

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Point-of-care tests for urinary tract infections: Protocol for a systematic review and meta-analysis of diagnostic test accuracy
AUTHORS	Fraile Navarro, David; Sullivan, Frank; Azcoaga-Lorenzo, Amaya; Hernandez Santiago, Virginia

VERSION 1 - REVIEW

REVIEWER	Jan Verbakel Department of Public Health and Primary Care, KU Leuven Nuffield Department of Primary Care Health Sciences, University of Oxford
REVIEW RETURNED	23-Sep-2019

GENERAL COMMENTS	<p>Dear editor,</p> <p>I appreciate the opportunity to review this manuscript by Navarro et al., entitled "Point-of-care tests for urinary tract infections: Protocol for a systematic review and meta-analysis of diagnostic test accuracy." The main aim is to systematically review and meta-analyse the diagnostic accuracy of currently available point-of-care tests for urinary tract infections. The authors have described their methodology in sufficient detail but there are a few issues which need to be resolved before publication can be recommended.</p> <p>GENERAL:</p> <ul style="list-style-type: none">- I have made a few suggestions regarding methodology to ensure you capture all the relevant articles. <p>INTRODUCTION:</p> <ul style="list-style-type: none">- "Point-of-care testing is defined as 'diagnostic testing, performed at or near the site where clinical care is delivered' [17]" There is a better definition available: https://academic.oup.com/fampra/article/35/4/475/4828050- you sometimes use POC instead of POCT, e.g. page 6 line 60, page 7 line 14 <p>METHODS:</p> <ul style="list-style-type: none">- "Randomized clinical trials (RCTs), cluster RCTs, evaluation studies, observational studies and regulatory or approval evidence reports (if available), evaluating point-of-care diagnostic tests for UTI in symptomatic patients versus urine culture (gold standard)": I would suggest to avoid the term gold standard as no test has perfect sensitivity and/or specificity, use "reference standard" instead.- "only those tests that could be categorised as 'point-of-care test'
-------------------------	--

	<p>will be included, defined as the tests that can be carried out in close proximity to the patient, without involvement of laboratory facilities [25].": As mentioned above, there is a more recent definition of POCT for primary care available, which could facilitate your review process. Consider revising.</p> <ul style="list-style-type: none"> - "inception to 8th October 2018": Is this date correct? You seem to mention a later date in the rest of your document. - I am not convinced your search strategy captures all articles that could be of interest. Not every paper on the diagnostic accuracy of point-of-care tests for UTI may mention the terms sensitivity and specificity. Your search now limits the results to articles that contain these terms. Consider revising your search strategy as many of the studies you seek might not be indexed accordingly. - "A standardised data extraction form will be developed. The review team (DFN, AAL and VHS) will independently extract the data from all studies. Study authors will be contacted if no data is available.": how will you evaluate discrepancies? Will all articles be double-extracted? - "If index tests use different threshold values, the hierarchical summary ROC model will be used instead, to obtain summary sensitivity and specificity for each threshold value.": there are now other techniques available to deal with multiple thresholds per study. See: https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-016-0196-1 - "Subgroup analysis will be explored and performed depending on the heterogeneity found": I would suggest to specify your subgroup analyses beforehand and perform them if sufficient data is available. Meta-regression could be considered if so.
--	--

REVIEWER	<p>Caroline Schneeberger Amsterdam UMC - The Netherlands</p> <p>1. Working on a systematic review (Cochrane) on point-of-care test for urinary tract infections in elderly.</p> <p>2. Part of the research team of the PROGRESS study. See publication Kuil SD, Hidad S, Fischer JC, Harting J, Hertogh CM, Prins JM, van Leth F, de Jong MD, Schneeberger C. Sensitivity of point-of-care testing C reactive protein and procalcitonin to diagnose urinary tract infections in Dutch nursing homes: PROGRESS study protocol. <i>BMJ Open</i>. 2019 Aug 10;9(8):e031269. doi: 10.1136/bmjopen-2019-031269. PubMed PMID: 31401614; PubMed Central PMCID: PMC6701568.</p>
REVIEW RETURNED	01-Nov-2019

