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)	Women's leadership in academic medicine: A systematic review of extent, condition, and interventions.
AUTHORS	Alwazzan, Lulu; Al-Angari, Samiah S.

VERSION 1 – REVIEW

REVIEWER	Stavroula Kalaitzi, PhD Maastricht University Care and Public Health Research Institute (CAPHRI) Department of International Health
REVIEW RETURNED	31-Jul-2019

GENERAL COMMENTS	This article engages with a topic which is of great interest. Unfortunately, as it currently stands, it does not meet any of the criteria which would make it suitable for publication. Neither am I convinced that it has the potential to meet the necessary standard, even following substantial revisions. I would advise the author(s) to revisit their research design, set out very clearly what they aim to achieve in this piece of research and be consistent to their aim throughout the paper (e.g. p. 9 vs Discussion, Conclusion parts). For example, it is hard to see how the Conclusion part addresses the research question 1 and 2; maybe only the research question 3. Even though the study on women's leadership in academic medicine has already received enough attention, there is still room for further research. However, the paper misses important publications in the explored field are missing (e.g. Eagly's, Carli's) and, thereby, ignores important aspects which would certainly enhance the depth and the quality of the research. As it currently stands, the article makes reference to a wide range of publications in the field. Yet, the referenced publications could have been read and explored in a more systematic and critical thinking way for the purposes of the study. The theoretical contribution is not very strong, as the authors address a quite broad topic and the contextualization of the research may benefit from further elaboration. Furthermore, authors may need to be far more precise with the language and the concepts that they use [eg. "the remaining barriers' cultural nature" (p. 5, line 21): what do you mean "remaining"? what are the barriers that left out and why? How does the study approach this? Does it impact the research topic?] The section entitled Methods and Results would benefit from a more clear presentation and description since the research is multi-dimensional and it is hard for the reader to follow properly;

yet, the annexed tables helped to better understand the research design and findings.
Overall, the author(s) need to substantiate better their many claims with academic literature and provide pages if they use quotations. The article could benefit from professional language editing.

In conclusion, this is a very interesting area of research, and I would strongly encourage work in this area and make a strong contribution to the field. Yet, the manuscript needs to be considerably reworked and therefore I suggest rejecting it.

REVIEWER	Elizabeth Ellinas, MD Medical College of Wisconsin, USA
REVIEW RETURNED	01-Aug-2019

GENERAL COMMENTS	I thank the authors for their thorough work on an important topic: the inclusion of women in leadership in academic medicine. My overall critique is that the questions you ask are truly broad. And then the conclusions/discussion do not appear to seamlessly emerge from the results. The results read a bit like a catalog of leadership articles and could use more connections. However, when you get to the discussion, you segue from the article description into an incredibly interesting discussion of the underpinnings of the articles themselves. It's a very interesting commentary – what's your evidence that it is true? For example, what's the evidence within the individual articles that indicates the old-mold thinking and positivist paradigms? Or that articles are institute centric? Further, are they all just count data or do some of them address "sustained success?" Can you, rather than catalog articles, grade them on a metric of count data through sustained success? Same idea for self-nomination, assignment to leadership, length of time in leadership. Rather than telling me the results of the study, can you tell me which studies contained those characteristics? That sort of data would lead me to have more faith in your conclusions. I find your ideas regarding the state of women's leadership literature incredibly interesting and insightful. That said, the overall article reads like a commentary on the current state, rather than conclusions drawn from results. I think that this could be fixed by either changing the article to a commentary, or, changing the type of results that are presented as indicated above.

REVIEWER	Kirsten Fiest University of Calgary, Canada
REVIEW RETURNED	17-Sep-2019

GENERAL COMMENTS	Thank you for the opportunity to review this paper. The topic is relevant and timely. I have a few comments for the authors to consider, split into minor and major. MAJOR 1) There are areas of the PRISMA guidelines that have not been followed completely. For example, the eligibility criteria for study inclusion/exclusion is not well described, and a copy of the search strategy from one database is not included as an appendix (a table is included, but there is not details of the search. 2) My main concern relates to the search strategy itself: i) there is not enough detail provided to allow replication of the search (e.g.,

were the terms searched as keywords, in titles and abstracts only, using subject headings?); ii) details on how the search was developed should be included (e.g., did a librarian help develop? did a subject matter expert help?); iii) there are a number of important synonyms (e.g., sex, gender) in each of the three 'buckets' that were not included in the search, but likely should be to ensure comprehensiveness. In general the search is very narrow.
3) Clarify whether the term "women" is referring to sex at birth or gender identity.

