PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The cost and cost-effectiveness of point-of-care testing and treatment for sexually transmitted and genital infections in pregnancy in low- and middle-income countries: A systematic review protocol
AUTHORS	saweri, olga; Batura, Neha; Adawiyah, Rabiah al; Causer, Louise; Pomat, Willie; Vallely, Andrew; Wiseman, Virginia

VERSION 1 – REVIEW

REVIEWER	Remco Peters University of Pretoria, South Africa
REVIEW RETURNED	01-Apr-2019

GENERAL COMMENTS	The main measure(s) of interest of this study are not clearly specified. 'Adverse pregnancy outcomes' is a very broad term. Moreover, the timing of testing plays a role, e.g. ectopic pregnancy cannot be averted by screening during pregnancy. I think that the fact that the authors plan to combine several STIs in their analysis and link these to a combination of adverse pregnancy outcomes is problematic and will generate results that are very difficult to interpret. There a large differences of the effect of each STI tested for on pregnancy outcomes, e.g. a neonate with congenital syphilis is different from a very-low birth weight infant attributed to maternal chlamydial infection during pregnancy. The authors should elaborate a lot more, possibly include a full paragraph, on how they plan to analyse cost-effectiveness (as mentioned in their title) since the manifestations averted are very broad. In addition to the previous comment: how do the authors plan to do costing of adverse outcomes averted? I think they should include the proposed values in the manuscript. Moreover: will this be done in a stratified manner, i.e. per STI/POC test and/or per pregnancy outcome averted? Despite its limitations, I think it would be very interesting to include the effects of 'syndromic management when adequately implemented' in the cost-effectiveness analysis since it would provide clear context to introducing diagnostics. In- and exclusion criteria: why are studies excluded in languages other than English? –In particular in the French literature there would be some relevant references.

	In- and exclusion criteria: is microscopy for Trichomonas vaginalis considered at POC in this study? Or only NAAT- or antibody-based tests? Would specify. Search strategy: I assume the reviewers are blinded to each other's results? How are discordances resolved? The manuscript requires serious language editing.

REVIEWER	Jennifer Pillay University of Alberta, Canada
REVIEW RETURNED	21-Apr-2019

GENERAL COMMENTS	Thank you for offering me the opportunity to review this protocol. I think the review could contribute important information for others considering implementation of a screening program using POC tests for curable STIs in pregnant women. I hope my comments below will be helpful to strengthen the protocol and hopefully systematic review reporting and methods. Introduction: The introduction in the abstract states that early detection and treatment reduces the risk of adverse pregnancy and birth outcomes; this should probably be revised to "may" or "aims to" reduce because of the lack of high quality evidence to support this which is highlighted more in the text introduction. Some of the references used in the full text's introduction are not highly appropriate, for instance a study on prevalence and incidence of the infections (ref 4) is used to support their association with poor outcomes. The objective does not fully encompass the aim of the protocol for a systematic review. It could either just state to describe the methods for a systematic review aimed to identify drivers of the costs and cost-effectiveness, or if the authors want to include individual aspects such as synthesize and appraise, they should also include "search for and identify studies" which are large components of a systematic review. I would think "appraise" is the same thing as assessing the quality of the studies, which will not be "explored" - I would suggest removing from the abstract "and the quality of the economic evaluations will be explored". See below for possibility to re-assess whether or not the quality of the studies is being assessed with CHEERS. Syndromic management would not be considered a screening program (p.4) Literature search & inclusion criteria: What is the comparison of interest to the authors? Are they including only "no screening/intervention" or also evaluations comparing 2 or more different interventions? Is the search being conducted independently by 2 reviewers, or is one person (if a librarian please state this) running the search and 2 reviewers screening and selecting? I don't see any value of having 2 people run the search. I am not a research librarian, although have helped develop many searches and am aware of several search filters for economic evaluations (e.g https://www.cadth.ca/resources/finding-evidence/strings-attached-cadths-database-search-filters#eco), and it appears to me that the search terms provided could miss many studies. For example "economic evaluation" or "cost-utility" is not applied, nor are terms specific to costs/costing/pricing/budget, especially if the aim of the review includes cost and budget impact studies as well as economic evaluations. If the search has already been ran (with completion date in May I would assume so) the limited sensitivity of the search could be listed as a potential limitation. Other limitations include the limit to English (when considering LMICs) and the non-inclusion of non peer-reviewed

