# *Comparative Effectiveness Review Disposition of Comments Report*

**Research Review Title:** *Diagnosis of Gout*

Draft review available for public comment from November 4, 2014 to December 2, 2014.

**Research Review Citation:** Newberry SJ, FitzGerald J, Maglione MA, O'Hanlon CE, Han D, Booth M, Motala A, Tariq A, Dudley W, Shanman R, Shekelle PG. Diagnosis of Gout. Comparative Effectiveness Review No. 158. (Prepared by the Southern California Evidence-based Practice Center under Contract No. 290-2012-00006-I.) AHRQ Publication No. 15(16)-EHC026-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2016. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

## Comments to Research Review

The Effective Health Care (EHC) Program encourages the public to participate in the development of its research projects. Each research review is posted to the EHC Program Web site in draft form for public comment for a 4-week period. Comments can be submitted via the EHC Program Web site, mail or E-mail. At the conclusion of the public comment period, authors use the commentators' submissions and comments to revise the draft comparative effectiveness research review.

Comments on draft reviews and the authors' responses to the comments are posted for public viewing on the EHC Program Web site approximately 3 months after the final research review is published. Comments are not edited for spelling, grammar, or other content errors. Each comment is listed with the name and affiliation of the commentator, if this information is provided. Commentators are not required to provide their names or affiliations in order to submit suggestions or comments.

The tables below include the responses by the authors of the review to each comment that was submitted for this draft review. The responses to comments in this disposition report are those of the authors, who are responsible for its contents, and do not necessarily represent the views of the Agency for Healthcare Research and Quality.

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| Commentator & Affiliation | Section | Comment | Response |
|---|---|---|---|
| TEP 1 | Quality of Report | Good | **Thank you.** |
| TEP 2 | Quality of Report | Superior | **Thank you.** |
| TEP 3 | Quality of Report | Good | **Thank you.** |
| Peer Reviewer 1 | Quality of Report | Superior | **Thank you.** |
| Peer Reviewer 2 | Quality of Report | Good | **Thank you.** |
| TEP 4 | Quality of Report | Good | **Thank you.** |
| TEP 5 | Quality of Report | Good | **Thank you.** |
| Peer Reviewer 3 | Quality of Report | Fair | **Thank you.** |
| TEP 6 | Quality of Report | Good | **Thank you.** |
| Peer Reviewer 4 | Quality of Report | Good | **Thank you.** |
| TEP 7 | Quality of Report | Good | **Thank you.** |
| Peer Reviewer 5 | Quality of Report | Good | **Thank you.** |
| TEP 1 | General Comments | This is a thorough review with appropriate key questions but it leaves this reviewer convinced only that we have a very long way to go before we have anything useful for decisive clinical decisions for primary care doctors not doing SF analyses or newer Imaging. Current specificities may be adequate for an impression that can justify treatment of an acute attack (with continued observation) but not for committing an individual to lifetime treatment with urate lowering agents which many will need. Would you please consider making that conclusion for readers? | Thank you. We agree that the evidence supporting current methods for diagnosing gout is not as strong and convincing as we would like. We believe the conclusions make that point. . Long term management of gout patients with urate lowering therapy and their monitoring is currently under investigation and is being addressed in another review. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| TEP 2 | General Comments | The report is clear and descriptive. Its length can be intimidating. KQ are appropriate and well stated. | Thank you. |
| TEP 3 | General Comments | This systematic review addresses the diagnosis of gout. Overall the document is clear and well written. The conclusions are generally well supported by the data presented. There are some issues that require further consideration: | Thank you. No response is needed. |
| TEP 3 | General Comments | (page ES-3, ES-10 thru ES-14, 3, 12-41) The key question 1 relates to analysis of the performance of tests or clinical features 'compared to synovial fluid analysis'. However, at various times throughout the analysis, data are included that use clinical criteria as a gold standard. This is inconsistent with the stated key question and is also methodologically concerning, as clinical classification criteria have relatively low sensitivity and specificity for gout, when compared with synovial fluid crystal analysis as the gold standard. | The aim of the present review is to examine the evidence for the accuracy of diagnostic tests for gout, with the emphasis on tests that would be used in a primary or emergent care setting. And given that the accuracy of the gold standard of crystal analysis is, itself, affected by a number of factors, we were asked to examine that evidence as well. Unfortunately, and partly due to the difficulties with crystal analysis, study participants do not always undergo this test, leaving us to either accept some studies that used clinical (classification) criteria as the standard or to reject most studies. Also, our reading of the original studies on the various sets of criteria indicates that a number of those designed for classification are used for diagnosis. To further emphasize these issues, we have added text to the Intro, Methods, the discussion of Limitations and the conclusions, both in the Executive Summary and in the main text of the report. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| TEP 3 | General Comments | (page ES-16 and 14) There also appears to be some confusion about the function of the various classification criteria (Rome, ARA and NY). These are NOT diagnostic criteria, but are classification criteria. Classification criteria are used to ensure a relatively homogenous but representative group of patients with a specific condition are identified for research purposes. It is not recommended that classification criteria are used for diagnosis of gout for clinical purposes. | Referring to the response on line 17, we are aware that most of the algorithms were conceived as classification criteria, not diagnostic criteria, and that classification criteria should not be used for diagnosis, but the literature makes it clear that at least some of these schemes are being used in clinical practice for diagnosis of individual patients (and published studies have assessed their validity against synovial fluid MSU assessment). We have now addressed this point in the introduction and in the descriptions of the algorithms in the Results, as well as in the section on Limitations. |
| TEP 3 | General Comments | (page ES-17 and 45 among others) A further related comment relates to the 2006 EULAR gout diagnosis paper that does not provide a validated diagnostic algorithm. | Thank you. The lack of validation was noted. |
| TEP 3 | General Comments | (page ES-20) In the executive summary, there is reference to assessing the role of imaging using ‘studies that enrolled only patients not previously diagnosed with gout’. This is the ideal situation to test the true diagnostic accuracy of tests and other diagnostic approaches. However, within the rest of the document, it is not clear that this approach has been taken or whether this presentation was included within the search strategy or subsequent analysis. I would favour a consistent approach throughout the document on this point. I would also like to see further information about the disease duration and whether only people with first presentations were included for each study in the tables describing the search results. | We do state in the Methods section that we have included studies with enrollment limited to individuals without prior gout diagnoses, and we were able to adhere to this criterion. We have noted studies that reported on average disease duration. We did not identify any studies that specifically employed as an inclusion criterion that patients had to present with their first attack of gout, as we note in the discussion of limitations of the literature. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| TEP 3 | General Comments | Minor comment: The description of monosodium urate crystals is quite inconsistent throughout the document (incorrectly called ‘UA’ crystals at times, and also called ‘MSU’ and just ‘crystals’ at other times). I recommend that the crystals are termed ‘monosodium urate crystals’ throughout the document. | Thank you. We have revised our reference to monosodium urate crystals to the spelled out form or MSU crystals throughout. |
| Peer Reviewer 1 | General Comments | In this report on the Diagnosis of Gout, the authors conduct an exhaustive systematic review of the literature to answer several key questions relevant to this process. The questions are appropriate and explicitly stated. | Thank you. |
| Peer Reviewer 2 | General Comments | There are 2 main problems with this report. Firstly, the timing is poor given that the new ACR-EULAR endorsed classification criteria were presented at the ACR ASM earlier this month and will be published during 2015. These criteria will immediately render the review out-of-date. Secondly, the report fails to explicitly consider the difference between diagnosis and classification, which confounds the inclusion of the 'diagnostic algorithm' studies - which are clearly not studies of diagnostic algorithms. | This report was requested for an assessment of methods used in primary care practice to *diagnose* gout. We have incorporated a discussion of these new classification criteria into our report however because the criteria for diagnosis and classification differ, updating of classification criteria for research purposes will not render the findings of this review on diagnostic criteria obsolete. The confusion on our part between tests for classification and tests for clinical diagnosis in the draft report seems to have arisen from the lack of explicit distinction in the literature (and especially in descriptions of inclusion/exclusion criteria in published studies) between diagnostic criteria and classification criteria, e.g., studies stating that patients were determined to have gout based on ACR criteria. In the introductory chapter , we now include a brief discussion about diagnostic and classification criteria. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| Peer Reviewer 2 | General Comments | (page ES-4, ES-5, ES-10-ES-14, 5, 6 and 12) In addition, I do not agree with the way that some of the key questions are framed: KQ1b seems circular since the diagnostic items will likely also include the location of the affected joint and the number of joints involved; KQ1c for most diagnostic settings, the duration of the current episode is not as important as the disease duration (ie time since onset of first ever symptoms). | Thank you. The questions were extensively revised by multiple technical experts' sand also posted for public input. However we have now addressed these concerns in the Discussion section. Regarding the circularity of KQ1b, the test whose accuracy was of primary interest (or at least how the accuracy of the test was affected by number and identity of affected joints) was synovial fluid MSU analysis, rather than the clinical algorithms. |
| TEP 4 | General Comments | The report provides a review of the evidence available to date on the diagnosis of gout from a primary care perspective. As stated in the report, gout has been insufficiently studied overall and the limited data have almost exclusively generated from the field of rheumatology. Yet, care for the disease (particularly the initial care) is provided by primary and urgent care providers. To that effect, this report provides meaningful information and a refreshing perspective to the field. The target population and audience appear to be defined, although this can be more explicitly stated as for primary and urgent care providers. However, the key questions are clearly stated and appropriate from a primary care perspective (which may be different from a rheumatology perspective that most of the reviewers are familiar with, including myself). | We have revised the introduction to increase emphasis on the fact that this report is focused on diagnosis of gout in primary/urgent/emergent care settings. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| TEP 5 | General Comments | As framed and organized, the report is difficult to put into a clinical context. Diagnostic accuracy is only meaningful in a specific clinical context, and the sensitivity and specificity of any tests is strongly dependent upon the population to which the test is applied.  The most widely applicable population would be “all patients presenting with symptoms of acute joint pain.”  But no gold standard would be available in such a population, since it would not be reasonable to obtain synovial fluid on all comers.  Note that the population of patients in whom synovial fluid can be sent for analysis is in itself a highly selected population.   All this deserves a nuanced discussion at the very start of the report.   It would be okay for the report to take on a variety of different diagnostic scenarios, but it would be much more digestible and useful if the evidence supporting each diagnostic test (or the lack of evidence) were described separately for each scenario.   Imagine, for example, a report that was entitled “the diagnosis of heart disease” covered everything from CAD in asymptomatic patients to ACS or CHF in patients presenting with acute symptoms but intertwined the conditions. | Thank you. The scope of the work and presentation of the evidence needed to create such clinical scenarios would be beyond the scope for this project but can be considered for a future study. However, we have substantially revised the introduction and Discussion to cover the points you raise about the specific populations that the report is meant to address. |
| TEP 5 | General Comments | Almost all of what I described in the aforementioned paragraph is touched upon at some place in the report, but these points are not used to frame the review as I believe they should be. | Thank you. We have revised the description of Diagnostic methods in the Introduction to address these points. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| TEP 5 | General Comments | There seems to be inconsistency with regard to how different tests are evaluated.    For example, the DECT studies that you cite seem to be simple evaluations of one diagnostic modality versus a reference standard; they don't evaluate the incremental diagnostic value as compared to H&P, for example.  If that is so, then uric acid or elements of the clinical history, for example, should be treated the same way.    But you say that there is essentially no evidence that relates the uric acid test to the gold standard of joint aspiration.  Surely the distribution of uric acid has been compared to patients who do/don't have crystals in their joints?? | Unfortunately, very few studies have attempted to assess the incremental value of, say, using ultrasound or DECT over that of using a clinical algorithm alone or the value of a clinical algorithm over that of MSU alone. In our discussion of study limitations, we now summarize the findings of the two studies that do address this question.  .<br>As for the question of whether studies have compared serum uric acid levels to the presence of crystals in synovial fluid, in the Discussion, in a section on research gaps, we review the evidence on this relationship in the context of gout diagnosis. |
| TEP 5 | General Comments | Consider reorganizing the report in the following way:<br>1. Clearly delineate each clinical scenario that your report covers. 2. For each of these scenarios, review the evidence for the diagnostic performance of particular historical features, clinical signs, laboratory, or imaging tests, in a univariate way. 3. Then, for each scenario, review the evidence for the additive diagnostic value of laboratory tests (primarily uric acid) and/or imaging studies compared to elements, compared to history and physical exam alone. Regarding the key questions, I would argue that KQ1 could also benefit from a more specific clinical context, recognizing that the evidence is likely to be imperfect. | Thank you. We have reorganized the Discussion and we now begin with a sketch of the kinds of patient scenarios to which this report applies. However, it is beyond the scope of the report to devise a set of detailed scenarios and solicit expert judgment on the appropriate diagnostic work up. |