GENERAL COMMENTS	<p>Improving diagnostic certainty of UTI is an important step to reduce unnecessary antibiotic use. Providing an overview of the point of care tests available can be helpful.</p> <p>Comments</p> <ol style="list-style-type: none"> 1. Abstract line 15: The word "rapid" is introduced, however not further defined or explained. 2. Based on the protocol it seems that solely studies on "uncomplicated" – "lower" – "nonfebrile" UTI are included. It may be helpful to further clarify the type of UTI studied.
-------------------------	--

	<p>3. The authors have chosen to compare the accuracy of current available POC-tests to the gold standard – the urine culture. In my opinion it would also be worthwhile to compare the accuracy of the identified POC tests to the most commonly used method namely based on clinical symptoms. POC-test may reduce unnecessary antibiotic use but are not free of charge. This brings me to the next comment:</p> <p>4. The authors do not mention the costs and barriers (e.g. training, validation of the test(s)) related to the introduction and use of POC-test.</p> <p>5. Introduction line 49: the authors could consider elaborating more on the reasons why urine cultures are often not performed.</p> <p>6. The authors decided to include adults and not make difference between adults and elderly even though the diagnostic process in elderly is consider significantly different. Asymptomatic bacteriuria (ASB) is very common in elderly. Moreover, elderly often do not present with the typical UTI symptoms – currently the basis for diagnosis of UTI in healthy adults. I would recommend to either exclude elderly from the analysis or perform separately analysis for healthy adults and elderly.</p> <p>7. The existence of asymptomatic bacteriuria and the potentially disruptive effects on the diagnostic process of UTI are not discussed at all.</p> <p>8. Overuse of antibiotics for UTI and rise of antimicrobial resistance are important drivers for this review despite that the effect on number of (unnecessary) antibiotic prescriptions is not part of the (secondary) outcomes.</p>
--	---

VERSION 1 – AUTHOR RESPONSE

Reviewers' comments:

Reviewer 1:

1. Definition of 'point of care test': This has now been changed in line to the definition proposed by the reviewers (<https://academic.oup.com/fampra/article/35/4/475/4828050>)
2. Concomitant use of POC and POCT: These have now been amended and all changed to POCT.
3. 'Gold standard' has been changed to 'Reference standard' on both abstract and main text.
4. The date has now been corrected and changed to 1st June 2019.
5. "I am not convinced your search strategy captures all articles that could be of interest. Not every paper on the diagnostic accuracy of point-of-care tests for UTI may mention the terms sensitivity and specificity. Your search now limits the results to articles that contain these terms. Consider revising your search strategy as many of the studies you seek might not be indexed accordingly"

Thank you very much for your suggestion and input. When designing our search strategy, we carefully considered the implications of using different terms, including those of sensitivity and specificity, and discussed the impact of using different strategies on the search strategy sensitivity and precision. Cochrane defines Sensitivity as 'the number of relevant records identified by a search strategy divided by the total number of relevant records on a given topic', whereas Precision is a measure of the ability of a search to exclude irrelevant articles. Achieving an appropriate balance between sensitivity and precision is challenging, but important. We performed a series of preliminary searches using a range of search terms to detect text words and subject headings. These searches and previous literature (1-3) informed the final search strategy. In order to maximize the sensitivity of

our search strategy, without losing scope and still having a good precision we added the mentioned terms to our search strategy. Our strategy was able to capture all articles referenced by the NIHR HTA Horizon Report, which had a broader scope than we had designed for our review. We have also checked previous diagnostic test accuracy systematic reviews and it is common practice to use such terms in the search strategy (4, 5). There are also methodological papers that recommend using these to maximize the signal/noise ratio of relevant papers (6). Nonetheless, we have carried out again our search strategy without “sensitivity” OR “specificity” as suggested, and the number of papers to this date increases from 1,011 papers to 8,894 in the PUBMED database alone. Given the previous considerations and the feasibility of the project, we have considered that the likelihood of missing relevant information is small. Given these reasons, we are keen in keeping the current search strategy as such, however we will discuss the points raised above in the discussion section of the main results paper in future.

