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

* Please include the study design in the title.

*We have updated the title to,*

*“Reassessing the evidence for universal school-age Bacillus Calmette Guerin (BCG) vaccination in England and Wales: re-evaluating and updating a modelling study”*

* In the patient and public involvement statement, please clarify that patients and the public were not involved in the design or planning of the study.

*We have revised the patient and public involvement statement to the following,*

*“We did not involve patients or the public in the design or planning of this study.”*

* Please complete and include a TRIPOD checklist, ensuring that all points are included and state the page numbers where each item can be found. The checklist can be downloaded from here: <http://www.equator-network.org/reporting-guidelines/tripod-statement/>

*We feel that the TRIPOD checklist is not completely appropriate for our study because we use an infectious disease model rather than a statistical model designed to be used as a tool for prediction, thus several of the checklist items are not relevant. However, we would be happy to fill in this or any other checklist if required.*

## Authors’ response to Reviewer #1:

This study revisited the model used for driving the change of BCG vaccination policy in 2005 in the UK. Assumptions on the declining trend of TB burden were evaluated and found to be an important source of uncertainty in estimating the impacts of BCG vaccination. This work will contribute to the understanding of appropriate evidence needed for planning BCG vaccination strategies. My comments for the manuscript are listed below.

*Thank you for your helpful comments on this paper.*

### Introduction

1. Page 4, Line 21: BCG efficacy has been known with great heterogeneity and it affects the assessment of vaccine impact on TB burden. It is therefore commonly considered in decision-making for BCG vaccination policy. I suggest that the authors include the uncertainty of BCG efficacy in this analysis and compared the relative importance to the assumption of how TB burden declines over time.

*We agree with the reviewer that this is an important issue. We have included some of this uncertainty by re-estimating incidence rates for both those BCG vaccinated and unvaccinated. These rates are then used to estimate vaccine efficacy. This approach makes use of the available data on the effiacy of the BCG vaccine in England and Wales. Including additional uncertainty observed globally would be inappropriate here as there is little evidence for this amongst the White UK born population. This does mean that the findings from this model reflect the impact in these countries only and cannot be directly generalised.*

1. Page 4, Line 36: The abbreviation should be ‘JCVI’ rather than ‘JVCI’. Note that the same typo appears in other paragraphs of the manuscript as well as the figure captions.

*Thank you. We have corrected this throughout*

### Methods

1. Page 5, Line 6-7: Is the ‘transmission chain model’ included in the Sutherland et al.’s model or developed by the authors? Please state it here.

*We have extensively reworded and extended this section (please see the marked up version). As part of ths we have added the following sentence for clarity,*

*“The TB transmission model is defined as follows (reproduced from Sutherland et al.)”*

1. Page 5, steps for defining the transmission model: Presenting equations of the model is certainly helpful for understanding the estimation methods. In addition to the descriptions of equation components, I encourage the authors to explain more about the rationales behind. For example, how to interpret ‘initial generation size (x)’? What is the purpose of relating step 1 to step 2? This kind of additional information will allow readers from a wide range of background to understand the model easily.

*Whilst this information was present in the original Sutherland et al. paper we agree that more detail was needed here. We have therefore added greater explanatory detail for each step (please see the marked up paper).*

1. Page 5, last paragraph: Is the ‘distribution of secondary cases (N)’ estimated by the three age groups (of 15-19, 20-24, and 25-29 year old)? The power term ‘5-z’ of the decay percentage (d) needs additional explanations in the text.

*Yes the distribution of cases is estimated by age group. We have reformulated these equations for greater clarity (please see the marked up paper). Note that due to paper changes this section of the model has been moved to the SI.*

1. Page 5, Line 54-55: Please revise the sentence fragment.

*Revised to:*

*“We validated the fitted model by comparing the results with those from the original implementation using the mean absolute percentage error, normalised by the original estimate, as the performance metric.”*