| | | | |
|---|---|---|---|
| Peer Reviewer 3 | General Comments | Generally I think the structure and readability of the report could be improved. I have a number of suggestions for improvements. Following methods for DTA reviews suggested by Cochrane would improve this report. There are a number of statements that are sufficiently vague that I do not understand what you have done. | We appreciate your suggestions and will follow them to the extent we can while still adhering to the AHRQ EPC guidelines for reviews on medical diagnostics. |
| Peer Reviewer 3 | General Comments | Scope of the review: What do you mean by validity? I would not say “compared to” because you are not evaluating the accuracy of the gold standard which “compared to” suggests. I would suggest adopting the Cochrane format of “‘To determine the diagnostic accuracy of [index test] for detecting [target condition] in [participant description]’.” This also applies to all your key questions. | Thank you. We can't change the wording of the key questions at this point but we have provided a footnote explaining what was meant by "compared to," and we have replaced these words with the correct usage in the text. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| TEP 6 | General Comments | This is a systematic review of 16 original studies and five systematic reviews comparing the sensitivity/specificity of tests including clinical factors, radiographs, ultrasound and Dual-energy CT in the diagnosis of gout against the gold standard test of joint aspiration and synovial fluid assessment for monosodium urate (MSU) crystals in population of adults 18 years of age or older who are suspected of having gout. Additional outcomes of interest were the accuracy of the test results, clinical decision-making, short term clinical outcomes and the presence of any adverse events. The authors concluded that promising diagnostic algorithms such as the Diagnostic Rule needs to be validated in primary care settings, and that both DECT and US showed good sensitivity and specificity for gout diagnosis in high risk patients. Additionally they concluded that an algorithm with high diagnostic accuracy can ideally form part of a decision tree that combines clinical signs and symptoms with more invasive tests or imaging for clinically ambiguous cases. | Summary of the review is provided by the reviewer No response is needed. |
| TEP 6 | General Comments | Major strengths: 1) The study initially encompasses a broad number of studies, up to 235 potential background studies and a long chronological period was used for literature search. 2) The study had very clear tables depicting their findings from each of the studies that was utilized in their review. 3) The protocol and design of the study is clear and easily reproducible. | Summary of the review is provided. No response is needed. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| TEP 6 | General Comments | Major Weaknesses:<br>(page 11) The authors had 235 potential background studies, however, chose to use only 16 original studies and 5 systematic reviews. This is a significant number of potential studies, which were included. The authors did not specify why these studies in particular were excluded. | The 235 studies were not studies that were eligible for inclusion, but rather non-systematic reviews and other publications that included no original data. We have revised the "Flow" to clarify this point. |
| TEP 6 | General Comments | (page 12) The conclusions drawn regarding the utility of ultrasound and DECT for diagnosis of gout were based on very small sample sizes and the results were concluded as low evidence. The authors may potential want to expand their inclusion number of studies to strengthen their evidence level. | Unfortunately, we cannot expand our inclusion criteria at this stage of the review to include studies that don't meet our inclusion criteria. We set the criteria to replicate as well as possible the kinds of patients who would be seen in a primary/urgent care setting. |
| TEP 6 | General Comments | (page 12) The authors concluded that Ultrasound had a high sensitivity in diagnosis of gout when sensitivities were listed as low as 38%. This needs to be addressed in the results/discussion. | We removed wording “high sensitivity” for ultrasound and reassessed the study in question. We modified our conclusion about US, and discussed possible reasons for the wide variation in sensitivity. |
| TEP 6 | General Comments | As the summary of findings and strength of evidence ranged from low to insufficient, the authors should consider expanding the number of included articles to hopefully improve the strength of evidence. | Unfortunately the number of included studies is determined by our inclusion and exclusion criteria. The 235 studies referred to in the flow were non-systematic reviews that did not provide original data and could thus not be included. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| TEP 6 | General Comments | Advances in Knowledge: The authors assessed the different diagnostic methods in the suspicion of gout, and stated that their findings provide some evidence to support the further development and validation of diagnostic algorithms based on a combination of clinical signs and symptoms for the diagnosis of gout in the primary care setting, with the use of imaging modalities (US and DECT) in cases where a definitive diagnosis cannot be made from signs and symptoms alone. However, the strength of evidence for all their conclusions, were either low or insufficient. Perhaps the utility of this review would be improved, if the number of reviewed articles was expanded. | Unfortunately the number of included studies is determined by our inclusion and exclusion criteria. The 235 studies referred to in the flow were non-systematic reviews that did not provide original data and could thus not be included. |
| TEP 6 | General Comments | Implications for Patient Care: The implications the authors' state are valid. | Thank you. |
| Peer Reviewer 4 | General Comments | The target population and intended audience are clear. The key questions are appropriate and clinically relevant. | Thank you. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| TEP 7 | General Comments | Excellent work by the Authors.<br>There were two similar manuscripts presented for peer review in this PDF document, so the comments apply to both versions.<br>It would have been desirable to have included more studies that examine the gold standard joint aspiration (with a success rate of 50 % in [Khosla S Foot & Ankle Int 2009]) and polarizing microscopy in the lab [Schumacher HR Arthritis Rheum 1986; Hasselbacher P Arthritis Rheum 1987; McGill NW Aust NZ J Med 1991; Gordon C Ann Rheum Dis 1989] rather than having just one study on this comparator. Some statements in the manuscript text seem to be more subjective opinion than evidence based, as below.<br>If feasible, the authors may still try to include recent efforts in formulating algorithms by EULAR and the ACR, as below. Co-Author Dr. Fitzgerald may be a resource for this.<br>There have also been recently published efforts (ACR abstracts) by OMERACT in coming to a consensus regarding diagnosis of gout .<br>The second manuscript is improved over the first draft. | Thank you and we apologize if there was confusion about the structure of the report. What may have appeared to be two manuscripts or reports was actually an executive summary within the main report, which is the required structure for AHRQ EPC reports. We strive to make the executive summary a standalone document, to the extent possible. The 2009 article in Foot and Ankle would not have been included because the participants were cadavers and the outcomes related to intraarticular injection. We have added reference to the 1986 study by Schumacher and the 1991 study by McGill to the introduction, as they support one of the rationales for the review, namely that the accuracy of MSU analysis varies widely by institution. We have added the 1989 Gordon study, (none of these were identified in our searches as they did not include "gout" in their MeSH terms; the 1987 study by Haselbacher, also mentioned, is already included). We now include the OMERACT report and describe recent efforts to update diagnostic and classification algorithms in the Discussion chapter. |
| Peer Reviewer 5 | General Comments | The key questions are explicitly stated and the results of this report clinically meaningful. | Thank you. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| TEP 6 | Structured Abstract | 1) The purpose of the study is well articulated.<br>2) The methods section is well written and clearly states it is a retrospective study.<br>3) The results are clearly stated.<br>4) Conclusion is valid in relation to results from the study. | Thank you. |
| TEP 7 | Structured Abstract | (P5/v line 12) “dual emission computerized tomography” Would use consistent writing throughout the manuscript and tables. | Thank you. We have revised the term throughout in the manuscript and tables (dual-energy...) |
| TEP 7 | Structured Abstract | (P5/v line 26) “accepted” is subjective if is not clear by whom it is accepted, and how this acceptance is known (i.e. the term “accepted” cannot be proven or disproven without additional information.) | Thank you. We have replaced the word "accepted" with "validated," although we are not sure this description is correct, either. Space is limited in the abstract and didn't permit a full description. |
| TEP 7 | Structured Abstract | (P5/v line 29) “grey” It may not be clear to all what a gray search is. | We omitted the term, as we had already described our sources of data as “unpublished or non-peer-reviewed study findings.” |
| TEP 7 | Structured Abstract | (P5/v line 38) …crystals in joint aspirate | Thank you. We realize a word was missing. We revised the sentence slightly to "crystals in synovial fluid aspirated from affected joints." |
| TEP 1 | Executive Summary | (Background, page ES-1, line 39) This in general is excellent but there are errors on ES-1. Crystals are not UA but MSU. They do not "preferentially dissolve....".Perhaps you mean "deposit". I can't find where now, but I believe this ignored that there are probably microtophi in joints very early in the disease. | Thank you. We have revised our reference to monosodium urate crystals to the spelled out form or MSU crystals throughout and fixed the typographical error ("dissolved"-->"deposited"). |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| TEP 1 | Executive Summary | (page ES-6) Might it be important to report whether criteria were intended for Classification (for studies) or for Clinical Diagnosis. Features needed may differ. Also were studies for only Acute Gout or for Gout in general? Were data collected by direct exam or by history which may introduce recollection errors? | We have attempted to clarify in the Introduction that our intent was to evaluate criteria used for diagnosis, not classification. We realize several of the important sets of criteria were intended for classification, but as we note, they have been used for diagnosis as well and therefore merit consideration. We note which algorithms were intended for use in classification. |
| TEP 1 | Executive Summary | (Results, page ES-11) Results  are thorough but there can be questions. On ES-11 I am concerned about the emphasis on the Netherlands diagnostic rule as this includes hypertension, heart disease,& male sex which are not actually part of gout, but related points. Even if these test well so far I am concerned.<br>Also the strength of evidence of all these key points is identified as low. Might these findings be given too much credence?<br>Concerns about the limitations of SF study for crystals in general practice are appropriate but we might need more emphasis on the need to use arthrocentesis to exclude infection. | We have added the definitions of the criteria on which the strength of evidence is based and the meanings of the ratings. We have also added a brief discussion about the concerns associated with some of the criteria on which the Netherlands diagnostic rule is based, and we have added a discussion about differential diagnosis. |
| TEP 2 | Executive Summary | Well written. I like Figure A (page ES-3) that identifies the KQ. | Thank you. |
| TEP 2 | Executive Summary | (page ES-4) PICOTS will be new to most of the audience; this section is vital to understanding the paper. | We have defined the term, PICOTs, in the text and provided an explanation, namely that PICOTs, which is an abbreviation for "Participants, interventions/exposures, Comparators, Outcomes, Timing, is a method often used in systematic reviews to categorize the important study-level characteristics that comprise the inclusion and exclusion criteria. |
| TEP 2 | Executive Summary | (page ES-8) The results section describes how well the searches were performed but are not the results that the normal [ie non gout specialist] would expect to see. | The AHRQ systematic review report format specifies we begin the Results section with a description of search results and disposition. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| TEP 4 | Executive Summary | (page ES-14) Minor comment: (PDF) Page 23, lines 14 and 16, I would put reference information there. It made me wonder which papers you meant. | We have added the references; that was an oversight. |
| TEP 6 | Executive Summary | A comprehensive summary of the entire document. | Thank you. |
| TEP 7 | Executive Summary | (P10/ES-1 line 15) would add: "…in joints, cartilage, tendons, bursae, bone and soft tissues" | We have revised the definition exactly as suggested by the commenter. |
| TEP 7 | Executive Summary | (P10/ES-1 line 27) "…e.g. thiazide diuretics, low-dose aspirin or their combination." | We have added the examples of drugs (thiazide diuretics, low-dose aspirin, or their combination) that increase the risk for hyperuricemia) suggested by the commenter. |
| TEP 7 | Executive Summary | (P10/ES-1 line 38) consider placing the metaphor "building blocks" in quotation marks. | In the course of revising the text, we actually deleted the term "building blocks." |
| TEP 7 | Executive Summary | (P10/ES-1 line 39) "dissolve" would use the opposite here, e.g. precipitate or deposit. | Yes, thank you! We have fixed the typo (it was supposed to be deposit). |
| TEP 7 | Executive Summary | (P23/ES-14 line 32) Would consider describing quality of the gold standard: Clinically guided aspiration is less than 100% [Khosla S Foot & Ankle Int 2009] and studies that show false positive or negatives in the lab: [Schumacher HR Arthritis Rheum 1986; Hasselbacher P Arthritis Rheum 1987; McGill NW Aust NZ J Med 1991; Gordon C Ann Rheum Dis 1989] | We have reviewed the studies you kindly cited: Hasselbacher is already included. Khosla, Schumacher, and McGill do not meet our inclusion criteria but we now discuss them in the Discussion chapter, in terms of how the current report fits into the framework created by those studies; and we have added Gordon in our assessment of MSU analysis. |
| TEP 7 | Executive Summary | (P25/ES-16 line 16) Would use "Dual energy" or "dual-energy" | We have corrected the typo in our definition of DECT as suggested. |
| TEP 7 | Executive Summary | (P26/ES-17 line 48) Could include the newly proposed ACR/EULAR algorithm that was introduced in November 2014 at the ACR annual meeting in Boston. Dr. Fitzgerald may know about it. | The newly proposed algorithm have been presented only as an abstract. Therefore we have added them only to the Discussion. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| TEP 7 | Executive Summary | (P26/ES-17 line 57) “not radio-sensitive” That sounds a bit vague. Would consider being more specific, e.g. use doses published by the manufacturer . The following manufacture information mentions 9.2 mGy for foot and ankle on page 42: https://www.healthcare.siemens.com/siemens_hwem-hwem_ssxa_websites-context-root/wcm/idc/groups/public/@global/@imaging/@ct/documents/download/mdaw/mtmz/~edisp/dual_energy_ct-00079047.pdf | Thank you. We reviewed the information in the suggested reference, and decided that for the intended audience, it would be better to use a more general term than “not radio-sensitive” rather than the more technical terms suggested by the reviewer. |
| TEP 7 | Executive Summary | (P27/ES-18 line 32) Consider mentioning the recently presented ACR/EULAR algorithm that has recommended use of a sonographic double contour sign for diagnosis of gout. | We were unable to find mention of the double contour sign (DCS) in the 2014 diagnostic algorithm. We don’t know if it was advocated for use in treatment follow-up (as opposed to initial diagnosis); but because we did not find it mentioned with reference to diagnosis, we did not include this point in the report. |
| TEP 7 | Executive Summary | (P27/ES-18 line 42) “…beyond what would be available from most radiology centers…” That appears to be more speculation than an evidence based statement, unless the authors surveyed all radiology centers, or a representative proportion thereof, regarding this question. | Thank you. We qualified the statement ("may be"); however it is based on the input of our subject matter expert as well as the practitioners whom we consulted. |
| TEP 7 | Executive Summary | (P27/ES-18 line 44) “…focus on …single joints” The AIUM (American Institute of Ultrasound in Medicine) guidelines, that are relevant for most insurers, recommend assessment of several joints at a time per joint area. http://www.aium.org/resources/guidelines/musculoskeletal.pdf | Our topic expert assessed the content of this guideline for inclusion in the report and decided that for the intended audience, the information was probably too detailed. The information pertained to the dose and other details of the imaging procedure for individual joints and combinations of joints. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| TEP 7 | Executive Summary | (P27/ES-18 line 52) “availability, cost…” Availability and cost are drastically different between DECT and US. | Regarding the use of DECT and ultrasound for gout diagnosis, the 3e Recommendations note, the “availability, cost, and the need for trained personnel and specific equipment...” might limit the use of these modalities in routine clinical practice. We simply provided this quote to support the point that DECT and ultrasound were probably not likely to be first line diagnostic methods in primary care settings; we couldn’t revise the statement as it was actually a quote from the 3e guidelines. |
| TEP 3 | Introduction | Clear and well outlined | Thank you. |
| Peer Reviewer 1 | Introduction | The introduction is well-written. On line 39, page 1 of the introduction (and ES-1 of the executive summary), it is incorrect to say that UA crystals preferentially dissolve in joints, tendons, and bursae. I believe that the authors intended to write "preferentially deposit". | Thank you. We have fixed the typo. It is now "preferentially deposit". |
| Peer Reviewer 2 | Introduction | None of the algorithms mentioned were designed or evaluated with diagnosis in mind, except for the Netherlands study. This study was mainly designed to identify patients who needed SF analysis for diagnosis by virtue of having neither so many or so few features of gout that SF analysis was unnecessary. This important point was not mentioned anywhere in the report. | We have now made the point in the introduction and again when we discuss the individual algorithms that most were not designed for diagnosis, however we also point out that several of the ones designed for classification are used for diagnosis. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| Peer Reviewer 2 | Introduction | Furthermore, these so-called 'clinical' algorithms are not wholly clinical since most of them include sUA and some contain radiographic imaging. It seems to me that the authors have forced existing literature into a framework that doesn't align adequately. | We have now explained our usage of the term "clinical algorithms" and provided examples of the elements included. An excerpt of the revised text now reads as follows:" Instead of analyzing MSU crystals in synovial fluid, PCPs and emergency medicine physicians tend to rely on clinical algorithms comprising some combination of clinical signs and symptoms to diagnose an acute episode of gout. **These clinical signs and symptoms include rapid development of inflammation and pain, erythema, monoarthritis, response to administration of the drug colchicine, and symptoms in the first metatarsophalangeal (MTP) joint, among others (with synovial fluid culture sometimes used to rule out septic arthritis and other potential causes for inflammatory arthritis).** Attempts to standardize and validate such clinical diagnostic algorithms—**some of which were developed not for diagnosis but for the purpose of classification of gout for enrollment in research studies**—date back to the 1960s." |
| TEP 4 | Introduction | I find the introduction to be reasonable. Minor comments are below. | Thank you. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| TEP 4 | Introduction | Page 10, first para: would revise the 2nd line as below.<br>"(…) that may progress to a chronic and persistent condition, with development of tophi (solid deposits of monosodium urate [MSU] crystals in joints, cartilage, and bones), a condition called chronic tophaceous gout."<br>There is no clear distinction between acute intermittent and chronic intermittent conditions, while the advanced stage of gout is characterized by more persistent joint manifestations and tophi (either clinically evident or hidden within the joint). | Thank you. We have revised the text exactly as the reviewer suggests. "The condition may progress to a chronic and persistent condition, with development of tophi (solid deposits of monosodium urate [MSU] crystals in joints, cartilage, tendons, bursae, bone, and soft tissue), a condition called chronic tophaceous gout." |
| TEP 4 | Introduction | Page 11, line 29: Would replace "physical findings" with "clinical signs and symptoms", as you mean symptoms as well here. | Thank you. We have made the suggested change. We replaced "physical findings" with "clinical signs and symptoms. |
| TEP 5 | Introduction | As described in the general comments, I would suggest that the introduction provide a nuanced discussion of (1) the clinical scenarios that this report means to address, (2) the importance (or not) of definitive gout diagnosis in each of these scenarios versus exclusion of other conditions (e.g. septic arthritis) which require specific treatment, (3) the performance of tests in isolation versus incrementally, (4) fundamental problems with the gold standard (e.g. on one hand you are studying the accuracy of clinical criteria compared to MSU, and on the other hand some studies use clinical criteria as the reference standard and other studies describe substantial inter-person variability in the diagnosis of MSU from synovial fluid aspirates). | At the beginning of the Discussion and briefly in the Introduction, we provide a short profile of the patients likely to be seen in the primary/urgent/emergency care settings, to put the diagnostic challenges into perspective. Unfortunately, it is beyond the scope of the project to provide clinical scenarios and to conduct a RAND (Delphi-like) appropriateness panel (a process in which a group of experts, similar to a TEP, is given a matrix of clinical scenarios and is asked to rate the appropriateness of a particular treatment or in this case, to rate appropriateness of diagnostic regimens, for each of the clinical scenarios, based on a combination of clinical expertise and experience and sometimes a newly completed evidence review). |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| TEP 5 | Introduction | (PDF Page 36)  What do you mean by prevalence of gout?  The proportion of the population who has ever had an attack?  Something else? | We used NHANES data. According to the NHANES, prevalence is measured by asking participants if they have ever been told by a doctor that they have a particular condition. |
| TEP 5 | Introduction | (PDF page 36) Annual costs $933 billion, annual ambulatory costs $1 billion?  Please reconcile. | Although the two sets of costs appear discrepant, we were actually describing two separate studies that looked at two different sets of costs. We have added some text to that portion of the introduction to clarify that point. |
| Peer Reviewer 3 | Introduction | I did not have time to review this section of the report. | No response needed. |
| TEP 6 | Introduction | 1) Good explanation of the background, etiology, pathophysiology, and risk factors regarding gout and its diagnosis.<br>2) Good explanation of the scope of the review with the key questions presented.<br>3) Figure 1 demonstrates a clear approach to the analytic framework of the review. | Thank you! |
| Peer Reviewer 4 | Introduction | The introduction would benefit from making the explicit statement that gout is the most common inflammatory arthritis -- this would help put gout into an appropriate perspective in terms of public health burden. | Thank you. We have added this exact statement to the introduction. |
| Peer Reviewer 4 | Introduction | As well, hyperuricemia drives not only the acute episodes, but also the development of tophi as well as bony destruction leading to joint abnormalities. Focusing only on the acute aspect of gout may unfortunately compound the already poor understanding of gout -- gout is not just the acute attacks; without appropriate management of the underlying hyperuricemia, thereby allowing the body's urate burden to rise, longer term consequences of gout arise. | Thank you. We have revised the introductory paragraphs to include chronic tophaceous gout and bony destruction leading to joint abnormalities. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| Peer Reviewer 5 | Introduction | (Pg 37/2 line 40) It's unclear as to what authors mean by "...evidence on the comparative validity and safety of tests used for the diagnosis of gout, ..". In particular what safety aspect of tests of gout diagnosis are authors looking for? This needs further clarification, if it is not a typo. | Thank you. We realize we may have been confusing in our usage of the term "safety" regarding diagnostic tests. We typically assess the evidence for the efficacy and safety of treatments. We have added an explanation to the introduction of what we mean by "safety," namely short term pain and discomfort that results from a test, long term consequences such as the effects of radiation exposure, and the potential harms of misdiagnosis, either missing a diagnosis of gout, or misdiagnosing another inflammatory arthritis as gout, |
| Peer Reviewer 5 | Introduction | PICOTS are very reasonable, I agree with them. Appropriate databases were used. | Thank you. |
| TEP 3 | Methods | See above re comments re synovial fluid analysis | We have addressed the issues regarding synovial fluid analysis and other comparators in the description of the inclusion and exclusion criteria. |
| Peer Reviewer 1 | Methods | The methods employed to complete this systematic review were exemplary. The inclusion and exclusion criteria are justifiable and the search strategies were explicitly stated and logical. The outcome measures and statistical methods were appropriate. | Thank you. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| Peer Reviewer 2 | Methods | (ES-4-ES-6, 5-7) There are 2 main problems with the inclusion and inclusion criteria. Firstly, the suspicion of having gout versus an established diagnosis of gout is a false distinction. Suspicion can range from no suspicion (established diagnosis of a different disease) to certainty (established diagnosis of gout). The rationale of excluding patients with established disease is unclear to me. The population of interest is those to whom the diagnostic test would be applied to - this includes people with very low and very high suspicion. As long as the index test is determined without knowledge of the clinical diagnosis, I do not see why it is necessary to exclude patients with established diagnoses from a review of diagnostic tests. | This review will be used as part of the basis for formulating guidelines for diagnosis of gout in primary, urgent, and emergent care settings, where many individuals with gout are first (or exclusively) seen. To examine studies with participants who resembled this patient population to the extent possible, the scope of this review included studies that enrolled patients who did not have an established diagnosis and who were identified in primary care; such patients might also likely be in an earlier or less advanced stage of the disease. This decision was reinforced by the sponsors. We would have considered including studies of patients with established diagnoses if the assessors were blinded, but these patients present another problem: the duration of their disease and the extent of symptoms is likely to be different than that of primary care patients with no prior diagnosis. |
| Peer Reviewer 2 | Methods | (page ES-12, 13, 14, 17, 22...) Secondly, the choice of the reference standard as the 1977 Wallace ARA criteria is inappropriate for at least 2 reasons: the criteria are inaccurate when compared against a real reference standard; the criteria are amongst the list of 'clinical algorithms' that the review evaluates. | We realize that inclusion of a study that used the ARA criteria as a reference standard created a methodological problem (because we were evaluating the validity of the ARA criteria themselves). However, in order to present a complete picture of the gout diagnostic methods, we accepted studies a very small number of studies that used the ARA criteria, and only if no other studies could be identified assessing the diagnostic method in question, and we noted that these criteria were used. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| Peer Reviewer 2 | Methods | (throughout) Choosing PPV and NPV as indicators of test performance needs to be accompanied by the caveat that these indices are also strongly affected by disease prevalence, so they effectively are indices of the test in a particular context. Without knowing more about that context, the values are very difficult to interpret. I suggest that they add little unless the authors wish to restrict their evaluations to particular clinical situations or populations. | Thank you. We have now briefly addressed the problem with interpreting NPV and PPV in the Discussion section on Limitations.. |
| Peer Reviewer 2 | Methods | I would like to see a PRISMA checklist included to show explicitly how guidelines for systematic literature review was followed. | As part of the requirements for AHRQ systematic reviews, we complete and submit a PRISMA checklist. The information included in that list is the same information we include in our introduction and methods section, so we don't include the checklist itself in the report. |
| Peer Reviewer 2 | Methods | (page 8) AUC for imaging tests is not very meaningful since imaging tests are usually reported as feature present or not (only one data point in the ROC curve). The SROC described in the methods was not presented in the results, nor pooled estimates of test performance. Why not? | We abstracted ROC when reported by the original studies. We didn't construct ROC curves ourselves for the reason the reviewer states. Because we did not pool studies, we did not calculate any SROC. We no longer refer to SROC in the Methods. |
| Peer Reviewer 2 | Methods | (page 8) The Applicability section is not easy to understand - can this be expanded and explained more fully. | We have added an explanation for our own conception of applicability as it applies to the patient and provider groups who are the focus of this report; we added this text to the subsection of the Discussion entitled “Applicability,” and have provided a clearer description of how we assessed applicability for this report. The report was intended for providers and patients in primary care, urgent care, and emergency care settings, rather than the secondary or tertiary care settings (e.g., academic rheumatology departments) that are usually the focus for gout diagnostic studies. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| TEP 4 | Methods | (page 30?) I was initially struck by the major discrepancy of included articles for DECT and ultrasound between this report and Ogdie et al's 2014 systematic review (ref 37). Of note, the latter was done in preparation for the development of new ACR-EULAR gout classification criteria, which have just been reported at the 2014 ACR meeting in Boston. Despite the major discrepancy, the employed inclusion and exclusion criteria seem to serve the specific purpose of systematically reviewing the evidence for the context of primary care providers encountering "gout suspects". While I was able to understand the reason behind the differences, you may want to further discuss the differences behind these reports, as readers (particularly from rheumatology) may wonder about them as I did initially. | We have added text throughout the Introduction and methods to call attention to our exclusion criteria. When we described the findings of the Ogdie review, both in the sections on US and on DECT. We noted the discrepancy in the numbers of studies they included compared with our review and explained why we had excluded studies they included. |
| TEP 4 | Methods | The search strategies are explicitly stated and seem logical, and the definitions and diagnostic criteria for the outcome measures are appropriate, although the gold standard itself has it own many limitations as discussed in the report. To that effect, I feel that there is room for further emphasis (in more explicit and methodologic terms, where relevant) of the suboptimal nature of the employed gold standard. | We have added a number of references and augmented the description of the limitations of MSU analysis in the Introduction and Discussion. We summarized original studies and systematic reviews that found limitations in the accuracy of needle placement, and in accuracy of synovial fluid crystal analysis across laboratories, practitioners, and patients. We have also included an additional study on interrater reliability of MSU assessment in the Results and added an abstract to the Discussion. |
| TEP 4 | Methods | There were no notable statistical methods used, given insufficient findings for the key questions. | The reviewer notes that because we did not identify sufficient numbers of similar studies, we did not consider conducting a random effects meta-analysis. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| TEP 4 | Methods | Both ‘algorithm’ and ‘clinical algorithm’ are used throughout the report to mean ‘clinical diagnostic algorithm’. I would consider using the full terminology, particularly for readers who are more clinically-oriented (as opposed to methodologically oriented). | Thank you. We changed to the full terminology throughout as clinical diagnostic algorithm’. |
| TEP 5 | Methods | (PDF page 6 and 94) Regarding the search strategy - Appendix A: Please clarify where you are searching for a title word, text word, MESH heading, or something else. | The search strings, that is, the individual lines of code or steps in the search that list the key terms, described in the Appendix A, specify where the key words are being searched. |
| TEP 5 | Methods | All of your literature searches require “gout” or “gouty.” It is unclear whether this will identify all articles of interest. Consider the possibility of papers focused on the exclusion of septic arthritis, or clinical strategies for patients presenting with mono-arthritis or oligo-arthritis. Are we sure that those will all be identified with the “gout” or “gouty” requirement? | Our search was aimed at identifying studies on gout diagnosis. Searches that identified studies on gout would identify studies on the differential diagnosis of gout, septic arthritis, CPPD, and other such conditions. If a study was aimed at diagnosing patients with a mono- or oligo-arthritis, there is a nearly 100% chance that the word gout would appear as that would be one possible diagnosis. In the report review process, one reviewer did, in fact, identify two studies on differential diagnosis of inflammatory joint diseases in general, that did not include the term “gout,” and we checked the reference lists for those studies, to ensure we had not missed anything else. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| Peer Reviewer 3 | Methods | (page 5) Inclusion criteria: This section is very confusing and needs to be improved. I would suggest not mentioning anything about the search in this section. I would suggest moving all the information to the subheadings and having separate inclusion criteria for the different review questions. However, PICO is not appropriate for DTA reviews In particular the use of comparators to refer to reference standard is misleading. You need to use appropriate categories for diagnostic data. I would suggest following the Cochrane format: Participants, index test, target condition, reference standard and possibly also outcome, with an additional category for study design (unless you want to cover this under the broader heading of inclusion criteria in which case make sure you just talk about the design features and do not also talk about other features e.g. participants and index tests that will be covered under specific categories). | AHRQ evidence reviews follow a set of publication guidelines that include the structure of the Methods descriptions. We do use the PICOTs framework to describe the inclusion and exclusion criteria for treatment reviews, and we believe its utility extends to the description of inclusion and exclusion criteria for diagnostic reviews; however, based on your suggestion, we have added the terms "index test" and "reference standard" to our framework. In the past, when there were many key questions and each had different PICOTs, we would construct a table with one column for each key question, but since this report has only two key questions, we thought it made sense to describe the questions to which each element applied as we did. |
| Peer Reviewer 3 | Methods | (page 6-7) Search: The “Cochrane collection” is not a database. Please be clear about which databases within the Cochrane Library you searched. Please replace “present” with the actual end-date of the searches. Present is not informative for someone reading the review after the searches were conducted. You have included end dates in the results – these should be moved to the methods. | Thank you. We have corrected the name of the Cochrane Library and provided the correct beginning and end dates for the update searches. |
| Peer Reviewer 3 | Methods | (page 7, line 26) “Full-text review was conducted in duplicate using the predetermined inclusion and exclusion criteria.” What do you mean by this? That two reviewers independently reviewed full text articles? | Thank you. We have revised the wording as suggested: “two reviewers independently reviewed full text articles” |
| Peer Reviewer 3 | Methods | (page 7, line 31)Why did you exclude “did not include or identify control groups (patients who | The intended audience for the report is a guidelines committee who will be setting |

