**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)** | Gender-stratified analyses of symptoms associated with acute coronary syndrome in telephone triage: a cross-sectional study |
| **AUTHORS** | Wouters, Loes; Zwart, D.L; Erkelens, Daphne; De Groot, Esther; van Smeden, Maarten; Hoes, Arno; Damoiseaux, Roger; Rutten, Frans |

**VERSION 1 – REVIEW**

| | |
|---|---|
| **REVIEWER** | Presbitero, Patrizia<br>Istituto Clinico Humanitas, Clinical and Interventional Cardiology |
| **REVIEW RETURNED** | 22-Jul-2020 |

| | |
|---|---|
| **GENERAL COMMENTS** | I think that it shoul be pointed out that even if the women population with ACS is older and has a higher prevalence of diabetes (41%) symptoms are similar to the one of men.<br><br>Probably a comparison between the population with / without diabetes should be considered, because another common idea is that diabetec patients present with different symtpoms due to the lack of pain. |

| | |
|---|---|
| **REVIEWER** | Cho, Dong-Hyuk<br>Wonju Severance Christian Hospital, Division of Cardiology, Department of Internal Medicine |
| **REVIEW RETURNED** | 03-Jul-2020 |

| | |
|---|---|
| **GENERAL COMMENTS** | In the manuscript by Wouters et al entitled " Gender differences in patients with acute chest discomfort: a cross-sectional study", the authors describe 1,795 patients who called OHS-PC for acute chest discomfort. Patients were divided by gender and compared based on detailed aspects of chest pain. They found that some characteristics (severe pain, a pale face, and radiation to the jaw) predicted ACS only in women, which can be distinguished from men.<br><br>This is an interesting topic since symptomatic difference by initial conversation has not been investigated among a large sample of patients calling the OHS-PC because of chest discomfort yet and finding of this study can help evaluate chest pain in general practice.<br><br>1. The authors should empower the publishing track record for gender differences of symptoms in CAD. There are various studies from ACS to stable CAD, and from a prospective multicenter to a retrospective single-center study. |

| | |
|---|---|
| | (Rubini Gimenez M, Reiter M, Twerenbold R, et al. Sex-specific chest pain characteristics in the early diagnosis of acute myocardial infarction. JAMA Intern Med. 2014;174(2):241–9. This study demonstrates that gender differences in chest pain profile are not enough to be used in the diagnosis of AMI)<br>(Cho DH, Choi JM, Kim MN, et al. Gender differences in the presentation of chest pain in obstructive coronary artery disease: results from the Korean Women's Chest Pain Registry. Korean J Intern Med 2020;35:582-592. This multi-center study demonstrated that symptoms with high probability for obstructive CAD were different between sexes in patients with suspected angina.)<br><br>2. The authors compared chest pain characteristics between women with and without ACS, and men with and without ACS. However, chest pain characteristics should be compared between men and women to investigate the gender difference. Consider p-for interaction between men and women.<br><br>3. This study seems to be quite similar to their previous work (BMJ open. 2019;9:e031613.) except for a sample size. Please explain the common and different findings between this and previous studies.<br><br>4. Appendix-Table 2 (Relation of the caller to the patient in women and men with ACS) is interesting. The difference of caller may explain the symptom difference. A discussion about this issue is recommended. |

| | |
|---|---|
| **REVIEWER** | Groen, Henk<br>Rijksuniversiteit Groningen, Epidemiology |
| **REVIEW RETURNED** | 19-Nov-2020 |

| | |
|---|---|
| **GENERAL COMMENTS** | General remark, althoug outside the scope of this statistical review:<br>Please check English language throughout the paper.<br>E.g. 'Retrosternally located chest pain ..' (line 20, page 9)<br>'Sweating was positively related ...' (line 22, page 10)<br>Redundancies in longer sentences.<br><br><br>Remarks on methodology and statistics:<br>No sample size calculation is presented. How was the sample size determined in relation to the desired precision of the outcome estimates?<br><br>I could not find the appendix-table 2 that is referred to in line 54 on page 8.<br><br>Some explanation of the ICPC codes is required to determine whether selection of the calls was adequate.<br><br>A flow chart with the numbers of initial calls and numbers of calls eliminated from analysis at subsequent steps would be useful. |

| | |
|---|---|
| | Is there any information about missed diagnoses? Were there cases where the ambulance was not sent and confirmation of abscense, or worse presence, of ACS was not obtained?<br><br>The presentation in Table 1 is not clear.<br>The heading is not well formulated. Numbers and percentages could be moved to the column headers in the tables.<br>The numbers for the respective symptoms are not clear. I assume the number pairs in parentheses are the numbers for men and women, for instance "(n=960;779)" for chest pain. However, the numbers in the respective columns for women and men with and without ACS do not add up to these totals. In the next line, severe pain appears to be only available for 412 women and 341 men), but again the numbers do not add up.<br>I can understand that these differences in numbers can be related to the triage algorithm that was used but it makes the results very difficult to interpret, since every line in the table concerns a different subset of the patients. If possible, present the triage algorithm or parts of it to indicate how the information was retrieved.<br>There is too much emphasis on the p-values. Instead of the p-values, present mean difference or mean difference of proportion with the respective 95% confidence intervals.<br>There is too much repetition of numbers from table 1 in the results section.<br><br>Finally, the statistics are too simplistic. These are observational data and as such are prone to confounding. The most apparent confounder here is age. Was the mean age of men and women different? This has to be addressed. In addition the overlap of symptoms, as the authors mention, is something to address. Is isolated presence of a specific symptom different from combined presentation with respect to presence of ACS.<br><br>The generalizability of the results is overstated. Other populations within the EU may have different compositions, for instance regarding ethnicity, which preclude generalization. |

