PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Deriving literature-based benchmarks for surgical complications in high-income countries: A protocol for a systematic review and meta-analysis
AUTHORS	Brindle, Mary E; Roberts, Derek; Daodu, Oluwatomilayo; Haynes, Alex; Cauley, Christy; Dixon, Elijah; LaFlamme, Claude; Bain, Paul; Berry, William

VERSION 1 - REVIEW

REVIEWER	Professor Ara Darzi Institute of Global Health Innovation Imperial College London United Kingdom
REVIEW RETURNED	05-Sep-2016

GENERAL COMMENTS	This is a study protocol applying standard meta-analytical techniques to assess surgical complication for specific procedures in high impact countries. It accommodates study quality and variation. The authors have not, however, commented on: (1) Potential loss of real-life data outcomes and when applying standard meta-analysis in combing national database results with that of a small local trial (standard meta-analysis is not very good at this) (2) Follow-up times in different studies (3) Management of overlapping data (i.e 2 studies reporting on the same patient cohort, e.g. at different times, or those reported concurrently on both national databases and studies) (4)Management of international multi-centre data when data according to country is not available

REVIEWER	David Watters Deakin University and Barwon Health University Hospital Geelong Victoria, Australia
REVIEW RETURNED	14-Sep-2016

GENERAL COMMENTS	I find the topic one that is laudable and I look forward to the results. I am not entirely sure of the value of papers that say what is going to be done - they whet the appetite but do not deliver. However, BMJ Open does publish study protocols and the value of reviewing their protocol is to give feedback on the protocol that will certainly influence the results.

	Personally I'd rather wait for the results of their systematic review. I agree this is an important topic Introduction:There is almost an implication in the introduction that complications are due to error. This is not true. Many procedures including those studies have a recognised risk of complications that form part of the informed consent process. Complications can occur as a result of the condition being treated, the comorbidities of the patient being treated as well as a range of errors that arise from commission, omission or inition. Methods: Within the text of the methods it is not clear to me how mortality is measured whether by 30 day mortality or death before discharge from hospital (WHO Indicator). The years chosen (2000-2016) represent a long period in which although complication rates may have changed little (to be determined by their results) mortality rates have certainly dropped and this is documented in the literature. There are issues with the 5 countries having quite different health systems, particularly in their proportions of surgical care provided by their private and public systems. How will private V public be resolved, particularly given the public system in ANZ tends to have less severe cases. I hope they will address the challenge of voluntary versus involuntary/independent reporting of outcomes of surgery Have they considered using the Dr Foster database (World Comparator) that is capable of comparing outcomes in a wide variety of countries and would address the problems of voluntary databases, selective entry and many biases. However what they plan to do will yield valuable and helpful data. It is just that Dr Foster would be more contemporaneous and free of publication bias.

VERSION 1 – AUTHOR RESPONSE

Below are the reviewers' comments and our responses as well as the location of the changes made in the manuscript.

1. Please shorten the 'Strengths and Limitations' section on page 4. This section should contain up to five short bullet points, no longer than one sentence each, that relate specifically to the methods of the study reported
This has been shortened to five bullet points. These points have been edited for brevity.

2. Potential loss of real-life data outcomes and when applying standard meta-analysis in combing national database results with that of a small local trial (standard meta-analysis is not very good at this)
We agree that there are limitations to database studies as well as meta-analyses that could be better addressed in specific, question-based trials. We have added this as one of our limitations: "A meta-analysis of database-related publications does not provide the same real-life data that exists in trials. Results of this review will be interpreted alongside that of landmark trials."

3. Follow-up times are different in different studies
Follow-up may indeed be different within these different databases. When we compare results across

countries and databases, we will use defined follow-up periods and stratify and report our comparisons according to these follow-up periods (eg. Mortality rate comparisons will be made for 30-day mortality and hospital admission mortality separately). We have now clarified this in our publication in the first two sentences of the second paragraph of “Statistical Analysis”: “Different follow-up durations can be anticipated depending on the database and study (eg. 30 days or total hospitalization). Outcome analyses will be stratified according to defined follow-up duration.”

