Reviewer Response

Dear Editor,

Thank you for sending review reports for our manuscript. The reviewer’s comments were useful and have greatly improved the paper.

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

Yours sincerely,

Sam Abbott

## Authors’ response to Editorial Requests

GENERAL

* In both the abstract and the main text please define all abbreviations at first use (e.g. UK; gradient boosting method – is this GBM?; 95%CI – is CI confidence interval? Credible interval?; MNAR in the discussion – is this missing not at random mechanism?). For Tables and Figures, please define all abbreviations used in a footnote below the Figure or Table.

*We have reviewed all abbreviations and added definitions where missing. Tables and footers have been checked for abbreviations without explanations and these have been added when missing. The abstract has had explanations for the abbreviations added where missing.*

* If possible, please use square brackets for the references with the punctuation following the reference.

*We have moved to using square brackets and moved references from following punctuation to punctuation following the reference.*

* Tables should not have any empty cells as design element or because information is not available (‘NA’ (subsequently defined as not available or not applicable) can be used for example).

*We feel that the table design we have used clearly imparts the information that it contains - part of this design is empty cells to space information. NA has not been used as information is not missing in the conventional sense - the results tables combine two different models with different confounders (hence not all fields contain data). Having this information in a single table allows for easier between model comparisons which we feel is essential to the presentation of our results. We have also included all results in csv form for readers who wish to use our results programmatically. We have clarified this in the data availability section. However, we are happy to alter the table design if we can preserve the clarity of the results presentation and welcome your suggestions.*

* For all Figure and Table titles, when possible, please indicate at the end of the title, the geographical area, period and total number of items covered by the illustration. E.g. Table 1: Summary of relevance and eligibility criteria for each cohort, United Kingdom, 2000-2010 (n = xxx individuals)

*We have added location and temporal information to all tables/figures. As this was a retrospective whole population based study we are not clear that sample sizes are useful in this instance (as they are the whole population meeting the cohort critiera which is age based). We are happy to add this if your feel this is meaningful.*

* Please check the overall text versus the tables for the name of the intervals (e.g. 95% confidence intervals vs 95% credible intervals). To give an example: In the abstract and in the main text subsection called ‘Adjusted estimates of the effects of the change in policy on school-age children’ there is ‘IRR: 0.74 (95%CI 0.61, 0.88)’ but in the table 2 there is ‘0.74 (0.61, 0.88)’ defined as IRR (95% CrI) CrI: Credible interval.

*We have standardised all usage across the text.*

ABSTRACT

* While keeping the word count below 250 words, please add an ‘aim’ subsection, describing briefly the aim of the paper, and please specify the study type (e.g. observational study?) also in the abstract (either in the aim or the methods).

*We have added an aim section that includes the aim of the paper and the study type. We have also made the required cuts to the abstract in order to keep within the word limit.*

METHODS

Data sources - Does the Labour Force Survey record only the number of adults of working age (employed or unemployed) in the country, or does it record more people than these (e.g. their family members such as children)? Please describe the population and demographics recorded in the Labour Force Survey in more detail.

*The LFS is the largest household study in the UK. Whilst it is used primarily to estimate unemployment rates it is also designed to be used by local governments etc. to understand the structure of their population. It therefore provides some of the best available yearly population estimates that are stratified by ethnicity, UK birth status, and age. A caveat to this is that it may not capture the older adult poplulation well. However, this was not relevant to this study. We have added these additional detials to the manuscript.*

Statistical methods

* While it is mentioned in the methods that 95% confidence intervals were calculated around the IR values, it is not described what the intervals are around the IRRs. Please check the earlier editorial comment about the intervals that are presented in text and tables around the IRRs.

*We have added this and checked all tables and figures to make sure we are consistent.*

Ethical statement

* An ethical statement should be included (<https://www.eurosurveillance.org/for-authors#ethical%20statement>); if ethical approval was not needed, this should be stated in the manuscript. In this case, if possible, it should be please explained why.

*We have added an ethical statement*

RESULTS

= ‘Of notifications with a known birth status 27% (29096/106765) were UK born, in comparison to 33% (2634/8055) in cases with an imputed birth status.’ Just for editor’s understanding: does this mean that ‘Among notifications with a known birth status, 27% had the status ‘born in the UK’, while among notifications with an imputed birth status, 33% had the status ‘born in the UK’?