| | |
|---|---|
| **REVIEWER** | Boehnke, Jan Rasmus<br>University of Dundee, School of Health Sciences |
| **REVIEW RETURNED** | 16-Dec-2020 |

| | |
|---|---|
| **GENERAL COMMENTS** | The manuscript "Gender differences in patients with acute chest discomfort: a cross-sectional study" presents the analysis of 1795 calls to out of hours primary care services in The Netherlands. Calls were coded for a number of characteristics, including presenting symptoms and within males and females separately it was evaluated for each characteristic whether it is associated with a later diagnosis of ACS. The manuscript is well-written and easy to follow, I nevertheless have a number of questions to the authors |

| | |
|---|---|
| | regarding the intent of the use of their results; aspects of STROBE reporting not sufficiently covered (missing data, sample selection); the justification of only presenting within-gender groups by-characteristic analyses; and regarding a more detailed discussion of their results and methods. I have listed more detailed comments on these below.<br>NB to editor: I was not able to verify the International Code for Primary Care codes used in this study as the WHO seems to have taken their online materials on this away from the web, i.e. I cannot vouch for the relevance of these codes or particular omissions. And I am not an expert on cardiovascular pathophysiology and have therefore only assessed the implemented aspects of design, methods and statistical analyses.<br><br><br><br>MAJOR<br>1) The manuscript, as many cross-sectional studies, could be clearer regarding its research question and consequently the framing of the results.<br>1a) Page 6: "We aimed to assess symptoms predictive of ACS in women and men separately, among patients presenting with acute chest discomfort to OHS-PC based on analyses of recorded telephone triage conversations." Based on the introduction and the rest of the paper it seems worthwhile being clear about in which context the "predictiveness" is relevant. It seems to be meant to be _diagnostic_, but I may be wrong. For example, this paper from the seasonally appropriate 2016 BMJ Christmas Issue provides the precise terminology to use (doi: https://doi.org/10.1136/bmj.i6536: table 1).<br><br>1b) The authors do not analyse whether the call characteristics are predictive of ACS or differentially predictive of ACS across gender groups. The authors present only by-characteristic significance tests within gender groups (male/female). There are certainly reasons to do this, but it is difficult to form a coherent picture of (i) which of these characteristics are actually predictive of ACS, (ii) whether they are differentially predictive across gender groups, (iii) or whether this actually matters. Whether such questions and accompanying analyses are relevant for the authors' research depends on their framing and intended use of the results (e.g., again doi: https://doi.org/10.1136/bmj.i6536). At the moment, as a reader I feel a bit left hanging with a list of call characteristics that may or may not be relevant for the gender-ACS question and neither an analysis what makes for example the diagnosis of ACS more reliable, nor am I given any reason why this was not investigated more formally.<br><br>1c) Without sorting this issue of how to address this question of potentially differential predictive association with the ACS diagnosis more formally, the first two |

| | |
|---|---|
| | paragraphs of the discussion section are not fully backed up by the analyses since (i) no interactions between gender and characteristics were investigated (staying with the by-characteristic approach currently implemented; only within-gender analyses are described); and (ii) just because by-characteristic associations exist, does not mean they all matter.<br><br>2) The introduction and discussion could benefit from a more differentiated take on what the authors mean when they use the term gender in this case? The introduction should cover this as this is one of the key variables and it is crucial for understanding about what exactly it is the team wanted to learn and about what we are learning as readers. It seems from the nature of the data covered it would be self-identified/-signified (how many of the callers did not fall into the male/female categories?), but this is only what I can infer from the presented information at the moment.<br><br>3) A bit more detail on the process of selecting the interviews could be useful.<br><br>3a) It seems the calls are recorded? Is this standard practice? Are all calls recorded?<br><br>3b) Since the calls could be filtered for ICPC code and keywords before the sampling, who is assigning those codes and keywords?<br><br>3c) How was the random sample drawn with Excel? Were all calls listed and then random numbers assigned? etc pp.<br><br>3d) The sample selection is also not described in sufficient detail (e.g., STROBE guidelines). How many calls were available? How many of these had the relevant codes and key words? How large was the random sample? How many of the selected were below the age of 18, didn't live in the area, or were calls of low quality? (The results section reads to me as if the 1795 are only the last number of this process, but this is not clear either, it is simply the number of calls analysed.)<br><br>3e) No justification for the size of the random sample was provided.<br><br>3f) The abstract states 2014-2017 as the sampling frame, while the methods section states 2014-2016.<br><br>4) Missing data: While the submitted STROBE statement points to page 7 for an explanation of much missing data was observed and how it was handled (i.e. with this sample |