	literature when many economic evaluations are conducted by governmental organizations that may not publish. Some rationale for the language and exclusion of grey literature could be provided. Data extraction & quality assessment: The CHEERS checklist is meant to be a reporting guideline, and while very suitable to guide data extraction and helping to determine differences between methods/models and thus results of studies, it does not incorporate a formal assessment/judgement of the methodological quality or risk of bias per se. The "and risk of bias" in the abstract should be removed I think. If the checklist is being used to assess quality the authors could describe how this is being done (.e.g. is poor reporting on some specific aspects considered to indicate poor methodological quality?). The protocol does not "list and define all variables for which data will be sought" or the process for data extraction (who, duplicate vs verification). The heterogeneity will likely be very large and systematic review methods would guide authors to base this assessment on the methodological and clinical (e.g. model inputs and structure etc) variability rather than only on the differences in outcomes/findings or a test for heterogeneity. This would require extraction and comparison of several variables: what costs are included, screening strategies (universal vs targeted, timing in pregnancy, frequency of screening), time-horizon/outcomes modeled (e.g. maternal and infant or longer term), assumptions used about natural history of infections (e.g. clearance rates of chlamydia) and re-infection rates, inputs related to effectiveness for outcomes, rates of partner treatment, estimates used for utility weights for QALYs in cost-utility analyses, screening coverage rates, etc. It is likely that the most informative findings from the review will not be a pooled estimate of the cost-effectiveness but rather what variables ("drivers") most impact the cost-effectiveness (i.e. from study sensitivity analyses or indirectly from between-study differences); the authors state they will look at "context-related factors" but could outline specific variables they are extracting and considering. I am not aware of any evidence on how to detect publication bias or selective reporting in economic evaluations; if these are being considered perhaps the authors could provide details on how they are assessing these. Analysis: The authors do not describe the method and software they will use if a quantitative synthesis is appropriate. As above I would suggest not only looking to differences in outcomes/findings when making decisions on conducting a quantitative synthesis. Other comments: Some of the headings in the manuscript are not accurate, e.g the text below "data analysis" does not just describe analysis but also study selection, quality assessment, and data extraction, which could each have their own subheadings. If poor reporting via the CHEERs is being used as a "quality" criteria, the authors should state how they are incorporating this into their analysis. The manuscript is fairly well written although a thorough read for minor errors would improve the readability e.g "Globally, the....(STIs) is alarming", and "However, few studies, most based in high income countries, have investigated".

VERSION 1 – AUTHOR RESPONSE

Reviewer 1's comments to the author:

1. The main measure(s) of interest of this study are not clearly specified. 'Adverse pregnancy outcomes' is a very broad term. Moreover, the timing of testing plays a role, e.g. ectopic pregnancy cannot be averted by screening during pregnancy. I think that the fact that the authors plan to combine several STIs in their analysis and link these to a combination of adverse pregnancy outcomes is problematic and will generate results that are very difficult to interpret.

RESPONSE:

Thank you, our introduction was summarising the risks associated with STIs. However, to avoid any confusion, we have removed the text related to ectopic pregnancies (see page 3 lines #79-82).

It is expected that the studies captured by this review will measure costs against a range of different health outcomes and that cost estimates will be influenced by many factors including the timing and cost of testing. This is consistent with our primary objective, which is to systematically review methods of economic evaluations of point of care testing to detect STIs in pregnancy and to establish the evidence base for cost-effectiveness. Specifically, the review will synthesise studies that have been published to date, assess their methods against reporting guidelines, identify broad drivers of costs/cost-effectiveness and finally, establish priorities for future research in this area. Given the sudden rise in economic evaluations in this field in the past 5 years, we feel there is value in taking stock of those studies already published and to review and assess the quality of their reporting.

It is now clearly noted on page 7 (lines #197-199) that there will be no restrictions on study outcomes and that combining the results of the studies for a metanalysis is unlikely to be feasible or useful in this context where there is diversity in diseases, outcomes and interventions. The purpose of this review is also made clearer on page 3 on lines #114-123.

