Author’s response to reviews

Title: Diagnostic accuracy of cardiac MRI, FDG-PET, and myocardial biopsy for the diagnosis of cardiac sarcoidosis: A protocol for a systematic review and meta-analysis

Authors:

Dominik Roth (dominik.roth@cms.ox.ac.uk)

Nikolaos Kadoglou (nikoskad@yahoo.com)

Mariska Leeflang (m.m.leeflang@amc.nl)

Rene Spijker (r.spijker-2@umcutrecht.nl)

Harald Herkner (harald.herkner@meduniwien.ac.at)

Marialena Trivella (marialena.trivella@csm.ox.ac.uk)

Version: 1 Date: 02 Mar 2020

Author’s response to reviews:

Reviewer reports:

Reviewer #1: The manuscript by Roth et al is the protocol for systematic review and metaanalysis of diagnostic accuracy for cardiac sarcoidosis. This is topic of intense conversation in this field, as clinical pathways have evolved with the advent of advanced cardiac imaging techniques in the absence of firm data on which to rest conclusions about diagnostic probability.

A few concerns should be address:

R1.1 While the Japanese Ministry of Health and Welfare criteria were formerly the most widely used, the 2014 Heart Rhythm Society guidelines (Birnie et al Heart Rhythm 2014) have been far more central to the contemporary conversation and have supplanted the Japan criteria, in part because of greater use of advanced imaging. This should be discussed accordingly.

>>>Thank you for addressing this important point. We are aware of the multitude of different diagnostic guidelines for cardiac sarcoidosis, including different iterations of the Japanese criteria, as well as the Heart Rhythm Society consensus statement, among others. Whilst most of those different recommendations follow a similar general approach, relevant differences, especially regarding the use of imaging modalities as part of the guidelines, exist.

Whilst we do agree that the 2014 HRS expert consensus statement has gained increased interest over the last years, this statement by definition focuses mainly on diagnosis and treatment of arrhythmias associated with CS. Most published studies within the last decade still seem to use the Japanese Criteria, especially in their most recent iteration as published in 2016 by the Japanese Circulation Society. This might partly be explained by the fairly high prevalence of CS in Japan, and hence the major role of Japanese scientists in global research on CS. We therefore decided to focus our current project on the Japanese criteria, but are very open to expand on this by studying other possible reference standards in the future (see also comments below).
R1.2 Similar reference and commentary should be made about the new EANM, EACVI and ASNC going procedural position statement on imaging in cardiac sarcoidosis. The theme here is that anchoring on the biopsy as the elusive gold standard confounds traditional approaches to test performance -- playing this up in the "why" behind your research will be essential to make this more impactful.

>>>Thank you for pointing at this valuable position statement. We wholeheartedly agree with the paper’s dismissal of biopsy as a useful reference standard, and we are very happy to cite the statement as a valuable source of background information. It has to be noted that this paper (in contrast to the recommendations discussed above) aims to “standardize imaging for cardiac sarcoidosis”, instead of providing a definition for diagnosis. It hence won’t serve as a possible reference standard. We added the following reference at the end of the ‘Index tests’ section:
“A recent joint procedural position statement by the European Association of Nuclear Medicine, the European Association of Cardiovascular Imaging, and the American Society of Nuclear Cardiology aims to standardise imaging for cardiac sarcoidosis(30).”

R1.3 The authors methodology assumes known or suspected cardiac sarcoid and proposes a rational flow diagram; yet most cases encountered are when a patient has a clinical event (syncope, heart block, abnormal imaging finding obtained for another reason) that the "arrows" in the flow diagram can be bidirectional and the probability for posterior testing can be greatly influenced by prior results or clinical suspicion. This will need to be carefully discussed when the final product is available for review.

>>>We totally agree. Diagnostic pathways in CS can be rather varying and challenging in general. To reflect this, we changed the introduction to the pathway to “[Diagnosis commonly follows a multi-step approach] and could be rather variable and challenging. Patients might enter diagnostic workup either due to known extra-cardiac sarcoidosis, without any cardiac symptoms, or due to a clinical event of possible cardiac origin (e.g. syncope, arrhythmia), without previous suspicion of CS. One possible clinical pathway might be as follows:” We also changed the wording for the provided figure to “Possible clinical pathway[...]”

R1.4 There are multiple typos, principally in the first 5 pages whereby two separate words have no space between them e.g. "found predominantly" and "epicardial areas" on page 5. Please correct.