| | |
|---|---|
| | size likely implying imputation or other means of modelling), no such information is presented.<br><br>5) The limitations section is very brief and does not pick up a number of key points, and instead uses wordage on likely not sustainable generalisability claims.<br>5a) p. 12: "Another strength is that our results are generalizable to comparable primary care settings, e.g. UK and Scandinavian countries, and some other European countries." There is no evidence presented to back this up, therefore this sentence should either be deleted or effort should be undertaken to provide a justification as to why the results would be relevant in countries with different triage and health care systems, as well as potentially other gender expectations and relations.<br><br>5b) p. 12: " Our results may even be generalizable to EMS settings, since the prior probability of having an ACS is comparable in EMS setting as in OHS-PC settings." No evidence or justification presented for this statement either. As this claim in particular could be researched by modelling implications of symptom priors as well as using comparable data for symptom characteristics in EM settings, it seems to be a particular long shot.<br><br>5c) As the authors present in the introduction themselves, women are also generally less frequently (correctly) diagnosed with ACS. The criterion against which the call characteristics are evaluated is nevertheless a practice-based diagnosis of ACS, which may be wrong – and differentially so across gender groups. I wonder whether the authors could discuss this limitation a bit more in detail.<br><br>5d) The study is investigating the role of gender in a diagnostic process. The team is using a resource that could have been masked to gender by both filtering voices as well as removing for example names and other information. This could have helped to further reduce the effect of pre-conceptions when coding call characteristics (blinding coders to both relevant variables for this analysis, gender and ACS status). This seems to be an important point to consider for future studies.<br><br>6) Conclusion & abstract: "Symptoms predictive of ACS were rather similar for women and men with chest discomfort…" As no comparative analyses across gender groups are presented, no statement regarding differences across gender groups can be made. Additionally, even if interactions were tested (by-characteristic or in a multivariate model), no statement about similarity could be made as the (at least the so far chosen methods) could only support differences, but they do not provide evidence for similarity/ equality. |

| | |
|---|---|
| | MINOR<br>1d) page 7: "An ACS was based on the cardiologist's diagnosis, including information on levels of (high-sensitivity) troponin and electrocardiography results." This statement could benefit from clarifying that (I assume at least) the diagnosis of ACS was made some time after the analysed phone calls. |

## VERSION 1 – AUTHOR RESPONSE

| Comments reviewer 1 | Response from Author |
|---|---|
| I think that it should be pointed out that even if the women population with ACS is older and has a higher prevalence of diabetes (41%) symptoms are similar to the one of men. | We agree with the reviewer this is an important aspect, but not so much for the clinician who is facing the patient suspected of ACS. He/she wants to know whether he/she should be suspicious for differences between women that eventually show to have an ACS and those who have not, and similarly for men.<br>Nevertheless, we have pointed out that diabetes and older age more in the result and discussion section, the more because the large majority of previous publications directly compare women with ACS to men with ACS.<br><br>Results:<br>"…women with ACS had more often a history of diabetes (41.4% vs. 14.6%, p<0.001)."<br><br>Discussion:<br>"In our study women with ACS had more often a history of diabetes and were older, which is in line with other studies. [8, 12] Some studies claim that patients with diabetes more often have atypical symptoms of ACS, however a review of eight studies concluded that the evidence of these studies was conflicting. [24] We showed that both women and men with diabetes had more often shortness of breath than those without diabetes, but shortness of breath in patients with diabetes was not helpful to diagnose ACS." |
| Probably a comparison between the population with / without diabetes should be considered, because another common idea is that diabetic patients present with different symptoms due to the lack of pain. | We thank the reviewer for this suggestion. In a subgroup analysis, we compared men and women with diabetes vs. no diabetes regarding presence of chest discomfort (pain, pressure or tightness) and shortness of breath and its relation with ACS diagnosis.<br><br>We added in the revised manuscript in the results section:<br><br>"Subgroup analyses in 56 women and 58 men with diabetes showed that both women (85.7% vs. 58.3%, p<0.001) and men with diabetes (67.2% vs. 51.5%, p=0.033) more often had shortness of breath than those without diabetes, but as often chest discomfort (women 90.9% vs. 95.0%, p=0.193, men 89.2% vs 94.1%, p=0.162). Shortness of breath in patients |