2. There a large differences of the effect of each STI tested for on pregnancy outcomes, e.g. a neonate with congenital syphilis is different from a very-low birth weight infant attributed to maternal chlamydial infection during pregnancy. The authors should elaborate a lot more, possibly include a full paragraph, on how they plan to analyse cost-effectiveness (as mentioned in their title) since the manifestations averted are very broad.

RESPONSE:

As indicated above, the studies are expected to be heterogenous (in terms of outcomes measured, costs included, settings, populations, etc.) This does not however, undermine our primary objective, which is to systematically review methods of economic evaluations of point of care testing to detect STIs in pregnancy and to establish the evidence base for cost-effectiveness We are not 'analysing cost-effectiveness'. This approach is consistent with other reviews of cost-effectiveness studies published in BMJ Open - see for example Sheikh A, Nurmatov UB, Cresswell K, et al Investigating the cost-effectiveness of health information technologies: a systematic review protocol BMJ Open 2013;3:e003737. doi: 10.1136/bmjopen-2013-003737

It is now stated more clearly on page 7 (lines #199-203) that in view of the anticipated heterogeneity in interventions, outcomes, study designs and health system contexts, we will most likely undertake a narrative synthesis of results.

3. In addition to the previous comment: how do the authors plan to do costing of adverse outcomes averted? I think they should include the proposed values in the manuscript. Moreover: will this be done in a stratified manner, i.e. per STI/POC test and/or per pregnancy outcome averted?

RESPONSE:

Please see our previous response – we will not be analysing costs or cost per adverse outcome averted. The purpose of this review is to systematically review methods of economic evaluations of point of care

testing to detect STIs in pregnancy and to establish the evidence base for cost-effectiveness. The text on page 4 (lines #118-123) has been revised to reflect this more clearly.

4. Despite its limitations, I think it would be very interesting to include the effects of 'syndromic management when adequately implemented' in the cost-effectiveness analysis since it would provide clear context to introducing diagnostics.

RESPONSE:

Our only restriction to testing strategies is that testing and treatment of STIs are at POC. This means that we will include all comparators in the analysis which may include, but is not restricted to, syndromic management. For clarity we have amended the manuscript to reflect this (see page 5, lines #147-148).

5. In- and exclusion criteria: why are studies excluded in languages other than English? –In particular in the French literature there would be some relevant references.

RESPONSE:

There will be no restriction on the language of publications during the initial search. During title and abstract screening, non-English and non-Dutch papers will be excluded. This reflects the languages spoken by our study team. The number of papers removed will be noted and detailed as a limitation in the discussion section of the systematic review. This is now explained on page 5 on lines #143-144.

6. In- and exclusion criteria: is microscopy for Trichomonas vaginalis considered at POC in this study? Or only NAAT- or antibody-based tests? Would specify.

RESPONSE:

In this study we do not consider microscopy as a POC test, we have defined POC testing on page 5 on lines #128-130.

7. Search strategy: I assume the reviewers are blinded to each other's results? How are discordances resolved?

RESPONSE:

Yes, reviewers are blinded. The text on pages 5-6 (lines #154-169) has been updated to reflect this. Discordances will be resolved by the third researcher - see pages 6, lines #18-183).

8. The manuscript requires serious language editing.

RESPONSE:

Thank you for this comment, we have closely edited the revised document.

Reviewer 2's comments to the author:

Please leave your comments for the authors below

Thank you for offering me the opportunity to review this protocol. I think the review could contribute important information for others considering implementation of a screening program using POC tests for curable STIs in pregnant women.

I hope my comments below will be helpful to strengthen the protocol and hopefully systematic review reporting and methods.

RESPONSE:

Thank you for your positive feedback.

1. Introduction: The introduction in the abstract states that early detection and treatment reduces the risk of adverse pregnancy and birth outcomes; this should probably be revised to "may" or "aims to" reduce because of the lack of high quality evidence to support this which is highlighted more in the text introduction.

RESPONSE:

In response to this comment, we have amended the abstract text to reflect that early detection and treatment may reduce the risk of adverse birth outcomes (see page 2, line# 32-33).