1. Page 6, Line 10-11: Is the serial interval defined as between linked ‘notifications’, instead of ‘infections’?

*Yes, thank you - corrected.*

1. Page 6, Line 28-32: This part is too long to follow and with sentence fragments. Please re-organise it.

*Reworded to:*

*“Data collection for the ETS began in 2000 and prior to this notification data was only available in years with notifications surveys (1973, 1978, and 1983). We therefore estimated incidence rates between 1984 and 1999, and for the years between notifications surveys (1974-1977 and 1979-1982), using locally estimated scatterplot smoothing (LEOSS) regression fitted to incidence rates published in [11] and the estimated incidence rates from 2000 on-wards. LOESS is a local regression method that combines multiple regression models in a k-nearest neighbours meta-model.[@FoxSnow:2019] This approach allows nonlinear trends to be fitted using a series of linear models. For years prior to 1973 the annual decreases were assumed to be the mean of the annual decreases from the previous 3 years. For both proxy measures the annual decreases in incidence rates post 2016 were assumed to be the average of the estimates in 2013, 2014, and 2015.”*

### Results

1. Page 8, Line 8-10: ‘The large degree of uncertainty’ was related to the different assumptions of the annual TB decrease. However, here it sounds like the uncertainty is in the range of a particular scenario estimate.

*The range of uncertainty here is within the scenario. This is due to the incorperation of uncertainty detailed in the Updating model parameter estimates section of the methods. We have reworded this section for clarity.*

### Discussion

1. Page 8, Line 40-41: ‘A scenario with a 1.9% annual decrease in incidence rates was most comparable to our results based on notifications.’ As the medium annual decrease estimated using notifications was 3.13% (Page 7, Line 20), why the scenario of a 3.9% annual decrease did not give the most comparable results?

*Unfortunately, the summary statistics do not capture changes in the annual decrease in notifications over time. Notifications first increased and then decreased and it is these trends that led to the overall 1.9% decrease being a better fit. To help clarify this we have moved a figure from the SI into the main body of the paper.*

1. Page 9, Line 37-38: Do you consider to report the primary impacts of ending the BCG programme and the additional cases due to transmission separately? This will show the importance of including transmission structure in assessing the BCG vaccination policy.

*We have separated the reporting of primary and secondary additional notifications to highlight the role of the transmission model.*

## Authors’ response to Reviewer #2:

### OVERVIEW

This paper revisits an analysis that was used to support the decision to discontinue the UK policy of BCG vaccination in schools. It considers updated evidence on parameters as well as incorporating uncertainty in model inputs. Re-evaluating modelling around decisions and considering the impact of uncertainty are both under-explored and worthy topics.

At the moment, I think this article needs major revisions before publication as it is extremely difficult to follow exactly what has been done.

*Thank you for your comments - especially the identification of the flaw with the original Sutherland et al. model. Based on them we have corrected the transmission chain model, extended our discussion of the methods used, and added further clarification where possible. Please see our detailed responses below.*

### MAJOR ISSUES

The idea of the approach is that information about the rate of decline of TB incidence over time combined with knowledge of the generation time should inform on the net reproduction number, and hence the scope for indirect benefits from preventing cases. However, I really can’t follow the model description on page 5. Part 1: T (and in part 2) is described as the sum over all generations of resulting cases, but R0^z (presumably R rather than R0) and (1-d)^z seem very much like the expected size of the z-th generation of resulting cases? Part 2: x is described as an initial size, but it seems in fact to be a net reproduction number from the way it is used in a geometric progression (resulting cases should scale linearly with initial size under branching process type models). If x=R and eqn 1 is true, things seem over-determined: one can solve R(z) without need for data on d?? Part 3 makes sense if x is R. Then I also don’t understand how this relates to the 3 equations at the bottom of the page. I also don’t remember the eligibility for BCG vaccination being introduced (line 37), and am not clear what it means to ‘reproduce the distribution of cases due to transmission over time’. I think all this needs much more careful explanation.

*We agree that this needed further explanation and that the definition used in 1. of Sutherland et al. was incorrect. We have corrected the Sutherland model and extended the model explanation to include this.*

I also found it harder than it should have been to understand exactly what was different between this analysis and the original analysis, and which differences mattered most (which is important as it is the main aim of the work as I understand it). I wonder if a table would be useful detailing the differences: any implementation/modelling differences; differences in point estimates of d (both from methods and use of more recent data?); differences in parameters from literature; inclusion of uncertainty in each. I think it’s important to differentiate between differences that could have been included in the analysis at the time (eg uncertainty) vs things that have emerged since (eg updated evidence on serial intervals). I also think it’s important to be clear whether methods, central estimates, or uncertainty matter most from the point of view of decision-relevant changes in the evidence generated.