| | |
|---|---|
| | with diabetes was not related to ACS diagnosis (women 81.8% vs. 86.7%, p=0.680, men 75.0% vs. 66.0%, p=0.615).” |
| **Comments reviewer 2** | |
| The authors should empower the publishing track record for gender differences of symptoms in CAD. There are various studies from ACS to stable CAD, and from a prospective multicenter to a retrospective single-center study.<br>(Rubini Gimenez M, Reiter M, Twerenbold R, et al. Sex-specific chest pain characteristics in the early diagnosis of acute myocardial infarction. JAMA Intern Med. 2014;174(2):241–9. This study demonstrates that gender differences in chest pain profile are not enough to be used in the diagnosis of AMI)<br>(Cho DH, Choi JM, Kim MN, et al. Gender differences in the presentation of chest pain in obstructive coronary artery disease: results from the Korean Women’s Chest Pain Registry. Korean J Intern Med 2020;35:582-592 PubMed . This multi-center study demonstrated that symptoms with high probability for obstructive CAD were different between sexes in patients with suspected angina.) | We thank the reviewer for the suggestions and we have added the first reference to the discussion of our revised manuscript.<br><br>“Our finding that radiation of pain to the arm and retrosternal chest pain was discriminative for ACS in both sexes was also reported in a study among 2,475 patients with acute chest pain in a multicentre ED-study. [23]”<br><br>We have not added the second reference to our manuscript, because this study focuses on *stable* coronary artery disease, which requires a different approach than patients (suspected of) *acute* coronary syndrome. |
| The authors compared chest pain characteristics between women with and without ACS, and men with and without ACS. However, chest pain characteristics should be compared between men and women to investigate the gender difference. Consider p-for interaction between men and women. | We disagree with the reviewer on this point. As pointed out in the introduction, for the clinician or telephone triage nurse it is crucial to differentiate ACS from other causes of chest discomfort. For that, studies are needed that include women and men presenting with chest discomfort, in which separately women and men who turn out to have ACS are compared to those who do not have an ACS. |
| This study seems to be quite similar to their previous work (BMJ open. 2019;9:e031613.) except for the sample size. Please explain the common and different findings between this and previous studies. | Indeed, there are similarities with that previous work. However, that study was executed in a single OHS-PC, while this is a multicenter study with data from 9 OHS-PC locations, with a prevalence range of ACS of 9 to 15% in those with suspected symptoms.<br>The design of the study (comparing women to women and men to men) is similar, because we consider this is the best comparison from the perspective of a clinician. Also the setting (OHS-PC) is similar. We therefore clearly referred this previous work in our paper. The more because it was one of the few studies that made this comparison, and the first ever in the domain OHS-PC. |
| Appendix-Table 2 (Relation of the caller to the patient in women and men with ACS) is interesting. The difference of caller may explain the symptom difference. A discussion about this issue is recommended. | We are pleased to clarify this item. The relation of the caller concerns the person who makes the *initial* telephone contact. However, Dutch triage nurses are obliged - by protocol- to ask the patient him/herself to the phone, in order to obtain the most complete information on symptoms.<br><br>We have now added in appendix-table 2 how often eventually the patient self took over the phone conversation. |

| | |
|---|---|
| | In the revised manuscript we added to the discussion:<br>"According to protocol in OHS-PC, triage nurses ask the patient to the phone, this to prevent loss of (paralinguistic) information from the patient him/herself. In our study, in about 50% of the conversations the patient took over the phone call." |
| **Comments reviewer 3** | |
| General remark, although outside the scope of this statistical review:<br>Please check English language throughout the paper.<br>E.g. 'Retrosternally located chest pain ..' (line 20, page 9)<br>'Sweating was positively related ...' (line 22, page 10)<br>Redundancies in longer sentences. | Thank you. We have double checked our English language throughout the paper. |
| No sample size calculation is presented. How was the sample size determined in relation to the desired precision of the outcome estimates? | We thank the reviewer for giving us the opportunity to clarify this aspect.<br><br>For an observational study<br>with descriptive univariable statistics, and without multivariabl e analyses, formal sample size calculation methods are lacking. We therefore have selected a convenient sample and aimed at least 80 patients with ACS in each sex category to achieve robustness of information.<br><br>We have added information to the method section:<br>"For a descriptive observational study, a method for sample size calculation is lacking. We therefore included a convenient number of patients, that is, at least 80 patients with ACS in each sex category." |
| I could not find the appendix-table 2 that is referred to in line 54 on page 8. | The appendix-table 2 can be found at the last page of the manuscript |
| Some explanation of the ICPC codes is required to determine whether selection of the calls was adequate. | The International Classification of Primary Care (ICPC) code-system is a worldwide-applied diagnosis coding system used in 45 countries and accepted by the WHO as an International Classification system for primary care. The ICPC code-system relates to the International Classification of Diseases (ICD) code-system and the Read system used by general practitioners in the UK.<br><br>In the revised manuscript we have added information about how ICPC-codes and keywords were assigned and selected, supplemented by an additional appendix-table.<br><br>Method section:<br>"General practitioners who work at the OHS-PC assign the ICPC codes to the call (see also appendix-table1). We combined ICPC-codes and keywords to achieve a sample with a broad variety of symptoms to capture the entire domain of patients suspected of ACS." |
| A flow chart with the numbers of initial calls and numbers of calls eliminated from analysis at subsequent steps would be useful. | We have added a flowchart of the study population (Figure 1). |
| Is there any information about missed | The diagnosis of ACS was based on a cardiologist' diagnosis |