MINOR
1) The abstract does not include all elements that should be included (see PRISMA)
2) The aims/objectives could be more clear and concise.
3) The inclusion/exclusion criteria need to be more clearly described.
4) The methods section requires definitions of key terms (e.g., women [biological sex or gender? both], leadership, and academic medicine).
5) Could the authors describe the "methodological issues not covered by the MERSQI", perhaps by providing an example?
6) What do the authors mean by "original data" in the results section?
7) Some of the statements on Page 11, Lines 42-50 would be better placed in the discussion (e.g., "which may be outdated or inaccurate" "compromising the validity of the findings").
8) The limitations section only discusses the limits of the included studies, not the limitations of the study itself. Please expand on the limitations of this systematic review.

VERSION 1 – AUTHOR RESPONSE

- Reviewer 1 feedback (Stavroula Kalaitzi, PhD):

1. This article engages with a topic which is of great interest. Unfortunately, as it currently stands, it does not meet any of the criteria which would make it suitable for publication. Neither am I convinced that it has the potential to meet the necessary standard, even following substantial revisions. I would advise the author(s) to revisit their research design, set out very clearly what they aim to achieve in this piece of research and be consistent to their aim throughout the paper (e.g. p. 9 vs Discussion, Conclusion parts). For example, it is hard to see how the Conclusion part addresses the research question 1 and 2; maybe only the research question 3.

Our response:

We thank the reviewer for her comments. We choose a systematic approach, as opposed to other review research designs e.g. scoping, narrative, for its robustness and to cast a wide net which would help us find out what has been done in the area of women's leadership in academic medicine. A

systematic approach helped us create research aims at the outset, which we phrased as two broad aims. The first aim has the research objectives embedded in it and phrased as questions (p. 8 line 21-23): "It is our aim to first synthesize work done in this area. We ask: 1) What is the extent of women's leadership? 2) What factors influence women's leadership? 3) What is the impact of leadership development programs on women's individual careers and on medical schools' environments?"

In doing our cursory reading, and prior to officially commencing this research project we knew that it would be difficult to fully commit to a deductive research paradigm, as is common in systematic reviews. Grappling with this, we chose to use, in addition to a systematic review, a narrative approach to document our methodology, results, and discussion. Such an approach is not uncommon in medical education as stated on (p. 12 line 17 - 18): "We draw on a strategy suggested by the Best Evidence Medical Education Collaboration, to provide a narrative of the results".

Our second research aim was (p. 9 line 2-4): "to present an analysis of such works, we concern ourselves, not only with what was done thus far, but how produced knowledge helps or hinders women's leadership.". Although we see the reviewer's point of view; that is the conclusion does address research questions 1 & 2, we believe that the conclusion is aligned with our second aim; in that we synthesised al findings of the 3 research questions.

Revised text:

Not applicable

2. Even though the study on women's leadership in academic medicine has already received enough attention, there is still room for further research. However, the paper misses important publications in the explored field are missing (e.g. Eagly's, Carli's) and, thereby, ignores important aspects which would certainly enhance the depth and the quality of the research. As it currently stands, the article makes reference to a wide range of publications in the field. Yet, the referenced publications could have been read and explored in a more systematic and critical thinking way for the purposes of the study.

