PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Medical leadership - a systematic narrative review: Do hospitals and healthcare organisations perform better when led by doctors?
AUTHORS	Clay-Williams, Robyn; Ludlow, Kristiana; Testa, Luke; Li, Zhicheng; Braithwaite, Jeffrey

VERSION 1 – REVIEW

REVIEWER	Thomas Andersson University of Skövde, Sweden
REVIEW RETURNED	14-Oct-2016

GENERAL COMMENTS	I appreciate the amount of work that is put into this review. FUrthermore, medical leadership is an important issue in the increasingly complex healthcare organizations. Especially considering the conflict between professionalism and managerialism that appears in most hospitals after New Public Management. However, I have some major concerns regarding the paper: - The relevance of the research question - The databases used in the literature review - Lack of definitions of concepts - Results The relevance of the research question The question whether doctors should lead hospitals or not is frequently debated. There is consequently good motives for performing such a literature review. However, is it really possible to answer this research question? I would say no, and therefore, the literature review maybe should have been performed with another focus. There are several reasons why the question is not possible to answer: 1. What is hospital performance and how should it be measured? Healtchare organization performance is complex and means balancing several different and often competing measures. What should be included in such a measure? What is possible to say about healthcare organisation performance based on one isolated measure? Furthermore, does it concern short-term or long-term performance? 2. Healthcare performance is dependent on many different variables. Who leads the organization is only one of many factors. How is it possible to relate the performance to who is leading the organization? Furthermore, what does "led by doctors" mean? The organization is led by a doctor? Almost all top managers are doctors? A majority of top managers are doctors? Doctors are represented among top managers?

	3. How is it possible to relate this to medical training? If it would be possible to the above research question, how is this result then possible to relate to medical training? How to ensure that the succes of a doctor manager is based in medical training and not experience of manager positions, networks, political skills, personality, gender, interpersonal skills, economical skills etc etc. The databases used in the literature review Medline and Embase are focused on medical/healthcare research, but there is a lot of relevant research within the general fields of organization studies, leadership, management, professions, quality etc that are excluded by your choice of databases. Several of the most cited and most influential articles in the field are thereby excluded in your literature review. Lack of definitions of concepts The most important concepts are never defined in your review, which confuses your contribution. Medical leadership is never defined. Leadership is not defined, and based on how you treat it in the text, I am worried that you confuse leadership with managerial work. Results It is not clear to me how a description of the five studies are relevant to fulfil your aim? What contribution do they enable? Considering the results I would have suggested a re-focus of the research question. Otherwise, the only result is that there are very few studies, and that these studies provide only partial contributions to the research question, but enable little comparison. Minor comments: p. 3 r. 3 I would say that this assumption is common among doctors, but among others the opposite is just as common. p.3 r. 8 Is this aim possible to fulfill? p. 3 r. 15 This excludes the fields of organizations studies, leadership, management, profession, quality etc. p. 4 r. 10-18 Considering these limitations, shouldn't you have changed the focus of the review already here? p. 6 r. 17 A too simple question for a too complex problem p. 9 r. 8 Which are these predetermined exclusion criteria? p. 9 r. 48 Healthcare leadership and medical leaderhip seems to used interchangably in the paper. Is there a difference? p. 13 r. 59 Research on management training does not confirm your analysis since the link between management training and succesful managers have proven very weak, if ever consisting at all. Managers seem to learn by practice (e.g. Watson, 2001; Glouberman & Mintzberg, 2001) I appreciate the amount of work that is put into this literature review, and I hope that you can re-focus your research question and still have use of the amount of work you have put into this review and paper. I wish you the best of luck in revising the paper! References Glouberman, S. and Mintzberg, H. (2001), “Managing the care of health and the cure of disease – part 1: differentiation”, Healthcare Management Review, Vol. 26 No. 1, pp. 56-69. Watson, T. (2001), In Search of Management: Culture, Chaos and Control in Managerial Work, 2nd ed., Thomson Learning, London.

REVIEWER	Noordegraaf Utrecht University The Netherlands
REVIEW RETURNED	09-Nov-2016

GENERAL COMMENTS	This is a text that generates mixed feelings. On the one hand it is very clear. There is a staigthforward reseach question: does medical training lead to better leadership? A literature review is performed. Only a few relevant papers are found. And there is no conclusive evidence. Although this is disappointing, it is an outcome - in fact, that it is disappointing might be seen as the most crucial outcome. We have high expectations or strong assumptions regarding medical leadership, but there are no grounds for this. It is hardly studied, and the studies that are available show little. On the other hand, it is too simple and 'empty'. Why this RQ? Is this really important, the link between leadership and medical training. If so - I think it is - why? This requires better explanation. There must be more stress on the expectations and/or assusmptions we (and the reader) might have. Medical leadership, for example, might be important because of expertise and authority dynamics - we might expect medical leadership to be effective, because of this. But conversely, we migth also assume that this produces leadership weaknesses. In other words, there might be various hyotheses, in order to clarify the relevance and substance of the research, and to work towards more tangible results. Moreover, the authors need to clarify their stress on leadership, and how this differs form executive and managerial perspectives. I assume that they focus on executive leadership, at the apexes of health care organizations, which is about leadership but also about executive work. Or do they also focus on health care managers, throughout hospitals (at some places it seems they do). This is all very confusing. Moreover, it leads to a further remark: the text would not only become better when there would some theoretical focus (expertise, authority), as well as expectations/hypotheses, but also when conditions would be introduced. Do we expect medical training to affect leadership/executive work/management (?) in a generic sense, or in specific senses. Does the nature and size of a hospital matter? Does the nature of a division/directorate matter? Does experiece matter? But also, does medical training itself matter? If medical training is renewed, with new competencies, does this lead to better management/leadership? Etc. Finally, set against all of these remarks, did the authors capture everything? E.g. recent work by Kirkpatrick & Veronesi, such as: Clinicians on the board: what difference does it make? G Veronesi, I Kirkpatrick, F Vallascas - Social Science & Medicine, 2013 This work poses and answers exectly the same question, but with more nuance and outcomes. In that respect, I am inclined towards 'reject', but in the light of the relevance of the theme and the potential of the text, I opt for 'major revision'.

REVIEWER	Josie Athens Department of Preventive and Social Medicine, University of Otago, New Zealand
REVIEW RETURNED	14-Dec-2016

GENERAL COMMENTS	1. My main concern is the stated aim of the paper in the abstract. The last sentence of the objectives says: "The aim of this systematic review is to determine whether medical training is associated with superior management performance in terms of organisational performance or patient outcomes". A systematic review, will not answer that question, a meta-analysis may, or an experimental study. By fixing this sentence, the paper will make more sense. 2. The authors considered papers in which both groups (medical trained and not medical trained) are compared. Another approach would be to look at articles that reported patient outcomes and management skills for only one of the groups and compare with results from the other group. They would have had more information to answer the proposed research question. I make this comment, is to emphasise the need to be clearer on the aim of the paper. 3. In their conclusions, authors make statements like "despite considerable interest in the topic". That seems to be an opinion with no references, and I do not have any idea what is their definition of "considerable". We avoid that kind of qualitative adjectives in the scientific literature. 4. Regarding the statistics, the authors report Cohen's Kappa coefficients with corresponding p-values. It would be better if the 95% CIs of the coefficients were also reported. 5. I recommend for the conclusions to be rewritten. The first sentence does not derive from any analysis presented in the paper. Furthermore, the conclusions lack implications for further research.

