Author's response to reviews

Title: Value of signs, symptoms and plasma heart-type fatty acid-binding protein (H-FABP) in evaluating patients presenting with symptoms possibly matching acute coronary syndrome: background and methods of a diagnostic study in primary care

Authors:

Robert TA Willemsen (robert.willemsen@maastrichtuniversity.nl)
Frank Buntinx (Frank.Buntinx@med.kuleuven.be)
Bjorn Winkens (bjorn.winkens@maastrichtuniversity.nl)
Jan F Glatz (glatz@maastrichtuniversity.nl)
Geert Jan Dinant (geertjan.dinant@maastrichtuniversity.nl)

Version: 3
Date: 23 October 2014

Author's response to reviews: see over
Your reference number:
MS: 150695692123317

Article title and authors:
Value of signs symptoms and plasma heart-type fatty acid-binding protein (H-FABP) in evaluating patients presenting with symptoms possibly matching acute coronary syndrome: background and methods of a diagnostic study in primary care

Robert TA Willemsen
Frank Buntinx
Bjorn Winkens
Jan F Glatz
Geert Jan Dinant

Maastricht, October 23th, 2014

Dear editor, dear Magda,

thank you very much for all your effort to have our manuscript peer reviewed, after our correspondence earlier. And thank you for the detailed and constructive review. In this letter I will respond to all comments and I will make clear what changes we made to the manuscript. Point by point, the original comment is stated in italic, my answers are in bold type.

Referee 1, dr Body:

- This work addresses a scientifically important and topical question. Positive findings could have significant clinical impact. I wish the authors the very best of luck with this good work. My sincere thanks!

Major compulsory revisions

- This is a derivation study, designed to derive a diagnostic algorithm for patients presenting to primary care with suspected acute coronary syndromes. It is important to acknowledge this, as the algorithm will require validation (and subsequent evaluation in practice) before it could be considered for clinical
implementation. Without validating the algorithm, there is the potential that the model may be over-fitted to the data and therefore that diagnostic accuracy is over-estimated.

I fully agree and have tried to solve the possible misunderstanding (derivation or validation of the algorithm) by clearly re-formulating this issue in the first paragraph of the 'study design and objectives' section. Indeed, currently we are deriving, not validating an algorithm.

- It would be useful to know the covariates that are being considered for inclusion in the model. How many covariates are there? (This will influence sample size). How were they identified? How will they be recorded? Including a copy of the case report form as an appendix would be very helpful.

  On our case report form, 23 possible covariates of history taking and 8 possible covariates of physical examination are registered. Whether or not each variable can be assessed depends on the incidence of each variable in our population. Therefore, only variables that are reported with a certain frequency can be used in the algorithm. We cannot predict yet which signs and symptoms will meet this criterium, this will be part of our outcome. A copy of the CRF (in Dutch, or an English translation) can indeed be included as an appendix, if the editor wishes so.

- The sample size calculation should be revisited. The authors don't really provide justification for the proposed levels of precision. As the primary goal is to derive a diagnostic algorithm incorporating point of care H-FABP levels, the sample size calculation ought to be based on this.

  I agree. On page 3, paragraph 2, I have added the current PPV, NPV, sensitivity and specificity of an assessment by a GP. An algorithm leads to more effective referral policies if false negative patients (i.e. not referred but ACS positive) and false positive patients (referred but ACS negative) are reduced. However, if false positive cases are importantly reduced, with a false negative count that remains the same, cost reduction without an increase of missed ACS-cases would be the case. In order to be cost effective, we aim to reduce the number of false positive patients, thereby at least keeping sensitivity at the same level as is currently seen in general practice. Therefore, we want to improve PPV and specificity by using the to be created algorithm instead of a sole judgment by the GP. Furthermore, NPV should not decrease. We based the power calculation on these demands. On page 8, 'sample size and power calculation', I've added information on the current diagnostic values that need improvement to reach our study goal.

Minor essential revisions

- The proposed statistical analyses could be described in some more detail. There are published methodological criteria for the derivation of clinical decision rules (e.g. Stiell et al, Ann Emerg Med April 1999;33:437-447; Steyerberg on Clinical Prediction Models, ISBN-13 978-1441926487). This work would benefit from providing some details in accordance with those publications.

  A very useful addition to our manuscript. I have added Stiell and Wells to
the reference section and I have added comments on relevant parts of their work regarding our study. You can find these in the first paragraph of the 'study design and objectives' section, as well as in the 'discussion' section at the end of the manuscript.

- There are a few minor grammatical errors, e.g. page 2, Background line 1: "Patients presenting chest complaints". This ought to be changed to, "Patients presenting with chest complaints" Also Page 2, Background lines 9-10: "Therefore, referring any patient with chest complaints to secondary care facilities is not applicable" should be re-phrased. The issue is perhaps that this would overwhelm secondary care resources.