| | |
|---|---|
| diagnoses? Were there cases where the ambulance was not sent and confirmation of abscense, or worse presence, of ACS was not obtained? | (97.1%) with electrocardiography and troponin levels. Patients without ACS were diagnosed in 45.4% by a cardiologist and in 5.5% by another hospital specialist (e.g. pulmonologist or internal medicine specilist). The remaining patients -who were not referred to the hospital- were diagnostically labeled by a GP based on information in the 30-days following the contact with the OHS-PC.<br>We had not a single case of missed ACS among the 1,795 patients.<br><br>In the revised manuscript we added the following to the method paragraph:<br>“We used medical information up to 30-days following the contact with the OHS-PC, to allow us to include diagnoses of ACS that was initially missed because the patient was not referred to the cardiologist the same day of the OHS-PC contact. In none of the patients in the study we had evidence of a missed diagnosis of ACS.”<br><br>We added the following in the results paragraph:<br>“In nearly all cases (97.1%) the ACS diagnosis was made by a cardiologist based on symptom presentation, troponin levels and electrocardiography. Three patients died before arrival of the ambulance (they were classified as acute cardiac death) and one patient died after resuscitation at the ED. Two patients were classified as ACS by the GP; they were not referred to the hospital because of short life expectancy due to cancer. Of the patients who were diagnosed with non-ACS diagnoses, 45.4% were assessed by a cardiologist, 5.5% by another hospital specialist (e.g. pulmonologist or internal medicine specialist) and the remaining patients were diagnosed by a GP.” |
| The presentation in Table 1 is not clear.<br>The heading is not well formulated. Numbers and percentages could be moved to the column headers in the tables.<br>The numbers for the respective symptoms are not clear. I assume the number pairs in parentheses are the numbers for men and women, for instance "(n=960;779)" for chest pain. However, the numbers in the respective columns for women and men with and without ACS do not add up to these totals. In the next line, severe pain appears to be only available for 412 women and 341 men), but again the numbers do not add up. I can understand that these differences in numbers can be related to the triage algorithm that was used but it makes the results very difficult to interpret, since every line in the table concerns a different subset of the patients. If possible, present the triage algorithm or parts of it to indicate how the information was retrieved.<br>There is too much emphasis on the p-values. Instead of the p-values, present mean | We agree that the presentation of table 1 could be improved.<br><br>Missing values on items result in different denominators per item.<br>For simplicity, we have now added up the numbers of complete data from women and men in the table 1 of the revised manuscript. |

| | |
|---|---|
| difference or mean difference of proportion with the respective 95% confidence intervals.<br>There is too much repetition of numbers from table 1 in the results section. | |
| Finally, the statistics are too simplistic. These are observational data and as such are prone to confounding. The most apparent confounder here is age. Was the mean age of men and women different? This has to be addressed.<br>In addition the overlap of symptoms, as the authors mention, is something to address. Is isolated presence of a specific symptom different from combined presentation with respect to presence of ACS. | We used the statistics appropriate for answering the research question. We aimed to answer whether each of the symptoms differed between patients with ACS and without ACS, and this for both men and women in the domain of suspected ACS in OHS-PC.<br>Confounding is only an issue in observational studies in which one is interested in evaluating a single variable in relation to an outcome, corrected for known and available (possible) confounders. A confounder is a variable that is related to both the determinant and the outcome. Thus, it is used in observational studies in which there is an etiological question. To assess the independent value of each variable indeed a multivariable analysis would be the best option, but that is if one strives to build the most optimal lean prediction model for ACS. This is not useful to assess differences in each of the characteristics between women with and without ACS and for men with and without ACS. |
| The generalizability of the results is overstated. Other populations within the EU may have different compositions, for instance regarding ethnicity, which preclude generalization. | UK and Scandinavian countries have a comparable population as the Netherlands, and a similar OHS-PC setting. Ethnicity is important for prior chance of ACS, e.g. Hindustani, or treatment, e.g. less effect of ACE-inhibitors in Afro-American. Importantly, however, there is no evidence of substantial differences in symptom presentation of certain ethnic populations. We therefore consider our generalizability as fair.<br><br>We have added three references:<br><br>• Burman RA, Zakariassen E and Hunskaar S. Management of chest pain: a prospective study from Norwegian out-of-hours primary care. *BMC family practice*. 2014;15:51.<br>• Deakin CD, Sherwood DM, Smith A and Cassidy M. Does telephone triage of emergency (999) calls using Advanced Medical Priority Dispatch (AMPDS) with Department of Health (DH) call prioritisation effectively identify patients with an acute coronary syndrome? An audit of 42,657 emergency calls to Hampshire Ambulance Service NHS Trust. *Emergency medicine journal : EMJ*. 2006;23:232-5.<br>• King-Shier K, Quan H, Kapral MK, Tsuyuki R, An L, Banerjee S, Southern DA and Khan N. Acute coronary syndromes presentations and care outcomes in white, South Asian and Chinese patients: a cohort study. *BMJ open*. 2019;9:e022479. |
| **Comments reviewer 4** | |
| The manuscript, as many cross-sectional studies, could be clearer regarding its | We thank the reviewer for his extensive review of our manuscript. |