*We have introduced a subsection of the model explanation detailing our additions to the model. The Updating model parameter estimates and incorporating parameter uncertainty section of the methods contains a discussion of differences between the original and updated analysis. We have reworded this for clarity (please see the marked up paper). As you say, parameter uncertainty for incidence rates could have been included at the time but was not, whilst the updated value for the serial interval has only recently become available. This is now explicitly stated in the reworded methods paragraph.*

*We agree that it is important to be clear which findings are relevant to policy makers and in what context. We have reworded the 4th discussion paragraph to provide more context.*

I found the abstract confusing and not self-contained. “assess the quantitative evidence…” is rather vague and broad compared with my perception of the paper’s aim. I think the fact a model was used to predict impact needs introducing, as it enters rather implicitly. It would be good to say in what ways the analysis was updated (see previous comment). “best matched” is vague. “ended in 2001” - need to have explained when it did end. Why is 2004 important? The results presented here are all about different end-dates for the vaccination policy rather than comparisons with the original analysis’s predictions. “In 2016…notifications would have occurred.” - under what policy change? In the conclusions: it hasn’t be explained that decline in incidence rates is relevant to the modelling, it hasn’t been made clear that parameters have been udated and a comparison with previous results has not been presented. In other words, the conclusion introduces new things, which is a shock, and is not supported by the rest of the abstract.

*Thank you for these helpful insights - we agree that the abstract needed revisions. We have updated it based on your comments (see the marked up paper) and some of the other changes to the paper. Unfortunately, due to word limit constraints it was not possible to fully detail how the model was updated at the same time as detailing the results and conclusions.*

### MINOR ISSUES

* " BCG schools’ scheme" in many places. Is this what it was officially called? I didn’t understand why it had an apostrophe.

*Thank you - corrected throughout.*

* In a couple of places (ie strengths/limitations bullets & Discussion) it is stated this is not “the most accurate method” for assessing the impact of these changes. This does beg questions about what other methods would have been better, and in what respects (eg how is “accuracy” defined?). This is worth discussing. I think it also means that the rationale for the work is about assessing the imapct of updated/modified analysis for inputs within an existing framework, rather than trying to provide a better answer to the same question (otherwise, why use a suboptimal method?)

*We agree that this could have been better phrased. We have rephrased as follows,*

*“The simulation approach used here, although updated where possible, is not the most accurate method for estimating the impact of ending the BCG schools scheme as it relies on numerous assumptions based on the available knowledge in 1987 and does not account for the role of non-White and non-UK born cases. However, the strength of this work is that the estimates are based on the framework used to inform policy making. This allowed the strength of the model used in the decision-making process to be assessed once parameter uncertainty had been incorporated and for flaws in the model to be identified. This would not have been possible if the impact had been assessed using only the observed data or with a alternative model. It also allowed estimates based on updated data to be compared to historic estimates within the same framework. This would also not have been possible if a different framework had been used.”*

*We discuss alternative approaches for assessing the impact in the final paragraph of the discussion but you are correct in saying that the rationale of this work was to assess the impact of updating an existing framework. We have clarified this by adding the following sentence to the introduction,*

*“Re-evaluating this work allows for the strength of the evidence used in decision making to be assessed and may highlight any issues with the approach used.”*

* page 4 line 36: missing apostrophe in “papers”

*Corrected.*

* last 2 sentences of intro are Discussion

*Removed.*

* “TB notification surveys” - could you explain what these are please?

*Added detail to the methods. Please see the marked up paper.*

* ref 17: I thought there had been a systematic review of serial intervals for TB (Ma et al, Epidemiology & Infection). Justify why using this study in particular please. Choice of evidence an important part of paper since arguing to update & compare.

*Thanks for pointing this out. Whilst there has been a systematic review the authors identifed only two papers with estimates (the one used here and one other older paper). Both papers made use of similar datasets in the Netherlands, however, the cited paper is most recent (by 10 years) and has an improved sample size and improved methods compared to the older paper.*

* pg 6 paragraph on methods to quantify decline: again, I found this confusing and it is a key part of the analysis. Also, what is “local regression” please?

