

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)	Influence of Hospital Volume on Nephrectomy Mortality and Complications: A Systematic Review and Meta-Analysis Stratified by Surgical Type
AUTHORS	Hsu, Ray; Salika, Theodosia; Maw, Jonathan; Lyratzopoulos, Georgios; Gnanapragasam, Vincent; Armitage, James

VERSION 1 - REVIEW

REVIEWER	Omar Aboumarzouk Queen Elizabeth University hospital, Glasgow, UK
REVIEW RETURNED	23-Mar-2017

GENERAL COMMENTS	<p>This is an Excellent paper in concept: A need for such a paper exists. Comments: The methodology section is lacking: need more detail on inclusion exclusion criteria. you need to define parameters, ie LV HV: simply stating you used what two other similar papers used is not adequate. Was Cochrane methodology employed here? a lot of the sections include its methodology; yet mixed with nonCochrane methodology techniques. justify this please. Results: meta-analysis from 2-3 studies is a thin line to tread on.... I would not take those results to heart. Meta-analyzing 2 studies for thrombectomy! You can not draw conclusive evidence for this from 2 studies. be cautious. Funnel plots are usually dont for analyses with 10+ studies. (See cochrane guidelines)! discussion and conclusion: very well written and summarized, but my concern is you emphasize that there is evidence, where is that i would argue there is lack of evidence for not only partial but thrombectomy nephrectomy too. while the nephrectomy data, with only a handful of studies included: minimal evidence. but this does not reflect this review, but the lack of published data. Nonetheless, centralization has only come to shine in the last 10-15 years with advent of robotics specifically. Hence warrants discussing.</p>
-------------------------	---

REVIEWER	Nele Brusselaers Karolinska Institutet, Sweden
REVIEW RETURNED	24-Mar-2017

GENERAL COMMENTS	I read with interest the meta-analysis of Ray Hsu et al on centralization in the field of nephrectomy. This study is well designed
-------------------------	--

according to the most recent recommendations, and discusses an important topic – the need for centralization of major surgery. The analyses appear to be robust, so most of my comments concern clarifications:

- Abstract: studies most likely to introduce misclassification bias: may be too vague. Misclassification of exposure or outcome?
- Introduction: you clearly describe the benefits for centralization of major surgery, but you remain vague about the disadvantages. Main disadvantages/problems are the distance for the patient to the hospital – and the huge organizational reform needed (which needs to be supported on governmental level since not easy to arrange if hospitals are not within the same network/group of hospitals). I don't immediately see why differences in disease biology or surgical complexity are barriers to implement a centralized model. If more complex surgery, or more differences in disease biology – centralization seems to be more appropriate since the surgeon needs more expertise.
- Intro: why is incidence of renal carcinoma increasing? If treatment techniques are advancing (less radical approaches, alternative treatment), and the disease may be detected in an earlier stage – there may not be a very dramatic rise in incidence of this type of surgery. Maybe good to mention something in the article about how common this type of surgery is. You do mention elsewhere that the centralization doesn't seem to have benefits for appendectomy – which is probably because it is a less complex surgery + a very commonly performed type of surgery (in other words, every surgeon can do it?).
- Language bias: you only selected English language studies – which makes perfect sense, since there is probably only little published in other languages on topics such as centralization + you did a broad literature search, so it will have reduced the workload. Did you check other language articles? Could be good to add that the risk of language bias is limited, since there were none in other languages (I would be surprised if you find anything on nephrectomy centralization with original data in another language).
- Methods: mention that you present results as odds ratios and 95%CI (not mentioned)
- Methods: did all studies define post-op mortality in a similar way? Number of days post-op?
- Methods: was list of complications pre-defined – or based on what you found? So if I understand correctly you just counted all reported complications (irrespective of severity) in each study?
- Methods: quality assessment could be explained in more detail. Minimum/maximum score, the higher the better etc. Quality assessment was also done in double and resolved by consensus?
- Methods: random effects model more conservative estimates... compared to fixed models.
- Methods: number needed to centralize. I struggle a bit with this number, since it doesn't reflect an annual number, and may be depending on the duration of the study period of the different studies? Or do I see this wrong? Since one centre may do 300 surgeries every year, while another may take 30 years to get to this number – so are both then considered high volume?
- Since the analyses are based on crude pooled odds ratios, I wonder if there are known confounders which should ideally be taken into account in the future –besides the well-described surgeon volume and oncologic factors.
- Methods: did you consider comparing the lowest volume group with the highest volume group – a method also done in some other meta-analysis. OK you do not use all information, and results are likely to

	be more extreme.
--	------------------

REVIEWER	Professor Shantini Paranjothy Cardiff University, UK
REVIEW RETURNED	01-May-2017