Our response:

We agree with the reviewer, the area of women's leadership in academic medicine requires further research. The work of Eagly and Carli are certainly important publications and are broad in their application (e.g. women in leadership). We problematize this in our paper, claiming that the study of women's leadership in academic medicine draws on the generic leadership literature, especially where leadership enactment is concerned. It is for this reason that we steered away from drawing on the broader leadership literature. Nonetheless, we elude to such works: p.5 line 13-14: "Such barriers are, of course, not unique to women leaders or to the context of academic medicine"

In the introduction p.5 lines 13- 22, we believe that we have situated our study within the works done on women's careers in academic medicine: "Such barriers are, of course, not unique to women leaders or to the context of academic medicine. The wealth of literature exploring women's careers, much of it reviewed in 2 systematic reviews, 1 narrative review, and 2 overviews conclude that broadly, women faculty face the very same hurdles as they join and progress in academic medicine in their roles as physicians, teachers, and researchers. Often, these barriers become reason enough for women to leave academia. Although these reviews, and the studies within them, broaden our understanding of women's experiences, they have treated women's leadership as ancillary to a bigger discourse on career progression, often coming to leadership as one solution to gender inequity. By doing so, these reviews ignore the centrality of leadership in shaping culture and the change needed to realize gender equity."

Revised text:

Not applicable

3. a) The theoretical contribution is not very strong, as the authors address a quite broad topic and the contextualization of the research may benefit from further elaboration.

Our response:

It was not the aim of this systematic review to make a theoretical contribution, rather a conceptual one. To that end and in the background section, we delve deeper, bringing attention to the connection between culture and leadership. A novel approach to women's leadership is examining it through the lens of leadership emergence p. 7 lines 16-19: "Such frameworks prompt us to ask question the emergence of women's leadership. For example, whether women self-nominate or are appointed to leadership positions in what Northouse[43] calls *assigned leadership*, how long they hold leadership positions, whether they go on to hold dual leadership appointments, and if they indeed have mentors or sponsors who support their careers?" and leadership enactment p.8 lines 8 -16: "A more nuanced conceptualization of leadership enactment, may offer new insights that would help us address stereotyping. For example, women may take on informal leadership roles, in what Northouse calls *emergent leadership,* referring to leadership that develops organically and is based on building alignments and fostering trust. Moreover, our understanding may be expanded by exploring the values that inform women's decisions, the behaviors they model, and the actions they take to improve the quality of medical education and practice whether formally or informally. Addressing these gaps situates women leaders as critical actors in culture change and begins to conceptually ground women's enactment of leadership in their lived experiences, rather than the broader generic leadership literature". These well known leadership constructs, and drawing on the work of Northouse, have not been used explicitly to understand women's leadership in academic medicine. Indeed, our work reveals that factors hindering women's leadership can be thought of as those which hinder

women's emergence and those which hinder their enactment of leadership (table 2). Moreover, such leadership constructs prompt us to shift perspective on women's leadership, asking questions that put the onus on the women rather than the institute. Such a shift also prompts a change in research design, currently most studies are cross-sectional and use websites as data sources. To ask about emergence and enactment requires that the methodology change as well. As we have commented in the discussion section p.20 line 1-6 for example, "Most studies in this theme neglected how women emerge as leaders. It is for this reason that we drew on Northouse's work and devised a metric (Table 1). Although the metric is not an exhaustive, it is an initial attempt to introduce the construct of *leadership emergence* into the discourse on women's leadership in academic medicine. For example, we found that only 2 studies commented on women being appointed, and no studies mentioned weather women self-nominated. Our intuition is that informal leadership is common amongst women but whether they self-nominate for formal leadership remains to be seen."

b) Furthermore, authors may need to be far more precise with the language and the concepts that they use [e.g. "the remaining barriers' cultural nature" (p. 5, line 21): what do you mean "remaining"? what are the barriers that left out and why? How does the study approach this? Does it impact the research topic?]

Our response:

After mentioning the cultural barriers we go on to enumerate them, by doing so we believe we have defined what we mean by the cultural barriers that remain (p. 5 line 8-11): "The difficulty, in part, stems from the remaining barriers' cultural nature: traditional models of work, implicit gender bias, limited access to support systems both mentors and sponsors, gender stereotyping, gendered views of leadership, and culture-abiding self-imposed constraints."