*We have reworded this paragraph (and made additional changes suggested by reviewer 1). Please see the marked up paper. We have also included a graph in the results from the SI that shows the changes over time in the annual decline depending on the data/assumptions used.*

*We have added the following clarification for local regression,*

*“LOESS (locally estimated scatterplot smoothing) is a local regression method. It combines multiple regression models in a k-nearest neighbours meta-model. This approach allows nonlinear trends to be fitted using a series of linear models.”*

* pg 7 paragraph on incidence rate changes: many medians presented do not lie between the corresponding 2.5% and 97.5% quantiles, which must be typos. Also there seems to be some confusion about whether to present decreases or negative increases.

*Unfortunately, our choice of “-” to separate estimates of the 2.5% and 97.5% quantiles made this paragraph unclear. We have moved to using “,” throughout the paper as this more clearly identifies when estimates are negative. Thank you for picking this up.*

* NNT vaccination results. I’m surprised by the factor of 10 change in the NNV going from 9% decline to 2% decline. Could this be double checked please? Sanity check: if TB incidence is of the order 10/100,000 and the vaccine better than 50% effective, I imagine NNV is ought to be less than 20,000 considering only individual-benefit (the magnitude calculated for 9% decline); to achieve the stated effect must involve many generations of transmission (perhaps exceeding the time-frame considered?). However, it would be useful to clearly ellaborate this: perhaps a table of these results, with proportion of benefit from individual vs transmission and estimates of R for each decline considered; so the way the numbers hang together can be more clearly understood.

*We agree that these estimates (from the original Sutherland et al. paper) are very unrealistic. Sutherland et al. assumed a year-on-year decline of 9% with an estimated incidence of ~ 23 per 100,000 in 1983 for the unvaccinated population. In 2013 this would give an incidence rate of ~ 2.6. These very small incidence rates mean that the difference between incidence rates in unvaccinated and vaccinated cases is also very small leading to the spuriously high number need to vaccinate. In comparison a 2% decline would give an inidence rate of ~ 12.5. Our additions to the methods section add extra detail on how these estimates are arrived at.*

* I think discounting needs to be discussed at some point. This is almost always included in decision analyses.

*We agree that discounting should be included for economic evaluation. However we have not seen it done widely when the study is focusing purely on the epidemiological impact of vaccination. We have added the following sentences to the discussion,*

*“Policy makers should consider these updated estimates when assessing the role of BCG vaccination in those at school-age. However, decisions regarding vaccine policy in the UK require economic evaluation, which discounts costs and benefits in the future; discounting has not been applied in this study which estimates the epidemiological impact of vaccination only.”*

* Figures 1 & 2: “Notifications” & “Incidence Rates” here are methods to derive a rate of annual decline. I think it would make more sense to label these with the derived rate of decline, order the y axis by these, and perhaps use color to label the method used to drive the rates of decline.

*The derived rates are variable by year and hence cannot be used as a label. We have included a figure from the SI in order to clarify this.*

* I wonder whether it is possible to compare the predicted effect on TB notifications made at the time of the policy change to actual observations and comment on this? If not, why?

*The model presented here is based on a series of assumptions from the 1980s that have not been updated (such as the population size) and does not account for notifications in the non-UK born and non-White populations which are likely to have a significant impact on incidence rates in the White UK born. For these reasons we have only compared between model estimates in this study as the primary aim was to assess the strength of the model used in decision making. As mentioned above we have clarified the aim of the study in the introduction and abstract.*

## Authors’ response to Reviewer #3:

This manuscript reports the results of a mathematical modeling analysis that updates a previous analysis done in 1987 to project the potential effect of ending universal school-age vaccination in the UK. It uses the same model framework but incorporates uncertainty estimates (which the previous model did not) and explores the effect of changing one of the original model assumptions about rate of decline in TB incidence to what has actually been observed in the intervening decades.