REVIEWER	Erwin Loh Chief Medical Officer, Monash Health Clinical Professor, Monash University
REVIEW RETURNED	31-Jan-2017

GENERAL COMMENTS	Overall, a very interesting paper, attempting to provide further evidence for an under-researched area of medicine, leadership and health services management. Although the paper does not provide an answer, it does provide a summary of what was found through the systematic review, and discusses them adequately. Although it does not say this, it does also show that this is an under-researched area in need of further examination, funding and focus. Some specific comments as suggestions for improvement: First sentence under Background - should the author define clinician and doctor, as they are being used interchangeably but may mean different things.

	Second sentence - should business-bureaucratic be in quotations as it's not a generally used term, and I assume it is quoted from the cited reference. Same with "organisational-operational" - the hyphenated word is not usually used. I am not certain I understand the significance of the 1983 Griffiths Report with the new model, and how it links to the US clinical directorates, unless it's pure timing - what prompted the clinical directorate idea in the US? Because the author then says "these ushered in the move" but you cannot say the idea of clinical directorates ushered in clinical directorates, which would be a tautology. The statement "Doctors did not always consider such roles attractive" deserves a citation or is this an anecdotal observation? Second paragraph, the author starts with clinicians but ended with doctors - as noted, do they mean the same thing, or is the author saying doctors are a subset of clinicians. Also, is a doctor in full-time management technically still a clinician if they don't do clinical practice? (this is a rhetorical question, but is an interesting one to explore). Also, the statement: "medical departments are normally led by doctors, and these medical leaders are typically a member of the executive team" is not always true - medical departments are usually led by clinical directors, who traditionally are not members of the executive team, only the medical director/director medical service/chief medical officer is (although this is changing in more recent times). There is also a minor typo - "onengaging doctors". In the third paragraph, the author refers to physician executives, which is unique term restricted to the US - no other country has this designation - in Australia, the term specialist medical administrator is used. The author may want to clarify for the readers who may not be aware or familiar. This is a strong statement: "articles in favour of medical leadership (invariably written by those with medical qualifications)" especially with the use of the term "invariably" - okay for an editorial or opinion piece, but unless it's back by empirical evidence, I know that not every single piece of literature supporting medical leadership is written by a doctor. This statement: "The prevailing opinion in some quarters" - needs qualification. What quarters? The author should explain based on the references cited to support this statement. Great explanation of the search strategy and eligibility criteria. In relation to synthesis of results "through inductive interpretive analysis of extracted data", was it done using software, or through human analysis? I assume the later? When the data was coded, was that entered into a computer for software or done manually to find connections? Great job narrowing down the review to five studies. Good description of how that occurred.

	Great analysis and description of each of the five studies. Very clear. Good discussion. As noted, it would be very difficult to determine if there was causation or just correlation, as there may be a lot of other factors at play. The Additional Evidence section is interesting but I am not sure how it helps in answering the research question, or even provide clues. The author needs to be clearer as to what clues each point that has been raised provides. What is the relevance of barriers, the point about training for doctors, the multidisciplinary approach, or doctors who combined clinical and management practice, with the research question? For example, the final section about clinician managers as competing is a good observation, but the author should explain how this is a clue so that it is obvious for the reader. Typo in the conclusion: "Despite considerable interestin healthcare" Overall, I believe the paper explores an important area in medicine and management, and even though there was not a conclusive answer, deserves publication subject to the minor amendments suggested above.

REVIEWER	Dr Yiannis Kyratsis City, University of London United Kingdom
REVIEW RETURNED	07-Feb-2017

GENERAL COMMENTS	Below is a list of comments for the authors to consider and reflect 1. Search strategy: you have included in your systematic review two medical databases, namely Medline and Embase but you have omitted existing health management / health policy specific databases such as ProQuest and HMIC (Healthcare Management Information Consortium). Given the topic of your review (leadership, management and healthcare organisational performance) this is a substantial shortfall as relevant studies reporting management research findings might not have been captured by your search strategy. 2. Search strategy: The terms you have used in your search strategy are apt but perhaps limited. Additional key terms, which in my view are relevant, include “medical”, “practitioner” (to also capture general practitioner-GP), “administrator”, “director”, “head” [lead]. This is important as you claim (p.7) that you included studies that report on both middle and upper level healthcare managers. Your strategy would not have captured for example clinical directors or heads of departments; it does not also capture the published work on similar topics i.e. medical engagement (see the work published by Peter Spurgeon from Warwick University). 3. If you (decide to) amend and re-run the search strategy it would be advisable to extend the time period to the end of 2016 to make it up to date.

	It would be also helpful to articulate your rationale for restricting your search for the period 2005-15 (16?). 4. On page 5 the claims made of the strengths and weaknesses of medical leadership need to be supported by appropriate referencing to relevant published work, as it currently looks rather arbitrary. The overall background discussion is mainly located to the USA and the UK with two almost random references to studies reporting from Nepal and Turkey. You could improve the consistency in presenting the broader global story on this topic. 5. Your research question: 'do hospitals and healthcare organisations perform better when led by doctors?' This implies that you only include doctors who take up designated senior roles in healthcare organisations; mainly CEOs, Chair of Boards or Partners/owners in GP practices. Does this capture the middle level managers you mention in the inclusion criteria? What would be the middle management roles for doctors and what would be the equivalent position for non-clinical managers? 6. Exclusion criteria: Since you ended up with a very limited pool of final papers the 8 studies that were excluded on the basis of criterion n.10 (in the table on pages 18-19) could be briefly presented in a summary or supplemented table and explain in more detail in the main text the parts of these studies you deemed relevant, though peripheral, to help you answer your question. On the same table, criterion n.8 is not very clear to me and I also think it overlaps with criterion n.10. Finally criterion n.6 states that you excluded systematic reviews and this need to be also mentioned explicitly in the main text. 7. Data collection process: the elements (h, data) and (I, methods) seem to be confusing and overlapping to a great extent; also what do you mean by "contextual factors" as stated on elements (i) and (j)? How these contextual factors relate to quantitative and qualitative results? Given the research question you use did you expect to find any qualitative study that would be relevant/appropriate to answer your question? 8. Synthesis of results: you state that the results were synthesised through inductive interpretive analysis; you provide no further detail or cite any relevant reference to allow the reader understand what exactly you did at this stage of your work. In the presentation of your results there is only a descriptive account of the reported findings from the five included studies in your final pool of articles. 9. Risk of bias: the critical appraisal of the studies using the categories by Hawker et al could be discussed in more depth rather than briefly presenting a short summary with your subjective rankings of the different components of the selected articles. Since you only have a limited number of included studies you could discuss the methodological strengths and weaknesses more explicitly and in more detail. 10. Discussion: Related to a previous comment in the synthesis of your results, the themes or sets of concepts are not presented to the expected extend anywhere in the paper; you only present a short list of topics examined by the studies at the beginning of the discussion section. 11. Discussion: what you report under the titled section 'additional evidence' although interesting and broadly associated to the topic it does not relate directly to the specific question you asked; why did you focus on this specific dimension of the characteristics of managers who were doctors and not other (e.g. leadership development for medical doctors, leadership practices etc.)? 12. In the limitations of your study you need to include the fact that your review is time and language limited (only English).