2. Some of the references used in the full text's introduction are not highly appropriate, for instance a study on prevalence and incidence of the infections (ref 4) is used to support their association with poor outcomes.

RESPONSE:

Thank you for pointing this out. Sources, including Ref #4 have been replaced with more appropriate sources, as suggested:

For example, Ref #4 is now: Mullick, S., et al., Sexually transmitted infections in pregnancy: prevalence, impact on pregnancy outcomes, and approach to treatment in developing countries. Sexually Transmitted Infections, 2005. 81(4): p. 294-302.

3. The objective does not fully encompass the aim of the protocol for a systematic review. It could either just state to describe the methods for a systematic review aimed to identify drivers of the costs and cost-effectiveness, or if the authors want to include individual aspects such as synthesize and appraise, they should also include "search for and identify studies" which are large components of a systematic review.

RESPONSE:

Thank you for this comment. The revised text in the abstract on page 2 (lines 37-39) and in the manuscript on Page 4 (lines # 114-123) describes the methodology more clearly.

4. I would think "appraise" is the same thing as assessing the quality of the studies, which will not be "explored" - I would suggest removing from the abstract "and the quality of the economic evaluations will be explored". See below for possibility to re-assess whether or not the quality of the studies is being assessed with CHEERS.

RESPONSE:

We have deleted the last sentence of the introduction of the abstract on page 2.

5. Syndromic management would not be considered a screening program (p.4)

RESPONSE:

In several low- and middle- income countries, sexually transmitted and genital infections are diagnosed using syndromic management as recommended by the WHO. Therefore, we include all comparators in the analysis, which includes syndromic management. For clarity we have amended the manuscript to reflect this (see page 5, line #147-148).

6. Literature search & inclusion criteria: What is the comparison of interest to the authors? Are they including only "no screening/intervention" or also evaluations comparing 2 or more different interventions?

RESPONSE:

We are interested in comparing ‘no screening’ or ‘an existing screening program’, which includes syndromic management, with a point-of-care testing intervention. There is no restriction on the type of comparator. The text on Page 5, (lines# 147-148) has been amended to clarify this.

7. Is the search being conducted independently by 2 reviewers, or is one person (if a librarian please state this) running the search and 2 reviewers screening and selecting?

RESPONSE:

Thank you for this comment. Two reviewers are conducting the search, screening the results and selecting papers for full text review independently. To clarify the search strategy, we have revised the text (pages 5-6, lines #153-169).

8. I don't see any value of having 2 people run the search. I am not a research librarian, although have helped develop many searches and am aware of several search filters for economic evaluations (e.g https://eur01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.cadth.ca%2Fresources%2Ffinding-evidence%2Fstrings-attached-cadths-database-search-filters%23eco&data=02%7C01%7Cn.batura%40ucl.ac.uk%7Cc0bec96f473045c89a7708d6ceef408d%7C1faf88fea9984c5b93c9210a11d9a5c2%7C0%7C0%7C636923924302421292&sdata=SLGJyhD9R0Cnk2ZGrLNZ%2Ff7NpL1LLQgIQy7DaS1aLsc%3D&reserved=0), and it appears to me that the search terms provided could miss many studies. For example "economic evaluation" or "cost-utility" is not applied, nor are terms specific to costs/costing/pricing/budget, especially if the aim of the review includes cost and budget impact studies as well as economic evaluations.

RESPONSE:

Thank you for this comment and for the helpful reference. The search terms are run by 2 researchers to improve the rigour of the search. This helps to ensure that articles are not missed and serves as an internal quality check for the search. The search terms were developed by the authors (health economists and infectious diseases physicians, epidemiologists) with the support of medical librarians from the University College London and the University of New South Wales with expertise in systematic reviews.

The terms are based on a combination of keywords (some are truncated) from key papers in the field as well as medical subject headings (MeSH) terms. The MESH term 'Cost-benefit analysis' captures keywords such as 'cost utility' and 'economic evaluation' (full list below). We reasoned that cost-utility analyses are a special form of cost-effectiveness analyses, therefore including the keyword 'cost-effectiveness' encompasses cost-utility analyses and ensures a more comprehensive search and have amended the text on page 6, lines #157-159).

