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)	The impact of physician sex/gender on processes of care, and clinical outcomes in cardiac operative care: A systematic review
AUTHORS	Etherington, Nicole; Deng, Mimi; Boet, S; Johnston, Amy; Mansour, Fadi; Said, Hussein; Zheng, Katina; Sun, Louise

VERSION 1 - REVIEW

REVIEWER	Sumayya Ahmad Icahn School of Medicine of Mt. sinai New York, NY
REVIEW RETURNED	17-Feb-2020

GENERAL COMMENTS	This is a systematic review that attempts to find evidence of outcome differences in patients receiving cardiac surgery by female and male physicians. Overall, the methodology is sound and the tables and results are clearly written. However, the discussion and introduction need some work. The second paragraph in the introduction (lines 23-33) sounds more like an opinion statement than facts, and would be cautious about presenting it the way it stands. The authors suggest that having male physicians in the OR inherently makes teamwork challenging. Later in the paragraph, they cite evidence stating that cooperation and communication decrease with male providers. I would add lines 32-33 toward the beginning of this paragraph, and elaborate further on the specific study referenced that suggest cooperation/communication has decreased. I also would like them to detail further why and how women providers may be challenged and less likely to speak up, and how this directly would impact patient outcomes. -Line 40-44 does not make sense in the context of the paragraph here. I would leave it out, because authors are not discussing resident training here. - The Tsugawa study had 10% female physicians and 90% female, but a sub-group analysis for cardiac surgery was not performed Was this study ultimately included because it did include cardiac procedures even though it did not specifically examine this outcome? If so, I would clearly state this in the Methods section "Eligibility criteria" section. (I.e. Line 25 say, "if a study explored surgical outcomes from various surgical specialties, including cardiac, then it was included in the analysis."

	"Explanation of Findings" section in discussion/results section (Lines 26-53). need to go into more details about the included studies here. Most of the discussion seems to be conjuncture about what is needed to really study this question, which is valid, but I want to know what has already been done. Starting in line 31, I would spend several sentences talking more specifically about the authors' take on the included studies and their limitations and strengths.Line 29-30 in the "Limitations" section briefly mentions this, but I would like more details about this. For instance, what do the authors think about having such a small percentage of female surgeons represented in the Tusgawa study (10%)? Were cardiac female surgeons even included in this-- was this published in the paper? Please elaborate. Regarding the Wallis study, how many female/male cardiac surgeons were included in this cohort? How many years of analysis of data were included, and what was the percentage of cardiac female surgeons included compared to male? (Was this lower then female other surgeons)? What did the authors of this study think about the "trend toward significance"? How do these findings support the results in the literature? - Finally, is there any literature anesthesiologist gender on outcomes of surgical procedures? Please add a sentence or two in the discussion about this. Overall, I think the paper is well executed and well written. With some more elaboration on the results themselves and their implications (as well as tweaking the introduction), I think this paper will contribute nicely to our knowledge about the implication of gender in cardiac surgery.

REVIEWER	Wendy Bernstein MD MBA FASA FAMWA University of Rochester Medical Center USA
REVIEW RETURNED	19-Feb-2020

GENERAL COMMENTS	Would expect a few more references from 2019 Interesting topic- but basically a meta analysis of 2 studies in which there were sex of surgeons identified....suggested idea about gender of cardiac anesthesiologist- but nothing discussed. Seems like alot of work in analyzing thousands of studies but generated only 2 to review closely...

REVIEWER	Elisabeth Svensson Statistics, Örebro University, Sweden
REVIEW RETURNED	03-Apr-2020

GENERAL COMMENTS	The impact of physician sex/gender on processes of care, and clinical outcomes in cardiac operative care: A systematic review This manuscript is a systematic review focussing at the role of physician's sex and gender on clinical outcomes in cardiac surgery care. The study is well motivated and reported according to PRISMA, but the manuscript has some weaknesses and could be improved.

