Dear Professor Haddad,

thank you very much for giving us the opportunity to improve our manuscript with the help of the excellent suggestions made by the two reviewers. Based on the suggestions and criticisms we have performed numerous changes, as indicated by underlined text in the manuscript, which we will refer to below by their corresponding line numbers.

We very much look forward for a new assessment of the manuscript!

Erik Bülow  
on behalf of the authors

=====================================================================================

Editor-in-Chief Comments to Author:  
A quick search of the recent literature highlighted the following papers:  
  
1. Bone Joint J. 2020 May;102-B(5):580-585.  
  
Modifiable risk factors for mortality in revision total hip arthroplasty for periprosthetic fracture  
  
Victoria N Gibbs, Robert A McCulloch, Paula Dhiman, Andrew McGill, Adrian H Taylor, Antony J R Palmer, Ben J L Kendrick  
  
2. Bone Joint J. 2020 Jul;102-B(7\_Supple\_B):11-19.  
  
2020 Frank Stinchfield Award: Identifying who will fail following irrigation and debridement for prosthetic joint infection  
  
Noam Shohat, Karan Goswami, Timothy L Tan, Michael Yayac, Alex Soriano, Ricardo Sousa, Marjan Wouthuyzen-Bakker, Javad Parvizi, ESCMID Study Group of Implant Associated Infections (ESGIAI) and the Northern Infection Network of Joint Arthroplasty (NINJA)  
  
Please review these papers and include in any resubmission if relevant.

---------------------- RESPONSE-----------------------------------------------------------------------------------------------------

We are grateful for those interesting references, which we have read and enjoyed. We have now inserted a citation to Noam et al. on line **250**. The paper by Shohat focuses on failed treatment after periprosthetic joint infection which is an interesting topic of its own, although slightly different from ours, and we chose not to refer to that article.   
=====================================================================================

Associate Editor Comments to Author:  
  
Associate Editor  
Comments to the Author:  
Thank you for your interesting manuscript. Unfortunately, there were a number of issues raised by the reviewers that need to be addressed before considering any resubmission. We hope you are able to address these

---------------------- RESPONSE-----------------------------------------------------------------------------------------------------

We are once again grateful for the chance to improve the manuscript and have tried to address each of the issues raised below.  
  
=====================================================================================

Reviewer(s)' Comments to Author:  
  
Reviewer: 1  
  
Comments to the Author  
Purpose - Development of 90-day mortality model following primary THA for shared preoperative decision-making using parsimonious and clinically accessible risk factors. Online risk calculator developed for clinical use.  
Methodology - Model building performed using Swedish registry from 2008-2015.Validated using National Joint Registry.  
  
Advantages - Advanced statistical methods to develop parsimonious prediction model  
-Excellent data collection on large scale for these registries  
  
Disadvantages - Mortality following primary THA is a rare outcome, would be more important to demonstrate predictive strength for pertinent THA related complications such as readmission, reoperation, dislocation, wound complications, length of stay, discharge to home, patient-reported outcomes.  
Major Flaws - As noted above, focus limited to rare outcome. Most shared decision making revolves around a review of the most likely expected benefits and risks. While patients should be aware of their particular mortality risk, as this is a catastrophic outcome following primary THA, shared decision tools would be more helpful to focus on the more likely outcomes and complications to make an informed decision without directing undue attention to this unlikely event.  
---------------------- RESPONSE-----------------------------------------------------------------------------------------------------  
We agree and have added a discussion on line **261**, where we also explain that we plan to develop risk estimators for the risk of periprosthetic joint infections and dislocations as a next step.  
------------------------------------------------------------------------------------------------------------------------------------------