Moreover, in the background section we go on to explain culture and the nuanced relationship it has with women's leadership. p. 7-8.

c) The section entitled Methods and Results would benefit from a more clear presentation and description since the research is multi-dimensional and it is hard for the reader to follow properly; yet, the annexed tables helped to better understand the research design and findings.

Our response:

We updated the Methods section according to PRISMA using subheadings for a more clear presentation and description.

d) Overall, the author(s) need to substantiate better their many claims with academic literature and provide pages if they use quotations. The article could benefit from professional language editing.

6

<u>Our response:</u>

We believe that we substantiate our claims with proper references. For example, we referenced the works of Carr, Morahan, Ellinas, pololi, notable scholars in the areas of women's leadership. We of course could not include all references in our work, as there are many, both in the women's literature and in the broader leadership literature. However, for the purposes of this systematic review, we believe we have substantiated our claims with suitable references.

We also had a professional language editor proofread the manuscript. Thank you for the recommendation.

- **Reviewer 2 feedback: (Elizabeth Ellinas, MD)**

1. My overall critique is that the questions you ask are truly broad.

<u>Our response:</u>

We thank the reviewer for her thought-provoking critique, which we drew on heavily to improve our paper. The first valid comment about the research questions being truly broad, is notion that we grappled with when conceiving, and again when writing, this work. We used the work already done in the area as a point of reference. Based on our cursory search and readings, we found previous reviews to report on leadership as an intervention. For example, the work of Laver et al (2018) a work published in the BMJ Open which looked at intervention that aimed to help women in their career progression, and where leadership was framed as an intervention. Such work encouraged us to pay attention to the nuances of women's leadership. To do so we had to cross-cut the outlined research objectives in order to present a well-founded conceptual argument.

As scholarly authors, it would have certainly been in our favor to slice the findings into 3 separate reviews, however, the examination we embarked on was not one of cataloguing studies in the field. Rather posing the 3 questions simultaneously allowed for conceptual findings to emerge.

In anticipation that the reader may also have the same concern. we added a note to the body of our manuscript.

<u>Revised text:</u>

Background Page 8 line 22 and page 9 Lines 1-4:
"..We concede that our research questions are broad in scope. We believe it is necessary to cross-cut through these interconnected areas to meet our second aim, which is to present an analysis of such works in the field and critique their collective conceptual framework. We concern ourselves, not only with what was done thus far, but how produced knowledge helps or hinders women's leadership."

2. And then the conclusions/discussion do not appear to seamlessly emerge from the results. The results read a bit like a catalog of leadership articles and could use more connections. However, when you get to the discussion, you segue from the article description into an incredibly interesting discussion of the underpinnings of the articles themselves. It's a very interesting commentary – what's your evidence that it is true?

<u>Our response:</u>

We agree the discussion and conclusions do not seamlessly emerge from the findings. We have reworked the manuscript considerably to address this very important feedback.

<u>Revised text:</u>

Considerable changes highlighted in the manuscript.

3. For example, what's the evidence within the individual articles that indicates the old-mold thinking and positivist paradigms? Or that articles are institute centric? Further, are they all just count data or do some of them address "sustained success?" Can you, rather than catalog articles, grade them on a metric of count data through sustained success? Same idea for self-nomination, assignment to leadership, length of time in leadership. Rather than telling me the results of the study, can you tell me which studies contained those characteristics? That sort of data would lead me to have more faith in your conclusions.

<u>Our response:</u>

We determined that positivist thinking underpinned most studies through the studies':
   1) Main objective, which was to document gender distribution.
2) The use of cross-sectional study designs and self-reported questionnaires which are aligned with positivist thinking. We added a paragraph in the discussion section to address this point and is outlined in the "Revised text" below.
3) Most studies neglected contextual data, understandably to present focused research objectives and to meet positivist standards.

We believe such study questions and methodology are sound in their own right. However, in reading the leadership literature, specifically the work of Northouse, we have come to recognize that such questions and methods do not serve the conceptual growth of women's leadership studies in the field of academic medicine. It is for this reason, we present the constructs of leadership emergence and enactment in the background section p. 7-8

Drawing on the work of Northouse, we have also appraised each article against an initial metric of 4 items in leadership emergence (Table 1).