GENERAL COMMENTS	<p>This was a generally well-written paper, although there are a few grammatical errors and varying tenses in some sentences, which can be sorted out with a proof-read.</p> <ol style="list-style-type: none"> 1. The second and third bullet point in the strengths and limitations section could do with re-phrasing. Bullet point two refers to 17 studies when 19 and 16 are mentioned in the abstract and full-text of the paper. Do you mean that you have stratified your analyses according to the types of nephrectomy, rather than the results? Bullet point three - - it is not clear from the description of the methods and results if this was done adequately (see points below re analysis). 2. Description of methods - - PRISMA guidelines provide a framework for reporting systematic reviews, it is unusual to see it quoted as a framework for the methods of a systematic review. 3. There is no mention of the types of study design that were included in the review - - was this a consideration in the search criteria? It would be helpful also to have some description of the type of population studies and outcomes of interest - - particularly the complications - - were these defined a-priori? How was postoperative mortality defined? Were there differences in this definition between studies? 4. The first mention of nephrectomy with venous thrombectomy is in the quantitative data synthesis section - - there is no mention of this earlier when discussing the types of studies that were included. 5. Given the considerable heterogeneity between studies and different levels of adjustment for case-mix and other confounding factors, it is unclear whether or not it is actually appropriate to combine these results in a meta-analysis. I also wonder whether an analysis using individual patient data may be more appropriate and if the feasibility of this was considered. More discussion about the types of study design and appropriateness of a meta-analysis in the light of the considerable heterogeneity is warranted. 6. What are the domains of study quality included here - it is not clear what the scoring is based on. 7. It is unclear how the overall mortality rates shown on page 14 were calculated - is this a weighted average across the studies?
-------------------------	--

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: Omar Aboumarzouk

Institution and Country: Queen Elizabeth University hospital, Glasgow, UK

Competing Interests: None

This is an Excellent paper in concept:

A need for such a paper exists.

Comments:

The methodology section is lacking: need more detail on inclusion exclusion criteria. you need to define parameters, ie LV HV: simply stating you used what two other similar papers used is not adequate. Was Cochrane methodology employed here? a lot of the sections include its methodology; yet mixed with nonCochrane methodology techniques. justify this please.

Response: Thank you very much to the reviewer for his comments. We have now elaborated on our inclusion and exclusion criteria. The methodology for evaluating volume-outcome relationship can be complex and little guidance currently exists for this kind of analysis. We followed the Cochrane methodology in many respects, for example in our calculation of pooled effect estimates, quantification of heterogeneity, subgroup analyses, sensitivity analyses, and meta-regressions for exploring potential source of heterogeneity. To the best of our knowledge, there is no consensus in how best to treat the various hospital volume categories presented by studies, and we therefore employed the methodology used in multiple previous meta-analysis examining volume-outcome relationships in other procedures, which is by dichotomising volume groups into two. Noting the potential misclassification bias of exposure that this may produce, sensitivity and secondary analyses were performed with consistent results. We also used an alternative scoring system for quality assessment to account for the unique nature of volume-outcome analysis and to better reflect the limitations presented in the current published studies. Details of the scoring system is now included in the methodology section and we have also added another table in the supplementary material outlining the breakdowns in score for each study.

Results:

meta-analysis from 2-3 studies is a thin line to tread on....

I would not take those results to heart. Meta-analyzing 2 studies for thrombectomy! You can not draw conclusive evidence for this from 2 studies. be cautious.

Funnel plots are usually dont for analyses with 10+ studies. (See cochrane guidelines)!