| | |
|---|---|
| research question and consequently the framing of the results.<br>1a) Page 6: "We aimed to assess symptoms predictive of ACS in women and men separately, among patients presenting with acute chest discomfort to OHS-PC based on analyses of recorded telephone triage conversations." Based on the introduction and the rest of the paper it seems worthwhile being clear about in which context the "predictiveness" is relevant. It seems to be meant to be _diagnostic_, but I may be wrong. For example, this paper from the seasonally appropriate 2016 BMJ Christmas Issue provides the precise terminology to use (doi: [https://doi.org/10.1136/bmj.i6536:](https://doi.org/10.1136/bmj.i6536)table 1).<br><br>1b) The authors do not analyse whether the call characteristics are predictive of ACS or differentially predictive of ACS across gender groups. The authors present only by-characteristic significance tests within gender groups (male/female). There are certainly reasons to do this, but it is difficult to form a coherent picture of (i) which of these characteristics are actually predictive of ACS, (ii) whether they are differentially predictive across gender groups, (iii) or whether this actually matters. Whether such questions and accompanying analyses are relevant for the authors' research depends on their framing and intended use of the results (e.g., again doi: [https://doi.org/10.1136/bmj.i6536](https://doi.org/10.1136/bmj.i6536)). At the moment, as a reader I feel a bit left hanging with a list of call characteristics that may or may not be relevant for the gender-ACS question and neither an analysis what makes for example the diagnosis of ACS more reliable, nor am I given any reason why this was not investigated more formally.<br><br>1c) Without sorting this issue of how to address this question of potentially differential predictive association with the ACS diagnosis more formally, the first two paragraphs of the discussion section are not fully backed up by the analyses since (i) no interactions between gender and characteristics were investigated (staying with the by-characteristic approach currently implemented; only within-gender analyses are described); and (ii) just because by-characteristic associations exist, does not mean they all matter. | 1a) We thank the reviewer. The aim of our study was to describe whether symptoms of women and men differed for those with an ACS and no ACS in the domain suspected of ACS. Indeed the term predictive is not the best. In the revised manuscript, we therefore changed 'predictive' in either 'diagnostic' or 'association'.<br><br>1b) For this comment we want refer to our response to reviewer 2. As pointed out in the introduction and discussion, it is crucial for clinician and telephone triage nurse to know what symptoms are associated with the diagnosis ACS, and whether this is different in men than women. Such studies are rare. Most previous studies *selectively* compared women with ACS to men with ACS, but that is clinically irrelevant.<br><br>1c) We agree with the reviewer that the development of a multivariable prediction model is worthwhile, e.g. with an interaction term for sex to more precisely estimate the risk of ACS. However, with such a study another research question is answered. Namely, what is the optimal (multivariable) prediction model for diagnosing ACS in men and women suspected of ACS.<br><br>We have added a sentence to the strength and limitation section of the discussion<br>"Future research could focus on developing a multivariable prediction model useful with telephone triage to estimate the risk of ACS in men and women suspected of ACS."<br><br>1d) Indeed, the diagnosis of ACS or other diagnosis was made after the telephone calls.<br><br>Method:<br>"The diagnosis was made after the phone call, which was in the case of ACS nearly always done by the cardiologist (97.1%) in the hospital based on (i) symptom presentation, (i) levels of (high-sensitivity) troponin and (iii) electrocardiography results." |

| | |
|---|---|
| 1d) page 7: "An ACS was based on the cardiologist's diagnosis, including information on levels of (high-sensitivity) troponin and electrocardiography results." This statement could benefit from clarifying that (I assume at least) the diagnosis of ACS was made some time after the analysed phone calls. | |
| 2) The introduction and discussion could benefit from a more differentiated take on what the authors mean when they use the term gender in this case? The introduction should cover this as this is one of the key variables and it is crucial for understanding about what exactly it is the team wanted to learn and about what we are learning as readers. It seems from the nature of the data covered it would be self-identified/-signified (how many of the callers did not fall into the male/female categories?), but this is only what I can infer from the presented information at the moment. | We thank the reviewer for giving us the opportunity to clarify more about our thoughts about gender and sex.<br><br>We deliberately choose for gender (and used the terms women and men) because we consider symptom presentation to be part of how women and men behave and express themselves. Of course this can be debated. Outside the diagnostic domain, e.g. the pathophysiology of myocardial ischaemia, differences between males and females are (mainly) biological sex related.<br><br>We have no data about self-identified gender or callers who felt not into the men/women category. Worldwide the prevalence of transgenders is estimated at approximately 0.58%. Such a small subgroup will not affect our results. |
| 3) A bit more detail on the process of selecting the interviews could be useful.<br><br>3a) It seems the calls are recorded? Is this standard practice? Are all calls recorded?<br><br>3b) Since the calls could be filtered for ICPC code and keywords before the sampling, who is assigning those codes and keywords?<br><br>3c) How was the random sample drawn with Excel? Were all calls listed and then random numbers assigned? etc pp.<br><br>3d) The sample selection is also not described in sufficient detail (e.g., STROBE guidelines). How many calls were available? How many of these had the relevant codes and key words? How large was the random sample? How many of the selected were below the age of 18, didn't live in the area, or were calls of low quality? (The results section reads to me as if the 1795 are only the last number of this process, but this is not clear either, it is simply the number of calls analysed.)<br><br>3e) No justification for the size of the random sample was provided.<br><br>3f) The abstract states 2014-2017 as the sampling frame, while the methods section | 3a) It is indeed standard practice in the Netherlands to record the calls, and archive them for five years before they need to be destroyed.<br><br>In the revised manuscript, we added clarification to the method section:<br>"All telephone calls to the OHS-PC are routinely recorded and archived for five years for training and quality control purposes."<br><br>3b) The GP in charge at the OHS-PC assigns the ICPC codes and the keywords are extracted from what the triage nurse wrote down in the electronic medical file at the OHS-PC. We also want to refer to our response to reviewer 3.<br><br>3c) We used the RAND function (a built-in function) in Excel that assigns random number between 0 and 1.<br><br>3d) We have added a flowchart of the study population (Figure 1).<br><br>3e) For this comment we want refer to our response to reviewer 3. There is no formal power calculation for an observational study comparing variables between two groups.<br><br>3f) The correct time frame is 2014-2016. We changed the abstract accordingly. |