| | | | |
|---|---|---|---|
| | | tested negative for gout using the gold standard test)." | guidelines for diagnosis in primary and urgent care settings. To ensure that the included studies enrolled participants as similar to as the target patients as possible, we sought to exclude studies that enrolled only patients with a previous definitive diagnosis of gout. Ideally, the "control" populations would be individuals with presumed gout or some monoarthritis who ended up with a negative test using the gold standard but we had to accept studies in which the controls were people who met the criteria for other forms of inflammatory arthritis. |
| Peer Reviewer 3 | Methods | (page 7, line 40) "For studies of apparent interest reported in meeting abstracts, we searched for peer-reviewed articles before determining whether to accept the studies." What does this mean? Does this mean that you excluded conference abstracts unless you could find a full text report? | Yes, we included data only from peer reviewed studies. If we identified an abstract of interest for which no peer reviewed article was or has been published, we would describe it in the Discussion. |
| Peer Reviewer 3 | Methods | (page 7, line 38) "We also searched accepted studies for additional references and screened any articles of apparent interest." What is an "accepted study"? Do you mean an included study, or just one that had passed the initial title and abstract screening stage or something else? What was an article of "apparent interest"? | Yes, we searched our included studies for references whose titles suggested they might fit our inclusion criteria. We have revised the wording. |
| Peer Reviewer 3 | Methods | (page 7, line 46) Data abstraction: "dually abstracted" – what does this mean? Two reviewers independently performing data extraction? | Thank you. We have revised the wording "two reviewers independently performed the data extraction" |
| Peer Reviewer 3 | Methods | (page 8, line 19) Synthesis: "a bivariate model proposed by Rutter and Gatsonis (2001)." Rutter and Gatsonis proposed the HSROC model not the bivariate model. Although essentially mathematically the same model the outputs from these are different. Please correct to either the HSROC model if used or attribute the bivariate model to the correct reference. | Thank you. We have removed this sentence, as we did not ultimately derive ROC curves. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| Peer Reviewer 3 | Methods | (page 8) Generally the synthesis section is rather short and lacks details on exactly what analyses were conducted. When the various measures of accuracy were used and when pooling was considered appropriate. There is no information on how heterogeneity was assessed and what was done where heterogeneity was found. | Because we identified only small numbers of studies that met our inclusion criteria, we did not do any pooling. |
| Peer Reviewer 3 | Methods | (page 8, line 53) Applicability: You have used QUADAS-2 to assess study quality. If performed following the guidance for this tool then you will also have assessed applicability as part of the quality assessment. Why do you therefore also have a section on applicability? This section lacks clarity and needs more detail on exactly what this involved. | Guidelines for AHRQ EPC systematic reviews suggest that we assess applicability explicitly and separately from risk of bias for individual studies or the overall strength of evidence for conclusions, because the reports are used by a variety of audiences, each with different needs and criteria for applicability of the information. A particular target audience for this report is the primary/urgent/emergency care practitioner. We have revised the sections that describe how we assessed applicability to make the criteria clearer. These sections appear in the Methods section of the Executive Summary and the Methods chapter of the main text, both under the heading, Applicability and in our description of the components of strength-of-evidence assessment, because applicability is often considered in assessing strength of evidence. |
| Peer Reviewer 3 | Methods | (page 8, line 37) Grading the body of evidence: Why did not you not assess directness and precision when GRADING the evidence? Tests used for publication bias are not appropriate for DTA data. | The AHRQ EPC guidance for assessment of strength of evidence for diagnostic test evaluation suggests that assessing directness and precision for diagnostic test questions is challenging and not always applicable. Regarding "directness," we chose to refrain from assessing that domain because of the lack of clarity about the use of synovial fluid uric acid crystals as the reference standard. We did not evaluate precision because of the inadequate numbers of studies. However, had we included these domains, our ratings of strength of evidence would not have changed. |