Response: Thank you very much for the comments. We have now highlighted in our discussion section regarding the limited number of studies included in synthesising the pooled effect estimates for partial nephrectomy and nephrectomy with venous thrombectomy and that results should be interpreted with that taken into consideration. We have also removed the funnel plot and statistical analysis of publication bias for radical nephrectomy complications as only six studies were included. The funnel plot for radical nephrectomy mortality was kept as nine studies were included and we felt that the statistical analysis is still of value in giving the readers an understanding of the potential publication bias or lack there of in our study.

discussion and conclusion:

very well written and summarized, but my concern is you emphasize that there is evidence, where is that i would argue there is lack of evidence for not only partial but thrombectomy nephrectomy too. while the nephrectomy data, with only a handful of studies included: minimal evidence. but this does not reflect this review, but the lack of published data. Nonetheless, centralization has only come to shine in the last 10-15 years with advent of robotics specifically. Hence warrants discussing.

Response: We agree with the reviewer that the evidence to support nephrectomy centralisation is limited at present particularly for partial nephrectomy and nephrectomy with venous thrombectomy. We have added more detailed discussion regarding this in the discussion section and amended our conclusions and abstract to better reflect this.

Reviewer: 2

Reviewer Name: Nele Brusselaers

Institution and Country: Karolinska Institutet, Sweden

Competing Interests: None declared

I read with interest the meta-analysis of Ray Hsu et al on centralization in the field of nephrectomy. This study is well designed according to the most recent recommendations, and discusses an important topic – the need for centralization of major surgery. The analyses appear to be robust, so most of my comments concern clarifications: Abstract: studies most likely to introduce misclassification bias: may be too vague. Misclassification of exposure or outcome?

Response: We would like to thank the reviewer for her comments and for noting the robust design and analyses of our study. We have now clarified in the abstract and main text that the misclassification bias is that of exposure where a study's high volume group may be considered as low when compared to the high volume groups in other studies.

- Introduction: you clearly describe the benefits for centralization of major surgery, but you remain vague about the disadvantages. Main disadvantages/problems are the distance for the patient to the hospital – and the huge organizational reform needed (which needs to be supported on governmental level since not easy to arrange if hospitals are not within the same network/group of hospitals). I don't immediately see why differences in disease biology or surgical complexity are barriers to implement a centralized model. If more complex surgery, or more differences in disease biology – centralization seems to be more appropriate since the surgeon needs more expertise.

Response: We agree with the reviewer that there are many disadvantages associated with surgical centralisation. We have therefore elaborated on some of the problems associated with such a healthcare model in our introduction. Disease biology and surgical complexity are not necessarily barriers to the implementation of a centralised healthcare model, but rather the natural disease progression and associated rates of adverse outcomes from surgery may limit the effect of centralisation. The perceived benefits of centralisation may also require a much larger patient cohort to appreciate which may be impractical in less populated countries, regions or hospital networks.

- Intro: why is incidence of renal carcinoma increasing? If treatment techniques are advancing (less radical approaches, alternative treatment), and the disease may be detected in an earlier stage – there may not be a very dramatic rise in incidence of this type of surgery. Maybe good to mention something in the article about how common this type of surgery is. You do mention elsewhere that the centralization doesn't seem to have benefits for appendectomy – which is probably because it is a less complex surgery + a very commonly performed type of surgery (in other words, every surgeon can do it?).

Response: Thank you for this comment. Increase in use of radiological imaging and prevalence of obesity are two major factors contributing to the rising incidence of renal carcinoma. Despite earlier detection and advancements in therapy, surgical resection is still the only potentially curative form of treatment with strong recommendation by international guidelines, such as those from the European Association of Urology. As a result, the incidence of nephrectomy is likely to increase as the incidence of renal cancer. There is however little population-based data published on the actual incidence of the operation. We have now included some of these points in our introduction.

- Language bias: you only selected English language studies – which makes perfect sense, since there is probably only little published in other languages on topics such as centralization + you did a broad literature search, so it will have reduced the workload. Did you check other language articles? Could be good to add that the risk of language bias is limited, since there were none in other

languages (I would be surprised if you find anything on nephrectomy centralization with original data in another language).