4. How will we manage overlapping data (i.e 2 studies reporting on the same patient cohort, e.g. at different times, or those reported concurrently on both national databases and studies)?

We agree that overlapping data presents an interesting challenge within this study. When different studies report on the same population within the same timeframe, this will be of particular interest as differences in findings may point to important differences in how data is collected in these national databases. In these cases, we will examine how the differences between database design effect the results and potential interpretation. We have added two sentences describing this in the “Data Synthesis section”:

“When different studies report outcomes for the same population within the same time frame using different databases, differences in population selection, data collection and data handling will be compared between these databases. How these differences impact apparent complication rates will be explored.”

We see the differing study time frames as an opportunity to explore changing rates of complications over time. This data will be examined in a descriptive fashion which we have now outlined within our “Data synthesis” section:

“Within subgroups (countries, procedures, outcomes) data will be presented both visually (timelines and graphs) as well as within tables and narrative text as appropriate to describe the outcome results obtained from databases at different time periods. Quantitative comparisons within subgroups will be restricted to obtainable time frames that are similar (within a two-month range both at start and stop of data collection period).”

5. How will we manage international multi-centre data when data according to country is not available?

Results that are a synthesis of multiple countries will be reported descriptively including a breakdown of participating sites. If the majority of participants (>80%) are from a single country, these results will be considered as representative of that country within analyses. We have now included this description in our “Data Synthesis” section:
“Results that are a synthesis of multiple countries will be reported descriptively including a breakdown of participating sites. If the majority of participants (>80%) are from a single country, these results will be considered as representative of that country within the comparative subgroup analyses. If the proportion from any single country is less than 80% or country-specific breakdown is unavailable, these results will be excluded from the comparative analyses.”

6. There is almost an implication in the introduction that complications are due to error. This is not true. Many procedures including those studies have a recognised risk of complications that form part of the informed consent process. Complications can occur as a result of the condition being treated, the comorbidities of the patient being treated as well as a range of errors that arise from commission, omission or inition.

We agree with this statement. There are both preventable and intrinsic risks at play for many complications. Understanding complications purely related to intrinsic risk are still important in medical decision-making. We have added an additional sentence in the introduction to clarify that adverse events are complex and include both intrinsic and preventable risks:
"Complications occur because of both intrinsic risks and medical error."

7. Within the text of the methods it is not clear to me how mortality is measured whether by 30 day mortality or death before discharge from hospital (WHO Indicator).
The years chosen (2000-2016) represent a long period in which although complication rates may have changed little (to be determined by their results) mortality rates have certainly dropped and this is documented in the literature.

These are excellent points and touch on some of the issues raised by another reviewer (see above). We will consider both measures of mortality and report these separately. Some of our descriptive analysis will specifically look at how complication rates may have changed over time. We will use timelines and graphs to visually represent follow-up periods and changing rates of mortality and complications over time. Please see our description above as to how we have addressed these issues.

8. There are issues with the 5 countries having quite different health systems, particularly in their proportions of surgical care provided by their private and public systems. How will private vs public be resolved, particularly given the public system in ANZ tends to have less severe cases?
Also, how will we deal with the challenge of voluntary versus involuntary/independent reporting of outcomes of surgery?

These differences are certainly worth exploring. We are interested in comparing systems of care such as public and private (including risk stratification when available). As well we would like to see how different reported outcomes may relate to different methods of data acquisition (database differences). Interpretation will be particularly important as we anticipate that comparisons will almost never represent equivalent populations or data (mentioned in our limitations). We have added an additional sentence and the end of our "Data Synthesis" section: "This synthesis will examine how health systems differ between and within countries in regards to risk-adjusted complication rates and population differences."