| | | | |
|---|---|---|---|
| TEP 6 | Methods | 1) Good explanation that this is a systematic search for prospective or cross sectional studies.<br>2) Good explanation of the research question (PICO), data abstraction and management. | Thank you. |
| Peer Reviewer 4 | Methods | (ES-4 and 5) A specific subgroup not mentioned in the PICOTS list of population(s) is CPP-related arthritis, which is a common DDx for gout (and is much more prevalent than septic arthritis). This subgroup is often misdiagnosed as having gout due to lack of synovial fluid aspiration confirming the diagnosis; treatment with urate-lowering therapy in this group is obviously inappropriate. Many studies lack sufficient #s of subjects with CPP-related arthritis, which also has important implications for interpreting the results for US in particular (and to a certain extent for DECT). | Thank you. We have added a discussion of the need for differential diagnosis to rule out conditions such as CPPD and septic arthritis to the introduction and have added it to the PICOTs list. |
| Peer Reviewer 4 | Methods | Minor point: in the Methods section, the authors should correct their characterization of the ARA criteria -- these are classification criteria, not "for gout diagnosis" (ES-4, lines 17-18 and main manuscript page 5, lines 17-18 ). | We have added the word, "classification" to describe the ARA criteria, and we added a statement to the introduction, noting that certain criteria were developed for classification, not diagnosis. |
| Peer Reviewer 5 | Methods | The inclusion and exclusion criteria are justifiable and the search strategies are clearly described. | Thank you. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| Peer Reviewer 5 | Methods | I am unclear about assessing synovial fluid crystal both as test standard and as a "gold standard". The challenge is that the construct of the disease is often described with MSU crystals as the gold standard. The authors point out studies that have questioned that. I think this is a clear dilemma, which may be beyond the scope of this review to address. | Relevant to the reviewer's concern, we have added a new study by the SUGAR group (Taylor et al., 2014) that compares the sensitivity and specificity of the diagnostic criteria (algorithms) that usually include MSU analysis as one element, between MSU being included as a criterion and MSU being omitted. We have also added mention of the issue the reviewer raises to the Discussion: that comparing the accuracy of an algorithm that includes MSU as one element to MSU itself is questionable. We also note in the discussion that the new two diagnostic algorithms we cite as showing good sensitivity and specificity (the Janssens and the CGD) don't include MSU as an element. |
| TEP 3 | Results | Aside from my comments above, there are no additional concerns in the results section. | Thank you. |
| Peer Reviewer 1 | Results | (throughout document) The results are well-written, summarizing the key findings in multiple studies relevant to the question being studied. There is superb use of tables to summarize these studies. I could not identify any missing studies. The authors should specify whether the DECT studies examined only the clinically affected joint or whether they examined both upper and lower extremities in a patient with suspected gout. This obviously influences the reported sensitivity of the test. The same applies to ultrasonography. | Thank you. We have added a description to each of the DECT and ultrasound studies noting whether only the affected joint or multiple joints underwent imaging. |
| Peer Reviewer 2 | Results | (page 12-13) citations do not seem to match references. I found this at various other parts of the report which was annoying. | Thank you. We have checked the references on pages 12 and the top of page `13 and we cannot identify the mismatched references but we have tried to make sure throughout the report that descriptions of references are not ambiguous. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| Peer Reviewer 2 | Results | (page 14)  The authors refer to the New York and Rome criteria as "subsequent briefer variations" of the 1977 ARA criteria. This is clearly not true: the NY and Rome criteria were developed by consensus in the 1960's. | Thank you. We have corrected this error. |
| Peer Reviewer 2 | Results | (page 14) The paragraph about Janssens studies in fact only concerned one study with different analyses. | Thank you. We have corrected the descriptions of the studies. |
| Peer Reviewer 2 | Results | (page 14) The Mexico criteria were a simplified version of ARA criteria - this could be made clearer. | Thank you. We have clarified it in the description, which we moved down in that section. |
| Peer Reviewer 2 | Results | (page 15) There is no mention of the bias that is introduced by recruiting patients who had undergone SF fluid analysis (rather than all patients who might have gout) eg as in the French and German studies. There is also no mention of the bias that occurs when the test performance is calculated from the same patient sample as was used to derive the criteria/algorithm. There is no mention of the implications of using co-morbidities and demographics in the diagnostic algorithm, rather than intrinsic features of the disease. | Our reason for excluding studies of patients with prior gout diagnoses (and even prior tests for gout) was based on the first issue the reviewer raises, as we mention in the Methods chapter; patients with a prior diagnosis or who have been suspected of having gout and received a prior test are not likely to represent the average patient who will be seen in the primary care setting with a first episode of gout and could skew the apparent sensitivity of the test. However we had not considered the additional sources of potential bias the reviewer mentions. We have now addressed them (e.g., comorbidities such as CVD and demographics such as sex) in the Discussion section in our discussion of the limitations of the literature. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| Peer Reviewer 2 | Results | (page 24) DECT and US review: The studies that used classification criteria for the reference standard should be excluded since this introduces unnecessary classification error. Only MSU identification should be used as the reference standard. I think that studies that include patients with established gout should be included (see above). There are studies included in the Ogdie paper that could have been included in this review. Overall, it is difficult to see how this report adds to the Ogdie paper and in my view is less informative. There are no pooled estimates or reasons given for not pooling results.<br>The US review does not discuss particular features that may be observed with this technique. This is important since some features may be more or less accurate than others. This is not so important with DECT since DECT reports only a single feature. | We did screen each of the studies included in the Ogdie and Chowalloor reviews. With the TEP's guidance, we excluded those that enrolled patients with hyperuricemia and prior diagnosis of gout because we were trying to replicate the kinds of patients most likely to be seen in primary care, rather than to repeat Ogdie's analysis We explained our exclusion and inclusion criteria in the sections on DECT and US. We did not pool studies because we regarded them as too small in number and heterogeneous, as we now state. We have added detail to the review of US studies to clarify relative accuracy, if reported by the authors. |
| TEP 4 | Results | I find the level of detail in the results section appropriate, and the characteristics of included studies are well-described in the text and tables. I find the figures, tables, and appendices to be reasonable. | Thank you. |
| TEP 4 | Results | In terms of potentially overlooking studies that ought to be included, the authors may want to look into and incorporate the ACR-EULAR criteria for gout that was presented at the Boston ACR meeting, if possible. This latest gout classification algorithm incorporates DECT or ultrasound as well as serum uric acid levels. Notably, the latter information was excluded from the NY criteria based on the conjecture that there was no lower level below which gout was not a possibility, as the report states in (PDF) page 80 line 35 (refs 45 and 46). | We have obtained the abstract for the 2014 EULAR classification criteria presented at the ACR meeting. It is now discussed in the Discussion section (Findings in Relation to What is Already Known). However we have been unable to obtain the full evidence review that was conducted to support the new classification criteria; therefore, we could not assess whether we missed any studies we should have included that the EULAR reviewers included. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| TEP 5 | Results | It would be much easier to digest the individual study results if they were presented in a more systematic manner (e.g. describe patient population {presenting where? consecutive patients or convenience sample? etc}, reference standard, etc., in the same order, for every study). More information regarding alternative diagnoses in patients determined not to have gout would also be helpful in understanding the population and the importance (or not) of specific diagnosis. In the evidence table, consider providing a separate column for reference standard rather than combining with diagnostic test. In the inclusion criteria, provide a more precise description of the patient population in terms of how selected it is. | Thank you. We have added methodological detail and information describing differential diagnoses to the descriptions of the individual studies in the text. We strive to design the evidence tables so that the entry for each study can be seen in its entirety on one page, so we have not added more columns. |
| TEP 5 | Results | (PDF Page 47, line 33) The sensitivity and specificity of US/DECT are not meaningful without the exact clinical context. | We have tried to present the clinical contexts to the extent that they were presented in the reports of the studies. In reality, the PPV and NPV are more difficult to understand without knowing the prevalence of gout in each country in which a study was conducted. |
| TEP 5 | Results | (PDF Page 47, line 41) Note that “clinical utility” is an unfair standard here if that is not what you were using for imaging. Please consider whether any study provides evidence for uric acid concentrations in patients without a previous hx of gout who do/don’t have crystals in their joints when aspirated. | Thank you. We agree "clinical utility" was an inappropriate choice of words. We have modified it to "validity." Also, we have added data from one of the studies on associations of imaging findings with sUA in the Discussion and we review the evidence on the use of serum uric acid to diagnose gout in the section of the Discussion on gaps in the evidence. |
| TEP 5 | Results | (PDF Page 47, line 44) Looking at table 2, it would appear that several studies have related the # of joints and site to the likelihood of gout. | We have now added descriptions of the algorithm components to our descriptions of the studies that assessed the validity of the algorithms in the text, however none of the studies provided enough data to determine whether sensitivities or specificities were affected by the number or identity of joints. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| TEP 5 | Results | (PDF Page 48, line 7) “the predictions based on these tests …” Which tests? | We changed "these tests" to "5 clinical algorithms." |
| TEP 5 | Results | (PDF Page 48, line 54) Note that the majority of patients in this landmark study were “presumed to have gout.” This highlights the gold standard problem and speaks to a certain amount of circularity. | The patients in the Wallace study had no definitive diagnosis of gout (as was the case in many imaging studies, for example). In fact they were patients referred by PCPs for signs and symptoms of gout. The fact that these signs and symptoms were likely some of the same ones that were then tested as part of the algorithm presents the same circularity problem the reviewer mentions. We now address this potential shortcoming in the assessment of test validity in the discussion of biases in the report in the Discussion chapter. |
| TEP 5 | Results | (PDF Page 50, line 34) “the Hosmer-Lemeshow goodness of fit was 0.64.” Rephrase as HL statistic was non-significant at p=0.64, or something similar. | Thank you. We rephrased it as. “HL statistic was non-significant at p=0.64”, |
| TEP 5 | Results | (PDF Page 51, lines 19-27) This study deserves a more thorough discussion. What was in the joints of patients without gout? Were other kinds of crystals identified? How did they know that the taps were performed correctly in patients with negative taps? | We agree and added as much additional detail as was provided in the article. In our response to KQ1b, we added a description of the tests the patients underwent to rule out a differential diagnosis and how the patients were followed over time; unfortunately, the authors did not report on how they ascertained that taps were done correctly. |
| TEP 5 | Results | (PDF Page 59) Did any patients have other kinds of crystals? Are DECT (and US, for that matter) expected to distinguish MSU from other kinds of crystals? | We have expanded the descriptions of each of these studies: DECT does in fact discriminate MSU from CPPD crystals, and we provided additional information in the text. |
| TEP 5 | Results | (PDF Page 65) You say that the combination of US and a “clinical algorithm” was studies but do not discuss the incremental value of US in addition to the clinical algorithm. Please do so. Again, what of the joint aspirations that did not show MSU? | None of these studies actually directly addresses the incremental value. We have provided the additional information.in the Discussion section. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| TEP 5 | Results | (PDF Page 75, lines 13-19)  Here you are addressing only who does the analysis, not who does the aspiration.  Since both are part of the key question, both parts of the question should be addressed here and these two very separate issues should be discussed separately on the rest of this page.  With regard to analysis, it  would be helpful to clearly state which studies compared same-sample readings by different people. | We identified no studies that addressed the role of the person performing aspiration, We have provided the additional information We have moved the studies on institutions to the introduction as they do not really address the question and we have added a description of the results of a study that actually compares the performance (detection of MSU and other crystals) of a group of various practitioners on the same samples of synovial fluid. |
| TEP 5 | Results | (PDF Page 77, lines 19-22)  Consider being more specific here … something along the lines of “failure to diagnose gout was reported in a single-center trial to lead to unnecessary surgical procedures for treatment of presumed septic arthritis.” | Thank you. We made a wording change to increase the specificity of the study findings, as suggested ”Missed diagnosis or delayed diagnosis of acute gout (failure to find MSU crystals in synovial fluid) was reported in a retrospective two-center study to be associated with a longer interval between the onset of attack and joint aspiration. A negative MSU finding was associated with higher risk for undergoing arthroscopic drainage, longer hospital stay, and delays in anti-inflammatory treatment.” |
| Peer Reviewer 3 | Results | (page 10) I don’t think you need to repeat what databases were searched in the results section. | Thank you. We have removed this information. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| Peer Reviewer 3 | Results | (page 10) Excluding studies for not having an abstract seems rather a strange reason for exclusion.  I would also question the decision of not including conference abstracts; these can be a valuable source of information for systematic reviews.  You state that you excluded 3 studies as not all patients received the reference standard but I do not think that this was specified as an inclusion criteria.<br>Please ensure that reasons for inclusion and exclusion criteria match up and that no post hoc exclusions were made i.e. that inclusion/exclusion criteria match those specified in the protocol.<br>I did not have time to review the results further. | 1) We screen a subset of full-text publications for which no abstract is posted online (usually about 20% of all such titles). Typically they are all letters, conference proceedings, or editorials/commentaries. If we find that none of the pieces meet our inclusion criteria, we don't screen additional ones in full text. 2) We specified in the first paragraph of our methods section (Criteria for Inclusion/Exclusion) that we excluded studies if patients did not receive a reference standard test; we have added some additional wording for emphasis. 3) We excluded conference proceedings because they do not undergo peer review; however, new, noteworthy conference abstracts are cited in the Discussion chapter. It is our practice that if we identify conference proceedings that report on a study that would appear to meet our inclusion criteria, we search for a subsequent peer-reviewed publication that reports the data, as we describe. |
| TEP 6 | Results | 1) The results are presented in a logical sequence with appropriate subheadings.<br>2) Good explanation of the initial exclusion criteria. | Thank you. |
| TEP 6 | Results | (page 11)  Please clarify how from the 235 potential background articles, only 17 original studies and 5 systematic reviews were included. What were the exclusion criteria for those articles? | We have revised the flow diagram to resolve the misunderstanding: the 235 articles described as potential background were non-systematic reviews that we examined to ensure we had not missed original studies and then excluded as they did not contribute to our analysis. |
| TEP 6 | Results | (page 12) Please explain how the authors were able to state that ultrasound demonstrated good sensitivity when their reported studies demonstrated sensitivities down to 37%? | We reviewed the study in question and changed our conclusion accordingly. In reviewing the study, it appeared that the low specificity might have been related to the inclusion criteria the authors used. Thus, although we think the 37% may be an outlier, we can't discount it. |