- Analyses, Cost-Benefit
- Analysis, Cost-Benefit
- Cost-Benefit Analyses
- Cost Benefit Analysis
- Analyses, Cost Benefit
- Analysis, Cost Benefit
- Cost Benefit Analyses
- Cost Effectiveness
- Effectiveness, Cost
- Cost-Benefit Data
- Cost Benefit Data
- Data, Cost-Benefit
- Cost-Utility Analysis
- Analyses, Cost-Utility
- Analysis, Cost-Utility
- Cost Utility Analysis
- Cost-Utility Analyses
- Economic Evaluation
- Economic Evaluations
- Evaluation, Economic
- Evaluations, Economic
- Marginal Analysis
- Analyses, Marginal
- Analysis, Marginal
- Marginal Analyses
- Cost Benefit
- Costs and Benefits
- Benefits and Costs
- Cost-Effectiveness Analysis
- Analysis, Cost-Effectiveness
- Cost Effectiveness Analysis

9. If the search has already been run (with completion date in May I would assume so) the limited sensitivity of the search could be listed as a potential limitation.

RESPONSE:

Please see previous response. Date of completion has been amended and the search will be run using the terms shown on page 7.

10. Other limitations include the limit to English (when considering LMICs) and the non-inclusion of non peer-reviewed literature when many economic evaluations are conducted by governmental organizations that may not publish. Some rationale for the language and exclusion of grey literature could be provided.

RESPONSE:

The rationale and limitations for excluding non-English/Dutch papers and grey literature are raised on page 5 (lines #141-145).

11. Data extraction & quality assessment: The CHEERS checklist is meant to be a reporting guideline, and while very suitable to guide data extraction and helping to determine differences between methods/models and thus results of studies, it does not incorporate a formal assessment/judgement of the methodological quality or risk of bias per se. The "and risk of bias" in the abstract should be removed I think.

RESPONSE:

In the abstract on page 2 line #48, we are referring to the 'risk of publication bias', we have made this clearer by amending the manuscript on page 7 lines #191-192.

12. If the checklist is being used to assess quality the authors could describe how this is being done (.e.g. is poor reporting on some specific aspects considered to indicate poor methodological quality?).

RESPONSE:

We have amended the manuscript to make this point clearer (see page 7 line #187-190)

13. The protocol does not "list and define all variables for which data will be sought" or the process for data extraction (who, duplicate vs verification). The heterogeneity will likely be very large and systematic review methods would guide authors to base this assessment on the methodological and clinical (e.g. model inputs and structure etc) variability rather than only on the differences in outcomes/findings or a test for heterogeneity. This would require extraction and comparison of several variables: what costs are included, screening strategies (universal vs targeted, timing in pregnancy, frequency of screening), time-horizon/outcomes modeled (e.g. maternal and infant or longer term), assumptions used about natural history of infections (e.g. clearance rates of chlamydia) and re-infection rates, inputs related to effectiveness for outcomes, rates of partner treatment, estimates used for utility weights for QALYs in cost-utility analyses, screening coverage rates, etc. It is likely that the most informative findings from the review will not be a pooled estimate of the cost-effectiveness but rather what variables ("drivers") most impact the cost-effectiveness (i.e. from study sensitivity analyses or indirectly from between-study differences); the authors state they will look at "context-related factors" but could outline specific variables they are extracting and considering. I am not aware of any evidence on how to detect publication bias or selective reporting in economic evaluations; if these are being considered perhaps the authors could provide details on how they are assessing these.

RESPONSE:

It is anticipated that the following data will be extracted: data on total cost of the intervention, cost per outcome, cost effectiveness and incremental cost ratios such as cost per outcome and cost per DALY averted, cost saving to the health system, and budget impact (affordability) estimates. As mentioned in the manuscript on page 7, lines #194-196 we will conduct a narrative synthesis of results as we anticipate a heterogeneity in outcomes that will identify and discuss the drivers of costs and cost effectiveness. In doing so, we will explore whether studies report on policy or contextual factors that could explain unusually higher or lower costs than anticipated. This has been clarified on page 8, lines #196-199.

