PEER REVIEW HISTORY

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

ARTICLE DETAILS

TITLE (PROVISIONAL)	Medical consumption compared for TIMI and HEART score chest pain patients at the emergency department: a retrospective cost analysis
AUTHORS	Nieuwets, Astrid; Poldervaart, Judith; Reitsma, Johannes; Buitendijk, Susanne; Six, Alfred; Backus, Barbra; Hoes, Arno; Doevendans, Pieter

VERSION 1 - REVIEW

REVIEWER	Edd Carlton North Bristol NHS Trust, UK I have received support from Abbott Diagnostics for related research and carried out research with Randox under a collaborative agreement.
REVIEW RETURNED	18-Dec-2015

GENERAL COMMENTS	Many thanks for asking me to review this interesting and well written paper. The authors have addressed an important issue in the costs of further testing in low-risk patients with suspected acute coronary syndrome. However, I have some comments which need addressing before this paper warrants publication. The authors have made some assumptions that could be questioned, these may explain the improved performance of the HEART score over TIMI. The first, which requires further explanation is the safety of each approach. Whilst the authors use cumulative frequency of MACE to assume safety this may not hold true and some measures of diagnostic accuracy (sensitivity/NPV) should be reported (I am aware that this has been published previously). When looking at the results it is evident that TIMI 0 missed no MACE, however the HEART score <3 missed 3. There is no mention of the potential medicolegal costs of missed MACE and this should at least be included in the discussion. There should also be mention of the missed MACE threshold that is acceptable to clinicians-probably around 1% (Than et al. What is an acceptable risk of major adverse cardiac event in chest pain patients soon after discharge from the Emergency Department?: a clinical survey. International Journal of Cardiology. 2012). The HEART score may not achieve this threshold in this population and this needs to be discussed. Following on from this issue, secondly, I am unclear as to why a MACE threshold of <5% was chosen for low risk patients. There is no evidence base for this assumption and it may be open to criticism as it appears to favour the HEART score over TIMI. This assumption means that a TIMI score of 0/1 has not been included in the analysis and I would suggest removing it. A TIMI score of 0/1 has recently been shown to have favourable diagnostic accuracy in the era of high-sensitvity troponin (Carlton et al. Identifying Patients Suitable

	for Discharge After a Single Presentation High-Sensitivity Troponin Result- A Comparison of Five Established Risk Scores and Two High-Sensitivity Assays. Annals of Emergency Medicine 2015) in comparison to HEART and results in a higher proportion of patients being identified as potentially suitable for discharge (Cullen et al. Validation of high-sensitivity troponin I in a 2-hour diagnostic strategy to assess 30-day outcomes in emergency department patients with possible acute coronary syndrome. Journal of the American College of Cardiology. 2013) (yes-I am aware that one is my own work). Both references are relevant and should be included as the references pertaining to a TIMI score of 0 (Page 5, line 46) are outdated and a more thorough discussion needed as to why a TIMI score of 0/1 was not analysed. Alternatively you could analyse the performance of a TIMI cut-off of 1 in addition to the current costing analysis-this would be of great interest. Thirdly, the issues regarding type (e.g. high-sensitivity vs contemporary) and time of troponin testing used should be discussed in more detail. The paper by Thokala et al. (Cost-effectiveness of presentation versus delayed troponin testing for acute myocardial infarction. Heart. 2012) may assist with this. A minor point is the use of the word "conventional" when describing troponin assays used; I think this is better understood in the literature as "contemporary." Finally, I think it would be of interest to include more discussion around the TIMI and HEART scores and their subjective elements. For example how does the history element of the HEART score relate to overall gestalt and does this have a reasonable inter-rater agreement? I am also a little uncertain from the methods as to how the subjective elements of TIMI recent; recent severe angina and known CAD were calculated-these are often open to interpretation. I hope you find these comments useful.