*Thank you for your comments.*

I think that the authors have executed their stated goal well, but I am having a hard time seeing this as an independently publishable piece of work. Making minor tweaks to the model parameters and adding uncertainty estimates did not produce fundamentally new conclusions. The more important change in the assumption of rate of decline in TB incidence did change the results, but the conclusions still do not seem independently publishable; the conclusion seems to be simply that the projected impact depends very much on what you assume is happening with the epidemic. I wonder if maybe I am not understanding the results completely? It seems like since 15 years have passed since the policy change, the more useful analysis would be to try to assess what the impact has been in those 15 years rather than repeat the projection exercise that led to the policy change with varying assumptions about declines in TB incidence when in fact it should be knowable what has happened in those 15 years. Maybe this is what the authors were trying to do, but I find myself getting to the end of the article with the unanswered question about whether the policy change was the right decision or not.

*We agree that this study does not fully answer whether or not the policy change was the correct decision and agree that analysing recent data may be a more valuable approach for this (see here for a preprint by the authors doing this:* [*https://doi.org/10.1101/567511*](https://doi.org/10.1101/567511)*). The aim of this study was instead to assess the strength of the evidence used for decision making at the time of the policy change - we have updated the abstract and introduction to clarify this. As mentioned by the second reviewer models are rarely evaluated once the policy decision has been made. In this study we show that the model used for decision making was methodologically flawed, was spuriously precise, and made several assumptions that led to what we now know were entirely unrealistic estimates. These findings are important for decision makers to understand when evaluating quantitative evidence for future policy decisions.*

I think I found this paper difficult to understand because it referred so heavily to the older piece of work. I understand that this paper is updating the previous results, but it is my opinion that every scientific paper has to stand alone. Without a conventional diagram of the model structure, list of parameters with sources, and definitions of key terms, I had a hard time piecing together what was done, and I am still not sure that I fully understand. Some of this may reflect my own lack of expertise in building these kinds of models, so I realize that to a reader more in the authors’ field, this may not have been as much of an issue. However, in the interest in making this work more accessible to the general reader (and reviewer), some specific key terms and procedures that I did not find well described are listed below:

*We have greatly expanded on the explanation of the model given in the methods, incorperating details from the SI, and improved the discussion of the Sutherland et al. model.*

* Introduction paragraph 2: Does “BCG schools’ scheme” mean universal BCG for school children? What age were children vaccinated? What was coverage assumed to be – that is, were any children assumed to be unvaccinated during the old policy? After the policy change, did anyone get vaccinated (e.g. high-risk children)?

*We have added the age at vaccination (13-14 years old) to the introduction and also discussed vaccination policy post 2005. We have also moved a detailed description of the Sutherland et al. model from the SI into the main paper which covers coverage assumptions etc.*

* Methods, paragraph 1: Who are the vaccinated and unvaccinated populations? Is vaccinated = hypothetical situation of BCG school vaccinations having continued, and unvaccinated = situation in which policy changed to stop vaccinations? Or in the two situations are different proportions of children assumed to be vaccinated and unvaccinated?

*We have incorporated details from the SI in order to expand on this - please see the marked up paper.*

* Methods, paragraph about assumptions: “Firstly, as incidence rates for those ineligible for the BCG schools scheme are not published…” Who was ineligible if it was a universal program as stated in the introduction? The actual policies need to be better described.

*See above.*

* It seems that somewhere in the model the estimated protective effect of BCG should appear, unless I am not understanding the model structure at all (which is possible).

*See above.*

* Methods, section on updating model parameter estimates: Were the data used for estimating changes in incidence obtained only for the white non-foreign-born population or for the entire English population? Also, it is unclear what these two data sources are (i.e. what they capture) and how they differ. The limitations suggest that maybe the notification data were not race/origin stratified and something was done to account for this, but the methods do not describe this.

*We apologise for the confusion. Data on incidence rates in the White UK born was not available. Therefore, we used two proxy measures notifications in England and Wales and incidence rates in the UK born in England only (data for Wales was not available). Based on your comments (and those of the other reviewers) we have reworded this paragraph.*

* Results: the statement is made that “We found that updating parameter values, and incorporating uncertainty, did not alter the number of vaccines required to prevent a single notification…” But changing the assumed decline in incidence rate does have a major impact. Isn’t this one of the parameters of the model? Or am I not understanding how the model works?

*We apologise for the confusion - hopefully our updates to the methods section has helped clarify the model structure. We considered the annual decrease in a scenario analysis but all other parameters were otherwise updated. It is variation in these parameters that we are referring to here.*