-Please provide age range and tabulation among patients that died within 90 days. May be inappropriate to extrapolate data from 35-year-olds if nearly all cases of mortality occur in 60+ group.  
---------------------- RESPONSE-----------------------------------------------------------------------------------------------------  
We agree and have now present this data in a new table 2 referenced on line **195**. We have also changed the recommended age range for usage of the prediction model to 57-95 years (i.e., the observed age range for patients who actually died; line **237**). We have also changed the two examples of actual patient characteristics and their effect on 90-day mortality on line **241-244** in accordance with this age range. Finally, we have imposed a new age limit to the online web calculator.  
------------------------------------------------------------------------------------------------------------------------------------------  
  
Minor Flaws -   
-Statistical method beyond the scope of most readers, a diagram explaining the process may make the description of methods more accessible.  
---------------------- RESPONSE-----------------------------------------------------------------------------------------------------  
We agree that the applied methods are hard to describe verbally. We hope that a visual representation, as now provided in figures 3 and 4 (referred to from line **147**) might provide a more intuitive understanding. We still fear, however, that the methods might still be perceived as rather complex by most readers. We therefore hope that the exact details, as provided through the referenced source code repository (<http://doi.org/10.5281/zenodo.3732852>), will contain enough details for readers with an additional interest in the technical details.  
------------------------------------------------------------------------------------------------------------------------------------------  
  
-Can you delineate the specific inclusions for diseases of the CNS. And cutoff for diagnosis of obesity? (BMI > 30?) If you have BMI data it seems unwise to rely on the diagnosis of obesity, rather than including the specific patient level data either as a dichotomous variable or as a linear BMI variable  
---------------------- RESPONSE-----------------------------------------------------------------------------------------------------  
The definition of each comorbidity (including CNS disease and obesity) is presented in table 1. We acknowledge that the description of this table was rather scarce and we have now expanded the caption with a reference to Quan et al. 2005 for the individual ICD-10-codes underlying each sub-category.

It is seen from the table that the comorbidity “obesity” is equivalent to obesity as defined by Elixhauser et al. (codified by Quan et al). This corresponds to patients with ICD-10 = E66 (as also specified on line **257**), thus with the conditions "drug-induced obesity, morbid (severe) obesity with alveolar hypoventilation, overweight, other obesity, unspecified obesity”. It would be possible to provide such details in the text but if so, similar details should preferably also be provided for most other comorbidities as well, but at least for the comorbidities “cancer” and “kidney disease” since these were also used in the final model. Cancer alone (“malignancy” and “metastatic solid tumours” according to Charlson, and “lymphoma”, “metastatic cancer” and “solid tumours” according to Elixhauser, as defined in table 1), entail more than 1,000 individual ICD-10-CM codes. We fear that descriptions at this level of granularity would be of limited use to most readers and hope that table 1 (with its improved caption) might provide enough details for readers with interest in the underlying diagnoses.

We have tried to describe in the text that both BMI, as well as diagnosed obesity (as indicated by ICD-10 = E66), were included as potential predictors in the model. Only the latter, however, was included in the final model after variable selection using LASSO-regression and repeated bootstrap samples. This is an empirical result which indicates that the inclusion of diagnosed obesity in the model improves the predictive power of the model, whereas BMI (treated as a continuous variable) did not. This is the reason to include diagnosed obesity but not BMI. Here, the distinction between traditional inference modelling and prediction modelling is relevant. In the first setting, we would value interpretability more, and the best understanding of association on group level might very well be provided through the study of BMI, preferably categorized along the WHO recommendations rather than diagnosed obesity. Prediction modelling is different, with the goal to filter out the most important variables, but where causal relations might be less intuitive.  
------------------------------------------------------------------------------------------------------------------------------------------  
  
-Addressed in discussion: extrapolation to countries or practices where cemented THA is less common, especially in younger patients. In the United States, for example, cemented THA is typically reserved for elderly and osteoporotic patients and this would impact the generalizability of your findings.  
---------------------- RESPONSE-----------------------------------------------------------------------------------------------------  
We agree that it is unfortunate to only investigate patients with cemented THA. We did try to develop a model on Swedish patients with cementless/hybrid/reverse hybrid THA as well, but we only observed 28 deaths among those patients, yielding a wide and therefore non-informative 95 % confidence interval for the area under the curve (AUC; 0.54-0.76; as now mentioned on line **281**). We tried to stratify this analysis for the subgroups of patients with cementless/hybrids/reverse hybrid THA as well, but this was impossible due to non-convergence of the regression models. We also considered to include all patients and to use fixation as a possible predictor in the model. This information is not chronologically known prior to surgery, however, and can therefore not be included in a model used for shared decision making.

