Reviewer Response

Dear Dr. Orenstein,

Thank you for sending review reports for our manuscript. The reviewer’s comments were very useful, and we have implemented the majority of them.

We enclose a revised version of the manuscript (clean and with tracked changes), which addreses the reviewer’s comments and suggestions. The reviewer’s comments are also shown below with our response to each comment.

Yours sincerely,

Sam Abbott

## Authors’ response to Reviewer #1:

Thank you for your detailed comments on our manuscript. We have addressed your concerns below.

### Summary

1. “The data set used has a very large amount of missing data that has high potential to introduce bias into the results. About 1/3 of people studied had unknown BCG vaccination status and those with unknown vaccination status had much higher mortality than people in whom vaccination status was known (11% vs 3%). Of those that died and had known vaccination status, a further 38% of people did not have a known cause of death. Because this study only analyzed cases with complete data, conclusions are based on a subset of TB cases that may not be representative of the population as a whole.”

We agree that missing data was an issue in this study, particularly missing BCG vaccination status. To address this we conducted all analyses in this study using multiple imputation (imputing all outcomes and exposures). We also discussed results from the complete analysis in the context of the results from the analysis with multiply imputed data. However, we were unable to fully account for the impact of a missing not at random mechanism for BCG vaccination. This was therefore listed as a limitation in the discussion:

“Variable data completeness changed with time, with both BCG vaccination status and year of vaccination having a high percentage of missing data, which may not be missing completely at random. We therefore checked the robustness of our results with multiple imputation including regional variability, however an unknown missing not at random mechanism, or unmeasured confounding may still have introduced bias. We found a greatly increased risk of all-cause mortality for those vaccinated more than 10 years ago in the analysis with multiply imputed data, compared to the complete case analysis. This is likely to be driven by a missing not at random mechanism for years since vaccination, with older cases being both more likely to have been vaccinated more than 10 years previously and to also have an unknown year of vaccination.”

### Highlights

1. “It would be more helpful if this section listed exactly what was found in the study and how it contributes to the literature rather than make general statements about what was explored.”

Based on the requirments laid out on the Vaccine help page (<https://www.elsevier.com/authors/journal-authors/highlights>), we feel that this section lists the results as presented in this study.

### Introduction

1. “I suggest the introduction section discuss in more precise terms exactly what is known about the efficacy of BCG vaccination and what questions remain unanswered that this study is trying to address.”

We acknowledge that the efficacy of BCG vaccination is important to include as context for our study. We updated paragraph two of the introduction to explicitly state that the BCG vaccine has been shown to be highly effective in England, and that its effectivness may be correlated with distance from the equator. This now reads as follows.

“BCG’s primary mode of action is to directly prevent the development of active, symptomatic disease. Its efficacy in adults is context specific, with estimates ranging between 0% and 78%.[3] It has been shown to highly efficacious in England and there is some evidence that efficacy increases with distance from the equator.[7] Efficacy has been shown to be dependent on previous exposure, with unexposed individuals receiving the greatest benefit.[4] Unlike in adults, BCG has consistently been shown to be highly protective against TB and TB meningitis in children.[5,6] For this reason the majority of countries that use BCG vaccinate at birth.[7,8]”

We feel that the final paragraph of the introduction details which questions remain unanswered - this is included below in response to your next comment.

1. “There is frequent mention of in the introduction (and throughout this manuscript) of the”additional effects" of BCG aside from reducing the incidence of active TB. However, this term is very imprecise, and it makes it difficult to interpret exactly which additional effects the authors are referring to. There are numerous studies showing BCG has many such additional effects, particularly reducing the development of TB meningitis and disseminated TB in young children. The authors need to more clearly describe which “additional effects” of BCG vaccination are not known and how this study can help us understand them better."

We agree that this needed clarifiction. In point 4 below, we have added an explanation of additional effects in the context of TB outcomes. We feel that the following sentences (page 6, lines 19-23 and page 7, lines 1-3) complete this definition:

“There is some evidence that BCG vaccination induces innate immune responses which may provide non-specific protection,[11] TB patients with BCG scars were found to respond better to treatment with earlier sputum smear conversion,[12] and there is evidence to suggest that BCG vaccination is associated with reduced all-cause neonatal mortality[13,14] and both reduced TB[15] and all-cause[16] mortality in the general population. Given that the immunology behind TB immunity is not well understood these findings suggest that BCG may play a more important role in improving TB outcomes than previously thought.”