| | |
|---|---|
| states 2014-2016. | |
| 4) Missing data: While the submitted STROBE statement points to page 7 for an explanation of much missing data was observed and how it was handled (i.e. with this sample size likely implying imputation or other means of modelling), no such information is presented. | For univariable analysis imputation is not done, but instead complete case analysis. Indeed when the aim is to built a diagnostic model, multiple imputation and bootstrapping should be performed. |
| 5) The limitations section is very brief and does not pick up a number of key points, and instead uses wordage on likely not sustainable generalisability claims.<br><br>5a) p. 12: "Another strength is that our results are generalizable to comparable primary care settings, e.g. UK and Scandinavian countries, and some other European countries." There is no evidence presented to back this up, therefore this sentence should either be deleted or effort should be undertaken to provide a justification as to why the results would be relevant in countries with different triage and health care systems, as well as potentially other gender expectations and relations.<br><br>5b) p. 12: " Our results may even be generalizable to EMS settings, since the prior probability of having an ACS is comparable in EMS setting as in OHS-PC settings." No evidence or justification presented for this statement either. As this claim in particular could be researched by modelling implications of symptom priors as well as using comparable data for symptom characteristics in EMS settings, it seems to be a particular long shot.<br><br>5c) As the authors present in the introduction themselves, women are also generally less frequently (correctly) diagnosed with ACS. The criterion against which the call characteristics are evaluated is nevertheless a practice-based diagnosis of ACS, which may be wrong – and differentially so across gender groups. I wonder whether the authors could discuss this limitation a bit more in detail.<br><br>5d) The study is investigating the role of gender in a diagnostic process. The team is using a resource that could have been masked to gender by both filtering voices as well as removing for example names and other information. This could have helped to further reduce the effect of pre-conceptions | 5a and b) For this comment we want to refer to our response to question reviewer 3. In the revised manuscript we added two articles to back up our generalizability claim to European countries and the EMS setting.<br><br>We refer now in the revised Discussion:<br><br>• Burman RA, Zakariassen E and Hunskaar S. Management of chest pain: a prospective study from Norwegian out-of-hours primary care. *BMC family practice*. 2014;15:51.<br>• Deakin CD, Sherwood DM, Smith A and Cassidy M. Does telephone triage of emergency (999) calls using Advanced Medical Priority Dispatch (AMPDS) with Department of Health (DH) call prioritisation effectively identify patients with an acute coronary syndrome? An audit of 42,657 emergency calls to Hampshire Ambulance Service NHS Trust. *Emergency medicine journal : EMJ*. 2006;23:232-5.<br><br>5c) The final diagnoses was not 'practice-based', at least not that of the GP in charge at the OHS-PC and based on history taking only. We retrieved after the calls the diagnosis from the patients' GP and mainly based on hospital specialist discharge letters. Patients referred to the hospital had an ECG and (multiple) troponin testing. History taking, ECG (ST-T wave abnormalities, pathological Q waves) and elevated troponin levels are key for diagnosing ACS, essentially as good in women as in men (ESC guideline on definition of myocardial infarction). We agree with the reviewer that a 'diagnosis' based on only history taking would result in many false positive and negative diagnoses, and probably differently in males than in females.<br><br>We have added to the results:<br>"Of the 205 patients with an ACS (85 women, 120 men), 55 (26.8%) patients had a STEMI (women 18.8%, men 32.5%), 85 (41.5%) a NSTEMI (women 48.2%, men 36.7%), 50 (24.4%) unstable angina pectoris (UAP) (women 20.0%, men 27.5%) and 15 (7.3%) unspecified ACS (women 13.0%, men 3.3%), the latter also including two sudden cardiac deaths in women and one in men (Table 2). In nearly all cases (97.1%) the ACS diagnosis was made by a cardiologist based on symptom presentation, troponin levels and electrocardiography. Three patients died before arrival of the ambulance (they were classified as acute cardiac death) and one patient died after |