9. Have we considered using the Dr Foster database (World Comparator) that is capable of comparing outcomes in a wide variety of countries and would address the problems of voluntary databases, selective entry and many biases.
This is an excellent suggestion. We will look at this within a follow-up study to see how a primary study of international outcomes from the Dr. Foster database compares with what we have found in our meta-analysis.

Please let us know if there are any additional changes suggested to our protocol.
Our team appreciates the opportunity to revise our protocol.

VERSION 2 – REVIEW

REVIEWER	Professor David A Watters Deakin University and Barwon Health University Hospital Geelong, Victoria, Australia
REVIEW RETURNED	27-Nov-2016

GENERAL COMMENTS	I think this is a laudable and much needed project and will be of international benefit. It is highly relevant and ultimately the ability to derive the data from different datasets and make international comparisons is much needed. Also learning how to compare large national datasets, something which is adequately discussed in your protocol. The following are my concerns: I am not sure how you are going to be certain that you can capture the rate of unplanned reoperations, surgical site infection and pulmonary embolism without capturing unplanned readmission to hospital. Although a 30 day time period might be appropriate for general surgical procedures it is not for joint arthroplasty where deep seated infection may result in readmission and then re-operation. I would recommend you include unplanned readmission to hospital at least within 30 days of discharge (not operation). Many pulmonary embolisms are readmitted, at least in Australia and New Zealand, under a medical unit and thus one needs to use a State or National Database to capture this. How is mortality being measured - in-hospital mortality would be fine for me as this is the WHO definition but how will you differentiate between those that present 30 day mortality and those that present mortality before discharge from hospital? Those institutions that already have access to Dr Foster's Global comparators tool are able to make comparisons such as the ones you wish to use. For example there are 4 hospitals in Victoria, Australia, that have access and do use this tool which here in Australia is based on the Victorian Admitted Episodes Dataset (VAED). The ability to derive the outcomes from national datasets that are not published papers and therefore subject to publication bias will be far more believable. I realise you are only selecting papers based on national datasets but these may still be subject to publication bias and I would prefer to see the outcomes from the datasets over similar time periods. I assume you are going to present outcomes and complications data temporally. There has been a significant decline in post procedural mortality in Australia (possibly not in NZ) in the years of the Australia and New Zealand National Audit. Using early 2000s outcomes would not be comparable with 2011-14 outcomes for example or may not be. The reason why you are using outcomes from 2000 onwards, I would suggest is not because of the publication of "To Err is Human" - by the way there were contemporaneous publications in each of the countries at the same time or even earlier - but because you are basing the outcomes of papers published. Although this will be interesting deriving the outcomes from the databases themselves using an approved investigating team from each country would be even more believable. There is a risk that if there is only old data (ie early 2000s) for some countries then you will be inferring that the oucomes are worse than if there had been contemporaneous publications. This has some political/diplomatic reputational risks which need to be considered.

VERSION 2 – AUTHOR RESPONSE

QUESTION/CONCERN:
I am not sure how you are going to be certain that you can capture the rate of unplanned reoperations, surgical site infection and pulmonary embolism without capturing unplanned

readmission to hospital. Although a 30 day time period might be appropriate for general surgical procedures it is not for joint arthroplasty where deep seated infection may result in readmission and then re-operation.
I would recommend you include unplanned readmission to hospital at least within 30 days of discharge (not operation).
Many pulmonary embolisms are readmitted, at least in Australia and New Zealand, under a medical unit and thus one needs to use a State or National Database to capture this.

RESPONSE
This is an excellent point. Although not all databases include readmissions, some do and this is an important very point to capture. We will add readmission to our analysis and have now included readmission in our search strategy. A quick look suggests that this will not increase our citations by much. As with mortality (see below), different databases/studies may capture readmission rates differently. We will plan to capture all data on readmission that gives a specific time-frame for collection (eg 30-day readmission rate) and stratify our results according to the definition of readmission. We will exclude studies that do not provide a time frame for readmission.
Appendix 1 now contains readmission. Within the manuscript, reference to length of stay is now included at the end of the search strategy and the issues surrounding the follow-up period during which this is captured is included in the “statistical analysis” section”

QUESTION/CONCERN:
How is mortality being measured - in-hospital mortality would be fine for me as this is the WHO definition but how will you differentiate between those that present 30-day mortality and those that present mortality before discharge from hospital?