This was the only use of this phrase in the manuscript.

3.“The statement on line 7-9 that the emergence of drug-resistant TB has made BCG the best available vaccine needs further justification. I am unaware of any vaccine candidates that have differential vaccine efficacy based on MDR status.”

We agree that this sentance was unclear. We have reworded it to the following:

“However, the lack of a more effective vaccine and the emergence of drug-resistant TB strains means that BCG vaccination remains an important tool for reducing TB incidence and mortality rates.”

1. “I would suggest revising the statement on lines 18-19 that”Vaccination policy has been primarily based on reducing the incidence of active TB and 19 little attention has been given to any additional effects of BCG" be revised. Vaccination policy is based just as much on the attenuation of disease severity (less meningitis and disseminated TB in kids) as it is reducing the overall incidence of TB."

We agree that this was an omission. We have amended this sentence to the following:

“Vaccination policy has been primarily based on reducing the incidence of active TB, and mitigating disease severity, with little attention having been given to any additional effects of BCG vaccination on TB outcomes.”

1. “I would recommend the introduction section end with a clear statement of the study question.”

The primary aim of the study is detailed in the penultimate sentence of the introduction:

“We aimed to quantify the effects of BCG vaccination on outcomes for individuals with notified TB in England using routinely collected surveillance data to provide evidence for appropriate public health action and provision.”

The final sentence details the secondary aim of the study:

“Where we found an association, we additionally explored the role of years since vaccination, and age at vaccination.”

### Methods

1. “A much more detailed discussion is needed to understand how the outcome variables differ from one another and how the information was collected. Specifically, a clearer description of how”all-cause mortality" and “TB-related mortality” were collected is needed."

Asked Maeve for details. Alternatively review PHE report.

1. “How long were most people followed to assess for all-cause mortality? The table indicates people were”followed for 12, 24, and 36 months (or until treatment completion)“. Given patients with drug susceptible TB are usually only treated for 6 months, it seems few people would have been followed for this amount of time.”

Asked Maeve for details. Alternatively review PHE report.

1. “It is also unclear what criteria were used to determine whether the death was due to TB and who made the assessment.”

Asked Maeve for details. Alternatively review PHE report.

1. “There is no information included as to how information about BCG vaccination was collected. Was this entirely by self-report? If so, was there any effort made to ensure the information provided was accurate (e.g. cross-referencing vaccination records), particularly the age at vaccination?”

Asked Maeve for details. Alternatively review PHE report.

1. “For the multiple imputation procedure, which variables were imputed? Was BCG vaccination status imputed for this model when missing?”

We updated the methods section to highlight that all outcome and explanatory variables were included in the multple imputation approach:

“To mitigate the impact of missing data we used multiple imputation, with the MICE package.[27] We imputed 50 data sets (for 20 iterations) using all outcome and explanatory variables included in the analysis as predictors along with Public Health England centre. The model results were pooled using the small sample method,[28] and effect sizes compared with those from the main analysis.”

1. “I would consider discussing the outcome variables in the text rather than listing them in table format. As above, the information in the tables does not provide much clarity about the outcome variables.”

We have moved the outcome variable definitions from a table into the main text of the manuscript. Including the modifications from comments 1., 2., and 3. above this now reads as follows:

### Results

1. “This manuscript contains a large number of tables and many of them seem to be redundant. I would suggest combining some of the tables so more information can be included in the main manuscript rather than the supplement. For example, Tables S1 and S2 contain information that would normally be a part of”Table 1“. I would consider combining these tables to make a new Table 1 so readers can more clearly see how outcomes and covariates differed based on BCG vaccine status.”

We have moved both S1 and S2 into the main body of the text. The tables are presented seperately to make it clear that not all outcomes were considered in the same model.

1. “I suggest including the total n used for each analysis in all tables.”

We have included both the univariate and multivariate sample size for all analyses that use complete case analysis. For analyses that use mutltiple imputation, all samples were used. We have now highlighted this by adding the following to each table using multiply imputed data (stars indicating data that varies by table): “Sample size: \*\*\*“.

1. “In supplemental table S9, the post-hoc power calculation is not necessary. This is simply re-stating the primary analysis. By definition those analyses that did not reach statistical significance were not adequately powered to detect statistical significance given the observed effect size.”

We agree that this is not required and have removed it from the supplementary information.