| | | | |
|---|---|---|---|
| TEP 6 | Results | Also, the authors should consider including more studies to help increase the strength of evidence from low and insignificant. | As we stated above, we included the studies that fit our prespecified inclusion criteria, so we would have had to change the inclusion/exclusion criteria in our protocol to allow more studies to be included, and we would no longer be limiting the assessment to studies of individuals with no prior diagnosis of gout. Also, had we relaxed our inclusion criteria to include lower-quality studies would not have helped raise the strength of evidence. |
| Peer Reviewer 4 | Results | Main issue with results is related to the comments from above re: CPP-related arthritis. Understanding the prevalence of CPP-related arthritis in each study and how the specific test fared in differentiating gout from CPP-related arthritis would be important. While one could argue that treatment of the acute episode may be similar between the two conditions, that would unfortunately perpetuate the inadequate longer term management of these conditions which have different risk factor profiles to target and very different approaches to managing the underlying disease (hyperuricemia for gout). | Thank you. We realize differential diagnosis is an important issue. Both US and DECT can distinguish CPPD from MSU crystal deposition, but none of the studies we identified directly assessed differential diagnosis. We now address this point in the Discussion and refer to the studies that at least considered differential diagnosis, mainly from septic arthritis and CPPD. |
| Peer Reviewer 4 | Results | There was an abstract presented at EULAR 2014 that evaluated accuracy of synovial fluid evaluations by different types practitioners, but this was presented in June, after the evidence review cut-off dates in Apr/14. | Thank you for calling our attention to this abstract. Although we did not include conference proceedings as evidence, we do now discuss the conference abstract presented at EULAR 2014findings in the Discussion, in reference to our reported findings. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| TEP 7 | Results | (P48/13 line 4) "…identification…in synovial fluid." See references provided for Executive Summary comment above. | We have reviewed each of the suggested references. The 2009 article in Foot and Ankle would not have been included because the participants were cadavers and the outcomes related only to intraarticular injection. We have added a reference to the 1986 study by Schumacher and the 1991 study by McGill to the introduction, as they support one of the rationales for the review, namely that the accuracy of MSU analysis varies widely by institution. We have added the 1989 Gordon study, (none of these were identified in our searches as they did not include "gout" in their MeSH terms; the 1987 study by Haselbacher, also mentioned, is already included in our review). We now include the OMERACT report and include additional updated efforts on diagnostic and classification algorithms in the Discussion. |
| TEP 7 | Results | (P82/47 line 18) "not radio-sensitive organs" Vague. See comment provided for Executive Summary above. Could be more precise here with data from DECT manufacturer as in reference above. | Thank you. We reviewed the information in the suggested reference, and decided that for the intended audience, it would be better to use a more general term than "not radio-sensitive" rather than the more technical terms suggested by the reviewer. |
| TEP 7 | Results | (P82/47 line 42 and following) same comments as above. | Regarding the use of DECT and ultrasound for gout diagnosis, the 3e Recommendations note, the "availability, cost, and the need for trained personnel and specific equipment..." might limit the use of these modalities in routine clinical practice. We simply provided this quote to support the point that DECT and ultrasound were probably not likely to be first line diagnostic methods in primary care settings; we couldn't revise the statement as it was actually a quote from the 3e guidelines. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| TEP 7 | Discussion/Conclusion | (P83/48 line 6) “enthusiasm” The presence of enthusiasm does not invalidate or speak against a particular study by itself. | We have revised the wording conveying the intended idea, that looking for many different signs across many joints may not be practical diagnostically for the average primary care patient or physician. |
| TEP 7 | Results | (P83/48 line 7) “…marginal utility…in lieu of joint aspiration.” That sentence does not make immediate sense to the reader: All DECT or US studies evaluate the utility in lieu or beyond clinical data. Would consider rephrasing for clarity. | The marginal utility of using imaging over that of clinical criteria alone for the diagnosis of gout is of considerable interest to the sponsor. The term itself was used in the study being reviewed. However, we realize the statement was confusing and have reworded it to the following: “Furthermore, we did not find any studies that evaluated the marginal utility of using ultrasound data to diagnose gout, above that of clinical criteria or in lieu of joint aspiration.” |
| TEP 7 | Results | (P83/48 line 32) “…further development…of diagnostic algorithms” Combined EULAR/ACR efforts regarding such development have recently been presented (at the ACR meeting in Boston, MA in Nov 2014). Similarly, OMERACT has recently published (in abstract form) collaborative efforts in diagnosing gout by ultrasound and other means. | Because the new guidelines have appeared only in conference abstract form, we summarize them in the Discussion section. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| Peer Reviewer 5 | Results | (not sure if his comment refers to PDF page 25-29 or actual page 25-29) Key points indicated that "The strength of evidence for this conclusion is low based on the identification of only two studies that assessed this particular clinical algorithm." However, the summary/details of these 2 studies is missing on pages 25-29. Either these studies need to be added and described along with their strengths and limitations, or this statement and several similar statements mentioning this need to be corrected. I can only see one abstract by the same group listed in the reference list. | Thank you. We have added text at the beginning of the actual presentation of the findings (following the description of the results of the literature searches) that describes how the remainder of the chapter is laid out and that the details of the studies are described below the Key Points and overview of the included studies. “The findings are organized as follows. For key question 1a through c, we present first the key points, followed by a brief overview of the studies and then detailed narrative descriptions of each study and prior systematic review that addresses that question or subquestion. The studies that address Key question 1d and key question 2 are described separately, with the key points followed by the study details.” |
| Peer Reviewer 5 | Results | The figures, tables and appendices are adequate. | Thank you. |
| Peer Reviewer 5 | Results | I did not perform a literature search to evaluate if any studies were accidentally excluded. | No response needed. |
| Peer Reviewer 5 | Results | Study characteristics have been adequately described. | Thank you. |