*Yes that is correct - we are open to rephrasing this if additional clarity is required.*

## Authors’ response to Reviewer #1:

This is an interesting paper in which the authors address a key TB control issue which is measuring the impact of BCG policy change in the incidence of tuberculosis in children. Whether and when targeted BCG vaccination should replace universal vaccination depends on several factors and may be a matter of discussion within national TB control programs, although in Europe most countries have a targeted BCG policy. The question is worthwhile investigating and discussing. I suggest publication in Eurosurveillance provided that the minor changes suggested are taken into account. HOWEVER, given the importance of the statistical methods in this analysis, which I do not feel confident to analyse, I recommend that a tough review of this part, and the results, be made by a statistician, which opinion is crucial to decide whether the paper is able to be published or not.

*Thank you for your helpful comments. We agree that understanding when/if targeted vaccination should be used rather than universal vaccination is a key TB control issue.*

Introduction

1. Paragraph 2. In France, universal vaccination was implemented in all children before entering “community care” and primary school at the latest, and not in neonates.

*We have corrected this.*

Methods

1. Paragraph 1. The geographical scale of TB local incidence is not specified and is an important piece of information that would be worth mentioning at this stage.

*We have added that TB incidence is highly heterogeneous in England with the majority of cases occurring in urban, non-UK born, populations.*

1. Paragraph 2. Not clear whether the LFS which reports “employment circumstances” of the UK population reports also data in children, who are not employed. We assume the answer is yes which should be probably clarified.

*We have added additional details to this section. Please see the responses to the editors comments for an expanded discussion of the LFS.*

1. Definition of cohorts. Not totally clear to me. Example for comparison cohort 2: a child born in 2000 and who is diagnosed with TB at 2 years old (2002): the vaccination programme in 2002 is universal not targeted, as seen in Table 1. Maybe further clarify this crucial point.

*We agree that this is somewhat difficult to interpret as the underlying cohort structure is complex. The cohorts are first split into targeted and universal vaccination based on age. Eligibility is then used to indicate if a given cohort was actually enrolled on a given programme. This gives a cohort who were enrolled and a comparison cohort who were not (resulting in a binary outcome). We have added additional detail outlining this to the paper.*

Statistical Methods

This is a key point that should be reviewed by a statistician; I do not feel confident on giving my opinion on this.

Results

1. No comments, except that I find some parts of the results difficult to understand for a “common” public health professional" with no deep knowledge on statistics and modelling methods.

Discussion

1. BCG protects mainly against severe disease (meningitis). This point is not discussed at all. An additional analysis could have investigated the impact of policy change on the incidence of meningeal TB in children, which could have broaden the consequences of policy change in the epidemiology of TB in children. Maybe a word on this should be said in Introduction or Discussion.

*We agree that considering disease severity stratified by age is an important issue and could be addressed in a follow up study. We have added comments detailing this to the discussion. We disagree on the evidence that the BCG protects mainly against severe disease in the UK where there is good evidence of strong protection against all disease with no evidence of protection weakening with increased age at vaccination. However, we do agree that this is relevant in other contexts.*

1. Other interventions may decrease TB incidence, such as a better screening, a change in screening policy in migrants, a better case management…This is not discussed. Could these potential interventions have also played a role in the results presented here?

*We agree that this is a potential issue and one that we could not fully account for. We have partially accounted for it by adjusting for between year cohort differences but this does not account for interventions that changed at the same time as vaccination policy. We mentioned this limitation in the discussion,*

*“Finally, we have assumed that the effect we have estimated for the change in BCG policy is due to the changes in BCG vaccination policy as well as other associated changes in TB control policy, after adjusting for hypothesised confounders. However, there may have been additional policy changes which we have not accounted for.”*

*We have now added additional details to the conclusions stressing this point further.*

1. No figures of BCG vaccination coverage trends in targeted groups are given (except the 68% mentioned in the Discussion). This determines to a large extent the impact of vaccination. A brief description of this could be given here, or in Introduction.