1. “Since there were an unusually high number of deaths determined to not be due to TB in this population, it would be helpful to include the other most common causes of death.”

Unfortunately causes of death that were not due to TB, or related to TB were not recorded in our dataset, so we are unable to add this information.

### Discussion

1. “Much of the discussion section is a re-statement of the results. I would suggest the authors focus more on interpreting the results from the study and describing how they relate to the published literature.”

We agree that a large part of the discussion is devoted to discussing the results of this study. This mostly relates to the limitations of the study, and the approaches that we took to mitigate them, that we felt needed to be discussed in order to put our results into context. That being said, we feel that paragraphs four and five of the discussion interpret the results from this study and relate to the published literature.

1. “This study had a large amount of missing data for the exposure and outcome variables of interest and this could be a source of significant bias. This is a major limitation of the study and warrants a thorough discussion (usually the 2nd to last paragraph).”

We agree that the proportion of missing data is a significant limitation of this study (as with many studies using routine surveillance data). As outlined above, to explore the impact of this we repeated all analyses using multiple imputation. This is discussed, along with the remaining limitations in the following sections:

“Variable data completeness changed with time, with both BCG vaccination status and year of vaccination having a high percentage of missing data, which may not be missing completely at random. We therefore checked the robustness of our results with multiple imputation including regional variability, however an unknown missing not at random mechanism, or unmeasured confounding may still have introduced bias. We found a greatly increased risk of all-cause mortality for those vaccinated more than 10 years ago in the analysis with multiply imputed data, compared to the complete case analysis. This is likely to be driven by a missing not at random mechanism for years since vaccination, with older cases being both more likely to have been vaccinated more than 10 years previously and to also have an unknown year of vaccination. The high percentage of missing data also means that we were likely to be underpowered to detect an effect of BCG vaccination on sputum smear status and deaths due to TB (in those who died), with years since vaccination, and age at vaccination likely to be underpowered for all outcomes”

## Authors’ response to Reviewer #2:

Thank you for your review of our manuscript. We have addressed your comments below.

1. “Given the recent revaccination results, it would be good to see a breakdown of the effect of primary vaccination and revaccination on outcome. It is not clear from S4, for example, whether the non-neonatal persons were revaccinated.”

We agree that it would have been interesting to explore the impact of revaccination on our findings. Unfortunately, we were not able to investigate this as we did not have data on revaccination.

1. “I am a bit baffled by examining the duration of effect data according to a discrete variable: more or less than 10 year since vaccination… why was age since vaccination not used as a continuous variable in analysis?”

We categorised duration of effect for two reasons. Firstly, to minimise recall bias. Our justification for this is that recall bias is likely to have an increasing impact with time since vaccination. This would mean that those vaccinated several decades ago would be more likely to report an incorrect vaccination date with greater error than those vaccinated more recently. Secondly, any relationship between outcomes and years since BCG vaccination is unlikely to be linear over the timescales considered in this study. We have now highlighted this in the manuscript with the following statement in the methods secton:

“Years since BCG vaccination was defined as year of notification minus year of vaccination and categorised into two groups (0 to 10 and 11+ years). This was based on: evidence that the average duration of BCG protection is 10-15 years;[15] concerns that recall bias maybe greater when vaccination took place over a decade ago compared to within the last few years; and any association between years since vaccination and TB outcomes may be non-linear.”

1. “The outcome measures used for this study appear randomly chosen - there could be many additional outcomes of TB disease. Such as extent of pulmonary disease on chest roentgenograms.

We agree that there could be additional outcome variables included in this study. Unfortunately, we could only include outcomes that were included in the routine surveliance dataset on which this study was based. Given the availability of outcome data, we selected outcomes for further evaluation based on evidence from the literature of prior associations with BCG vaccination, associations with TB infectiousness, or poor consequences for patients. We have now highlighted this in the manuscript with the following statement in the methods section:

“We considered five TB outcomes that were selected based on: their availability in the ETS; evidence from the literature of prior associations with BCG vaccination; associations with increased case infectiousness; or severe outcomes for patients”

1. “Similarly, was a specific analysis done to look at extrapulmonary, disseminated or tuberculous disease?”

This would be an interesting follow up study as it is plausible that some of the effects we have identified would differ by disease type. However, we did not pre-specify this in our analysis plan and therefore did not explore this in the current study. Additionally, given the low sample size for some outcomes, analyses on disease subset are likely to be underpowered.