	You also did not conduct any library hand search or followed up the references of the included pool of papers or other seminal papers/books in the field; you did not manage to retrieve the full text from 5 potentially relevant papers. I also do not understand the bullet point on p.4, which is repeated in the limitations section, and which states: “most studies examined only one or two aspects of leadership; this limited the generalisation of findings”. All empirical studies need to be focused on specific domains and any empirical article can only report findings that examine a limited empirical observation. The wording of your research question was limited to empirical studies that could show an association between medical leadership and organisational performance or patient outcomes rather than also explain how or why this association might work, in other words what is the specific mechanism leading to such a result. Your question focused on individual models of leadership rather than distributed or team-based / collective leadership approaches. 13. Conclusion: this is underdeveloped. You could at the very least suggest areas for future research and propose robust study designs that would better answer your research question. I hope you find this useful to help you improve and strengthen your paper. Wish you all the best with your revision.

REVIEWER	Beth Shaw NICE, UK
REVIEW RETURNED	12-Apr-2017

GENERAL COMMENTS	1. The overall research question is very broad - perhaps it would be helpful to the reader to expand on the more detailed objectives in the Methods section? I think this would also help in understanding more about the approach taken as I found this also to be less clear than it could be... 4. In terms of systematic reviewing, there are areas where I would usually hope to see more detail so that if wished, the review could be replicated. I have used your PRISMA statement responses to highlight some areas of specific concern. Please see attached... 5. As you will see in the PRISMA table I am not fully clear what the outcomes are - in some areas it reads as if any quantitative assessment of organisational or patient outcomes by type of manager are the key outcomes - but then the first sentence (p13 lines 16-18) seems to contradict this... 7. Please can full statistics be reported - even if only as reported in the papers - and noted if there is no additional information. For example, (p10 lines 1-3) "no sig difference between clinicians and clinical managers" - what was the measure for this or if not reported, this could be noted? 9. This is my key concern - and reasons for this are noted in other sections. 10. Similar concerns as for 9 - but there is also some need to improve consistency of language - I think your focus is on medical doctors - perhaps that could be clarified throughout. I also think the section on Additional evidence suggests the purpose of the review was not sufficiently clear at the outset hence the inclusion of 'other evidence' that is interesting? 11. Please see comments related to 5 and 10

	Overall, this is an interesting question and the authors make a strong case for its importance. However, in many areas it reads less like a true systematic review of effectiveness and more like a structured scoping review. There also may be other methods of review (such as realist synthesis or the Cochrane EPOC approach) that would allow more insight to be gained...

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

R1: I appreciate the amount of work that is put into this review. Furthermore, medical leadership is an important issue in the increasingly complex healthcare organizations. Especially considering the conflict between professionalism and managerialism that appears in most hospitals after New Public Management. However, I have some major concerns regarding the paper: the relevance of the research question, the databases used in the literature review, the lack of definitions of concepts, and the results.

A: Thank you for your support of our work and agreement as to the importance of medical leadership. We address each of your concerns below.

The relevance of the research question
R1: The question whether doctors should lead hospitals or not is frequently debated. There is consequently good motives for performing such a literature review. However, is it really possible to answer this research question? I would say no, and therefore, the literature review maybe should have been performed with another focus.
A: We agree that the question of whether doctors should lead hospitals is an important topic. However, it was not possible to know to what extent the literature provides an answer to this question without conducting the review. With hindsight, we can see that there is little evidence published that enables us to answer this question, but the paucity of evidence is an important finding in itself.

R1: There are several reasons why the question is not possible to answer: 1. What is hospital performance and how should it be measured? Healthcare organization performance is complex and means balancing several different and often competing measures. What should be included in such a measure? What is possible to say about healthcare organisation performance based on one isolated measure? Furthermore, does it concern short-term or long-term performance?
A: We agree that there is no single accepted measure to determine hospital performance, but we do not agree that this is sufficient reason to abandon our review. Governments across the world make healthcare funding and other decisions based on organisation performance measures, and one of the key roles of healthcare executives is to improve organisation performance. In recognition of the variability of performance indicators in use, we collected the methodological or statistical approach used to identify performance and/or outcomes from each study. We have now included this information for each included study (Supplementary Table 1) thereby enabling the reader to interpret the relevance of the findings to their own circumstances.

R1: 2. Healthcare performance is dependent on many different variables. Who leads the organization is only one of many factors. How is it possible to relate the performance to who is leading the organization? Furthermore, what does "led by doctors" mean? The organization is led by a doctor? Almost all top managers are doctors? A majority of top managers are doctors? Doctors are represented among top managers?

A: The assumption of those who appoint healthcare executives to lead a hospital is that there is a relationship between leadership and organisation performance. We included in our review all studies that compared performance of healthcare organisations where a doctor was a member of the leadership team with those healthcare organisations that did not have a doctor in the leadership team.

R1: 3. How is it possible to relate this to medical training? If it would be possible to the above research question, how is this result then possible to relate to medical training? How to ensure that the success of a doctor manager is based in medical training and not experience of manager positions, networks, political skills, personality, gender, interpersonal skills, economical skills etc etc.
A: We did not intend to suggest that it was only medical training that might affect leadership of healthcare organisations, but rather that it was the total package (training, experience, attitudes, and so on) that a doctor might bring to the leadership table that might affect performance. We have modified the language in our review to clarify this point.

The databases used in the literature review

R1: Medline and Embase are focused on medical/healthcare research, but there is a lot of relevant research within the general fields of organization studies, leadership, management, professions, quality etc that are excluded by your choice of databases. Several of the most cited and most influential articles in the field are thereby excluded in your literature review.

A: In view of your comment, and that of other reviewers, we have added the Emerald Management database to our search strategy. In particular, this database indexes two key health services management journals: Leadership in Health Services and the Journal of Health Organization and Management. The choice of database was made in consultation with a professional research librarian. In addition we have extended our strategy to cover studies published after our original submission and up until 7th June 2017.
While there is potentially additional literature on management and leadership in databases other than Medline, Embase, and Emerald Management, we believe that our extended search, and inclusion of additional search terms, captures the relevant recent literature on medical/physician leaders. In illustration, we discuss the two examples given by the reviewer of papers that we missed as a result of our original Medline/Embase strategy:

References
1. Glouberman, S. and Mintzberg, H. (2001), “Managing the care of health and the cure of disease – part 1: differentiation”, Healthcare Management Review, Vol. 26 No. 1, pp. 56-69.
This reference is actually indexed on Medline, and would likely have been identified by our strategy had our search gone back to 2001. The studies included in this paper, while interesting from an historical perspective, were completed prior to the mid 1990s (the discussion paper underpinning this journal article was published in 1996). In preparation for our systematic review, one of the authors of our review (RCW) completed a scoping review that covered this timeframe. The scoping review identified 30 potentially relevant studies from a pool of 1323 references. During discussion among the co-authors, we felt that the rate at which health care organisations have changed over the last decade meant that any findings dating back to the last century are of limited relevance to how healthcare is organised and managed today, so elected to only include articles from the last decade in our review.
2. Watson, T. (2001), In Search of Management: Culture, Chaos and Control in Managerial Work, 2nd ed., Thomson Learning, London.
This reference, while again interesting, is a book not a study, so would not have been captured regardless of our database choice or search strategy.