REVIEWER	Bruce Ling Stanford University, USA
REVIEW RETURNED	03-Jan-2016

GENERAL COMMENTS	This paper addressed a very interesting question of identifying low risk patients such that less resource will be utilized. However, I would like the authors to address what would be the impact on the identification of intermediate and high risk patients when switching from one risk metric to the other. Comparative analysis is needed. In the end, it is also important to identify high risk patients for targeted care.

VERSION 1 – AUTHOR RESPONSE

REVIEWER 1, Dr. Carlton

• “The authors have made some assumptions that could be questioned, these may explain the improved performance of the HEART score over TIMI.”

• “[…] the safety of each approach. Whilst the authors use cumulative frequency of MACE to assume safety this may not hold true and some measures of diagnostic accuracy (sensitivity/NPV) should be reported […]. “

AUTHORS’S RESPONSE: Sensitivity and specificity are indeed commonly used measures of diagnostic accuracy. However, they are defined in a more technical sense (i.e. conditional on the true disease status), while our interest is in the downstream consequences in practice when using the test

in a particular way. These consequences are best expressed by reporting the absolute numbers of interest.

Furthermore, we chose not to calculate these diagnostic accuracy measures for our own cohort of 635 patients, as it is only a selection of a considerably larger cohort of 2,388 patients, and the sensitivity and NPV of the full cohort will be much more accurate. However, we addressed this comment partially by including the sensitivity and NPV as reported by Mahler et al. in the introduction, and the NPV of Carlton et al. to give the reader a sense of these measures reported by others.

• "When looking at the results it is evident that TIMI 0 missed no MACE, however the HEART score <3 missed 3."
AUTHORS'S RESPONSE: Indeed the cut-off of HEART score of 3, results in 4 patients with MACE, compared to 0 patients in the low risk TIMI group of TIMI=0. In the revised version of the manuscript we do highlight that when using a cut-off of 0 for TIMI, no events would have been missed. However, only a small fraction of all patients received a TIMI score of 0, meaning that the estimate of safety is still imprecise and the efficiency as a triage test will be limited at this threshold given the low number of patients then defined as low risk.

Moreover, as mentioned in the discussion, we looked into these 4 patients with MACE (1 patient with a HEART score of 2, 3 with a HEART score of 3) and found that one was a CABG scheduled before ED presentation and the other three were diagnosed with ACS at the ED immediately, thus not missed. They received elective PCIs in a later stage, indicating mild severity of disease in these patients. However, had the HEART score been used blindly, these last three patients would possible indeed have been missed. Therefore, we do advise in our discussion that clinicians to not blindly follow the HEART score, but rather use it as a tool to strengthen their decision-making.

• "There is no mention of the potential medicolegal costs of missed MACE and this should at least be included in the discussion."
AUTHORS'S RESPONSE: In the Netherlands, medicolegal costs resulting from missed diagnoses are not prevalent. However, it is a valid point to could have included these costs in the cost analysis, when our data was to be extrapolated to other countries. Therefore, we have included this as a point of remark and mention a possible underestimation of these costs in our revised discussion.

• "There should also be mention of the missed MACE threshold that is acceptable to clinicians- probably around 1% [...].The HEART score may not achieve this threshold in this population and this needs to be discussed."
AUTHORS'S RESPONSE: We agree with the reviewer that a threshold of acceptable risk should be included in the discussion. We have now addressed in our discussion the article by Than et al., as the reviewer suggested. We also mention the 2% (as suggested by Kline et al.) and what this would have meant for the number of patients that could have been discharged from the ED immediately.

• "I am unclear as to why a MACE threshold of <5% was chosen for low risk patients. There is no evidence base for this assumption and it may be open to criticism as it appears to favour the HEART score over TIMI. This assumption means that a TIMI score of 0/1 has not been included in the analysis and I would suggest removing it."
AUTHORS'S RESPONSE: We chose the cut-offs based on the previous literature by the original investigators of both the TIMI and HEART score. Furthermore, it is important to mention the difference between a 1% risk of missing MACE within all patients with a low risk for MACE (1-NPV) as compared to the 1% incidence of missing MACE in all patients (1-sensitivity). The <5% was based on the latter. However, we see this will be confusing for readers and furthermore, as the reviewer suggests, since the emergence of high-sensitivity troponin has such an impact on TIMI (as well as HEART) we have extended our discussion by the mention of high-sensitivity troponin, including both references the

reviewer suggested.

