## PEER REVIEW HISTORY



## ARTICLE DETAILS

| | |
|---|---|
| **TITLE (PROVISIONAL)** | Cost-effectiveness of Anatomic and Functional Test Strategies for Stable Chest Pain: Public Health Perspective from a Middle Income Country |
| **AUTHORS** | Bertoldi, Eduardo; Stella, Steffen; Rohde, Luis Eduardo; Polanczyk, Carisi |

## VERSION 1 - REVIEW

| | |
|---|---|
| **REVIEWER** | Khaled Alfakih<br>Lewisham University Hospital and King's College Hospital. London, UK. |
| **REVIEW RETURNED** | 18-Jul-2016 |

| | |
|---|---|
| **GENERAL COMMENTS** | This is a very interesting and very good piece of work and is timely for a UK as well as a world wide audience. The UK NICE is about to publish a guideline on chest pain where they recommend that all patients with chest pain should be investigated with a CT coronary angiogram in the first instance which would agree with the conclusions of these authors. World wide: this is just as relevent as the number of patients presenting with chest pain continues to increase and hence the importance of finding effective ways of timely investigation, without excessive cost to health services.<br><br>I am not a health economist. I am a clinician who happens to think that this is a very important subject as it impacts on a large number of patients. Hence i think this is an important piece of work and should be accepted.<br><br>I only have one important point to make: The authors take a hypothetical patient with pre test probability of between 20% and 70%. They should note that in Promise trial, the pre test probability was 53% overall, and the actual prevlance of disease was 10.7% in one arm and 11.7% in the other arm. So how does this low prevlance of CAD and this huge overestimation of risk by the risk scores, affect their model? |

| | |
|---|---|
| **REVIEWER** | Bart Ferket, MD, PhD<br>Icahn School of Medicince at Mount Sinai, NY, NY, USA |
| **REVIEW RETURNED** | 14-Aug-2016 |

| | |
|---|---|
| **GENERAL COMMENTS** | This is a decision-analysis of different functional and anatomical test strategies for diagnosing stable coronary artery disease. The authors chose not to use the conventionally used cost per QALY as outcome, which may be a defendable choice in this case as differences in QALYs are expected to be negligible. The manuscript |

| | |
|---|---|
| | is clearly written and reads very well.<br><br>Major comments:<br>1) The decision model as outlined in Figure 1 seems over-simplified for the medical problem at hand. Were costs of readmissions due to myocardial infarctions (MIs), treatment of MI, and post-MI costs included as well to take the economic consequences of false negative test results into account?<br><br>2) Generally it is much easier to primarily model the underlying “true” disease rate with test results conditional on disease status. It seems that the authors chose to first model test result (positive/negative) and subsequently disease status (present/absent). Performing sensitivity analyses of test accuracy then become cumbersome. If the latter is true, then I believe the authors should show how the probabilities in the tree were calculated for each strategy in an appendix to show the face validity of the model. A picture of the TreeAge tree as an appendix would be helpful as well.<br><br>Minor comments:<br>1) A PSA was done, but it is not clear for which risk scenario the results were presented. I would rather see the PSA per risk scenario. Because a WTP threshold is unknown, the diagonal line in Figure 3 seems not useful. In this case an acceptability curve plot could be useful, but I would suggest to include all strategies in such a plot and not only CTA and echo.<br><br>2) In Table 3 the average costs and correct diagnoses per patient with their 95% CIs should be included. Also other outcomes such as FNs, death, invasive CA and negative invasive CA should have confidence intervals.<br><br>3) I would appreciate to see figures presenting results of the one-way sensitivity analyses as an illustration of the text in the Results section.<br><br>4) Cumulative radiation dose per strategy should ideally be considered as well.<br><br>Thank you for providing me the opportunity to review your manuscript. |

**VERSION 1 – AUTHOR RESPONSE**

Response to reviewer #1:
Dear Dr. Alfakih,
Thank you for reviewing our manuscript and for your relevant comments. In our opinion, clinicians are a crucial audience for this kind of research, and we will be immensely pleased if our manuscript is able to provide information that can be useful in clinical practice. Please find our response to your question below.

The authors take a hypothetical patient with pre test probability of between 20% and 70%. They should note that in Promise trial, the pre test probability was 53% overall, and the actual prevlance of disease was 10.7% in one arm and 11.7% in the other arm. So how does this low prevlance of CAD and this huge overestimation of risk by the risk scores, affect their model?

Reply: In the model, the pretest probability inputted will be directly used for determining risk in our

hypothetical cohort. There is no need to estimate risk (using a score) because we are arbitrarily defining our patients' risk in the model, and so it is not affected by score performance. An imperfect risk-prediction score might influence the applicability of our model to real-world populations, since our recommendations are stratified by risk profile; in that sense, continuous improvement of risk-prediction scores using data from contemporary populations (such as the one from Promise trial) may further improve our ability to select strategies for our patients. Nonetheless, we must remember that, as is true in modelling, risk scores will always be a simplification of reality, and can never be perfect, but this should not discourage us from using them. An imperfect decision tool is better than no tool at all.