Response: Thank you very much for this comment. We did not check for other language articles during the literature search phase of the study. As suggested, the risk of language bias is likely to be limited and previous study have demonstrated that this exclusion generally have little effect on the summary effect estimates. We have now included these points in the methods section.

- Methods: mention that you present results as odds ratios and 95%CI (not mentioned)

Response: This is now included in the methods section under quantitative data synthesis.

- Methods: did all studies define post-op mortality in a similar way? Number of days post-op?

Response: The majority of the studies defined post-operative mortality as in-hospital mortality although four studies used 30-day mortality or did not otherwise specify. We have clarified this in our results section including updating Table 1 to outline which mortality measure was used by individual study.

- Methods: was list of complications pre-defined – or based on what you found? So if I understand correctly you just counted all reported complications (irrespective of severity) in each study?

Response: We took the complications as defined by individual authors and did not otherwise attempt to extract and standardise complications across the included studies as this would not have been possible from the data published. The summary of what constituted complications in each study is outlined in Table 1. We highlighted this limitation in our discussion and suggested that this may be a source of the high heterogeneity observed.

- Methods: quality assessment could be explained in more detail. Minimum/maximum score, the higher the better etc. Quality assessment was also done in double and resolved by consensus?

Response: Thank you very much to the reviewer for this comment and suggestion. We have now included a more detailed explanation regarding the quality assessment including the domains assessed. Quality assessment was also performed using two independent reviewers with any disagreement resolved via a third reviewer. We have also included an appendix outlining the scores for each domain for individual studies.

- Methods: random effects model more conservative estimates... compared to fixed models.

Response: Thank you for pointing this out. We have clarified in the methods section that the random-effects model provide more conservative estimates when compared to fixed-effect model.

- Methods: number needed to centralize. I struggle a bit with this number, since it doesn't reflect an annual number, and may be depending on the duration of the study period of the different studies? Or do I see this wrong? Since one centre may do 300 surgeries every year, while another may take 30 years to get to this number – so are both then considered high volume?

Response: Many thanks for this comment. Number needed to centralise suggests the number of nephrectomy that needs to be moved or centralised from low-volume hospitals to high-volume hospitals in order to reduce one potentially avoidable adverse event. It should not be interpreted as the minimum number of nephrectomy that a hospital needs to perform to be considered high-volume. For example, our results showed that the NNC for radical nephrectomy mortality is 234. This suggests

that in order to avoid one radical nephrectomy mortality, 234 patient will need to be treated by high volume hospitals instead of by low volume hospitals. From our analysis it was not possible to define the precise case number required for hospitals to be considered high volume and this limitation is described in the discussion section.

- Since the analyses are based on crude pooled odds ratios, I wonder if there are known confounders which should ideally be taken into account in the future –besides the well-described surgeon volume and oncologic factors.

Response: Thank you for this comment. Other well known confounders that have been demonstrated to be associated to patient outcomes include patient age, sex, ethnicity, and number of comorbidities. Some studies have also found the type of hospital (eg. Teaching status) or type of insurance that a patient hold to be significant predictors. We were unable to account for this potential confounders as part of our meta-analysis and have now discussed about this. Future meta-analysis should ideally adjust for these case-mix variables, although this may be methodologically challenging.

- Methods: did you consider comparing the lowest volume group with the highest volume group – a method also done in some other meta-analysis. OK you do not use all information, and results are likely to be more extreme.

Response: Thank you very much to the reviewer for this suggestion. This was an alternative method for calculating the pool effect estimates which was considered by the investigators, but as suggested, not all information would be used and likely to exaggerate the differences in outcomes, which may be harder to interpret or appreciate from a clinical and practical perspective as few hospitals (for example in the UK) would have such a low nephrectomy volume. An attractive benefit of this method would be that we could use an inverse-variance approach whereby the intervention effects (odds ratio) between the lowest and highest volume group and their standard errors could be used in synthesising the pooled effect estimates. This would also accounted for some of the case-mix adjustments performed in multivariable logistic regressions which many studies utilised in their analyses. However, many studies did not report the standard errors and some studies also did not quantify the intervention effects making the practicality of using inverse-variance and its benefits less obvious. We therefore did not proceed with this methodology of comparing the lowest volume group with the highest volume group, but would consider this as part of the secondary analysis if reviewers feel strongly about this analysis.