• " […] references pertaining to a TIMI score of 0 (Page 5, line 46) are outdated"
AUTHORS'S RESPONSE: We agree with the reviewer and removed most of these references.

• " […]and a more thorough discussion needed as to why a TIMI score of 0/1 was not analysed. Alternatively you could analyse the performance of a TIMI cut-off of 1 in addition to the current costing analysis-this would be of great interest."
AUTHORS'S RESPONSE: We have included an extra analysis that looks into the a more conservative approach for HEART (low-risk defined as 0-2, MACE miss-rate <1%), and a broader TIMI cut-off (low-risk defined as 0-1, MACE miss-rate: 6.4%). It should be noted, that our cohort was included in 2008, it was analysed using contemporary troponin, which might explain the high miss-rate of 6.4% in the broader TIMI 0-1 category.

• Thirdly, the issues regarding type (e.g. high-sensitivity vs contemporary) and time of troponin testing used should be discussed in more detail.
AUTHORS'S RESPONSE: As stated before, this section in our discussion has been broadened.

• A minor point is the use of the word "conventional" when describing troponin assays used; I think this is better understood in the literature as "contemporary."
AUTHORS'S RESPONSE: We have adjusted the manuscript accordingly.

• Finally, I think it would be of interest to include more discussion around the TIMI and HEART scores and their subjective elements. For example how does the history element of the HEART score relate to overall gestalt and does this have a reasonable inter-rater agreement?
AUTHORS'S RESPONSE: We agree and now address this in our limitation section in the Discussion.

• I am also a little uncertain from the methods as to how the subjective elements of TIMI recent; recent severe angina and known CAD were calculated-these are often open to interpretation.
AUTHORS'S RESPONSE: We have included the relevant information regarding the methodology of calculating the TIMI score, the original study used a syntax to calculate the TIMI score out of the original data collection at the ED. This was done without interpretation of the researchers. Both scores include subjective elements, we address this in the limitation section. Indeed this leaves room for interpretation, although we tried to use the definition of the elements as mentioned by the original investigators of the two scores.

REVIEWER 2, Dr. Ling

"This paper addressed a very interesting question of identifying low risk patients such that less resource will be utilized. However, I would like the authors to address what would be the impact on the identification of intermediate and high risk patients when switching from one risk metric to the other. Comparative analysis is needed. In the end, it is also important to identify high risk patients for targeted care."

AUTHORS'S RESPONSE: We should have stated more explicitly that the identification of high-risk patients was beyond the scope of this particular research. The main reason for this was the fact that for the TIMI score, there is no intermediate or high-risk group defined, which complicates comparison with the defined risk categories of the HEART score. Therefore, we chose not to incorporate these risk categories in our research question, because we felt this would not do right to the original TIMI categories. Another reason was that we believe the most benefit at the ED is to be gained in the low-risk group and this group has the most clear proposed policy, i.e. early discharge from the ED, whereas the intermediate and high-risk group are to be debated which exact policy should be

appropriate (for example, exercise testing, or a CT-angiography in intermediate risk patients, and a CT-angiography or coronary angiography in high-risk patients). To address the comments of the reviewer, we now have revised the discussion to include a paragraph of possible applications of the HEART or TIMI score to determine diagnostic pathways for intermediate- or high-risk patients and have stressed that these pathways might be a good topic for future research.