---

Response to reviewer #2:
Dear Dr. Ferket,
We appreciate the time and effort put into reviewing our manuscript and thank you for your very relevant contributions. Issues raised by the reviewer are addressed below.

Major comments:

1) The decision model as outlined in Figure 1 seems over-simplified for the medical problem at hand. Were costs of readmissions due to myocardial infarctions (MIs), treatment of MI, and post-MI costs included as well to take the economic consequences of false negative test results into account?

Our analysis is focused on the short-term performance of the diagnostic tests at hand, and aims to provide a thorough and precise description of immediate implications of selecting one testing strategy over another. This is the reason why we felt it relevant to report results in a cost-per-diagnosis (as opposed to cost-per-QALY) format. Truly, it is reasonable to assume that the diagnosis established by these strategies (especially a false-negative diagnosis) will impact future probabilities for MI, revascularization and death; however, the magnitude of this influence is not easily measured, and to include it in the model requires some major assumptions and adds uncertainty to the model. We have performed such analysis in a separate, long-term model, applying a lifetime horizon timeline, and reports results in a cost-per-QALY format. The core results (in terms of strategy selection) are similar to the ones we describe in the current manuscript. Considering that the impact of diagnostic test on long-term prognosis is questionable and difficult to fully capture in model analysis, we opted to present the result focusing on short-term results, per case correctly diagnosed.

2) Generally it is much easier to primarily model the underlying "true" disease rate with test results conditional on disease status. It seems that the authors chose to first model test result (positive/negative) and subsequently disease status (present/absent). Performing sensitivity analyses of test accuracy then become cumbersome. If the latter is true, then I believe the authors should show how the probabilities in the tree were calculated for each strategy in an appendix to show the face validity of the model. A picture of the TreeAge tree as an appendix would be helpful as well.

We felt that a full representation of our model structure would be excessively complex and possibly overwhelming for those not accustomed to economic modelling. For that reason, we chose to schematically represent the inner workings of the model in a concise "concept" figure. In the actual model, each strategy starts with a definition of "true" disease state, because, as the reviewer correctly states, this makes definition of downstream probabilities and sensitivity analysis much more straightforward. We understand that readers who are more familiar with economic analysis may find our figure to be over-simplified, and have now submitted a more thorough depiction of our model as

supplementary material.

---

Minor comments:

1) A PSA was done, but it is not clear for which risk scenario the results were presented. I would rather see the PSA per risk scenario. Because a WTP threshold is unknown, the diagonal line in Figure 3 seems not useful. In this case an acceptability curve plot could be useful, but I would suggest to include all strategies in such a plot and not only CTA and echo.

As suggested by the reviewer figure 3 now shows PSA per risk scenario. The diagonal line has been suppressed. Thank you for the correction.

2) In Table 3 the average costs and correct diagnoses per patient with their 95% CIs should be included. Also other outcomes such as FNs, death, invasive CA and negative invasive CA should have confidence intervals.

We have included a column with average costs per patient. Correct diagnosis per patient is represented by the overall accuracy, already on the table. Regarding 95% CIs for all the outcomes, we feel that their inclusion would leave the table looking rather cluttered, and suggest leaving them out in favor of conciseness. If the reviewer feels their inclusion is essential for submission, we will conform, so please let us know.

3) I would appreciate to see figures presenting results of the one-way sensitivity analyses as an illustration of the text in the Results section.

Properly showing results of one-way sensitivity analyses with charts would require the use of a large number of figures, and might undermine the paper's readability. We feel that figure 3 is a better, more concise depiction of the uncertainty surrounding the comparison of ECHO- and CTA-based strategies.

4) Cumulative radiation dose per strategy should ideally be considered as well.

We have included data on radiation exposure in the manuscript text, and also in a table (to be provided as supplementary material). Thank you for the suggestion.

**VERSION 2 – REVIEW**

| | |
|---|---|
| **REVIEWER** | Khaled Alfakih<br>Lewisham University Hospital<br>King's College Hospital,<br>London, UK |
| **REVIEW RETURNED** | 13-Dec-2016 |

| | |
|---|---|
| **GENERAL COMMENTS** | I have nothing to add to my previous review and I accept their response to my comment.<br><br>This is a health economics paper and I guess you did show it to a statistician. OR you can simply accept that the health economic modelling they did is accurate. In fact the fact that they conclude that CTCA and stress echo were the most cost effective tests is very believable as they are both low cost tests. This work is topical considering that NICE have just recommended that all patients with chest pain in the UK should be investigated with CTCA. So NICE agree with them. |

| | |
|---|---|
| **REVIEWER** | Bart Ferket<br>Icahn School of Medicine at Mt Sinai, NYC, NY, USA |
| **REVIEW RETURNED** | 21-Nov-2016 |

| | |
|---|---|
| **GENERAL COMMENTS** | The authors have addressed my concerns satisfactorily. |