

We had to find a way to code, and we couldn't follow our preferred pathway of following protocol. Other researchers might have chosen differently, but there is nothing flawed or biased in use choosing MedDRA. An expert coder coded blinded. This is a much higher quality of coding than in the vast majority of RCTs published in the BMJ and elsewhere.

It is not consistent with any practice that we are aware of to 'ask completely independent investigators to code the AEs, report inter-rater agreement'. Is this a requirement for other BMJ authors?

4. You are unclear as to why we do not do any statistical tests on the AEs.

Once again we ask you to review what we have written that we think clearly justifies that decision:

3. Filtering data on AEs through statistical techniques

For instance, Keller et al. (and GSK in subsequent correspondence) ignored unfavourable harms data on the grounds that the difference between paroxetine and placebo was not statistically significant. In our opinion, statistically significant or not, all relevant primary and secondary outcomes, and harms outcomes, should be explicitly reported. Testing for statistical significance is most appropriately undertaken for the primary outcome measures. We have not undertaken statistical tests for harms, since we know of no valid way of interpreting them. To get away from a dichotomous (statistically significant/non-significant) presentation of evidence, we opted to

present all original and recoded evidence to allow readers their own interpretation. The data presented in Appendix 2 and related worksheets lodged at www.xxx will, however, readily permit other approaches to data analysis for those interested, and we welcome other analyses.

However, we will publish statistical significance figures, with a note that we are doing so at the editor's request. This note is necessary because some of us have published elsewhere in support of authorities who argue that such analysis is inappropriate.

5. You say that you are uneasy about any conclusions we made because of the fact that we did not examine all case report forms (CRFs).

We disagree that this should reduce confidence in the findings and that it is a major limitation. We were careful not to draw strong conclusions, and those conclusions that we did draw were firmly based in the CSR rather than the CRFs. At your suggestion, we removed any inferences beyond those CRFs that we did examine.

6. You ask that we make fewer claims about AEs.

Please identify any claims that we make that are not supported by the data, and we will be happy to review.

In our opinion, the point of the exercise is that making the data available makes it possible for others to have the kinds of concerns you may have and to argue your point of view. We do not want to stifle debate.

If we can reach common ground on the AEs then I am confident we can resolve the other issues, as outlined below.

With regard to imputation, we continue to hold that it is inappropriate to publish these results. However, if you insist, we will carry out multiple imputation and report it (preferably in an appendix), adding a note that this departs from our RIAT methodology and has been done at the editor's request.

1. The following points from your review are accepted and will be changed by us:

* We could not find a clear statement about whether paroxetine and imipramine were to be compared with each other or just placebo.	The comparison was to placebo only. We can add this.
* Page 6, add 20 April 1994. Page 10, OC = observed case. Page 19 and abstract, what is LS MEAN?	This can be done
* The numbers in Table 3 (and page 21) have far too many decimal places / significant figures (up to 5).	This can be done
* Table 11 needs to include group denominators.	This can be done.
* Page 32 should mention the CRFs before	This can be done.

referring to the periscope.	
* We do not think the abstract makes sufficiently clear for readers who may not be familiar with the RIAT initiative that this is a reanalysis of a trial published years ago. Perhaps the objectives could start by saying: 'This is a reanalysis of data from GSK's Study 209 (originally published in xxx) done as part of the RIAT initiative. The objective was to see if reanalysis led to similar...' You might also mention in the abstract that registration in a trial registry was not required at the time the study was done.	We can modify the abstract
Table 11 A fresh statistician who reviewed the paper this round cannot tell from reading the paper why you are determined not to test the table for significant differences (top of page 31). He suggests this should be done and if not an explanation included as to why not.	We will comply with this request (see above).
* We were puzzled by the statement that 'This analysis contrasts with both Keller et al.'s published findings and the outcomes reported in the CSR.' My understanding is that the CSR was the source of the information.	Corrected to read: 'This analysis contrasts with both Keller et al.'s published findings and the WAY THE outcomes WERE reported AND INTERPRETED in the CSR.'
* It may be helpful to have an additional box listing where the authors deviated from the original plan/protocol or where their findings differ. While the information is provided throughout the paper, we thought it would help readers to see a summary.	This can be done.
* We think questions about this paper should be channeled through the BMJ's traditional rapid response feature, and ask that you remove the following from the paper: 'We invite readers to contact us for clarification of any ambiguities through a public Q&A forum at www.xxx.com [TBA], where we will respond to any queries about our data or analysis, with further follow-up as required.'	We agree to this.

2. For the following points, we have disagreements that we think can be easily resolved:

* Additional detail is needed in the methods section. This should be detailed enough that others could replicate what you have done.	We believe others could copy what we have done by reading our method and referring to the SKB Protocol. Tell us if there is something in particular you would like us to spell out.
* We recommend that the order of sections should be based on CONSORT, i.e. the RIATAR form.	We did have things in the CONSORT order until one of your earlier reviewers asked us to change it. Your call, tell us what you want.
* Can you discuss in the methods whether a change in the HAM-D of 4 points is clinically significant?	We already say in the method under 'sample size': *This effect size entailed a difference of 4 in the HAM-D Total change*

	from baseline scores at endpoint, specified in the protocol to be large enough to be clinically meaningful, considering a standard deviation (SD) of 10. We could add, 'consistent with NICE, which designates 3 as clinically significant reduction'.
* Was there a pattern to the missing data: 'At least 1000 pages were missing from the Case Report Forms reviewed with no discernible pattern to missing information'	The text you quote answers the question. Unless we are missing something?
* We also feel uneasy about the recategorisation of the lack of efficacy dropouts based on factors such as Adverse Events and HAM-D scores. These decisions seem very subjective and again, there may be a perception of potential bias given your involvement in litigation related to this matter.	We can make clear this makes us uneasy also – and it may reveal bias on our part but it seemed needed when placebo responders with Hamilton Rating Scale scores of 2 were categorized by GSK as lack of efficacy.
* We did feel this version of the paper is much more readable than the initial version. Thank you for all of your work on that. You make 2 clear points: using the prespecified primary outcomes, there is no significant difference between the 3 groups. We felt that you might be able to pare down the portion of the paper that discusses this. One of our editors noted, for example, that it only takes half a page to nicely summarise this at http://www.ncbi.nlm.nih.gov/pubmed/11437014.	We don't follow. Our report of the efficacy results is half a page and 2 figures. And we are perplexed that you refer us to the original Laden – GSK paper as a model.
* Finally, we remain concerned about the tone of the paper. It should be neutral. In several places you stray into editorial comments about the difficulties of doing the analysis and so forth. Those things detract from the presentation of the research itself.	. We think we have been extraordinarily neutral in the circumstances, but if you indicate each episode of non-neutral language, we are prepared to have a go at flattening the tone further – from pancake to Kansas. (Kansas is technically the flatter of the two).

We look forward to hearing from you.

Yours sincerely

Jon Jureidini
on behalf of the Study 329 team