Memorial Sloan Kettering Cancer Center

Tiago Villanueva MD
Associate Editor, The BMJ

9 January 2018

Dear Dr. Villanueva,

Thank you for your thoughtful re-review of our Manuscript (BMJ.2017.041528) entitled "Immune-related Toxicities in PD-1 and PD-L1 Immunotherapies: a Systematic Review and Meta-analysis." Below please find detailed responses to the comments from Reviewer 1.

Comment 1: *I was surprised the authors opted for the inverse-variance approach when this has been shown to be one of the most biased estimators with rare event data. Why not use the Peto odds method also implementable in Revman. The authors appear to be following a Cochrane approach – this is documented in section 16.9 of the Handbook.*

Response 1: We thank the Reviewer for the suggestion to consider using the Peto odds method to pool data. As the Reviewer observes, we did follow a Cochrane approach, but we revisited our approach and re-did all meta-analyses using the Peto method, though we had some concerns that this method might result in biased estimates of the odds ratios for the outcomes with large effect sizes (see Brockhaus AC, Bender R, Skipka G. The Peto odds ratio viewed as a new effect measure. Stat Med. 2014 Dec 10;33(28):4861-74). Results in the repeat analysis were largely similar to our original results, and when they differed, our original estimates were more conservative. The one

Deborah Korenstein MD, FACP

Director of Clinical Effectiveness

Chief, General Medicine Service

korenstd@mskcc.org

485 Lexington Avenue, New York, NY 10017 T 646.888.8139 F 929.321.1518

www.mskcc.org

NCI-designated Comprehensive Cancer Center

exception was the pneumonitis outcome. For this outcome, the overall pooled OR using the Peto method was no longer significant (OR 1.12, [0.81 to 1.54]) in contrast to our original result showing increased risk of pneumonitis overall (OR, 3.82, [1.27 to 11.45]). However, sub-results from everolimus control (1 study) and non-everolimus control subgroups were similar to results in our original analysis. Of note, everolimus is known to cause pneumonitis, so pooling a study that used everolimus as a control with other studies that used other control therapies may not result in an interpretable and meaningful result.

After considering these results, we opted overall to stay with our original approach, since we believe that more conservative estimates are appropriate for our target clinician audience (i.e. we want to be sure not to overstate our findings). However, the Peto method results informed our interpretation of the pneumonitis analysis and we made changes to the manuscript in response. While we still present the sub-analyses and the overall result in the Figure, since we do not think it is meaningful we no longer discuss the overall result and make the point that it is unlikely to be meaningful. We made the following changes:

- In the Results section, in the second paragraph under Organ-specific Immune-related Toxicities (page 13), we removed pneumonitis from the sentence reporting conditions that were more likely with anti-PD1 agents than with control and added specific language about pneumonitis as follows: "For pneumonitis, excluding the study in which the control treatment was everolimus, a drug known to causes pneumonitis, the risk of pneumonitis with anti-PD1 agents was also higher than control (OR 5.37; 95% CI 2.73, 10.56, P<0.00001). However, the risk was lower with anti-PD1 agents compared to everolimus (Figure 5); the overall pooled estimate may not be meaningful." This change resulted in a renumbering of the Figures.
- In the first paragraph of the Discussion (page 15), we changed the third sentence, which now reads: "We found that the risk of organ-specific irAE's such as hypothyroidism, colitis, and perhaps pneumonitis are increased with anti-PD1 agents compared to standard therapies although overall event rates remain low."
- In the second paragraph of the Discussion (page 15), we changed the third sentence, which now reads as follows: "Unlike the prior study, we found an increased risk of pneumonitis with anti-PD1 agents compared to control therapies other than everolimus, though our findings regarding risk of colitis were similar."

Comment 2: For transparency, can the funnel plots please be included as supplementary material?

Response 2: We thank the Reviewer for this suggestion and we now include the funnel plots as an Appendix in the Supplement.