	The variables of interest being extracted are mentioned on page 17. 1. However, the operational definition of the variable “risk of bias” that is assessed by the NIH Quality Assessment Tool for assessment risk of bias is not described. The items, the scale categories and the overall scoring and interpretation must be given. Study selection, Appendix 2. 2. Eight of the 19 full-text reviewed papers were excluded because of the predetermined exclusion/inclusion criteria. The reason for not excluding them in the first run is unclear. What does the reason “commentary” in the Appendix 2 mean? Results, Study characteristics and synthesis and Table1. 3. The two eligible studies found by the systematic review are described in Table 1 and in the text. A presentation of the reviewed studies in the same order as in the table is recommended. 4. The findings could be better structured and linked to the data concepts/variables of interest according to “Data items and abstraction”. 5. Page 18, lines 23-28: A p-value (p=0.04) is not a result. The measures of interest, according to the study aims, should be given together with 95% confidence intervals. 6. Write the 95% confidence interval as for example 95% CI (0.93 to 1.01). Table 1. Wallis: 7. Is there a link between the information given in the table regarding the text “Fewer patients treated by …..” and the percentages that are presented in the table? Were the proportions found in the Wallis-paper or calculated for this review? Page 18, lines 30-31: 8. The 95% CI(OR) from 0.82 to 1.01 is NOT a sign of “trend”, only a weak indicator of evidence. Despite large data sets, the measures of odds ratio and the confidence intervals are close to unity, indicating weak evidences of differences. Have the authors of the reviewed study considered, and adjusted for multiple tests? How about sub-group analyses? Page 18, lines 33-41: 9. The interpretation of the 95% CI(OR) ranging from 0.93 to 1.01 differs from the corresponding confidence interval on line 31! Page 18. Risk of bias. 10. As mentioned, the NIH Quality Assessment Tool must be described and the result of the assessments must be given referring to the items. The information in Table 2 is noninformative – what does quality 11 and 12 mean? The conclusion that the “quality of rating was relatively high” is not a scientifically correct expression.

REVIEWER	Ahmad Farouk Musa Monash University Malaysia
REVIEW RETURNED	05-May-2020

GENERAL COMMENTS	The authors were right in stating that the main limitation of such a study is the scarcity of literature on this topic.

VERSION 1 – AUTHOR RESPONSE

REVIEWER 1 COMMENTS

Comment 1: The second paragraph in the introduction (lines 23-33) sounds more like an opinion statement than facts, and would be cautious about presenting it the way it stands. The authors suggest that having male physicians in the OR inherently makes teamwork challenging. Later in the paragraph, they cite evidence stating that cooperation and communication decrease with male providers. I would add lines 32-33 toward the beginning of this paragraph, and elaborate further on the specific study referenced that suggest cooperation/communication has decreased. I also would like them to detail further why and how women providers may be challenged and less likely to speak up, and how this directly would impact patient outcomes.
Response: Thank you for these suggestions. We have now revised this paragraph as follows:
"Cohesive teamwork and effective communication are especially important in the cardiac OR given the high acuity of cases, frequent and sudden events of hemodynamic instability, critical moments of cardiopulmonary bypass initiation and separation, and the need for precise blood pressure control during key stages of operation. Moreover, the high-risk nature of cardiac surgery makes effective teamwork and communication even more critical in this operative setting. The predominance of male physicians in the cardiac OR compared to other surgical specialties 1 may carry implications for operative communication and teamwork related to gendered hiearchies. For example, studies on non-cardiac OR teams show that female staff anaesthsiologists are challenged more often by the respiratory therapist than their male colleagues when an incorrect clinical decision is made 2,3. This suggests that there are implicit gender hierarchies within the OR and a potential reduction in the professional hierarchy gradient associated with female leadership. Another study found that if the attending surgeon's gender differed from the primary gender composition of the overall surgical team, cooperation increased, and conflict decreased [33]. Specifically, cooperation and communication were observed to decrease when more than half of the providers in an OR were male 4. The highest percentage of conflict interactions was observed in the cardiothoracic OR, where over 95% of staff surgeons were male [33]. With increasing gender diversity in surgery, however, it is likely that team dynamics will also evolve."

Comment 2: Line 40-44 does not make sense in the context of the paragraph here. I would leave it out, because authors are not discussing resident training here.
Response: Thank you, this sentence has been removed.

Comment 3: The Tsugawa study had 10% female physicians and 90% female, but a sub-group analysis for cardiac surgery was not performed Was this study ultimately included because it did include cardiac procedures even though it did not specifically examine this outcome? If so, I would clearly state this in the Methods section "Eligibility criteria" section. (I.e. Line 25 say, "if a study explored surgical outcomes from various surgical specialties, including cardiac, then it was included in the analysis."
Response: That is correct. We have now clarified this in our eligibility section: "Studies that explored the implications of surgeon sex and/or gender in a variety of surgical specialties met inclusion criteria as long as cardiac procedures were included."