>>>We are terribly sorry for this mistake, which was caused by a glitch in MS Word. We carefully went through the whole manuscript and hope we were able to get rid of all the errors.

Reviewer #2: The authors have written a protocol for a systematic review and meta-analysis of MRI, PET, and biopsy in the diagnosis of cardiac sarcoidosis - a rare but potentially fatal disease. Please note that review is focused on the statistical aspects as I am not a clinician.

This looks like a very interesting review from a methodological point of view, and I hope the authors obtain the data for all their planned analyses. However, I do have some comments which I hope will improve the manuscript.

R2.1. The Reference Standard does not explicitly say that the Japan Criteria are the reference standard,
and I only know that the Japan Criteria are the reference standard as it is stated later on in the manuscript. Also, do the Japan Criteria include the index tests as components? The description suggests imaging but doesn't say whether it is MRI or PET. Also, the clinical pathway flowchart suggests that there would be at least two reference standards - long term follow-up for the patients with normal screening test results and more intensive testing for those with abnormal screening test results. Could the authors please spell out exactly what the reference standard is for all patients and does it include MRI or PET imaging? The authors will know that including the index tests as components in the reference standard is a cause of incorporation bias.

>>>Thank you very much for pointing this out. For clarification, we added the following paragraph at the end of the section ‘Reference standard’:
“For the purpose of this review, the Japanese criteria, as published from 2006 onwards, will serve as a reference standard. Methods to deal with different versions of this standard, as well as the fact that some of the index tests might be included in the reference standard, are outlined in the methods section.”
Furthermore, we added the following to the ‘Sensitivity analysis’-section:
“We will also assess the impact of different version of the reference standard used, including versions modified not to include any of the index tests, by performing separate analysis for each version.”

R2.2 There are the criticisms against the Japan Criteria, and this is important. I would not know how to interpret sensitivity and specificity from a study where the reference standard may not be fit for purpose. I can see that this systematic review is interesting from a methodological point of view, but a clinician might question its usefulness. The authors will know that an accurate index test will not have high estimates of sensitivity and specificity if the reference standard is inaccurate.

>>>We fully agree on the criticisms against the Japan Criteria, and as a matter of fact, this, together with the known inadequacies of endomyocardial biopsy, was a main reason for starting this project. When planning our review, we discussed this issue among both statisticians and clinicians. We decided to choose the Japanese Criteria as a reference standard, because in clinical practice, to our knowledge, they are used just as that. To deal with the possible problems of the reference standard, we planned a multi-step approach, including a Bayesian meta-analysis. We therefore believe that our findings would be helpful to clinicians, despite the known limitations. We acknowledge however, that this might just be a first step in exploring the complex relationship of reference standards and index tests within the field of cardiac sarcoidosis. Based on data and experience gained by this review, we plan to conduct further projects, including different takes on what could be considered a reference standard for CS.

R2.3. Are there two patient groups? Asymptomatic and symptomatic? Pre-existing extra-cardiac sarcoidosis and "established isolated cardiac sarcoidosis" [page 9 lines 28-29]? If they are including patients with established cardiac sarcoidosis then the sensitivity and specificity estimates will not be clinically applicable - to get a clinically applicable estimate you need a clinically applicable sample, i.e. people in need of a diagnosis, not those who already have one. I would have thought that difference patient groups could require separate analyses.

>>>We absolutely agree that findings based on patients with a known diagnosis might alter the results as compared to those only including patients without an already established diagnosis. Based on preliminary search results and our own experience, studies including patients with an already established diagnosis of CS however form a considerable portion of studies published in the field. One
possible explanation for this, among other reasons, might be the fact that even in clinical practice patients, in whom diagnosis for CS has already been established (e.g. by biopsy, or using the Japanese criteria without any of the index tests) still undergo CMR or PET to confirm previous findings. We therefore chose to include those studies as well, to represent a full picture of the evidence that has been published. To deal with the aforementioned problems, we will consider those differences in population as a potential covariate in our analysis. To reflect this, we added the following to the ‘Investigations of heterogeneity’-section:

“[To explore heterogeneity, patient demographics (for example age, sex, weight) and clinical setting (in-hospital vs. out-hospital)], as well as differences in population (pre-established vs. suggested CS) [will be considered as potential covariates.]”

R2.3. Types of studies - the authors haven't said that they will exclude case-control diagnostic studies (also sometimes called "two-gate" designs). These studies are known for their tendency to overestimate sensitivity and specificity (and I believe that one of the co-authors has written a paper on this very topic). Case-control diagnostic studies recruit patients known to have the disease already, so my comment above about using data from patients with an established diagnosis of cardiac sarcoidosis applies here.