| | | | |
|---|---|---|---|
| TEP 1 | Discussion/Conclusion | (page ES-16 thru ES-21 and page 45-50) As in my General Comments I think Conclusions might be changed a bit to use evolving criteria for initial handling of attacks but to recognize that they are often not specific enough for lifetime commitments that are needed.<br>Please reconsider if the criteria you seek are to diagnose gout in general or to diagnose acute gout.<br>Please also consider a series of very recent papers from the ACR/EULAR classification of symptomatic gout from Taylor et al. Although for Classification for research they may be worth attention. | We have entirely reframed the Discussion, which now includes a summary of the studies that compare clinical algorithms, the possible influence of patient population and duration of disease on the accuracy of the algorithms, and how they compare with synovial fluid MSU analysis.. We have clarified the typical patient profile in the Introduction and the Discussion. We have obtained and included the recent article by Taylor et al and the Study for Updated Gout clAssification cRiteria (SUGAR) group and have now included it. We also modified the bottom line Conclusion  as follows: "...An algorithm with high diagnostic accuracy can ideally form part of a decision tree that combines clinical signs and symptoms with—or refers patients to rheumatologists for—more invasive tests or imaging for long term management of clinically ambiguous cases. Research is needed to assess the incremental value of synovial fluid MSU crystal analysis and imaging over that of a diagnostic clinical algorithm." |
| TEP 2 | Discussion/Conclusion | (page ES-16 thru ES-21 and page 45-50) I think this section is 'wordy'.  The information is there but buried in a flood of other words. | Thank you. We have reorganized the Discussion so that we now summarize and discuss the individual issues with each of the different kinds of tests. |
| TEP 3 | Discussion/Conclusion | See comments about [above?] | No response is needed |
| Peer Reviewer 1 | Discussion/Conclusion | The Discussion is concise and well-written. The authors correctly identify gaps in our knowledge and areas in need of further study. | Thank you. |
| Peer Reviewer 2 | Discussion/Conclusion | The evidence, to the extent it was aligned with the authors search strategy was fairly presented but the limitations and scarcity of research findings did make it difficult to product any clear-cut recommendations for practice. | Thank you. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| Peer Reviewer 2 | Discussion/Conclusion | Given the poor evidence base, I think the authors could have described the design for further studies in a lot more detail. | Thank you. We have expanded and focused the discussion of future research needs. |
| Peer Reviewer 2 | Discussion/Conclusion | (page 46) The criteria foreshadowed by Dalbeth were presented at the recent ACR ASM and will be published during 2015. This will tend to render the report very quickly irrelevant. There is also a publication regarding criteria performance in early vs established disease that recently appeared online (Taylor WJ, Fransen J, Dalbeth N, et al. Performance of classification criteria for gout in early and established disease. Ann Rheum Dis 2014; doi:10.1136/annrheumdis-2014-206364) that is relevant to the review. | We have obtained this article and have now included it in the Results. We address the issue of new classification criteria in the introduction and discussion. |
| TEP 4 | Discussion/Conclusion | The first line of the conclusion in the Abstract (page v) points out 'the Diagnostic Rule', which I would avoid, as the algorithm is not widely accepted in the field given the controversy associated with the potential lack of face validity in some of the included components (e.g., risk factors such as hypertension or cardiovascular comorbidities, as opposed to a part of the disease features). The first conclusion line in (PDF) page 30 would be appropriate and safe. | Thank you. We have revised the Abstract, mentioning the two most recent algorithms but noting that they need much broader validation. |
| TEP 4 | Discussion/Conclusion | I think that the implications of the major findings and the limitations of the review/studies are described adequately. | Thank you. |
| TEP 4 | Discussion/Conclusion | As stated above, the investigators may want to discuss the aforementioned ACR-EULAR criteria for gout that was presented a few weeks earlier this month. | Yes, we now include the 2014 criteria in the Discussion. |
| TEP 4 | Discussion/Conclusion | I find the future research section clear and easily translatable into new research. | Thank you. |
| TEP 5 | Discussion/Conclusion | Please see general comments and comments regarding the introduction. Discussion should circle back to these points. | We believe we have now addressed the comments you raised about the Introduction in the Discussion. |