Lack of definitions of concepts

R1: The most important concepts are never defined in your review, which confuses your contribution. Medical leadership is never defined. Leadership is not defined, and based on how you treat it in the text, I am worried that you confuse leadership with managerial work.

A: Leaders of healthcare organisations are required to perform managerial duties as part of their leadership role, therefore we believe that 'leadership' in terms of medical leaders encompasses both leading and managerial components. We have added in definitions, and otherwise modified the text of our review to clarify this point.

Results
R1: It is not clear to me how a description of the five studies are relevant to fulfil your aim? What contribution do they enable? Considering the results I would have suggested a re-focus of the research question. Otherwise, the only result is that there are very few studies, and that these studies provide only partial contributions to the research question, but enable little comparison.

A: We agree that the only result is that there are very few studies, and that these studies provide only partial contributions to the research question, but enable little comparison. Our extended search strategy identified a further 11 papers, to bring the total up to 16 papers, but we agree that this number is still small. These are the main findings of our literature review, and we think that these are important findings as many people are not aware of the lack of evidence linking medical qualifications/background of the leader to improved organisational performance.

Minor comments:
A (note): The page numbers below do not align with the page numbers on our PDF as compiled by the BMJ Open submission site, so it has been difficult in some cases to determine which of our points the reviewer is commenting on. We believe that the page referred to as p.3 below is actually page 6 in our document, so have adjusted our numbering accordingly for the remaining comments.

R1: p. 3 r. 3 I would say that this assumption is common among doctors, but among others the opposite is just as common.

A: P6, r3. We agree: we stated in the text that this was the view in some quarters only, and provided references (see #12, 21-23), and also provided references in support of our statement that there was a contrary view (see #24).

R1: p.3 r. 8 Is this aim possible to fulfill?

A: P6, r8. We believe so.

R1: p. 3 r. 15 This excludes the fields of organizations studies, leadership, management, profession, quality etc.

A: p6, r15. We have extended our search strategy, including choice of databases.

R1: p. 4 r. 10-18 Considering these limitations, shouldn't you have changed the focus of the review already here?

A: P7, r10-18. We do not believe so.

R1: p. 6 r. 17 A too simple question for a too complex problem

A: p9, r17. We do not disagree about the complexity of the problem, but felt that it was necessary to ask the simple questions first, and to see what answers, if any, the literature provided.

R1: p. 9 r. 8 Which are these predetermined exclusion criteria?

A: P12, r8. We have been unable to determine what the reviewer is referring to here.

R1: p. 9 r. 48 Healthcare leadership and medical leadership seems to used interchangeably in the paper. Is there a difference?

A: The reviewer is correct in that we used the terms interchangeably in our original submission. On reflection, this may be confusing for some readers, so we have attempted to standardise our language or added definitions of terminology to clarify.

R1: p. 13 r. 59 Research on management training does not confirm your analysis since the link between management training and successful managers have proven very weak, if ever consisting at all. Managers seem to learn by practice (e.g. Watson, 2001; Glouberman & Mintzberg, 2001)

A: p16, r59. We think the reviewer may be referring to our statement on p14, r5 “Doctors who receive leadership or management training, however, may perform well in leadership roles.” If so, this finding was taken from the work of Xirasagar et al., see references #47-48 (see the following few sentences in the text for explanation).

R1: I appreciate the amount of work that is put into this literature review, and I hope that you can re-focus your research question and still have use of the amount of work you have put into this review and paper. I wish you the best of luck in revising the paper!

A: Thank you. We have spent considerable time revising our paper and trust that our response addresses the reviewer’s concerns.

Reviewer 2

R2: This is a text that generates mixed feelings. On the one hand it is very clear. There is a straightforward research question: does medical training lead to better leadership? A literature review is performed. Only a few relevant papers are found. And there is no conclusive evidence. Although this is disappointing, it is an outcome - in fact, that it is disappointing might be seen as the most crucial outcome. We have high expectations or strong assumptions regarding medical leadership, but there are no grounds for this. It is hardly studied, and the studies that are available show little.

A: We agree with the reviewer’s assessment here.

R2: On the other hand, it is too simple and 'empty'. Why this RQ? Is this really important, the link between leadership and medical training. If so - I think it is - why? This requires better explanation. There must be more stress on the expectations and/or assumptions we (and the reader) might have. Medical leadership, for example, might be important because of expertise and authority dynamics - we might expect medical leadership to be effective, because of this. But conversely, we might also assume that this produces leadership weaknesses. In other words, there might be various hypotheses, in order to clarify the relevance and substance of the research, and to work towards more tangible results.

A: We have amended the background to our paper, and included additional literature, to address these points.

R2: Moreover, the authors need to clarify their stress on leadership, and how this differs form executive and managerial perspectives. I assume that they focus on executive leadership, at the apexes of health care organizations, which is about leadership but also about executive work. Or do they also focus on health care managers, throughout hospitals (at some places it seems they do). This is all very confusing.

A: We have amended our paper to better define what we mean by leadership, and to clarify our research question.

R2: Moreover, it leads to a further remark: the text would not only become better when there would some theoretical focus (expertise, authority), as well as expectations/hypotheses, but also when conditions would be introduced. Do we expect medical training to affect leadership/executive work/management (?) in a generic sense, or in specific senses. Does the nature and size of a hospital matter? Does the nature of a division/directorate matter? Does experience matter? But also, does medical training itself matter? If medical training is renewed, with new competencies, does this lead to better management/leadership? Etc.

A: We have amended the background to our paper, and included additional literature, to address these points.

R2: Finally, set against all of these remarks, did the authors capture everything? E.g. recent work by Kirkpatrick & Veronesi, such as: Clinicians on the board: what difference does it make? G Veronesi, I Kirkpatrick, F Vallascas - Social Science & Medicine, 2013

A: On review of this paper, we felt it was relevant to our research question, but it was not captured by our original search strategy. On investigation, we determined that this was as a result of the keywords used by Veronesi et.al, which did not include words around medical leadership. Nevertheless, concern that we may have missed other, similar studies influenced us to extend our search strategy to that in our revised manuscript.

R2: This work poses and answers exactly the same question, but with more nuance and outcomes. In that respect, I am inclined towards 'reject', but in the light of the relevance of the theme and the potential of the text, I opt for 'major revision'.

A: Thank you. Our paper is different to the Veronesi manuscript, as it addresses a broader range of healthcare leadership roles than just those on governing boards. We have completed a major revision of our paper to addresses the reviewer’s concerns.

Reviewer 3

R3: 1. My main concern is the stated aim of the paper in the abstract. The last sentence of the objectives says: "The aim of this systematic review is to determine whether medical training is associated with superior management performance in terms of organisational performance or patient outcomes". A systematic review, will not answer that question, a meta-analysis may, or an experimental study. By fixing this sentence, the paper will make more sense.

A: We agree, and have amended the wording of our objective.

R3: 2. The authors considered papers in which both groups (medical trained and not medical trained) are compared. Another approach would be to look at articles that reported patient outcomes and management skills for only one of the groups and compare with results from the other group. They would have had more information to answer the proposed research question. I make this comment, is to emphasise the need to be clearer on the aim of the paper.