>>>We fully agree with your assessment of case-control studies. Because of the known problems of case-control studies, the use of such a design has been included into the QUADAS-2 tool as a major source of bias. Hence, the problem will partly be addressed by our planned sensitivity analysis based on QUADAS-2 domains, as well as by the methods for dealing with studies on patients with already known diagnosis, as outlined above. Having said this we would like to highlight that we will not use the ‘control’ arm of case-control studies, because those patients would not meet our inclusion criteria. Based on preliminary search results, the number of case-control studies will be very, very small. If there is, however, a relevant number of such studies, we will perform sensitivity analysis to assess the impact of excluding case-control studies.

To reflect this, we added an explanatory sentence in the ‘types of studies’ section

“For case-control studies we will only use available data from the ‘cases’ arm of the study, since the ‘control’ arm would not fulfil the participant inclusion criteria.”

We also added the following to the ‘Sensitivity analysis’-section

“[We will follow a standard approach by assessing the impact of excluding studies based on QUADAS2-domains(44)], as well as explicitly excluding case-control studies, if there is a relevant number of such studies.”

R2.4. I find the analysis plan ambiguous (please note that I am a native speaker of English). What do they mean by, "Part 1: A direct comparison of both CMR and 18F-FDG (separately) as index tests to the reference standard Japan Criteria." To me, the phrase "direct comparison" in this context means using data from studies with a head-to-head comparison of CMR and PET, i.e. studies with results from the same patients from the Japan Criteria, CMR, and PET in one statistical model. So why, "separately"? If they are analysing the data separately, how is this different to the next planned analysis, "An indirect comparison of both CMR and 18F-FDG as index tests to the reference standard Japan Criteria"? Re the analyses with biopsy, is this MRI and biopsy versus PET and biopsy, or MRI and biopsy versus PET, or MRI versus PET and biopsy? There is more detail for the analyses a few paragraphs later, but the whole description could be clearer.

>>>Thank you for pointing out the ambiguity of our wording here. Your interpretation of direct comparison meaning including only studies which provide data on both index tests is correct. This is
different to the indirect comparisons outlined in part 2 and 3, which will also include studies which do not provide data for all the index tests, but indirect comparison between results will still be possible, using meta-regression methods as outlined.

Regarding the word “(separately)”, we intended to express that CMR and 18F-FDG and CMR will be treated as separate index tests, and separate analysis will be performed, as opposed to a combination of index tests (i.e. combined accuracy of performing both CMR and 18F-FDG). The same is true for the part with biopsy, where, again, we will only consider separate tests, not combination of tests. For clarification, we modified the sentence on Part 1 to read as follows:

“Part 1: Direct comparisons of both CMR and 18F-FDG as index tests to the reference standard Japan Criteria”

We also added the following sentence at the end of the list:

“For all parts, only separate index tests, but no combination of index-tests (e.g. combined accuracy of performing both CMR and 18F-FDG) will be considered.”

R2.5. Re Part 4 (the Bayesian meta-analysis with informative priors), this looks interesting, but I would like the authors to present the prior and posterior distributions so that the influence of the prior over the posterior can be judged. Prior distributions for the variance and correlation parameters can be particularly problematic, so please include these and not just the priors and posteriors for the sensitivity and specificity parameters.

>>> We are grateful for this valuable suggestion and added the following accordingly:

“We will present prior and posterior distributions of sensitivity and specificity, as well as variance and correlation parameters.”

R2.6. How are they going to assess the adequacy of their statistical models? I can't find much on this.

>>> Thank you for this important comment. Most of the methods we plan to use make only very little unusual assumptions on the characteristics of the data. Especially for the bivariate meta-regression of sensitivity and specificity, the only specific assumption is that thresholds for individual index tests will be uniform among studies, which will be met in our case. Meta-regression model fit will be assessed using standard measures such as log likelihood, AIC and BIC, as well as by the Deviance Information Criterion (DIC), which might be more appropriate to complex random-effects models.

R2.7. How are they going to edit the QUADAS-2 tool for their review? E.g. instead of "Was there an appropriate interval between index test(s) and reference standard?", something like, "Was the interval between index and reference standard less than 1 month?"

>>> We found that the tool questions were adequately clear for our project. Should the need arise to modify any signalling questions, we will present a full list of the questions and any modifications (with justification), when publishing the results of our review, as outlined in the first paragraph of the ‘Assessment of methodological quality’-section.