Comment 3: The authors report "The median treatment time in the investigational arm ranged from 3 to 8 months". Can you present the median and range (as for follow-up time)?
Response 3: We appreciate this suggestion and have updated the manuscript accordingly. The manuscript now reads (page 12): "The median treatment time in the investigational arm was 3.9 months (range 3 to 8 months) and the median duration of follow-up across studies was 13.1 months (range 5 to 25 months)."

Comment 4: Table 1. I do not understand the new footnote "‡Applies to primary outcomes only and not to toxicity outcomes". Using this version of the Cochrane RoB assessment you are assessing the study as a whole (i.e. all outcomes) not just primary outcomes. However, the efficacy outcomes are not of interest to this review so you could define your selective reporting assessment RoB criteria (as a method) for the toxicity outcomes only because this will impact the analyses within this review. I believe an assessment can be done because you have defined very specific AE review outcomes of interest, e.g. colitis. Further because toxicity outcomes were not collected and reported differently we could not assess their selective reporting." I don't totally agree with this statement – how do you know toxicity outcomes were not collected? You cannot assume this if they were not reported, this is part of the purpose of the selective reporting bias domain. If you know the outcome was not collected (confirmed by the author) then bias is unlikely (but perhaps a missed opportunity to record data on this outcome)- but there is no suggestion of this confirmation"
Response 4: We appreciate the Reviewer's feedback about our RoB assessments. We have reconsidered our approach and now agree that we should assess RoB with regard to our outcomes of interest and not (as we had done before) with regard to the reported primary and secondary study outcomes. We have therefore repeated the RoB assessments in the domains of Blinding of outcome assessors, Incomplete outcome data, and Selective outcome reporting. Given that toxicity outcomes are assessed by clinicians caring for enrolled patients in all included studies, we now rate all studies at high RoB for Incomplete outcome data and Selective outcome reporting, and most at high RoB for Blinding of outcome assessors. These assessments were made by 2 authors (DK and SB, with differences resolved by consensus).Table 1 now reflects the new assessments. Please note that while kappa values were recalculated, they did not

change. In addition, we made appropriate changes to the manuscript to reflect the current approach, as follows:

- In the Methods section under "Data Extraction and Quality Assessment" (beginning on page 9), we now state "Two of three authors (D.K., A.Y., S.B.) independently assessed the quality…" and (page 10) "To optimize relevance to immune-related toxicities we evaluated risk of bias with regard to toxicity outcomes, not the efficacy outcomes the individual studies were primarily designed to assess."
- In the Results under "Quality of Included Studies" (beginning on page 12) we now note "Further, because included studies were not designed primarily to assess drug toxicities, collection of toxicity information was poorly described and we deemed all studies at high risk of bias with regard to incomplete outcome data and selective outcome reporting."

Comment 5: How was overall RoB judged?

Response 5: We appreciate the Reviewer's observation regarding overall RoB. The Cochrane tool does not include an assessment of overall risk of bias. We had previously offered a qualitative assessment, but we have now removed it. We have deleted the clause on page 9 that mentioned overall risk of bias and now simply offer the RoB assessments in each domain.

Comment 6: Table 2: use of dagger and double dagger. In Table 2, can confidence intervals also be provided for the incidence rates (same for all other rates)?
Response 6: We thank the Reviewer for the suggestion and for pointing out the error with the dagger. The typographical error is corrected in the current version. In addition, we now present incidence rates with confidence intervals for all events in Table 2 as well as in Table 3 (for consistency).