*We agree that vaccination coverage impacts the success/failure of vaccination programmes. Unfortunately, coverage estimates for the targeted vaccination programme are of very poor quality. This was part of the motivation for using a population based approach that looked at overall incidence rates and hence did not need to account for vaccination coverage. We agree that understanding coverage in targeted populations should be a focus for further work in this area.*

1. BCG has an impact mainly in severe forms of TB, and prevents about 50% of pulmonary TB in children. But children, to my knowledge, do not frequently transmit TB. So BCG has rather poor effect on TB transmission. The authors speak in their discussion on indirect effects and ongoing transmission. Is this really the case with BCG and TB?

*This is an interesting point and the motivator for some of our current research. Whilst the reviewer is correct that children rarely transmit TB this is not the case for adults. This means that school-age vaccination may lead to fewer overall cases in children than direct vaccination of children once onwards transmission has been accounted for due to increased protection from TB as young adults. This depends on the level of TB transmission in the population and therefore may not be relevant in the UK where transmission is generally thought to be low. As per our previous comment we disagree with the reviewer that BCG mainly impacts severe forms of TB in the UK. For areas where this is the case we agree that direct vaccination of children is more clearly beneficical. However, as effectiveness when given later in life varies so too will the thresold that school-age vaccination becomes more effective.*

## Authors’ response to Reviewer #2:

Thank you for providing an excellent overview on the effects of the policy changes in BCG vaccination in England.

*Thank you for this review.*

## Authors’ response to Reviewer #3:

This study is of high relevance, not just for the England, but for the global tuberculosis community: The UK (England and Wales notably) had been one of the few countries with a universal vaccination of adolescents and a clinical trial demonstrating high efficacy for at least 20 years in the age vaccinated groups and beyond to include young adults during follow-up.

From the switch from universal adolescent vaccination to selective (largely the descendant newborn of foreign-born ancestors) one would thus hypothetically expect an increase in the incidence rate among the former and a decrease in the incidence rate among the latter (among the respective cohorts in both groups), if confounding resulting from secular trends could be properly controlled for. This study aimed to do precisely that.

*Thank you for this detailed and helpful review. We agree that the findings from this study should be relevant to the global TB community. We particularly appreciate your suggestions for improving the discussion of the approach used.*

The authors created 8 cohorts, each for a 5-year follow-up, using tuberculosis surveillance data for the numerator and Labor Force Survey data for the population denominator. These procedures are all clearly explained. As it gets substantially steeper and difficult to follow for the more general user’s consumption is when the models are elaborated in the same paragraph. Perhaps it would help to split the elements of describing the basic epidemiology of incidence rates and ratios from the latter into two separate paragraphs, so that those unfamiliar with “negative binomial models”, “penalised by model complexity”, and the intricacies of Bayesian estimation with Markov Chain Monte Carlo, etc. This would clarify that here is a basic methodology, and here is a complex modeling description, the former for every epidemiologist, the latter only for the seasoned expert in statistical modeling using special techniques. I concur that it should and has to be there, but the more basic working techniques of calculation might be preferably “visually” separated from complex modeling techniques, notably when unexplained acronyms such as “GBM” are interspersed that do not help in easily following what the modeler actually did. Nor does it become that clear why exactly these models were chosen and not something else. The reader should probably have some inkling what the motivation for model selection is.

*We agree that it is important that all readers can easily follow the analysis. We have split the modelling section as you suggested into a non-technical overview and a more technical discussion of the approach. As per the editor comment we have carefully reviewed the use of abbreviations and added definitions where missing.*

In any case, it becomes again clear that the procedures resulted in obtaining adjusted estimates of the effects of the policy change from universal adolescent to selective neonatal vaccination. The effect was measurable but seemingly modest. Apparently, the model required adjustments by including the change in policy, age, incidence rates in the non-UK born, and year of eligibility as a random effect. To the non-initiated, all these elements make intuitively sense and reflect the complexity of interplaying variables that all played a substantial role in what was finally observable. Such complexity may also contribute to variability and thus uncertainty about the unbiased estimation of the magnitude of the observable effect. However, the authors later also state that all models (does this refer only to the non-UK born?) that included at least age in the adjustment resulted in comparable effects. As parsimony is king in all modeling, perhaps it could be stated more explicitly that always the model with the minimum required variables was used and perhaps why recourse was taken to retaining more complex models if a more simple one gave actually comparable results. Perhaps I am missing here something, but then it would indeed be useful to review the sum of the modeling exercises to the point.