Reviewer: 3

Reviewer Name: Professor Shantini Paranjothy

Institution and Country: Cardiff University, UK

Competing Interests: None declared

This was a generally well-written paper, although there are a few grammatical errors and varying tenses in some sentences, which can be sorted out with a proof-read.

1. The second and third bullet point in the strengths and limitations section could do with re-phrasing. Bullet point two refers to 17 studies when 19n and 16 are mentioned in the abstract and full-text of the paper. Do you mean that you have stratified your analyses according to the types of nephrectomy, rather than the results? Bullet point three - - it is not clear from the description of the methods and results if this was done adequately (see points below re analysis).

Response: Thank you very much to the reviewer for these suggestions. We have rephrased bullet point two and clarified the number of studies included in the study and ensured that this is consistent throughout the texts. We have also clarified and amended bullet point three to better reflect the

strength of our study.

2. Description of methods - - PRISMA guidelines provide a framework for reporting systematic reviews, it is unusual to see it quoted as a frameworks for the methods of a systematic review.

Response: We have now rephrased this sentence to reflect our reporting of the study using PRISMA guidelines.

3. There is no mention of the types of study design that were included in the review - - was this a consideration in the search criteria? It would be helpful also to have some description of the type of population studies and outcomes of interest - - particularly the complications - - were these defined a-priori? How was postoperative mortality defined? Were there differences in this definition between studies?

Response: Thank you very much for this comment. We have now clarified the inclusion and exclusion criteria in the methods section. Similar to comments from previous reviewer, mortality was defined by the majority of the studies as in-hospital mortality with four studies defining mortality as 30-day mortality or not otherwise specified. We took the complications as defined by individual authors and did not otherwise attempt to extract and standardise complications across the included studies as this would not have been possible from the data published. We highlighted this limitation in our discussion and suggested that this may be a source of the high heterogeneity observed. The definition of mortality and summary of what constituted complications in each study is now outlined in Table 1.

4. The first mention of nephrectomy with venous thrombectomy is in the quantitative data synthesis section - - there is no mention of this earlier when discussing the types of studies that were included.

Response: We have now mentioned about nephrectomy with venous thrombectomy as part of the inclusion criteria at the beginning of the methods section.

5. Given the considerable heterogeneity between studies and different levels of adjustment for case-mix and other confounding factors, it is unclear whether or not it is actually appropriate to combine these results in a meta-analysis. I also wonder whether an analysis using individual patient data may be more appropriate and if the feasibility of this was considered. More discussion about the types of study design and appropriateness of a meta-analysis in the light of the considerable heterogeneity is warranted.

Response: Thank you very much to the reviewer for this comment. Points raised in this comment was discussed and considered during the design of the study. Multiple studies have used patient level data to evaluate the volume-outcome relationship in nephrectomy, but no consensus has been reached about the efficacy of nephrectomy centralisation. We believe our study therefore offers the most contemporary review amalgamating the available data and evidences to date. Interpretation of the results however needs to take into consideration the study limitations such as the high heterogeneity and this was highlighted in our discussion. We note that other study designs may be more appropriate in testing our hypothesis for example by combining primary data from the published studies. However, ethical and practical limitations associated with these designs make them difficult to implement. This is now discussed in the discussion section. As part of our ongoing study, we will be using individual English patient data to test our hypothesis, but this will be with the limitation of a smaller patient cohort.

6. What are the domains of study quality included here - it is not clear what the scoring is based on.

Response: We have now elaborated in the methods section on how quality assessment was

performed including detailed descriptions of the domains assessed. We have also included an additional table in the supplementary material breaking down the scores given to each study in each domain.

7. It is unclear how the overall mortality rates shown on page 14 were calculated - is this a weighted average across the studies?

Response: Thank you very much for this comment. As the cohort size and the number of deaths or complications were known for each study, the overall mortality and complication rates were calculated by summing the number of deaths or complications from each studies and divided that by the total number of patients.

VERSION 2 – REVIEW

REVIEWER	Nele Brusselaers Karolinska Institutet, Sweden
REVIEW RETURNED	24-May-2017

The reviewer completed the checklist but made no further comments.