Comment 7: In some cases the subgroups being considered seems to be producing very different results. For example in Figure 4, the single study with everolimus control seems very different than studies with a different control – does it therefore make sense to provide an overall pooled effect estimate for these two groups? This is important as the main subgroup (containing most studies) seems to be exhibiting very little heterogeneity and hence a fixed effect MA maybe OK. Higher levels of heterogeneity are observed for some of the latter outcomes such as Fatigue and therefore random effects MA is probably the best approach if pooling these studies is deemed sensible. Why have fixed and random effects forest plots been shown? Statistical packages such as STATA allow both fixed and random pooled effects estimates to be

shown on one plot if this is what you want to show.
Response 7: We appreciate the Reviewer's comments about the meta-analysis and the subgroups and we agree that it is important to note the different comparators. We opted to present an overall pooled estimate since in many cases the difference in comparator may not impact the toxicity outcomes under consideration. However, we also present pooled results for separate subgroups based on comparator types (e.g. chemotherapy, targeted agents) in the interest of transparency and to partly address the Reviewer's concern. In addition, the Reviewer points out that some of the subgroups have low heterogeneity and thus we may be able to pool in some cases using a fixed effects, rather than a random effects, approach, or we might show it both ways. While we agree that these are excellent suggestions, we believe that given that our target audience for this paper consists largely of clinicians who may have limited understanding of meta-analytic methods, it may be confusing to present a more complicated analysis. Instead we have opted for a conservative approach to data pooling, although it may result in wider confidence intervals in some cases.

Comment 8: For data extraction – prioritisation was made of the data included in CT.gov. What was the rationale for this as a trial may have been registered in a different trial registry. Is it because all studies are not yet published? Also, if the data differed in the publication, how was this issue dealt with as it is possible that registry data is not up to date.
Response 8: We thank the Reviewer for bringing up this important issue. We prioritized data from ClinicalTrials.gov because, when available, CT.gov information was far more complete than what was presented in publications. Clinical trials routinely collect a large amount of toxicity information and present only a fraction of it in publications. One of the strengths of our study was our use of the rich data available on CT.gov. We did not look to other registries because all included trials were indeed registered on CT.gov, although some had not yet posted results. While we appreciate the Reviewer's point that registry data could theoretically be outdated, in practice the registry data tends to be more current and for trials included in our analysis, toxicity rates posted on CT.gov were either the same or higher than what was reported in publications, suggesting that it was more current.

To clarify these points, on page 10 we now state "we prioritized data from ClinicalTrials.gov over toxicity data from the publications, because it was more complete" and we mention that all trials participated in ClinicalTrials.gov by adding the

statement (page 12): "all studies were registered on ClinicalTrials.gov…"

Comment 9: "We also assumed that no events of a particular type occurred if none were reported and in our meta-analyses studies with zero events did not contribute to the pooled result. This may have led to errors in our pooled estimates, though the issue impacts the intervention and control arms equally." I do not agree with this statement. If the event rates truly are 0 in both arms then the m203448saeta-analysis is unaffected and this is not an error. However, it's possible that the non-reporting occurred because there were more harms in the intervention group and the non-reporting occurred as it masks the beneficial effect of the interventions. If this is the case then this would underestimate the harmful effect of treatment so the meta-analysis would be biased. So if the non-reporting is based on the results, I don't think it does effect both arms equally. How about just make a statement of this nature "A limitation of this analysis is that we assumed that the non-reporting of "AE" data was the result of either no events or the outcome was not measured. If a selective non-reporting mechanism was present, such that the AE data was measured but not reported based on the results, then there is the possibility that the harm of intervention maybe underestimated [cite Saini 2014 BMJ]

Response 9: We appreciate the Reviewer's clarification of this issue. We have removed the objectionable sentences and instead added a sentence that is nearly identical to what the Reviewer suggested (with a reference to Saini added as well), as follows (page 18): "Further, a limitation of this analysis is that we assumed that the non-reporting of "AE" data was the result of either no events or non-measurement of the outcome. If a selective non-reporting mechanism were present, such that the AE data was measured but not reported based on the results, then there is the possibility that we may have underestimated the harm of intervention. (45)"

Yours truly,

Deborah Korenstein, on behalf of the authors