| | | | |
|---|---|---|---|
| TEP 5 | Discussion/Conclusion | It would be helpful to summarize available information about the patients in the reviewed studies who did NOT have gout. What do/don't the data tell us about the assertion that a specific diagnosis of gout (versus the exclusion of other conditions) is truly the clinical question of interest. Under what circumstances is it important to be certain that joint inflammation is caused by uric acid crystals vs. another kind of crystal vs. anything else? | We have now added information on patients who did not have gout and how the tests behaved in those patients, when reported by their actual diagnosis. We also now address the issue of differential diagnosis in the Discussion, summarizing the findings of several studies that developed and tested new lab assays to differentially diagnose gout, septic arthritis, and CPPD. |
| TEP 5 | Discussion/Conclusion | (PDF Page 83) please say something about the need to study the incremental value of US and DECT above and beyond history and physical exam, which is always available. Ideally the threshold for incorporating imaging should be that it changes patient management, so ideally that would be studied as well. | We have now revised our summaries of the very small number of studies that mention incremental value, and we now address this point in the last paragraph of our discussion of suggestions for future research: “Finally, studies are needed that assess the incremental value of US and DECT imaging over the use of a clinical diagnostic algorithm or even MSU analysis alone. One study...” |
| Peer Reviewer 3 | Discussion/Conclusion | I did not have time to review the discussion/conclusions | No response needed |
| TEP 6 | Discussion/Conclusion | The conclusions appear to be supported by the data presented. However, again, with all conclusions being of low or insufficient evidence, again the utility of the conclusions and application to clinical practice, is questionable. | Thank you. |
| TEP 6 | Discussion/Conclusion | Good review of the recent guidelines and recommendations published regarding the diagnosis of gout. | Thank you. |
| TEP 6 | Discussion/Conclusion | (page 47, line 37) 3) The authors conclude that for patients with first inflammatory monoarticular attack due to gout, DECT may not be sensitive. Where is the data/literature supporting this conclusion? | We have noted in the report that the study by Bongartz, 2014 suggested sensitivity might be low in those with early gout, based on their findings. We also highlight that it was only one study. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| TEP 6 | Discussion/Conclusion | (page 47, line 45) 4) Authors state that ultrasound sensitivity and specificity were typically high but report sensitivities down to 38%. Please address. | We verified that for the DCS alone, Lai et al did report a low sensitivity. Therefore, we have revised the conclusion regarding ultrasound, as follows: "Sensitivity and specificity are **generally** good in patients with suspected gout; **however the sensitivity may be lower in patients with early disease,**" although we wonder if this is an outlier and we comment on it in the Discussion. |
| TEP 6 | Discussion/Conclusion | Good summary of findings and strength of evidence in Table 6. | Thank you. |
| Peer Reviewer 4 | Discussion/Conclusion | (ES-16 thru ES-21 and 44-50) A few points would benefit from further clarification. The DECT and US literature needs to be interpreted cautiously as CPP-related deposition may not have been adequately represented in the comparator groups. As well, sufficient urate deposition burden must be present before detectable on DECT. For US, the findings are highly operator-dependent. Finally, there are issues surrounding the identification of asymptomatic hyperuricemia with these imaging methods that is of unclear clinical significance since the prognosis of asymptomatic hyperuricemia has not been fully elucidated. | We have added data regarding differential diagnoses (e.g., CPPD or another inflammatory joint disease) to our descriptions of the imaging studies, when these data were reported. We also now address the issue of differential diagnosis in the Discussion, summarizing the findings of several studies that developed and tested new lab assays to differentially diagnose gout, septic arthritis, and CPPD. |
| Peer Reviewer 4 | Discussion/Conclusion | (ES-17 and 45) Response to colchicine has not been assessed in other rheumatic diseases, and therefore cannot be considered specific for gout necessarily. Differentiating gout-related tophi from other nodules isn't an issue for differentiating from other crystal arthritides (since they don't form nodules), but rather the difficulty is in differentiating from rheumatoid nodules in RA. | We now address the concern regarding the lack of specificity of the response to colchicine in the Discussion. We did not address the second part of the reviewer's question, for two reasons: differentiating tophaceous gout from rheumatoid nodules was never mentioned in any of the studies or reviews we identified, and when we spoke with the general internist and rheumatologist on the research team, we were fairly assured that physicians in primary/urgent care don't see undiagnosed tophaceous gout in those settings. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| Peer Reviewer 4 | Discussion/Conclusion | (Page ES-21 and 50) There wasn't a specific future research section, but there is a 'research gap' section. This section would benefit from explicit mention of CPP-related arthritis, one of the most relevant conditions with respect to misdiagnosis. It may be worthwhile to explicitly state that one of the impacts of misdiagnosing gout as something else is the longer term sequelae of untreated gout (i.e., unabated hyperuricemia). | We substantially expanded the Research Gaps section and discussed the issues related to misdiagnosis and differential diagnosis of gout and the other inflammatory joint diseases, at least the ones that were considered in studies we identified for the report. |
| Peer Reviewer 5 | Discussion/Conclusion | Study limitations are clearly recognized and listed. | Thank you. |
| Peer Reviewer 5 | Discussion/Conclusion | (Page ES-20, line 23 and 49, line 27) I don't agree with the following statement in the discussion. The lack of stratification could affect both, not one more than the other.<br>"The lack of stratification by duration of condition would likely affect the positive and negative predictive value of imaging techniques more than it would affect diagnostic tests based on clinical signs and symptoms, but not necessarily, as one criterion in the latter is almost always the presence of typhus.[tophi]" | Thank you. We have now revised this statement to the following: “The lack of stratification by duration of condition affects the sensitivity and specificity of both clinical diagnostic algorithms and imaging techniques.” |
| Peer Reviewer 5 | Discussion/Conclusion | (Page 49, line 34) I agree that the time since the onset of current flare may impact both aspects as mentioned by the authors. I think it might also impact the imaging, since some imaging criteria also depends on the presence of crystals.<br>"The time since onset of the current flare definitely affects the presence of crystals as well as clinical signs and symptoms." | Thank you. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