A: We did consider this approach, but found that it was not possible to control for the variables in such a way as to make a valid comparison between studies that did not include both groups (often because the studies did not report the study context in sufficient detail to enable such comparisons to be made). We therefore settled for the aim of identifying those studies that investigated the performance of both medical and non-medical qualified managers in the same healthcare setting. We have amended the manuscript to explain this decision.

R3: 3. In their conclusions, authors make statements like "despite considerable interest in the topic". That seems to be an opinion with no references, and I do not have any idea what is their definition of "considerable". We avoid that kind of qualitative adjectives in the scientific literature.

A: We feel that the large volume of published literature on the topic is indicative of a ‘considerable interest’; however, we have modified the sentence to be more specific.

R3: 4. Regarding the statistics, the authors report Cohen's Kappa coefficients with corresponding p-values. It would be better if the 95% CIs of the coefficients were also reported.

A: We have consulted with a bio statistician on this point, and have been advised to retain our current wording.

R3: 5. I recommend for the conclusions to be rewritten. The first sentence does not derive from any analysis presented in the paper. Furthermore, the conclusions lack implications for further research.

A: We feel that the implication of our findings i.e., that there is currently insufficient evidence to answer the question ‘do hospitals and healthcare organisations perform better when led by doctors?’ is contained in the conclusion, but have added a sentence to clarify that further research is required. We have also extended our conclusion, following the expanded review, to include the new finding on physician participation on governing boards.

Reviewer 4

R4: Overall, a very interesting paper, attempting to provide further evidence for an under-researched area of medicine, leadership and health services management.

A: Thank you.

R4: Although the paper does not provide an answer, it does provide a summary of what was found through the systematic review, and discusses them adequately. Although it does not say this, it does also show that this is an under-researched area in need of further examination, funding and focus.

A: We agree, and have added a sentence to the conclusion to more specifically advocate for further research in this area.

R4: Some specific comments as suggestions for improvement: First sentence under Background - should the author define clinician and doctor, as they are being used interchangeably but may mean different things.

A: We agree that these terms could be confused, and have amended our paper to use the words ‘doctor’, ‘physician’ or ‘medical’ to indicate that we are referring to doctors, and the term ‘clinician’ to refer to the group of clinical professionals that includes doctors, nurses and allied health professionals.

R4: Second sentence - should business-bureaucratic be in quotations as it's not a generally used term, and I assume it is quoted from the cited reference. Same with "organisational-operational" - the hyphenated word is not usually used.

A: We considered this suggestion, but feel that the meaning is clear as is.

R4: I am not certain I understand the significance of the 1983 Griffiths Report with the new model, and how it links to the US clinical directorates, unless it's pure timing - what prompted the clinical directorate idea in the US? Because the author then says "these ushered in the move" but you cannot say the idea of clinical directorates ushered in clinical directorates, which would be a tautology.

A: We have re-phrased to clarify our point.

R4: The statement "Doctors did not always consider such roles attractive" deserves a citation or is this an anecdotal observation?

A: It was from an earlier reference, and the citation has been added to clarify.

R4: Second paragraph, the author starts with clinicians but ended with doctors - as noted, do they mean the same thing, or is the author saying doctors are a subset of clinicians. Also, is a doctor in full-time management technically still a clinician if they don't do clinical practice? (this is a rhetorical question, but is an interesting one to explore).

A: As above, we have modified our language around ‘doctors/clinicians’ to make it less confusing.

R4: Also, the statement: "medical departments are normally led by doctors, and these medical leaders are typically a member of the executive team" is not always true - medical departments are usually led by clinical directors, who traditionally are not members of the executive team, only the medical director/director medical service/chief medical officer is (although this is changing in more recent times).

A: Unfortunately, the arrangement seems to differ by country and organisation. We have modified the language, however, to try to make our statement more generally applicable.

R4: There is also a minor typo - "onengaging doctors".
A: We have fixed this typo.

R4: In the third paragraph, the author refers to physician executives, which is unique term restricted to the US - no other country has this designation - in Australia, the term specialist medical administrator is used. The author may want to clarify for the readers who may not be aware or familiar.

A: We agree that the terms used for medical leaders differ by country and organisation, but there is no easy solution that will please all. We have tried to use variations on 'doctor', ‘medical leader’ and ‘physician executive’ throughout our manuscript so that a variety of international readers can understand our terminology.

R4: This is a strong statement: "articles in favour of medical leadership (invariably written by those with medical qualifications)" especially with the use of the term "invariably" - okay for an editorial or opinion piece, but unless it's back by empirical evidence, I know that not every single piece of literature supporting medical leadership is written by a doctor.

A: We have modified our statement to remove the generalisation.

R4: This statement: "The prevailing opinion in some quarters" - needs qualification. What quarters? The author should explain based on the references cited to support this statement.

A: We have modified this statement as part of our revision; however, it was supported by references #12 and #21-23 in the original version.

R4: Great explanation of the search strategy and eligibility criteria.

A: Thank you.

R4: In relation to synthesis of results "through inductive interpretive analysis of extracted data", was it done using software, or through human analysis? I assume the later? When the data was coded, was that entered into a computer for software or done manually to find connections?

A: We have added a more detailed explanation of our synthesis of findings.

R4: Great job narrowing down the review to five studies. Good description of how that occurred.

A: Thank you.

R4: Great analysis and description of each of the five studies. Very clear.

A: Thank you. We have tried to maintain that clarity in the addition of 11 further studies.

R4: Good discussion. As noted, it would be very difficult to determine if there was causation or just correlation, as there may be a lot of other factors at play.

A: Thank you. We agree about the difficulty of determining causation. There was insufficient information in the studies to tease out the reasons behind the findings. We have added discussion on this point our the revised manuscript.

R4: The Additional Evidence section is interesting but I am not sure how it helps in answering the research question, or even provide clues. The author needs to be clearer as to what clues each point that has been raised provides. What is the relevance of barriers, the point about training for doctors, the multidisciplinary approach, or doctors who combined clinical and management practice, with the research question? For example, the final section about clinician managers as competing is a good observation, but the author should explain how this is a clue so that it is obvious for the reader.

A: We agree about the relevance of the additional evidence, and have removed this section from the manuscript. Where relevant, some of the references have been used in our discussion to illustrate points around the identified themes.

R4: Typo in the conclusion: "Despite considerable interestin healthcare"

A: Thank you for pointing this out. This sentence portion has been removed as part of our revision.

R4: Overall, I believe the paper explores an important area in medicine and management, and even though there was not a conclusive answer, deserves publication subject to the minor amendments suggested above.

A: Thank you. We have revised our paper considerably to address the reviewer's points.

Reviewer 5

R5: Thank you very much for the opportunity to review the article, which reports findings from a systematic review on an important and timely topic. Below is a list of comments for the authors to consider and reflect

A: Thank you for your review and useful suggestions for improving our manuscript. Some of the points were also made by other reviewers. Where that is the case, we will only comment briefly here, and refer the reviewer to the additional information located above.