Furthermore, we clarified in the hindering factors sections that such factors can be categorized to what hinders leadership emergence and what hinders enactment (Table 2). The distinction is important and we believe it will influence future study designs.

Moreover, we have updated the leadership development section, pointing out that leadership development programs are an exception to the rule in terms of leadership enactment, examining the values, behaviors, actions, and style women embrace (another metric we devised to appraise leadership enactment study). Once more, the metrics we devised are an initial attempt to orient work in this area but should be helpful in generating future research questions.

4. I find your ideas regarding the state of women's leadership literature incredibly interesting and insightful. That said, the overall article reads like a commentary on the current state, rather than conclusions drawn from results. I think that this could be fixed by either changing the article to a commentary, or, changing the type of results that are presented as indicated above.

5. Our response:
   Thank you. As a result of the reviewer's feedback we have considerably reworked the following sections: background, results, and discussion. So much so, that we can not present any one quote to illustrate our changes in this document. We came to the conclusion that studies are institute-centric only after reviewing the literature and not before.
   We completely agree with the reviewer and have attempted to create more connection between the results and our conclusions.

-Reviewer 3 feedback: (Kirsten Fiest PhD)

MAJOR

1) There are areas of the PRISMA guidelines that have not been followed completely. For example, the eligibility criteria for study inclusion/exclusion is not well described, and a copy of the search strategy from one database is not included as an appendix (a table is included, but there is not details of the search.

Our response:

We thank the reviewer for her valuable comments. We considerably reviewed and re-worked the methodology section based on the comments and concerns provided. We included subheadings instead of a narrative

description in the methodology to align with the PRISMA statement as follows:

- Eligibility criteria
- Information sources
- Search strategy
- Study selection
- Data collection process
- Risk of bias assessment

Inclusion/exclusion criteria were revised and added to the text and an example of the search strategy was added to the text under "search strategy section".

<u>Revised text:</u>

- Methods, Eligibility Criteria, Page 9, Lines 7-20:
  "Search results were independently reviewed against a set of a priori inclusion criteria that included all peer-reviewed 1) English-language articles; with 2) quantitative methodologies (Prospective and retrospective cohort, cross-sectional, and pre-post interventions); of studies evaluating 3) the extent of women's leadership in academic medicine at a departmental, college, and medical graduate program level; 4) hindering and facilitating factors to women's leadership as perceived by women and men faculty members and leaders; 5) studies that document leadership interventions and their efficacy as reported by women participants of such programs and their home medical schools. We included a case study because it presented quantitative descriptive information on women in leadership across non-Western multinational settings. Although we recognize the interconnection, we excluded quantitative studies that explored women's leadership in professional societies, journal editorial boards, and journal editorships, focusing our examination solely on leadership within medical schools and graduate residency programs. In addition, qualitative study designs were excluded."

- Methods, Search strategy, Page 11, Line 2-21: "An example of a database search strategy is as follows:

"An example of a database search strategy is as follows:

"MEDLINE search: Ovid

1. (Women or woman or female or females or girl or girls).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
2. limit 1 to English language

3. (Leadership or Leader or leaders or leading).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

4. limit 3 to English language

5. (Medical education or academic medicine or health professions education or health profession education or professional development or faculty development).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

6. limit 5 to English language

7. 2 and 4 and 6"

An updated literature search followed the same search strategy from April 2018 to October 2019."

2) My main concern relates to the search strategy itself: i) there is not enough detail provided to allow replication of the search (e.g., were the terms searched as keywords, in titles and abstracts only, using subject headings?); ii) details on how the search was developed should be included (e.g., did a librarian help develop? did a subject matter expert help?); iii) there are a number of important synonyms (e.g., sex, gender) in each of the three 'buckets' that were not included in the search, but likely should be to ensure comprehensiveness. In general the search is very narrow.