6. All articles will be double-extracted, and risk of bias will be double-assessed. Discrepancies will be evaluated and solved by discussion, and if no agreement, a third reviewer will be involved. This information has now been added to main text ('Data collection process' section in Methods).

7. "If index tests use different threshold values, the hierarchical summary ROC model will be used instead, to obtain summary sensitivity and specificity for each threshold value.": there are now other techniques available to deal with multiple thresholds per study. See:

<https://bmcmmedresmethodol.biomedcentral.com/articles/10.1186/s12874-016-0196-1>

Thank you for highlighting the above reference. The hierarchical summary ROC model was chosen as recommended by Cochrane for dealing with multiple thresholds in systematic reviews of diagnostic test accuracy, (7) and more recently, specifically recommended in diagnostic accuracy systematic reviews in the infection and microbiology field (8). The comparison of tests on the basis of their ROC curves takes into consideration their accuracy across a range of thresholds, and is aided by single summary statistics. If varying thresholds were used in the studies, a summary receiver operating characteristic (SROC) curve is estimated to describe the change in sensitivity and specificity while varying the threshold. It is very interesting to see that the method proposed by Steinhauser S et al (9) is also able to determine an optimal threshold and is particularly useful if most studies report more than one threshold. However, they acknowledge increasing bias on sensitivity and specificity, and optimal threshold estimates with increasing heterogeneity, which is something we expect to encounter, due to the nature of our review.

8. "Subgroup analysis will be explored and performed depending on the heterogeneity found": I would suggest to specify your subgroup analyses beforehand and perform them if sufficient data is available. Meta-regression could be considered if so.

Thank you for suggesting this. We will perform subgroup analysis separately for different point of care tests, and we will also analyse separately adults and elderly patients. This information has now been added to main text.

Reviewer 2:

1. Abstract line 15: The word “rapid” is introduced, however not further defined or explained. The Word ‘rapid’ has now been changed to ‘faster’ (“faster than the reference standard”)

2. Based on the protocol it seems that solely studies on “uncomplicated” – “lower” – “nonfebrile” UTI are included. It may be helpful to further clarify the type of UTI studied.

Thank you for bringing this up to our attention. We don’t specify “uncomplicated”, “lower” or “nonfebrile” in our inclusion criteria nor in our search terms as the search is not restricted to those.

The search strategy uses broad terms for defining UTI with the aim of capturing all potentially relevant studies looking at POCT used in this context, and the exclusion criteria that will be applied are defined in main text. Classical symptoms include those mentioned in Table 1 above, and we will also explore different symptoms / combinations and UTI definition reported in each study. We have now clarified this a bit better in the Methods section.

3. The authors have chosen to compare the accuracy of current available POC-tests to the gold standard – the urine culture. In my opinion it would also be worthwhile to compare the accuracy of the identified POC tests to the most commonly used method namely based on clinical symptoms.

Thank you for your suggestion. The Cochrane Screening and Diagnostic Test Methods Group's guidelines for conducting systematic review of diagnostic test accuracy studies recommend the use of only one reference standard for all studies in this type of review, to avoid ambiguity and bias (10). It also defines 'reference standard' as 'the test, series of tests, or set of procedures that is used to determine the presence or absence of the target condition in patients'. Urine culture has been universally accepted as the reference standard for diagnosing urinary tract infections, and hence why chosen for this review. It is true that clinical algorithms and other forms of decision support have also been described as potential stewardship interventions that can improve antimicrobial use in certain conditions / settings, and their use could be an interesting aspect to examine in the context of UTIs. However the scope of this review was to look at available point-of-care tests for symptomatic UTIs, defining point-of-care tests as advised by Reviewer 1: "a test to support clinical decision making, performed in close proximity the patient (and on any part of the patient's body or its derivatives), to help the patient and healthcare professional upon the best management approach during or very close to the time of the consultation, with results available at the time of clinical decision making". Looking at clinical algorithms based on symptoms would be a different review.