Publication bias arises from a tendency from authors to publish studies with significant results. Economic evaluations do not report cost effectiveness ratios for

non-statistically significant study outcomes, as this is a zero numerator. Further, economic evaluations so not report cost effectiveness ratios for statistically significant outcomes that indicate a negative impact. So, in our case, we will explore whether studies report on interventions deemed not cost effective according to established thresholds.

14. Analysis: The authors do not describe the method and software they will use if a quantitative synthesis is appropriate. As above I would suggest not only looking to differences in outcomes/findings when making decisions on conducting a quantitative synthesis.

RESPONSE:

For data extraction from the included studies, we are using Microsoft Excel 2013. As indicated in the manuscript on page 7 (lines #185-187), if there is adequate homogeneity in the outcomes, it might be possible to conduct a meta-analysis, or sub-group analysis. This will be done using STATA v 14.0 see page 7, lines #202-203).

Other comments:

15. Some of the headings in the manuscript are not accurate, e.g the text below "data analysis" does not just describe analysis but also study selection, quality assessment, and data extraction, which could each have their own subheadings. If poor reporting via the CHEERs is being used as a "quality" criteria, the authors should state how they are incorporating this into their analysis.

RESPONSE:

We have reviewed the headings of the manuscript and amended.

16. The manuscript is fairly well written although a thorough read for minor errors would improve the readability e.g "Globally, the....(STIs) is alarming", and "However, few studies, most based in high income countries, have investigated".

RESPONSE:

Thank you for this comment, we have reviewed the language of the manuscript.

VERSION 2 – REVIEW

REVIEWER	Remco Peters University of Pretoria, Pretoria, South Africa
REVIEW RETURNED	25-Jun-2019

GENERAL COMMENTS	Thank you for the revised manuscript - this is well done.

REVIEWER	Jennifer Pillay University of Alberta, Alberta Research Centre for Health Evidence
REVIEW RETURNED	04-Jul-2019

GENERAL COMMENTS	Thank you for asking me to review this revised manuscript. The manuscript language has been improved and some aspects clarified, but there are still concerns, mainly with its appropriateness as a systematic review with a comprehensive search for studies and synthesis of the findings. As the authors responses' and revisions suggest, this review does not appear to attempt to undertake a full synthesis of the studies and their findings (e.g. narrative or interpretive synthesis with comparing and contrasting and exploring differences in findings etc), but more charting/mapping of the studies and their findings as well as determining research gaps ("systematically review methods of economic evaluations of point of care testing to detect STIs in pregnancy and to establish the evidence base for cost-

	effectiveness"). Even if a quantitative analysis (pooling) is not appropriate, a systematic review should still provide a formal synthesis of the results (e.g. see the authors' supplied example in Sheikh et al. BML Open 2013 for an example using an interpretive synthesis; also see http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.178.3100&rep=rep1&type=pdf for narrative synthesis guidance). Moreover, there is still concern about the limitations of their search (terms not comprehensive and it appears limited publication dates) to meet systematic review standards. I would suggest the authors refer to their review as a literature, rapid, or scoping review (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4491356/, https://www.ncbi.nlm.nih.gov/pubmed/26857112) with suitable references and methods/terminology provided. Other comments: I attached a copy of the manuscript with some additional comments on the abstract, background and methods, including but not limited to, • The inclusion date of study publication should be included in the abstract and study type sections, and please mention if there were (or not) study date or language limits added to the database search strategies. • Please state that you plan to extract data on all effectiveness outcomes reported if this is so, or otherwise. • Do you plan to extract data and report on the effects of sensitivity analyses conducted? • If keeping the possibility of quantitative analysis and adding more description of these methods, when describing the exploration of between study heterogeneity, if "tests of heterogeneity" are being used please describe these; I would suggest that other considerations, i.e. the methodological heterogeneity, that are not identified in any statistical tests, may be sufficient to avoid pooling. The reviewer also provided a marked copy with additional comments. Please contact the publisher for full details.