Our response:

We agree that the search strategy may not have reflected the thorough and systematic search process performed by the authors.

i)We opted to use key terms instead of subject headings as the latter revealed either too broad or out of context search results. Terms used were key words mainly extracted from the preliminary readings of the literature in the titles and abstracts of the searched literature. We added details to our search strategy to allow for replication, as recommended.

ii) The first author followed a rigorous and systematic search strategy plan according to best review practices as a librarian help was not available at the time. In addition, experts in the field were contacted for published studies not revealed through the database search. The point raised have been indicated in the search strategy.

iii) Yes, the terms were included to ensure comprehensiveness. Key terms were defined based on the preliminary readings of the literature to ensure the comprehensiveness of our search key terms.

We updated the text according to the reviewer's comments.

Revised text:

- Methods, Search strategy, Page 9, Lines 13-20:

"Systematic searches were performed on the selected 6 online bibliographic databases using a combination of key terms including, but not limited to, "women", "female", "females", "girl", "girls", "leadership", "leader", "academic medicine", and "medical education". The literature was searched in only the "title" and "abstract" search fields. The searches were filtered by applying the inclusion criteria and literature was identified by using keywords and applying Boolean operators 'OR' and 'AND'. Key terms were defined based on the preliminary readings of the literature to ensure the comprehensiveness of our search key terms."

-Methods, Search strategy, Page 10, Lines 5-6:

" In addition, experts in the field were identified and contacted for published studies not revealed through the database search…... The first author followed a systematic and rigorous plan according to best review practices. A librarian's help was not available. Following the PRISMA protocol, she then screened the compiled results, excluding irrelevant articles, and inductively developed a preliminary thematic framework."

-Methods, Search strategy, Page 10, Lines 19-22:

"Key terms were defined based on the preliminary readings of the literature to ensure the comprehensiveness of our search key terms. For example, "Women", in our search context included articles with a clear indication of participants being identified in the published studies as "females" regardless of their background and given that the literature was not distinctive between sex at birth and gender identity in women's leadership."

3) Clarify whether the term "women" is referring to sex at birth or gender identity.

Our response:

"Women", in our search context included articles with a clear indication of participants being identified in the published studies as "females" regardless of their background and given that the literature was not distinctive between sex at birth and gender identity in women's leadership.

Revised text:

Methods, Search strategy, Page 10 Line 19-21 &page 11, Line 1:

"Key terms were defined based on the preliminary readings of the literature to ensure the comprehensiveness of our search key terms. For example, "Women", in our search context included articles with a clear indication that the participants in the published studies identified as "females". The literature did not differentiate between sex at birth and gender identity in women's leadership. As a result, we do not differentiate sex and gender in this review."

MINOR

1) The abstract does not include all elements that should be included (see PRISMA).

Our response:

The abstract was modified according to PRISMA and highlighted in the manuscript in page 2-3.

2) The aims/objectives could be more clear and concise.

Our response:

We believe our aims are now clearly outlined in the background section Page 8 line 22 and page 9 Lines 1-4:
"..We concede that our research questions are broad in scope. We believe it is necessary to cross-cut through these interconnected areas to meet our second aim, which is to present an analysis of such works in the field and critique their collective conceptual framework. We concern ourselves, not only with what was done thus far, but how produced knowledge helps or hinders women's leadership."

Furthermore, our research objective phrased as a questions are clear on page 8. Lines 20-22

"We ask: 1) What is the extent of women's leadership? 2) What factors influence women's leadership? 3) What is the impact of leadership development programs on women's individual careers and on medical

schools' environments?"

<u>Revised text:</u>

Not applicable

3) The inclusion/exclusion criteria need to be more clearly described.

<u>Our response:</u>

We have modified and updated this important point both in the abstract and in the body of the manuscript.

<u>Revised text:</u>

Abstract page 2 line 12-16

Methods, Eligibility Criteria, Page 9, Lines 6-20.