We have now emphasized those difficulties of generalizability on line **285-287**.  
------------------------------------------------------------------------------------------------------------------------------------------

-Page 9 Line 165 "Figure 6 and ??" - Was there another figure being referenced?  
-Page 10, line 192 "patters" I believe is typo  
-page 11, line 218 "cphorts" should be cohorts  
-page 11, line 226 "BMIt" should be BMI  
---------------------- RESPONSE-----------------------------------------------------------------------------------------------------  
Thank you very much! We have corrected accordingly.

=====================================================================================

Reviewer: 2  
  
Comments to the Author  
Strengths:  
1)      A very interesting study and potentially is useful in guiding discussions with patients who have significant comorbidities. At present, there is no consistency in how we describe risk to patients in terms of mortality risk. It appears very much based on individual anaesthetic/surgeon bias and judgement of risk. A clinical tool that is easy to use and validated would help with this potentially difficult decision making process.  
2)      This has both internal and external validation using well known registry data. The authors have acknowledged some of the limitations in using such datasets but for the purposes of this sort of study, registry data is ideal and probably the only way such research questions can be answered.  
---------------------- RESPONSE-----------------------------------------------------------------------------------------------------  
Thank you very much! We are very grateful for all valuable input and the supportive review.  
------------------------------------------------------------------------------------------------------------------------------------------  
  
Limitations  
1)      I would like to know the level of missing BMI data especially in the NJR as historically, BMI data has been poorly completed and up to 50% of BMI data can be missing. If so, has imputation modelling been used to mitigate against such missing data in the statistical modelling.  
---------------------- RESPONSE-----------------------------------------------------------------------------------------------------  
Completeness of BMI in NJR was only 60 % in 2008 but has increased yearly since then. It was 80 % in 2018, as now reported on line **294-297** (<https://bit.ly/3hySnNF>). The reason why we previously only reported missing data for BMI in Sweden/SHAR but not in England and Wales/NJR (figure 1) was that BMI treated as a continuous variable was not included in the derived statistical prediction model. It was only considered as a potential predictor during variable selection in the derivation cohort from SHAR. We realize, however, that the difference between obesity due to reported BMI above 30, or diagnosed obesity due to the ICD-10-code E66, could be emphasized further. We have therefore made sure that the condition used in the statistical model is always referred to as "diagnosed obesity".

Imputation in the derivation cohort would admittedly be a better alternative than exclusion of patients without BMI. It does require that data is "missing (completely) at random", however. This, unfortunately, might not be the case for BMI, since; “all BMI data has to be viewed with caution as surgeons may be more likely to enter BMI data when the BMI is high, introducing an element of bias.” [NJR website <https://bit.ly/2Ex2rZ5>]. We have anecdotal evidence of a similar situation in Sweden. Hence, we cannot preclude association between the missing values and the probability of the values to be missing. We thus have data "missing not at random", which, unfortunately, disqualifies the use of imputation. The potential bias introduced by imputation relaying on miss-specified assumptions, might not exceed the bias likewise introduced by the alternate exclusion. Such practice, on the other hand, might artificially inflate the power, leading to false certainty in a sub-optimal model. We have therefore abstained from imputation.  
--------------------------------------------------------------------------------------------------------------------------  
  
  
2)      Why did the authors only look at cemented THA’s? And is this for cemented stems and acetabular components or were hybrid THRs also reviewed. Please clarify. In the latest NJR 2019, cemented THRs only accounted for 32% of all THRs compared to 37% for uncemented and 21% for hybrids. This means this study is representative of only 1/3 patients. Will the authors undertake further work looking at uncemented and hybrid THRs also?  
---------------------- RESPONSE-------------------------------------------------------------------------------------  
This issue is also partially discussed above were we extended the limitations section considering challenges of generalizability. We have now also clarified that hybrids and reverse hybrids were indeed excluded as well (line 103-104).