Comment 4: "Explanation of Findings" section in discussion/results section (Lines 26-53). need to go into more details about the included studies here. Most of the discussion seems to be conjuncture about what is needed to really study this question, which is valid, but I want to know what has already been done. Starting in line 31, I would spend several sentences talking more specifically about the authors' take on the included studies and their limitations and strengths.Line 29-30 in the "Limitations" section briefly mentions this, but I would like more details about this. For instance, what do the authors think about having such a small percentage of female surgeons represented in the Tusgawa study (10%)? Were cardiac female surgeons even included in this-- was this published in the paper? Please elaborate. Regarding the Wallis study, how many female/male cardiac surgeons were included in this cohort? How many years of analysis of data were included, and what was the percentage of cardiac female surgeons included compared to male? (Was this lower then female other surgeons)? What did the authors of this study think about the "trend toward significance"? How do these findings support the results in the literature?
Response: Thank you for these excellent suggestions. Strengths and weakness and comments about trend towards significance have been added to the section. In summary:
• For Tusgawa, subgroup analysis was not conducted for cardiac surgery as indicated in Table 1 and Study characteristics and synthesis 3rd paragraph.
• For Wallis, the study was conducted from 2007 – 2015; dates have been added to Study characteristics and synthesis 2nd paragraph. The number of female vs male surgeons in each specialty was not specified.
The paragraph now reads as follows:
"Tsugawa et al. included Medicare beneficiaries over 65 years of age undergoing a variety of non-elective procedures. In this study, only 10.1% of surgeons were female and it is unclear how many specialized in cardiac surgery [47]. Hence, sex and gender analysis may have been underpowered in the arena of cardiac surgery practices. Wallis et al. provided greater generalizability by considering all adult patients undergoing 25 common elective and non-elective procedures, with complete tracking of mortality and postoperative complications. Interestingly, Wallis et al. noted some degree of evidence (OR 0.91, CI 95% 0.82 to 1.01) for superior composite outcome of postoperative death, readmission, or complications in patients under the care of female cardiothoracic surgeons compared to male cardiothoracic surgeons. They attributed this finding to female surgeons' tendency to adhere to guidelines, provide patient-centred care, and attention to communication and teamwork [54,55]. Alternatively, this observation could also have be a consequence of effect modification, as female surgeons were more heavily involved elective surgeries, which were in themselves associated better postoperative outcomes as compared to urgent or emergent procedures [29]. Overall, the study by Tsugawa and colleagues did not provide subgroup analysis for cardiac surgery, while both the Tsugawa and Wallis studies failed to specify the proportion of male and female surgeons within each specialty. These studies were also limited by shorter lengths of postoperative follow-up (i.e., 30 days), as well as unmeasured confounders such as complexity of the operation and underlying disease severity."

Comment 5: Finally, is there any literature anesthesiologist gender on outcomes of surgical procedures?
Response: We have now added the following sentence (page 13): "Attention to anesthesiologist sex/gender, in particular, would be warranted given the lack of literature in this area in addition to the potential interaction between anesthesiologist and surgeon sex/gender."

REVIEWER 2 COMMENTS
Comment 1: Would expect a few more references from 2019
Interesting topic- but basically a meta analysis of 2 studies in which there were sex of surgeons identified....suggested idea about gender of cardiac anesthesiologist- but nothing discussed.

Seems like alot of work in analyzing thousands of studies but generated only 2 to review closely...

Response: We agree with the reviewer that the generation of only 2 studies for review after an exhaustive search was somewhat anti-climatic. For this very reason, our systematic review highlights an important knowledge gap in health services research in the field of cardiac surgery.

REVIEWER 3 COMMENTS

Comment 1: The variables of interest being extracted are mentioned on page 17. 1. However, the operational definition of the variable "risk of bias" that is assessed by the NIH Quality Assessment Tool for assessment risk of bias is not described. The items, the scale categories and the overall scoring and interpretation must be given.
Response: Thank you for your suggestion. We have now added to the risk of bias section (page 7) as follows:
"The NIH Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies was used to assess risk of bias 7. This tool includes 14 dichotomous items (i.e. yes or no), such as clarity of the research question, specification of the study population, sample size justification, and measurement of confounding variables. Studies are assigned a score of "1" if the criterion is present, for a total possible score of 14 (high quality). Reviewers assessed risk of bias independently and in duplicate, using consensus or third reviewer consultation to resolve disagreements (MD, FM)."