4. POC-test may reduce unnecessary antibiotic use but are not free of charge. This brings me to the next comment: The authors do not mention the costs and barriers (e.g. training, validation of the test(s)) related to the introduction and use of POC-test.

'Additional training required', as well as 'need for supplementary equipment' are already specified as data to extract in the original manuscript submitted (see 'Definitions for data extraction', page 7 of original manuscript). We will explore cost of each POCT if this is provided (added now to main manuscript).

5. Introduction line 49: the authors could consider elaborating more on the reasons why urine cultures are often not performed.

This has now been further elaborated on 3rd paragraph of 'Introduction' section.

6. The authors decided to include adults and not make difference between adults and elderly even though the diagnostic process in elderly is consider significantly different. Asymptomatic bacteriuria (ASB) is very common in elderly. Moreover, elderly often do not present with the typical UTI symptoms – currently the basis for diagnosis of UTI in healthy adults. I would recommend to either exclude elderly from the analysis or perform separately analysis for healthy adults and elderly.

Thank you for highlighting this point. We will perform subgroup analyses separately for different point of care tests, and we will also analyse separately adults and elderly patients. This information has now been added to main text.

7. The existence of asymptomatic bacteriuria and the potentially disruptive effects on the diagnostic process of UTI are not discussed at all.

Thank you for this comment. We agree asymptomatic bacteriuria is highly prevalent, particularly in elderly patients. However, this review is focused on symptomatic urinary tract infections, so we won't

include studies looking at asymptomatic bacteriuria. We acknowledge symptoms may be varied in the elderly population, and occasionally not as clear. Doing subgroup analysis separately for elderly patients will help us examine different definitions used for UTI in this age group, and potential implications for drawing conclusions regarding the use of POCTs in this group.

8. Overuse of antibiotics for UTI and rise of antimicrobial resistance are important drivers for this review despite that the effect on number of (unnecessary) antibiotic prescriptions is not part of the (secondary) outcomes.

The aim of the present study is to examine the currently available point-of-care tests for diagnosing symptomatic urinary tract infections, and to compare and meta-analyse their diagnostic test accuracy against reference standard, rather than examining how the use of a particular test reduces or not the use of antibiotics, which is important but a different research question.

VERSION 2 – REVIEW

REVIEWER	Jan Verbakel Nuffield Department of Primary Care Health Sciences, University of Oxford Department of Public Health and Primary Care, KU Leuven
REVIEW RETURNED	06-Jan-2020

GENERAL COMMENTS	<p>Dear editor,</p> <p>I appreciate the opportunity to review the revised manuscript by Navarro et al., entitled "Point-of-care tests for urinary tract infections: Protocol for a systematic review and meta-analysis of diagnostic test accuracy." The authors have responded to most of my comments in sufficient detail but there are still a few issues which need to be resolved before publication can be recommended.</p> <p>Search strategy: -----</p> <p>Although I understand the importance of balancing sensitivity and precision of your search strategy, some eligible articles will only mention likelihood ratios or predictive values, rather than sensitivity or specificity. Excluding these articles will result in biased results. The downside of performing systematic reviews of DTA studies is that you have to screen through a large amount of papers to finally end up with only a handful of relevant articles. Ignoring this by using a very selective approach as you suggest, inevitably leads to article selection bias. I would still encourage you to broaden your search strategy.</p> <p>Different thresholds of the index test per article: -----</p> <p>My suggestions was not to use the "optimal threshold"-approach (which is clinically unsound) as suggested in the Steinhauser paper, but rather use their method to allow the modelling of multiple thresholds per paper in your statistical analysis. Several other methods have been proposed over the past few years (after</p>
-------------------------	---