Thank you, I've corrected these errors.

- It would be useful to have more details about the proposed economic evaluation. How will costs be calculated? Which costs will be included? If ICERs are to be determined, how and when will health state be evaluated?

Costs will be based on medical costs. I have added this briefly in the 'primary and secondary' outcome section. Medical costs are the cost of a GP's consultation and the costs of a PoC H-FABP test versus the cost of a visit to a secondary care facility. The benefit, as stated, will be assessed by the decrease in referrals and the decrease in missed ACS cases.

Discretionary revisions.

- The lack of blinding to H-FABP results does present some challenges. It is probably naive to think that the judgement of GPs will not be influenced, at least to some extent, by the levels of a cardiac marker that are available immediately. There is no way of time stamping manual entry into a paper case report form to ensure that data were entered prior to the results being available. Therefore, this does still warrant some attention. Perhaps levels of H-FABP could be evaluated after the patients have left? (E.g. by collecting whole blood into sample tubes with venipuncture)

Whole blood is collected, enabling the analyses afterwards. Furthermore, our case report form is designed to first report a working diagnosis, then a PoCT result. GPs are instructed to pay all their attention to this essential order of working with the CRF. Since blinding is not possible, this is regarded as the next best option to deal with the lack of full blinding.

- Only patients who have provided full written informed consent will be included. What if patients die? This is an extremely important outcome. Would the ethics committee consider issuing a waiver (or relative assent) for these circumstances, given the scientific importance of this work in emergency settings?

In these cases, people die after having signed the short informed consent, but before signing the long informed consent. We can use their outcome measure (death) in calculations of diagnostic performance of the point of care test. However, preserving and using their blood
samples for later analyses - not directly related to the current protocol - will be impossible.

- **Point of care troponin tests are available. Have the authors considered including this in the protocol?**
  Yes, but we decided not to work with these tests. We have data to be published that reveal that H-FABP using a cut-off value of 4 ng/ml, has similar performance as hs-troponin with a cut off of 14 ng/ml, when measured in the traditional way. Available PoC troponin tests have cut-off values of 50 and higher and they lack diagnostic performance in primary care, as recently presented by Nilsson et al at the EPCCS in Brussels.

- **It is notoriously difficult to adjudicate a diagnosis of 'unstable angina' as there is no acceptable reference standard and clinical features are notoriously unreliable. Evaluating major adverse cardiac events after 30 days may act as a reasonable (and more objective) surrogate.**
  The UA diagnosis will be based on the initial judgment and the 30 day follow up. For example, should the initial diagnosis by a cardiologist be thoracic wall pain - based on clinical judgment and a negative troponin result - and should after thirty days an adverse cardiac event be the case, the expert panel will define the final diagnosis as UA.

- **Page 3. The authors state that the study by Bruins et al "lacked diagnostic strength". Presumably this refers to the statistical "power" of the study rather than "strength".**
  I cancelled this line, since on page 4 (in the first full paragraph of page 4), the Bruins Slot study is discussed in more detail.

- **There are existing alternative prediction models that have already been derived, although they do include additional parameters (e.g. troponin) and they werederived in other settings. (E.g. the MACS rule and the HEART score). It would be a very useful secondary objective to validate existing scores. If those scores can be successfully validated in the primary care setting and with the use of point of care tests, it could reduce the time taken to clinical implementation and therefore benefit patients sooner.**
  Very interesting, and we wish to do so. I think our current protocol leaves space to derive the data in a blank way, but also to test predefined combinations of symptoms.

- **Level of interest: An article of outstanding merit and interest in its field**
  Quality of written English: Needs some language corrections before being published
  Statistical review: Yes, and I have assessed the statistics in my report. Thank you very much.
Thank you for asking me to review this protocol for an interesting study that may have clinical importance. The authors have presented many of the issues clearly, particularly in relation to outcome definition. Potential risks of bias for this type of study have been addressed though some remain and need to be addressed as outlined in my comments below. It may be worth explicitly grouping risks of bias and how they are addressed using the McGinn criteria outlined in the reference below and used in the development of a CPR register (see Keogh et al, Ann Fam Med July/August 2014 vol. 12 no. 4 359-366). This would improve the reporting of the paper overall.

Thank you very much.

major comments:

• The authors describe this study as a “delayed type cross-sectional study”. I don’t understand what this means really. It seems to me that the study is what would be called a derivation study for a clinical prediction rule or algorithm and there is a substantial literature developing in this area. This study is essentially seeking to develop a CPR that combines discriminating signs and symptoms with the 4ng/ml cut-point PoC H-FABP test (see McGinn et al for one of the definitive papers on this topic - McGinn TG, Guyatt GH, et al; Evidence-Based Medicine Working Group. Users’ guides to the medical literature: XXII: how to use articles about clinical decision rules. JAMA.2000;284(1):79–84)
  
  I have dealt with this topic by responding to Dr Body’s first major comment and also his first minor comment, see above. I changed the manuscript on this topic as indicated above. I added the useful reference (paper by McGinn et al) that is proposed here in the ‘discussion’ section.