Comment 2: Study selection, Appendix 2. 2. Eight of the 19 full-text reviewed papers were excluded because of the predetermined exclusion/inclusion criteria. The reason for not excluding them in the first run is unclear. What does the reason "commentary" in the Appendix 2 mean?
Response: We have now revised the results section: "Nineteen studies proceeded to full-text screening by satisfying the inclusion criteria on abstract screening or the abstract did not provide information to confidently be excluded without full-text review." For Appendix 2, commentary was explicitly described as: "comment/opinion on a published article".

Comment 3: Results, Study characteristics and synthesis and Table1. 3. The two eligible studies found by the systematic review are described in Table 1 and in the text. A presentation of the reviewed studies in the same order as in the table is recommended.
Response: Thank you for pointing this out. We have made changes accordingly.

Comment 4. The findings could be better structured and linked to the data concepts/variables of interest according to "Data items and abstraction".
Response: Thank you for this suggestion. Given that only two studies were included, we believe it is appropriate to present the findings as is. We are happy to structure differently, however, if the reviewer could kindly provide more specific suggestions.

Comment 5: Page 18, lines 23-28: A p-value (p=0.04) is not a result. The measures of interest, according to the study aims, should be given together with 95% confidence intervals. 6. Write the 95% confidence interval as for example 95% CI (0.93 to 1.01).
Response: We revised to: "Overall, patients treated by female surgeons had a small but statistically significantly lower 30-day mortality (adjusted OR 0.88, 95% CI 0.79 to 0.99, p = 0.04)."

Comment 6: Table 1. Wallis: 7. Is there a link between the information given in the table regarding the text "Fewer patients treated by ....." and the percentages that are presented in the table? Were the proportions found in the Wallis-paper or calculated for this review?

Response: Table 1 presents the findings of the Wallis et al. study reported by the study authors. They were not calculated for this review.

Comment 7: Page 18, lines 30-31: 8. The 95% CI(OR) from 0.82 to 1.01 is NOT a sign of "trend", only a weak indicator of evidence. Despite large data sets, the measures of odds ratio and the confidence intervals are close to unity, indicating weak evidences of differences. Have the authors of the reviewed study considered, and adjusted for multiple tests? How about sub-group analyses?
Response: We have now revised this sentence: "Interestingly, Wallis et al. noted some degree of evidence (OR 0.91, CI 95% 0.82 to 1.01) for superior composite outcome of postoperative death, readmission, or complications in patients under the care of female cardiothoracic surgeons compared to male cardiothoracic surgeons." In Wallis et al, subgroup analysis was conducted for cardiothoracic surgery: adjusted OR 0.91 (CI 95% 0.82 to 1.01).

Comment 8: Page 18, lines 33-41: 9. The interpretation of the 95% CI(OR) ranging from 0.93 to 1.01 differs from the corresponding confidence interval on line 31!
Response: We are not sure what the reviewer is referring to here and kindly ask for clarification.
Comment 9: Page 18. Risk of bias. 10. As mentioned, the NIH Quality Assessment Tool must be described and the result of the assessments must be given referring to the items. The information in Table 2 is non- informative – what does quality 11 and 12 mean? The conclusion that the "quality of rating was relatively high" is not a scientifically correct expression.
Response: Thank you for this suggestion, please see revised paragraph: "The included studies were evaluated using the NIH Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies. The overall quality rating for the internal validity of each study was relatively high (Table 2). Both studies mitigated risk of bias by having a well-defined research question, pre-specified eligibility criteria, justified duration of follow-up, consideration for key confounding variables, and insignificant loss to follow-up, among others. Two deductions in quality rating were due to the inability in examining different levels of exposure as related to the outcome and in assessing exposure more than once over time, as gender was determined to be binary and fixed in both studies. Tsugawa et al received an additional quality rating deduction for failure to provide sample size justification, power description, or variance and effect estimates." We have also now added an appendix with the ratings for each individual item of the tool for each included study.

REVIEWER 4 COMMENTS
Comment 1: The authors were right in stating that the main limitation of such a study is the scarcity of literature on this topic.
Response: We agree with the reviewer and believe our systematic review identifies an important knowledge gap in health services research for cardiac surgery patients.

VERSION 2 – REVIEW

REVIEWER	Elisabeth Svensson Biostatistics Örebro University Sweden
REVIEW RETURNED	02-Aug-2020

GENERAL COMMENTS	Bmjopen-2020-037139.R1 The impact of physician sex/gender on processes of care, and clinical outcomes in cardiac operative care: A systematic review As far as I can see, most of my comments have been taken into account.