R5: 1. Search strategy: you have included in your systematic review two medical databases, namely Medline and Embase but you have omitted existing health management / health policy specific databases such as ProQuest and HMIC (Healthcare Management Information Consortium). Given the topic of your review (leadership, management and healthcare organisational performance) this is a substantial shortfall as relevant studies reporting management research findings might not have been captured by your search strategy.

A: We have added a management database, Emerald Management. See comments in preamble and also for Reviewer #1.

R5: 2. Search strategy: The terms you have used in your search strategy are apt but perhaps limited. Additional key terms, which in my view are relevant, include "medical", "practitioner" (to also capture general practitioner-GP), "administrator", "director", "head" [lead]. This is important as you claim (p.7) that you included studies that report on both middle and upper level healthcare managers. Your strategy would not have captured for example clinical directors or heads of departments; it does not also capture the published work on similar topics i.e. medical engagement (see the work published by Peter Spurgeon from Warwick University).

A: We agree and have added further terms. In total, seven terms were included for executive leadership and linked using the Boolean operator OR to maximise the sensitivity of the search: "executive", "leader", "leadership", "manager", "director", "CEO", or "board". For the role of medical practitioner, terms including "physician", "clinician", and "doctor" are used sometimes interchangeably in literature. Thus we searched for all three terms using the OR operator. Spurgeon's work was captured by our new strategy, but did not meet final inclusion criteria. We did, however, complete a citation review of Dickinson's 2008 literature review from the Warwick group, which led us to identify additional papers for inclusion.

R5: 3. If you (decide to) amend and re-run the search strategy it would be advisable to extend the time period to the end of 2016 to make it up to date. It would be also helpful to articulate your rationale for restricting your search for the period 2005-15 (16?).

A: We have extended the time period to bring it up to 7th June 2017, and explained our rationale for only seeking (relatively) recent research studies.

R5: 4. On page 5 the claims made of the strengths and weaknesses of medical leadership need to be supported by appropriate referencing to relevant published work, as it currently looks rather arbitrary. The overall background discussion is mainly located to the USA and the UK with two almost random references to studies reporting from Nepal and Turkey. You could improve the consistency in presenting the broader global story on this topic.

A: We have made some extensive revisions to our manuscript that we believe will address these concerns.

R5: 5. Your research question: 'do hospitals and healthcare organisations perform better when led by doctors?' This implies that you only include doctors who take up designated senior roles in healthcare organisations; mainly CEOs, Chair of Boards or Partners/owners in GP practices. Does this capture the middle level managers you mention in the inclusion criteria? What would be the middle management roles for doctors and what would be the equivalent position for non-clinical managers?

A: We have included senior leaders in our review, where there is comparison data with non-medical leaders available. The main justification for not including more junior leaders was that the leadership tasks tended to have a greater clinical focus, making it more difficult to compare performance with non-medical leaders.

R5: 6. Exclusion criteria: Since you ended up with a very limited pool of final papers the 8 studies that were excluded on the basis of criterion n.10 (in the table on pages 18-19) could be briefly presented in a summary or supplemented table and explain in more detail in the main text the parts of these studies you deemed relevant, though peripheral, to help you answer your question. On the same table, criterion n.8 is not very clear to me and I also think it overlaps with criterion n.10. Finally criterion n.6 states that you excluded systematic reviews and this need to be also mentioned explicitly in the main text.

A: Our revised version removes the additional material.

R5: 7. Data collection process: the elements (h, data) and (I, methods) seem to be confusing and overlapping to a great extent; also what do you mean by "contextual factors" as stated on elements (i) and (j)? How these contextual factors relate to quantitative and qualitative results? Given the research question you use did you expect to find any qualitative study that would be relevant/appropriate to answer your question?

A: We have revised the wording of this section. We were happy to accept any study that was based on empirical data, and did find one qualitative and one mixed method study in the revised search that met inclusion criteria.

R5: 8. Synthesis of results: you state that the results were synthesised through inductive interpretive analysis; you provide no further detail or cite any relevant reference to allow the reader understand what exactly you did at this stage of your work. In the presentation of your results there is only a descriptive account of the reported findings from the five included studies in your final pool of articles.

A: We have revised the wording of this section, and included additional information about the process we used for data synthesis.

R5: 9. Risk of bias: the critical appraisal of the studies using the categories by Hawker et al could be discussed in more depth rather than briefly presenting a short summary with your subjective rankings of the different components of the selected articles. Since you only have a limited number of included studies you could discuss the methodological strengths and weaknesses more explicitly and in more detail.

A: We have reviewed and revised our risk of bias section, to include more discussion on bias within and across studies.

R5: 10. Discussion: Related to a previous comment in the synthesis of your results, the themes or sets of concepts are not presented to the expected extend anywhere in the paper; you only present a short list of topics examined by the studies at the beginning of the discussion section.

A: We have revised our manuscript to address this point.

R5: 11. Discussion: what you report under the titled section 'additional evidence' although interesting and broadly associated to the topic it does not relate directly to the specific question you asked; why did you focus on this specific dimension of the characteristics of managers who were doctors and not other (e.g. leadership development for medical doctors, leadership practices etc.)?

A: We agree with this critique and have deleted the additional evidence section.

R5: 12. In the limitations of your study you need to include the fact that your review is time and language limited (only English). You also did not conduct any library hand search or followed up the references of the included pool of papers or other seminal papers/books in the field; you did not manage to retrieve the full text from 5 potentially relevant papers. I also do not understand the bullet point on p.4, which is repeated in the limitations section, and which states: "most studies examined only one or two aspects of leadership; this limited the generalisation of findings". All empirical studies need to be focused on specific domains and any empirical article can only report findings that examine a limited empirical observation. The wording of your research question was limited to empirical studies that could show an association between medical leadership and organisational performance or patient outcomes rather than also explain how or why this association might work, in other words what is the specific mechanism leading to such a result. Your question focused on individual models of leadership rather than distributed or team-based / collective leadership approaches.

A: We have revised the limitations section to address these points.

R5: 13. Conclusion: this is underdeveloped. You could at the very least suggest areas for future research and propose robust study designs that would better answer your research question.
A: We have attempted to address this point.

R5: I hope you find this useful to help you improve and strengthen your paper. Wish you all the best with your revision.

A: We found the review very helpful in strengthening our work – thank you.

Reviewer: 6 (plus PRISMA checklist with comments)

R6: 1. The overall research question is very broad - perhaps it would be helpful to the reader to expand on the more detailed objectives in the Methods section? I think this would also help in understanding more about the approach taken as I found this also to be less clear than it could be...

A: We have completed an extensive redraft of our paper, which we hope addresses these concerns.

R6: 4. In terms of systematic reviewing, there are areas where I would usually hope to see more detail so that if wished, the review could be replicated. I have used your PRISMA statement responses to highlight some areas of specific concern. Please see attached...

A: We have reviewed the points on the attached PRISMA statement, and believe we have addressed them in our revision.

R6: 5. As you will see in the PRISMA table I am not fully clear what the outcomes are - in some areas it reads as if any quantitative assessment of organisational or patient outcomes by type of manager are the key outcomes - but then the first sentence (p13 lines 16-18) seems to contradict this...

A: We have modified our wording to clarify the outcomes.

R6: 7. Please can full statistics be reported - even if only as reported in the papers - and noted if there is no additional information. For example, (p10 lines 1-3) "no sig difference between clinicians and clinical managers" - what was the measure for this or if not reported, this could be noted?