*We agree that parsimony is an important goal of modelling. However, it is also important to adjust for confounding in the data. Our approach used a hypothesis free (excepting our initial use of domain knowledge in selecting an overall pool of confounders), statistically rigorous metric (the leave one out information criterion) to select the model that best fit the data after being penalised for model complexity (to avoid overfitting). We hope that our discussion of this approach is clearer now that we have adopted your suggestion above of splitting the statistical methods. We included a discussion of other models as it is gives an indication of which variables greatly altered the results and may be used when interpreting the findings. It is also important to include to make it clear that we explored multiple models (doing otherwise might be considered p hacking).*

In table 3, among the other variables, is also the year of study eligibility (please correct spelling). The point estimate incidence rate ratio varies from 0.83 to 1.12 (UK born) with no discernible trend. Is this therefore an expression of the intrinsic uncertainty of the effect of the model? If so, then perhaps the authors would have to be more explicit that there remains substantial uncertainty in the modeled effect as small as it may be. Thus perhaps the variability in the estimated magnitude of effect for cases prevented in the following paragraph which is substantial is therefore not all that meaningful after all (CrIs not uncommonly overlapping zero).

*We feel that adjusting for year of study eligibility helps to account for potential differences between cohorts. We would not expect to see a trend and if we did this might indicate that we had missed adjusting for a variable that may be introducing systematic bias. We feel that we have satisfactorily captured uncertainty using our modelling approach and that therefore the CrIs are representive of the underlying uncertainty. We have corrected the spelling of eligibility.*

Thus, in the end in the discussion, the authors speak - statistically probably correctly - of “evidence of an association”, but perhaps disappointingly it is not as clear-cut as one may have hypothesized (including this reviewer’s pre-reading subjective gut feeling). Perhaps there is more not easily definable confounding that couldn’t be addressed in the model. Because it seems indisputable that at least the BCG / Vole bacillus UK trial had given unusually solid long-term protection against pulmonary tuberculosis among young adults and it remains baffling that stopping this scheme of adolescent vaccination should have left such little effect. Would the reasons for this phenomenon deserve a bit more extensive addressing in the discussion? That BCG vaccination has given various levels of protection (say, nothing in Chingleput) seems entirely irrelevant for the UK, here it gave a non-varying high-level of protection for about ten to 12 years before it precipitously dropped towards zero. This fits with what we know from effector cell immunity and it is not clear why this should not also be reflected more substantially in the period following policy change unless the models couldn’t really account for all factors that matter, but it is surely not variability in vaccine protection elsewhere that is an issue here, this is only about the UK after all.

*We agree that our conclusions were not as clear-cut as might be expected. However, part of this is due to comparing two vaccination programmes rather than using a baseline of no vaccination. We were in effect estimating the incremental difference between the two programmes and it is perhaps to be expected that this difference was small. In addition much of the uncertainty is due to incidence in neonates being very low in England throughout the study. It should also be noted that stopping universal school-age vaccination did result in increased incidence rates overall when only the UK born were considered (as expected). It required the inclusion of the non-UK born in order to alter these findings. The non-UK born were less clearly exposed to UK vaccination policy and as the reviewer has noted this impact may be due to other policy changes. We feel that the reviewer suggested changes to the discussion has helped highlight this. A final point is that we were only able to use 5 years of follow up. It may be the case that extending follow up alters these findings. We have included this point in the discussion.*

I appreciate the work of the authors with these thought-provoking study results!

*Thank you again for your useful comments.*

Minor comments:

Several references have capitalization vs lower case issue be in names or e.g. “bcg”, and also whether words in the title should be capitalized or not, etc.

*Resolved issue with BCG. Use of title case is a journal dependant decision and therefore we have not standardised.*

References 7 and 8 are duplicates

*Resolved.*

Reference 11 is a bit of a screw-up (2018 and January-2011)

*Resolved*