VERSION 2 – AUTHOR RESPONSE

Reviewer 1's comments:

No comments

Reviewer 2's comments:

2) The manuscript language has been improved and some aspects clarified, but there are still concerns, mainly with its appropriateness as a systematic review with a comprehensive search for studies and synthesis of the findings. As the authors responses' and revisions suggest, this review does not appear to attempt to undertake a full synthesis of the studies and their findings (e.g. narrative or interpretive synthesis with comparing and contrasting and exploring differences in findings etc), but more charting/mapping of the studies and their findings as well as determining research gaps ("systematically review methods of economic evaluations of point of care testing to detect STIs in pregnancy and to establish the evidence base for cost-effectiveness"). Even if a quantitative analysis (pooling) is not appropriate, a systematic review should still provide a FORMAL SYNTHESIS of the results (e.g. see the authors' supplied example in Sheikh et al. BML Open 2013 for an example using an interpretive synthesis; also see XX for narrative synthesis guidance).

RESPONSE: Thank you for your comment.

Our objective is to undertake a full synthesis of the studies and their findings and compare/contrast the differences in the findings to understand why costs and cost effectiveness of interventions may vary across settings and key drivers of cost-effectiveness. We have amended the text on page 4 to clarify our objectives.

Also, as recommended, we have followed a similar approach to Sheikh et al and now explain on page 7 that in lieu of inter-study heterogeneity, where a meta-analysis is not possible, a "descriptive summary and narrative synthesis" will be conducted. This will involve an interpretive synthesis of findings by exploring relationships in the findings (e.g. common drivers of costs and cost-effectiveness) and exploring the potential transferability of findings to other settings (e.g. contextual factors affecting findings). Due to the expected heterogeneity (e.g. in the interventions evaluated, methods, outcomes and study populations), the pooling of results is expected to be of limited value. As suggested by reviewer 2 below, we plan to explore methodological heterogeneity in selected studies.

3) Moreover, there is still concern about the limitations of their search (terms not comprehensive and it appears limited publication dates) to meet systematic review standards. I would suggest the authors refer to their review as a literature, rapid, or scoping review (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4491356/) with suitable references and methods/terminology provided.

RESPONSE: The literature search will have no restrictions. With respect to "limited publication dates", we will not limit the literature search nor exclude publications based on publication date and time. We have revised the text on page 6 to clarify this.

We would like to reiterate that the literature search will not restrict studies on the basis of language. However, we do acknowledge that during article screening, specifically, during the title abstract and screening we will exclude studies that are not published in English as discussed on page 5.

This is a systematic review which has a clear/structured question, systematic methods to search for and extract data, critically appraise methodological and reporting practices, as well as synthesize findings qualitatively (in our case). This is not a scoping review as we did not start with a concept, map language and then adjust the search method iteratively.

Other comments:

I attached a copy of the manuscript with some additional comments on the abstract, background and methods, including but not limited to,

4) The inclusion date of study publication should be included in the abstract and study type sections, and please mention if there were (or not) study date or language limits added to the database search strategies.

RESPONSE: Please see response to comment 3, there are no publication nor language limits to our literature searches in the pre-selected databases (MEDLINE, Embase, Web of Science and Google Scholar).

We will only exclude articles during screening stage of the review, specifically articles not written in English will be excluded during the title and abstract screening. We will not exclude articles based on publication date and/or time.

5) Please state that you plan to extract data on all effectiveness outcomes reported if this is so, or otherwise.

RESPONSE: We plan to extract data on all outcomes related to cost and cost-effectiveness as discussed on page 7 - we have revised the text to reflect this.

6) Do you plan to extract data and report on the effects of sensitivity analyses conducted?

RESPONSE: Yes, on page 7 it is indicated that we will extract data and report on the effects of the sensitivity analyses conducted in the included studies.

7) If keeping the possibility of quantitative analysis and adding more description of these methods, when describing the exploration of between study heterogeneity, if “tests of heterogeneity” are being used please describe these; I would suggest that other considerations, i.e. the methodological heterogeneity, that are not identified in any statistical tests, may be sufficient to avoid pooling

RESPONSE: Following on from response #2 above, we agree that due to the likely heterogeneity across interventions, study designs, and populations, the focus should be on methodological heterogeneity. The text on page 7 has been amended to reflect this point.