"Search results were independently reviewed against a set of a priori inclusion criteria that included all peer-reviewed 1) English-language articles; with 2) quantitative methodologies (Prospective and retrospective cohort, cross-sectional, and pre-post interventions); of studies evaluating 3) the extent of women's leadership in academic medicine at a departmental, college, and medical graduate program level; 4) hindering and facilitating factors to women's leadership as perceived by women and men faculty members and leaders; 5) studies that document leadership interventions and their efficacy as reported by women participants of such programs and their home medical schools. We included a case study because it presented quantitative descriptive information on women in leadership across non-Western multinational settings. Although we recognize the interconnection, we excluded quantitative studies that explored women's leadership in professional societies, journal editorial boards, and journal editorships, focusing our examination solely on leadership within medical schools and graduate residency programs. In addition, qualitative study designs were excluded."

4) The methods section requires definitions of key terms (e.g., women [biological sex or gender? both], leadership, and academic medicine).

We thank the reviewer for addressing this point. We updated the methods section to accommodate the meaning of the terms "women" and "academic medicine" literally and contextually. In defining "leadership in academic medicine", we point out that the examined leadership positions were on the departmental, college, and post-graduate program levels.

Revised text:

- Methods, Search strategy, Page 10, Lines 20-22:
 "For example, "Women", in our search context included articles with a clear indication that the participants in the published studies identified as "females". The literature did not differentiate between sex at birth and gender identity in women's leadership. As a result, we do not differentiate sex and gender in this review.

- Methods, Eligibility criteria, Page 9, Line 11:
" …leadership in academic medicine at a departmental, college, and medical graduate program level…"

5) Could the authors describe the "methodological issues not covered by the MERSQI", perhaps by providing an example?

Our response:

In addition to objectively appraising the included articles using MERSQI, specific methodological issues were assessed and addressed such as lack of contextual demographic or career data, limited population, lack of statistical adjustments, and lack of follow up. We updated the text based on the reviewer's comment to allow for clarification.

Revised text:

Methodology, Risk of bias assessment, Page 13, Lines 3-5:

"Furthermore, we point out specific methodological issues (e.g. lack of contextual demographic or career data, limited population, lack of statistical adjustments, and lack of follow up) not covered by the MERSQI assessment (Supplementary material 1, Methodological limitations)."

6) What do the authors mean by "original data" in the results section?

Original data in that context refers to primary source data and original research study. Two papers were reported in 2 different studies, making the number of articles 42 but the actual studies 40.

Revised text:

Not applicable

7) Some of the statements on Page 11, Lines 42-50 would be better placed in the discussion (e.g., "which may be outdated or inaccurate" "compromising the validity of the findings").

Our response:

We removed the highlighted sentences from the results section and better worked them in the discussion section page 20 line 7-10.

Revised text:

Discussion, Page 20, Lines 7-10:
"Furthermore, even within the parameters of positivist thinking, all studies are methodologically poor, having a MERSQI range of 7-12.5. Of the 40 studies, 62.5% were cross-sectional. Most studies used websites (which may be outdated or inaccurate, compromising the validity of the findings) or self-reported surveys for data. The median response rates where questionnaires were used was 60% (range, 22%-100%). Many studies failed to explain how their questionnaires were developed or if they were validated."

8) The limitations section only discusses the limits of the included studies, not the limitations of the study itself. Please expand on the limitations of this systematic review.

We outlined the following limitations which are of our study and not the individual studies which are already critiqued in the review. The limits include:

1)Restricting the review to quantitative literature and arguing for studying contextual organizational nuances, which might have been explored in qualitative studies, which we did not review.

2)Defining leadership as a process of influence between leaders and followers but have limited our discussion to the leader's perspective.

3) We found that all studies except one were conducted in North America and Europe. As a result, the presented evidence may not reflect non-Western contexts, but we have forgone discussion of this finding.

The third limitation may be understood as a critique of the studies context, and it is. However, this is a limitation of our study in that we did not discuss how this finding influences interpretation of our results.

Revised text:

Not applicable.

VERSION 2 – REVIEW

REVIEWER	Kirsten Fiest
	University of Calgary, Canada
REVIEW RETURNED	28-Oct-2019

GENERAL COMMENTS	Thank you for addressing the reviewer's comments.