Reviewer Response

Dear Dr. Orenstein,

Thank you for sending review reports for our manuscript.

We have revised the paper in light of the reviewers' comments. In particular, we have revised the tables for clarity, and improved their captioning.

We enclose a revised version of the manuscript (clean and with tracked changes), which addresses the reviewers' comments and suggestions. The reviewers' comments are shown below with our response to each comment in italics.

We hope that the manuscript is now suitable for publication in Vaccine.

Yours sincerely,

Sam Abbott

Authors' response to Reviewer #1:

The authors have addressed all of my comments. The additional detail about the study methodology and revisions in the presentation of results represent significant improvements to the manuscript.

Thank you again for your detailed comments on our manuscript. We are pleased that we have been able to address your concerns and agree that the revisions have improved the paper.

Authors' response to Reviewer #3:

The authors have done responded to the previous reviewers' comments as well as can be expected. Despite the use of multiple imputation techniques using MICE, the sheer amount of missing data that are so vital to the core questions being asked is worrisome in terms of outcomes. As the authors note, this is the nature of "real world" studies and there likely is no better way to get better answers.

Thank you for your helpful comments on our manuscript.

I have some issues that were not addressed in the original reviews:

1. This is a "pet peeve" of mine. Study authors act and speak as if there has been only one BCG vaccine. I know of no other vaccine that exists in multiple preparations that is spoken of or analyzed as a single vaccine. In fact, there have been dozens of BCG vaccines over the decades and detailed analyses and studies have shown that they vary greatly in many physical and immunological characteristics. The authors of this study also "speak" as if there is only one BCG strain. In fact, just within the U.K. there have been several different strains used over time and given at different ages. In addition, a large proportion of the subjects of this study received their BCG vaccine in another country. At a bare minimum, the authors need to recognize this fact and address it as a study limitation. I could not tell if any analyses were performed comparing country of origin; for several countries the "N" is quite large and the analyses would have some significant power to detect BCG strain differences. I realize that it is not possible to identify the exact strain of BCG vaccine used at different times in different countries, but this might give some insight into possible BCG strain-related differences.

We agree that this needed clarification as it is rarely discussed. We have addressed this as a study limitation by adding the following to the discussion:

"*We could also not adjust for the BCG strain each individual may have received, the BCG strain used may vary both temporally and geographically.*"

In this study we did not conduct any analyses comparing country of origin. However, it would be interesting to do a follow up analysis looking at between country variation as a proxy of BCG strain. We have added the following to the final paragraph of our discussion highlighting that this is a potential next step.

"Between country variations in the strength of associations could also be explored as a proxy to BCG strain using sub-group analysis."

2. I like to "follow the data" within the paper but found the tables to be nearly impossible to follow; in fact, I still don't really understand them because of how things are labeled and presented. I apologize if I am just missing something, but I have no idea to what the "Yes" and "No" refer. I have reviewed a lot of papers and if I have to spend this much time trying to decipher the tables, then there is likely a better way to present the data. For example, in Table 1, there is a heading "Death due to TB (in those who died)"; below that are the rows "Yes" and "No". I thought that the yes-no referred to whether they died of TB, but then I got to Table 3 where the numbers next to "Death due to TB (in those who died)" are used. Sorry but I just cannot follow it - makes no sense to me. Perhaps the problem is how things are labeled and formatted but please try to make this easier to understand. In addition the format {% all cases}(% complete within vaccine status)[complete within category] is also very confusing.

We are sorry that the table formatting was difficult to interpret. We have added the following to the table captions for tables 3 and 4 to clarify the cases and cases with outcome columns:

"Cases represents all notifications with complete data and a given BCG status, regardless of outcome. Cases with outcome is similarly defined but includes only cases with the specified outcome."

We have also clarified the definition of sputum spear status in Table 1 and UK birth status in Table 2.

We agree that the use of the {% all cases}(% complete within vaccine status)[complete within category] structure in Table 1 and Table 2 adds complexity but feel that this is justified as missing data is an important consideration for this study. We would welcome suggestions for improvements.

3. In Table 2, please explain to what "IMD rank" refers. This should be presented as part of the table.

Thank you. We have added the following definition to the table footer:

"Index of Multiple Deprivation (2010) categorised into five groups for England"