A: We have reported statistics where present in the original papers.

R6: 9. This is my key concern - and reasons for this are noted in other sections.

A: Please see above.

R6: 10. Similar concerns as for 9 - but there is also some need to improve consistency of language - I think your focus is on medical doctors - perhaps that could be clarified throughout. I also think the section on Additional evidence suggests the purpose of the review was not sufficiently clear at the outset hence the inclusion of 'other evidence' that is interesting?

A: These are good points, and were also made by other reviewers. We have modified our wording to address them.

R6: 11. Please see comments related to 5 and 10

A: Re. 5: There is no published protocol. The protocol is contained in the methods we have articulated in the manuscript.
Re. 10: We have added detail on how we checked the extracted data.

R6: Overall, this is an interesting question and the authors make a strong case for its importance. However, in many areas it reads less like a true systematic review of effectiveness and more like a structured scoping review. There also may be other methods of review (such as realist synthesis or the Cochrane EPOC approach) that would allow more insight to be gained...

A: We agree that the question of whether hospitals and healthcare organisations perform better when led by doctors is an interesting and important question. In view of the reviewers' comments, we have modified our strategy and reporting to align better with the PRISMA requirements for a systematic review.

R6: Thank you for the opportunity to review!

A: Thank you for your helpful suggestions to improve our work.

VERSION 2 – REVIEW

REVIEWER	Thomas Anderson University of Skövde, Sweden
REVIEW RETURNED	07-Aug-2017

GENERAL COMMENTS	I appreciate the amount of work that is put into revising the manuscript. It as improved in many ways,but there are some major problems that still are remaining: 1. Unprecise research questions that take the paper into many different directions. The paper is not guided by one overall research question, but three different (with further sub-questions). This takes the manuscript in many different directions. 2. It is hard to see how a literature review could answer the research questions considering the few studies that matched the inclusion criteria. 3. The result section include hospital boards, strategic decision-making, organizational management (director), and middle management level. The question is whether these levels are comparable at all since the leadership tasks are very different on different levels. 4. It is not clear to me what "a narrative synthesis of coded data" is and how it is performed? Especially based on quantitative data.

REVIEWER	Erwin Loh Monash University
REVIEW RETURNED	04-Aug-2017

GENERAL COMMENTS	Excellent paper on a very important issue. Very strong methodology for a systematic review, well-described, and reproducible. Strong outline and summary of the included cases, with a clear table explaining the strengths and weaknesses of each. Succinct description of the limitations. I'm familiar with this area of research, and the authors have been comprehensive in their review. The paper itself is well-written, using plain language, and nicely structured. I agree with the conclusion, and likewise believe that this is an under-researched area. The challenge has always been developing a study methodology that would be able show correlation, and if possible, even causation, while taking into account all the potential confounding factors. Congratulations on this final version of the paper.

REVIEWER	Dr Yiannis Kyratsis City, University of London United Kingdom
REVIEW RETURNED	10-Aug-2017

GENERAL COMMENTS	. I am pleased to see that the revised version has addressed many of the comments raised in the first review round. Below I outline some final relatively minor comments for your consideration as you finalise the manuscript: - Types of participants under the eligibility criteria: you state on page 7: “we included empirical studies on senior managers in healthcare organisations that involved participants who were both doctors and leaders, and participants who were non-medical leaders”; many of the studies you included in the literature review do not strictly meet this criterion as they do not involve participants who were doctors & leader & the study included participants who were non-medical leaders for comparison. You need to clarify and clearly frame this inclusion criterion. There is further confusion on page 10 lines 46-47 where you state: “After the additional exclusion criterion was added, allowing for a comparison of medical and non-medical leadership, ten more studies were eliminated” – connected to the above point, what was this additional exclusion criterion? Where these studies eliminated because they did not also involve a comparison of medical and non-medical leadership? This section requires revision and clarity. - Also under the types of participants, you state in your eligibility criteria that you included only empirical studies on senior managers. First, it would be helpful to define what you consider a senior manager in your study (director level and above? Participation in the executive team?); second, since you state that you only focus on senior managers why did the five studies that report on middle-level managers were included in the literature review and discussed on pages 15-16 [References 30, 33, 34, 38, and 43]? This point necessitates some explanation. - There is some repetition in the text, which with careful proof reading can be avoided. You also need to be consistent in the use of either “physician” or “doctor” in the text. - On bottom of page 6 – top of page 7 the objectives outlined need to be revised as the current sentence does not make sense: “The objective of the review was to determine whether there is an association between whether the leader has a medical background and management performance in terms of organisational performance or patient outcomes , and was framed by the research question: Do hospitals and healthcare organisations perform better when led by doctors?” - The synthesis of results using a narrative synthesis approach requires citation of a relevant methodology reference - The study characteristics on page 11 could be better summarised in a table

REVIEWER	Elizabeth Jane Shaw National Institute for Health and Care Excellence United Kingdom
REVIEW RETURNED	03-Aug-2017

GENERAL COMMENTS	Minor points are raised below: 4. You have included summary reasons for exclusion in Figure 1 - this is really helpful! Due to the challenges inherent in these types of reviews, it is helpful to see a full dataset would be available on request. 5. Should a sentence be added to note no ethics approval would be needed for this review? Just for completeness (apologies if I have missed this) 12. You very helpfully show the quality of each study and then state overall this is at 'moderate risk of bias'. I am not clear how this assessment was made nor how it may have influenced your confidence in the results. It might be helpful to add a brief explanation how this was judged (even if only qualitatively) or perhaps removed - as you do state an overview of the quality in p16 lines 34-36. Additionally, a reference to the study quality could be added to p20 lines 43-46... This is an interesting review and challenges researchers to answer this in a robust manner...

VERSION 2 – AUTHOR RESPONSE

Reviewer 1

R1: I appreciate the amount of work that is put into revising the manuscript. It as improved in many ways, but there are some major problems that still are remaining:
1. Unprecise research questions that take the paper into many different directions. The paper is not guided by one overall research question, but three different (with further sub-questions). This takes the manuscript in many different directions.

A: Our research was guided by the related questions: Do healthcare executives who are doctors perform better than those who are not; and, if so, in what ways do they perform better? What is it about a physician that might enhance their leadership ability: medical training, experience in the physician role, or something else? Is it important for organisational performance to have doctors as members of the executive leadership team, and if so, why? We believe that exploring these questions in the literature leads directly to answering our main question of ‘Do hospitals and healthcare organisations perform better when led by doctors?’ rather than in other, different directions.

R1: 2. It is hard to see how a literature review could answer the research questions considering the few studies that matched the inclusion criteria.

A: We agree, and have said in our conclusion that, with the exception of hospital or governing boards where modest evidence supporting physician leadership was found, there were insufficient studies to robustly answer our questions. Nevertheless, this is an important finding. This is a standard approach to systematic reviews: see, for example, these Cochrane Reviews where few studies were found that met inclusion criteria:

1. Reeves S, Zwarenstein M, Goldman J, Barr H, Freeth D, Hammick M, et al. Interprofessional education: effects on professional practice and health care outcomes. The Cochrane Database of Systematic Reviews. 2008(1):CD002213.
2. Reeves S, Perrier L, Goldman J, Freeth D, Zwarenstein M. Interprofessional education: effects on professional practice and health care outcomes (update). The Cochrane Database of Systematic Reviews. 2013.