VERSION 2 – REVIEW

REVIEWER	Dr Edd Carlton Emergency Department, Southmead Hospital, Bristol, UK Dr Carlton has received funding from Abbott Diagnostics for research in high-sensitivity troponin assays
REVIEW RETURNED	17-Feb-2016

GENERAL COMMENTS	The authors have addressed my points from initial review. The manuscript now reads well and better informs the reader. I thank them for their efforts in improving this manuscript which is now suitable for publication. I do, however have some minor points that should be addressed as follows: Page 4, Line 77: Please change "stress testing" to read further testing (some patients will undergo anatomical tests such as CTCA/angiography) Page 4, Line 86. Change "TIMI score is used to risk stratify..." to "was developed to risk stratify" (since its use has become broader over time). Page 5, Line 104. After "both risk scores have been validated..." please add "for use in a low-risk ED population" Page 6, Line 139: "The HEART score was developed in 2007" is repetitious (cf introduction) and should be removed Page 10, Line 231: "Within the low-risk TIMI group" please clarify whether you are referring to TIMI 0 or <=1, similarly Page 10 Line 235 Please include the TIMI<=1 costs in the abstract if word limit allows, this can be followed with a statement such as "but with a MACE rate of 3%" so as to justify your support of the HEART score. Many thanks for giving me the privilege to review this interesting manuscript.

REVIEWER	Bruce Ling Stanford University, USA
REVIEW RETURNED	13-Feb-2016

GENERAL COMMENTS	I think the authors made good revision this time. However, a few points still need to be addressed to make it acceptable to be in press: 1. Please discuss why 6 week window was chosen for the following up of the ED discharge? 2. Given both scoring metrics authors discussed have not used as the rule to discharge ED patients, therefore, the author should discuss why both metrics or one of the metrics would be suitable to

	be used to manage ED discharge. Please address these above two points such that it can be in press.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1
The authors have addressed my points from initial review. The manuscript now reads well and better informs the reader. I thank them for their efforts in improving this manuscript which is now suitable for publication.
I do, however have some minor points that should be addressed as follows:
• Page 4, Line 77: Please change "stress testing" to read further testing (some patients will undergo anatomical tests such as CTCA/angiography) Page 4, Line 86. Change "TIMI score is used to risk stratify..." to "was developed to risk stratify" (since its use has become broader over time).
We have changed the manuscript according to the suggestion
• Page 5, Line 104. After "both risk scores have been validated..." please add "for use in a low-risk ED population"
We have changed the manuscript according to the suggestion
• Page 6, Line 139: "The HEART score was developed in 2007" is repetitious (cf introduction) and should be removed
We have changed the manuscript according to the suggestion, and also changed this for the TIMI score, since here the same repetition was found
• Page 10, Line 231: "Within the low-risk TIMI group" please clarify whether you are referring to TIMI 0 or <=1, similarly
We have changed the manuscript according to the suggestion
• Page 10 Line 235 Please include the TIMI<=1 costs in the abstract if word limit allows, this can be followed with a statement such as "but with a MACE rate of 3%" so as to justify your support of the HEART score.
If the editor would allow the extra words in our abstract, we would wish to follow the authors suggestion of adding the information for TIMI <=1. The following lines were included in the abstract:

"With different cut-offs for low-risk, HEART 0-2 (miss-rate 0.7%), would have resulted in a total of €25,365 in savings, compared to €71,905 when an alternative low-risk cut-off for TIMI of TIMI ≤1 would be used (miss-rate 3.0%)."

Reviewer: 2
I think the authors made good revision this time. However, a few points still need to be addressed to make it acceptable to be in press:

• Please discuss why 6 week window was chosen for the following up of the ED discharge?
We have used this window for follow-up of the outcome of MACE since the original validation study also used this window in their outcome. Furthermore, the follow-up for use of health care resources was also set on these 6 weeks, because we argued the window for diagnostic procedures could not be too long, or otherwise there would be a higher risk that the diagnostic procedures did not pertain anymore to the initial ED visit with chest pain.
• Given both scoring metrics authors discussed have not used as the rule to discharge ED patients, therefore, the author should discuss why both metrics or one of the metrics would be suitable to be used to manage ED discharge.
Indeed, the TIMI score nor the HEART score were actively used to direct discharge from the ED, since data was used from a validation study, where the HEART score was calculated, but not adhered to actively, and the TIMI score was calculated retrospectively from admission data. Future studies

with implementation of the HEART score and/or the TIMI score will address if these potential savings can indeed be achieved.