• The overall presentation of the study protocol is a little unclear, partly due to the lack of fluency in English but also because of the lack of clarity in design. I hope to have solved this issue by my abovementioned changes in the manuscript, particularly on issues regarding design. Furthermore, I have changed several sentences for the sake of fluency. In case of remaining lack of clarity of the English writing I would appreciate to be noticed by the editor.

• Abstract need to report the comparison group
  
  We have no comparison group. GP’s indicate their regular diagnosis and decision whether or not to refer on the case report form, before the PoCT is performed. See subheading ‘data collection’. Therefore the control group is embedded in the study group.

• There are multiple instances throughout the text where wording is used that is not standard eg ‘beneficial’ causes of chest pain as opposed to benign causes.
  
  I solved this by using advantageous or benign.

• I think it would be interesting if the authors commented on the wide range in prevalence of 1.5% to 22% of cases in primary care. Does this vary at practice level as if so need to consider some kind of practice level analysis as well
  
  The wide range is reported in literature and is due to variation at different
levels: country, practice, methods of measurement and/or registration. We will be able to comment afterwards on differences between rural practices and urban practices, and between Dutch and Belgian practices, since all are represented in our study.

- How will you monitor fidelity to the protocol, particularly in relation to reading the test before or during the decision making process which may take longer than 5 minutes. Why not ask an independent person to read the test if it is not meant to play a part in the process. This may be important as GPs are expected to change their mind and then refer the patient if the test turns out to be positive, which might be difficult to do in front of a patient and in the specified time frame. GPs are instructed to fulfill this essential part of the protocol. We found no other solution here. GP’s see patients in different situations, sometimes without anyone able to perform the test and sometimes with persons that are only involved once in the study (medical personnel when GP is on call, drivers when GP is on call, etc). Therefore, a GP would have to instruct those persons which would be inconvenient and these persons would not have any experience with the test.

- Analysis: Can you comment on why you are not presenting a C Statistic to determine model discrimination. C-statistics will be presented, I added a sentence under 'statistical analyses'.

- Sample size: a rough guide for derivation studies is that there should be at least 10 outcome events for every variable in the final prediction rule – this seems likely to be the case but the previous published work will give you an idea of how many predictors remained in the model (see McGinn et al). See above for comment on this topic.

- Why are you presuming you will have 100% uptake in the 60 GPs identified – is there a larger pool from which they are being selected? No, but we will evaluate this process and take measures when necessary.

- How will you monitor selection bias in terms of inclusion of the ten consecutive cases? We interview GP’s and evaluate the study when running several times. We will report on this bias afterwards. I added one sentence in the paragraph under subtitle 'recruitment'.

Minor issues

- Protocols are usually written in the future tense. The authors use a mixture of present and future tense. I agree but did not find it disturbing. Since Dr Body had no comment on this either, I propose to wait for any further comments, if any, on this issue by the editor.
• Wording as above – specialist setting would be more accurate than hospital setting as some hospitals don't manage ACS
  I have changed the word hospital where appropriate into specialist / secondary care setting

• Please clarify if practices re offered any incentive to participate
  I assume that the question is whether practices refused further participation after initially taking part in the study? No, that hasn’t been the case. If so in the future, I will report this.

• If the ECG is one of the potential predictors – will it be a problem if they don’t all have one done
  Given the poor ability to exclude coronary problems and because of the suboptimal availability in primary care, our expectation is that ECG will not play a role in a decision rule. Therefore, we didn’t claim ECG’s in cases GP’s wouldn’t make an ECG in a regular situation. We decided to use the ECG data that we obtain by registering data from the ECG’s that are made anyway, from the GP’s that tend to do so in cases of chest complaints. We expect to have ECG’s of more than half of the population.

• What is meant by ‘coded data’ in the data management section?
  By coded, I meant anonymous. I cancelled the word since it does not add any information.

• Public disclosure means open access to the actual data not just published papers – is there a plan to make data accessible?
  I see, I didn’t use this term correctly. I changed this paragraph.

• Are there any conflicts of interest based on the fact that the device being tested is likely to have economic value if found to be useful?
  If the question is whether the researchers will benefit financially in case of an economic value, the answer is: no, that is out of the question.

• Level of interest: An article of importance in its field
  Quality of written English: Needs some language corrections before being published
  Statistical review: No, the manuscript does not need to be seen by a statistician.
  Thank you.