	<p>Cochrane suggested their default approach of choosing one threshold per study) to deal with this, as it is now known that ignoring these multiple thresholds results in bias.</p> <p>As an example, we have used another approach to deal with multiple thresholds in this paper (extension of the bivariate method of Dukic and Gatsonis): https://www.bmj.com/content/361/bmj.k1450</p>
--	--

REVIEWER	<p>Caroline Schneeberger Amsterdam UMC, the Netherlands Involved in the PROGRESS study: Point-of-care diagnostics to Guide appropriate antimicrobial therapy of urinary tract infections in nursing homes and currently working on a Diagnostic Cochrane review "Procalcitonin, C-reactive protein and erythrocyte sedimentation rate for the diagnosis of lower urinary tract infections in older people".</p>
REVIEW RETURNED	07-Feb-2020

GENERAL COMMENTS	<p>Looking forward to the results of the review. My only suggestion is to reformulate "Symptomatic UTIs (dysuria, polyuria, urgency and/or suprapubic pain)" in box 1 inclusion criteria to for example "Symptomatic UTIs (variously defined)". And describe the used definition of symptomatic UTI of the included studies in the characteristics.</p>
-------------------------	---

VERSION 2 – AUTHOR RESPONSE

Reviewers' comments:

Reviewer 1:

1. Search strategy: Although I understand the importance of balancing sensitivity and precision of your search strategy, some eligible articles will only mention likelihood ratios or predictive values, rather than sensitivity or specificity. Excluding these articles will result in biased results. The downside of performing systematic reviews of DTA studies is that you have to screen through a large amount of papers to finally end up with only a handful of relevant articles. Ignoring this by using a very selective approach as you suggest, inevitably leads to article selection bias. I would still encourage you to broaden your search strategy.

Many thanks for your comments and suggestions about the search strategy. We recognise the search needs to be wide for diagnostic test accuracy, and therefore we have now broadened our search strategy to take into account any possible terms referred to diagnostic test accuracy, and so we are now reviewing a much broader number of papers. The new search strategy is now referenced in main text, and also included in amended Appendix 1.

2. Different thresholds of the index test per article:

My suggestions was not to use the "optimal threshold"-approach (which is clinically unsound) as suggested in the Steinhauser paper, but rather use their method to allow the modelling of multiple thresholds per paper in your statistical analysis. Several other methods have been proposed over the past few years (after Cochrane suggested their default approach of choosing one threshold per study) to deal with this, as it is now known that ignoring these multiple thresholds results in bias. As an

example, we have used another approach to deal with multiple thresholds in this paper (extension of the bivariate method of Dukic and Gatsonis): <https://www.bmj.com/content/361/bmj.k1450>

Thank you for your suggestion, and attached reference. Similarly to Taylor K et al, we will also be using bivariate and hierarchical summary receiver operating characteristic meta-analysis methods. I can also see however some differences between the review referenced and our review. Taylor K focuses on just primary care / ambulatory care settings, reason why they wanted to select the lowest threshold only for each study if studies reported data at multiple thresholds. Our review however is not restricted to primary care only, and rather we are looking at any studies using point-of-care tests regardless site of care. Therefore, we are interested on looking at all thresholds reported within each study, and this may have informed choosing a slightly different method.

Reviewer: 2

My only suggestion is to reformulate "Symptomatic UTIs (dysuria, polyuria, urgency and/or suprapubic pain)" in box 1 inclusion criteria to for example "Symptomatic UTIs (variously defined)". And describe the used definition of symptomatic UTI of the included studies in the characteristics.

Thank you. This has now been changed in Box 1 as suggested.

VERSION 3 - REVIEW

REVIEWER	Jan Y Verbakel Nuffield Department of Primary Care Health Sciences, University of Oxford Department of Public Health and Primary Care, KU Leuven
REVIEW RETURNED	01-May-2020

GENERAL COMMENTS	Dear editor and authors, I would like to thank the authors fr their careful review of the paper and addressing the concerns raised, which I now believe are fully answered and accounted for. best wishes, Prof Dr Jan Verbakel
-------------------------	--