RESPONSE:
We agree that mortality will be reported differently within different databases. Fortunately, we anticipate that this will be one of the most common outcomes reported- whether it be 30-day mortality or in hospital mortality.
We will stratify studies by the definition of mortality. We anticipate that most will fall into either 30-day or prior to discharge but we will collect all information on mortality including time to event information and stratify accordingly.
We describe this with some further detail in the statistical analysis section

QUESTION/CONCERN:
Those institutions that already have access to Dr Foster's Global comparators tool are able to make comparisons such as the ones you wish to use. For example there are 4 hospitals in Victoria, Australia, that have access and do use this tool which here in Australia is based on the Victorian Admitted Episodes Dataset (VAED). The ability to derive the outcomes from national datasets that are not published papers and therefore subject to publication bias will be far more believable. I realise you are only selecting papers based on national datasets but these may still be subject to publication bias and I would prefer to see the outcomes from the datasets over similar time periods.

RESPONSE:
Understanding the difference in analyses that involve data from different time frames will be a crucial part of the synthesis of the study. All results will be stratified by time frame (we describe this within our data synthesis section). We are going to be very cautious and avoid making any assumptions that studies represent comparable populations. The strong likelihood of publication bias will be acknowledged in our study. We have included an assessment of publication bias through the traditional funnel plot and Eggers test in our methods (end of statistical analysis section) but we acknowledge that careful interpretation is still needed. The results of our review will provide us with the scope of published data that are currently used to inform our understanding of morbidity and

mortality rates but we have added a new element to our study based on your suggestion. We will use the results from our systematic review to target surgeries where there is significant variability in reported complication rates within a similar time frame and also specific surgeries where the complication rates are very high. We will examine these specific targets within existing clinical databases with a national scope with contemporary data. Databases will include the Dr. Foster database and NSQIP. This is now a new section entitled “Correlation between Systematic Review Findings and Contemporary National Data”.

QUESTION/CONCERN:
I assume you are going to present outcomes and complications data temporally. There has been a significant decline in post procedural mortality in Australia (possibly not in NZ) in the years of the Australia and New Zealand National Audit. Using early 2000s outcomes would not be comparable with 2011-14 outcomes for example or may not be.
The reason why you are using outcomes from 2000 onwards, I would suggest is not because of the publication of "To Err is Human" - by the way there were contemporaneous publications in each of the countries at the same time or even earlier - but because you are basing the outcomes of papers published.
Although this will be interesting deriving the outcomes from the databases themselves using an approved investigating team from each country would be even more believable.
There is a risk that if there is only old data (ie early 2000s) for some countries then you will be inferring that the outcomes are worse than if there had been contemporaneous publications. This has some political/diplomatic reputational risks which need to be considered.

RESPONSE:
This perspective is very helpful. We removed the reference to “To Err is Human” within the PICOS description and have emphasized in our introduction and discussion the international push for improved healthcare safety including representative references. We completely agree that stratifying results by comparable time frames is essential. We have expanded upon this in our data synthesis section. Most importantly, we agree that understanding the results will require extraordinary sensitivity to differences between both databases and countries. We will involve health systems experts familiar with databases from each of the countries involved to help frame the data and avoid direct comparisons when these are not relevant. We have added this element to our study within the “data synthesis” section. This study is part of a larger, international proposal to explore targets for improved outcomes across countries and our goal is to generate an international collaborative rather than divisive approach. As such, it is important that we do not approach this study as having “winners” and “losers”. We anticipate (and will seek out) broad participation from international investigators in the completion of the project.