We can confirm the proportions of cemented procedures from figure 2 in the manuscript (153,477 out of 460,209 = 1 of 3, as stated by the reviewer above). We can also see from figure 1 that the proportions are reversed in the Swedish derivation cohort (57,923 out of 85,168 = 2 of 3). This is a matter of national differences, making transportability, and even the external validation challenging (as discussed on line **X**). Even with those national differences, the discriminatory ability of the model, was maintained, also when applied to patients from the UK. This is an indication of a quite robust model where we minimized the risk of over-fitting by combining LASSO-regression and the bootstrap ranking procedure. The calibration curve did show slightly less coherence between observed proportions of deaths compared to predicted probabilities. This is a natural consequence of the national different concerning indications for fixations, as also discussed on line **285**.

As mentioned in a comment above, the possibilities to undertake any further prediction modelling considering uncemented /hybrid/reverse hybrid cases in Sweden is, unfortunately, very limited, since such data is too scarce.  
--------------------------------------------------------------------------------------------------------------------------  
  
3)      I would like to know if the risk of dying in 90 days was comparable to the risk if the patient had not had a THR? Ie, is there an additional risk in having a THR or is the risk the same? For example, in a 99 year old, their risk of dying may also be almost 9% just by the fact they are 99 years old. Can the authors provide comparisons to normal mortality rates? The authors have already stated (line 232) that THA was not regarded as an intervention.  
---------------------- RESPONSE-------------------------------------------------------------------------------------  
We agree that this is undeniably a very important and relevant issue. Unfortunately, we cannot simply compare the number of deaths within 90 days for patients with THA and without, however, since that would require a control group with patients in need of elective THA, but for whom such procedure is not undertaken at a certain date comparable to the date of surgery. What we can do, however, is to study relative survival, a concept otherwise most commonly used in cancer epidemiology; a comparison of cohort survival, related to the survival of the general population of the same age, sex and year of birth. Such studies were previously performed on the same cohort by Cnudde et al. (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6263594/>). They found that the relative one year survival for patients with THA due to osteoarthritis, was 101 %. This means that our patients are likely to have a better survival than the general population (due to "selection bias", hence that only relatively healthy patients undergo elective THA). If we take a closer look at only the cemented cases during the initial 90 days, however, as done by Cnudde in figure 23 of his PhD thesis expanding the topic of the published paper (<https://gupea.ub.gu.se/bitstream/2077/54531/6/gupea_2077_54531_6.pdf>), we can see that the relative survival is below 100 % during this specific period. This combined, is a strong indication that the patients under study are not likely to die in general, but that the THA surgery itself elevates this risk during a short time following surgery. We think this is a strong indication that most deaths occurring within 90 days are caused by THA-related issues. This is equally true for the 99-year olds, since those patients do not constitute a random sample from their age group, but are relatively healthy with a foreseen advantage of THA, outweighing the risk of near-by mortality (they would otherwise not have been offered this elective surgery).

We have now extended the discussion on line **302** to include this reasoning.  
---------------------------------------------------------------------------------------------------------------------------  
  
4)      Data accuracy and completeness in the NJR has been an issue historically. Can the authors provide some data on data completeness and the level of missing data in the statistical analysis.  
---------------------- RESPONSE-------------------------------------------------------------------------------------  
We agree that this is important information that we seem to have missed in the previous version. This data is now included on line **131** for NJR, and similarly on line **108** for SHAR. Exclusion due to missing data is depicted in figure 1 and 2.  
--------------------------------------------------------------------------------------------------------------------------  
  
  
Summary: I think this study could be really useful for clinicians.