R1: 3. The result section include hospital boards, strategic decision-making, organizational management (director), and middle management level. The question is whether these levels are comparable at all since the leadership tasks are very different on different levels.

A: We acknowledge the difficulty of controlling for the many confounders. Our review was designed to compare doctor and non-doctor performance in the same setting. Therefore, while it is true that leadership tasks may differ between roles, we compared differences in performance between those performing the same, or very similar, roles.

R1: 4. It is not clear to me what "a narrative synthesis of coded data" is and how it is performed? Especially based on quantitative data.

A: A narrative synthesis is an approach to reviewing and synthesizing data from multiple studies where data might be quantitative, qualitative or a combination of both. We have added a methodology reference to the text that directs the interested reader to guidance on how and why this method might be used in systematic reviews (Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the conduct of narrative synthesis in systematic reviews. ESRC Methods Programme, UK National Centre for Research Methods. 2006;1:b92). It is used widely: see also, Hinchcliff R, Greenfield D, Moldovan M, Westbrook JI, Pawsey M, Mumford V, et al. Narrative synthesis of health service accreditation literature. BMJ Quality & Safety. 2012:bmjqs-2012-000852.

R1: Good luck with reviewing the manuscript!

A: Thank you.

Reviewer: 2
R2: No comment

Reviewer: 3
R3: No comment

Reviewer: 4
R4: Excellent paper on a very important issue. Very strong methodology for a systematic review, well-described, and reproducible. Strong outline and summary of the included cases, with a clear table explaining the strengths and weaknesses of each. Succinct description of the limitations. I'm familiar with this area of research, and the authors have been comprehensive in their review. The paper itself is well-written, using plain language, and nicely structured.
I agree with the conclusion, and likewise believe that this is an under-researched area. The challenge has always been developing a study methodology that would be able show correlation, and if possible, even causation, while taking into account all the potential confounding factors.
Congratulations on this final version of the paper.
A: Thank you for your positive comments, and support of our work.

Reviewer 5

R5: Thank you very much for sharing a copy of the revised manuscript for review. I am pleased to see that the revised version has addressed many of the comments raised in the first review round. Below I outline some final relatively minor comments for your consideration as you finalise the manuscript:

A: Thank you for your support of our revisions.

R5: Types of participants under the eligibility criteria: you state on page 7: "we included empirical studies on senior managers in healthcare organisations that involved participants who were both doctors and leaders, and participants who were non-medical leaders"; many of the studies you included in the literature review do not strictly meet this criterion as they do not involve participants who were doctors & leader & the study included participants who were non-medical leaders for comparison. You need to clarify and clearly frame this inclusion criterion. There is further confusion on page 10 lines 46-47 where you state: "After the additional exclusion criterion was added, allowing for a comparison of medical and non-medical leadership, ten more studies were eliminated" – connected to the above point, what was this additional exclusion criterion? Where these studies eliminated because they did not also involve a comparison of medical and non-medical leadership? This section requires revision and clarity.

A: We have re-phrased the sentence: "After the additional exclusion criterion was added, allowing for a comparison of medical and non-medical leadership, ten more studies were eliminated" to clarify that the exclusion criteria was that they did not involve a comparison of medical and non-medical leadership. We also identified a numerical error in the text: eight studies were eliminated based on the final criterion, not 10. The PRISMA flowchart was correct, and we have now modified the text, to ensure that the text and PRISMA flowchart are both consistent and correct.

R5: Also under the types of participants, you state in your eligibility criteria that you included only empirical studies on senior managers. First, it would be helpful to define what you consider a senior manager in your study (director level and above? Participation in the executive team?); second, since you state that you only focus on senior managers why did the five studies that report on middle-level managers were included in the literature review and discussed on pages 15-16 [References 30, 33, 34, 38, and 43]? This point necessitates some explanation.

A: This is a vexing problem in healthcare, where there is a wide range of terminology for leadership positions that sometimes bears little relation to the types of tasks undertaken. We found, for example, that no common or distinct line could be drawn across all studies to separate 'middle' and 'senior' managers. We found that even terms such as 'director level' meant different things in different contexts (director of the board, director of a hospital department, director of a program, director of a project). Our intent was to include only senior managers, as we were looking for relationships between their performance and organizational outcomes. Some of the studies, however, reported on 'middle-level' leaders who actually performed a broad range of leadership tasks and provided comparative data with non-medical leaders. We elected to include these, as the managers were involved in a broad range of leadership tasks and the findings were therefore relevant to our review question.

R5: There is some repetition in the text, which with careful proof reading can be avoided. You also need to be consistent in the use of either "physician" or "doctor" in the text.

A: We have modified our manuscript to replace the word 'physician' with 'doctor', except where not appropriate (such as in the term 'physician executive', which has a distinct meaning in the USA). We have also proof-read the document and eliminated repetition, where possible.

R5: On bottom of page 6 – top of page 7 the objectives outlined need to be revised as the current sentence does not make sense: “The objective of the review was to determine whether there is an association between whether the leader has a medical background and management performance in terms of organisational performance or patient outcomes, and was framed by the research question: Do hospitals and healthcare organisations perform better when led by doctors?”

A: We have rephrased the objective, and broken it into two sentences to make the meaning more clear.

R5: The synthesis of results using a narrative synthesis approach requires citation of a relevant methodology reference

A: We have added a methodology reference, that explains this approach. (Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the conduct of narrative synthesis in systematic reviews. ESRC Methods Programme, UK National Centre for Research Methods. 2006;1:b92)

R5: The study characteristics on page 11 could be better summarised in a table

A: The study characteristics are already presented as a table (Supplementary Table 1).

Reviewer 6

R6: Thank you for all your work to address concerns. Minor points are raised below:
4. You have included summary reasons for exclusion in Figure 1 - this is really helpful! Due to the challenges inherent in these types of reviews, it is helpful to see a full dataset would be available on request.

A: Thank you.

R6: 5. Should a sentence be added to note no ethics approval would be needed for this review? Just for completeness (apologies if I have missed this)

A: Thank you for this suggestion (and you have not missed it), but we do not believe so. It would be the exception for a systematic review to require ethics approval, so we do not think that the sentence is needed in this case.

R6: 12. You very helpfully show the quality of each study and then state overall this is at 'moderate risk of bias'. I am not clear how this assessment was made nor how it may have influenced your confidence in the results. It might be helpful to add a brief explanation how this was judged (even if only qualitatively) or perhaps removed - as you do state an overview of the quality in p16 lines 34-36. Additionally, a reference to the study quality could be added to p20 lines 43-46...

A: We have amended the section on bias, to clarify that we did not ‘measure’ bias, but only identified that it may be present. We have also added a sentence to the limitations section to remind the reader that “It is also important to consider that risk of bias was evident across studies due to the majority of studies employing self-reported measures, and an absence of information concerning ethics approval, funding, or conflicts of interests in some studies.”

R6: This is an interesting review and challenges researchers to answer this in a robust manner...

A: Thank you.

VERSION 3 – REVIEW

REVIEWER	Thomas Andersson University of Skövde, Sweden
REVIEW RETURNED	01-Sep-2017

GENERAL COMMENTS	No further comments