| | | | |
|---|---|---|---|
| Peer Reviewer 5 | Discussion/Conclusion | (Page 49) It is unclear whether the authors consider use of synovial fluid urate crystals as the gold standard a good thing or a bad thing. They discuss this in a paragraph, but it's a bit unclear. Truly, like many chronic diseases, the disease construct can not be seen, and defined. However, over time synovial fluid MSU crystals have emerged as a specific construct which is difficult to argue against. Feasibility is always an issue and current clinical practice makes it difficult to become a common standard, as authors point out. However, when present, they are difficult to argue against. The authors should discuss crystals vs. clinical criteria issue vs. individual judgement a bit more in this paragraph. | We have now completely revised the discussion of use of MSU. We summarize the findings of several original studies and systematic reviews showing the variability in synovial fluid aspiration and MSU analysis by institution and by individual practitioners as well as factors that affect the likelihood of finding crystals in the fluid. These findings tend to support not undertaking MSU analysis in the primary care setting . |
| Peer Reviewer 5 | Discussion/Conclusion | (ES-17, line 20 and 46, line 14) At the beginning of the discussion authors state "This Diagnostic Rule has been shown to perform better than the ACR criteria in comparable populations.29". This reference is not about the comparison of ACR criteria against the diagnostic rule, it's only about the suboptimal performance of ACR criteria. Please either add the correct reference or make a correction here. | Thank you. We rechecked this reference and corrected the text. |
| TEP 1 | Clarity and Usability | Nicely organized. In my opinion the Conclusions should be (as I think they are) that there is actually little new help for practice decisions. | Thank you. |
| TEP 2 | Clarity and Usability | See Discussion/Conclusion above. | No response needed. |
| TEP 3 | Clarity and Usability | The document is clear and easy to follow. | Thank you. |
| Peer Reviewer 1 | Clarity and Usability | The review is well structured and organized. | Thank you. |

| | | | |
|---|---|---|---|
| Peer Reviewer 2 | Clarity and Usability | The report is clear and well laid out but the evidence base greatly limits any practice decisions. In-text citations and references were not well aligned at times, and need to be checked carefully. Possibly the only usable conclusion is the need for more research, but that is hardly novel. | We have checked the references and fixed any that are mismatched. |
| TEP 4 | Clarity and Usability | The report is reasonably structured and organized, and the main points are presented well. The conclusions are more about the large gaps in relevant knowledge that the authors are seeking, as well as the future research agenda; thus, it remains somewhat unclear to me how much this report would inform policy and/or practice decisions. | Thank you. |
| TEP 5 | Clarity and Usability | Please see the "general comments" section. | No reponse needed. |
| Peer Reviewer 3 | Clarity and Usability | I found this review difficult to follow. Generally I think the structure and readability of the report could be improved. | We have substantially revised the structure of the report and hope that its readability has improved. |
| Peer Reviewer 3 | Clarity and Usability | (page ES-1 thru ES-25) The executive summary is much too long – I thought I had read the whole report after reading this so was confused when this then lead into introduction. In fact having looked quickly at this and what I assume is the main document there does not appear to be any difference in the introduction or methods; it is only the synthesis that differs. A much more concise summary would be more useful. I would also not expect to find references in an executive summary. | Because we know that many readers will read only the executive summary, we try to make it a stand-alone document. It is customary to include references in AHRQ executive summaries. |
| Peer Reviewer 4 | Clarity and Usability | The document is relatively clear. Given the lack of sufficient data, policy and practice decisions are unlikely to be altered. | Thank you. |
| Peer Reviewer 5 | Clarity and Usability | I think most points are presented clearly. | Thank you. |
| Peer Reviewer 5 | Clarity and Usability | The conclusions have implications for decisions. | Thank you. |

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*

AHRQ
Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • www.ahrq.gov

Effective Health Care Program

*Source: https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=2188*
*Published Online: February 24, 2016*