| | |
|---|---|
| when coding call characteristics (blinding coders to both relevant variables for this analysis, gender and ACS status). This seems tobe an important point to consider for future studies. | resuscitation at the ED. Two patients were classified as ACS by the GP; they were not referred to the hospital because of short life expectancy due to cancer. Of the patients who were not diagnosed with an ACS, 45.4% were assessed by a cardiologist, 5.5% by another medical specialist (e.g. lung specialist, internist) and the remaining patients were diagnosed by a GP."<br><br>5d) Importantly, the researchers were blinded to ACS status as mentioned in the manuscript. Because our aim was to assess whether symptom presentation was different for women with compared to without ACS and similarly for men, the suggestion of 'blinding' sex would not be worthwhile. For our research question only blinding (preventing for pre-conceptions) to the final diagnosis was essential.<br><br>Additionally, an important strength of this study is that we analyzed the very first symptom presentation, and could assess the exact words patients used to express themselves, which is a major advantage above questionnaires. Questionnaires force the patient and researcher to simplify the symptoms to standard options, with loss of important information.<br><br>These considerations we mentioned in the introduction and discussion section of the manuscript. |
| 6) Conclusion & abstract: "Symptoms predictive of ACS were rather similar for women and men with chest discomfort..." As no comparative analyses across gender groups are presented, no statement regarding differences across gender groups can be made. Additionally, even if interactions were tested (by-characteristic or in a multivariate model), no statement about similarity could be made as the (at least the so far chosen methods) could only support differences, but they do not provide evidence for similarity/ equality. | We agree with the reviewer that the wording was not correct.<br><br>It should read: "There were more similarities than differences in symptoms associated with the diagnosis ACS for women and men. Important exceptions were severity, type, and radiation of pain, and in women a pale face, and in men sweating." |

## VERSION 2 – REVIEW

| | |
|---|---|
| **REVIEWER** | Cho, Dong-Hyuk<br>Wonju Severance Christian Hospital, Division of Cardiology, Department of Internal Medicine |
| **REVIEW RETURNED** | 21-Jan-2021 |

| | |
|---|---|
| **GENERAL COMMENTS** | The authors have thoroughly addressed most of the previous concerns raised. I have no further comments for the authors. |

| | |
|---|---|
| **REVIEWER** | Groen, Henk<br>Rijksuniversiteit Groningen, Epidemiology |
| **REVIEW RETURNED** | 12-Jan-2021 |

| | |
|---|---|
| **GENERAL COMMENTS** | The authors have addressed most of my comments adequately. |

| | |
|---|---|
| | However, I do not agree with the reasons why a sample size calculation was not performed. The fact that this is an observational study does not imply that sample size calculation is not possible. There are different ways to assess in advance whether the sample size will be sufficient to provide meaningful answers to the research questions.<br>One approach is to calculate the width of 95% confidence for proportions derived from the study in advance, assuming a desirable level of precision. The sample size required to estimate a proportion with a desired precision of 5% based on 95% confidence interval can be calculated.<br>A second approach is to determine the relevant difference in proportions between men and women with and without ACS that should be detectable. In the current manuscript, all statistically significant differences are implicitly assumed to be relevant and conversely all non-significant differences are dismissed as irrelevant. For instance, in line 37-40 on page 9, the following is written:<br>"Chest pain was the most common complaint, both in those with and without an ACS; in women with and without ACS 98.8% and 93.1% (p=0.055), and in men 92.4% and 94.5%, respectively (p=0.364). "<br>The difference in proportions for women is bigger than for men (5.7% vs 2.1%), but neither are significant. However, could the difference for women be clinically relevant?.<br>There are numerous other examples of the same approach, for instance in lines 8-10 on page 10. Here the authors write:<br>"Recognition of symptoms being similar to a previous cardiac event was associated with ACS in men (52.9% vs. 32.1%, p=0.004), but not clearly for ACS in women (32.5% vs. 21.4%, p=0.108). "<br>Admittedly, the difference in proportions for men is clearly bigger than for women (20.8 vs 11.1%) but could the difference be relevant in women?<br>The minimum difference in proportions that would be detectable with sufficient power could have been calculated before the study for a range of proportions to provide more robustness to the results.<br>Strictly speaking, all analyses in a study that is not supported by a sample size calculation are exploratory, as the authors admit in their reply. However, no mention of this is made anywhere in the manuscript, most obviously not among the weaknesses of the study. Meanwhile, p-values are ubiquitous in the results section. This is at odds with the descriptive nature of the study and also with the description "an observational study with descriptive univariable statistics" that the authors provide in their reply.<br><br>I am also not convinced about the reasoning behind the lack of confounding for age. The goal of the study is to look at gender differences in presentation of complaints of acute chest discomfort. If the type of complaints of acute chest discomfort depends on age, then finding the true association between complaints and presence of ACS requires adjustment for confounding by age (if age is |

| | |
|---|---|
| | significantly different between men and women irrespective of ACS or not, this is not specified).<br><br>Finally, I am a little confused about some results in the table 2. The first part of the table is about ACS yes or no and type of ACS. Are these types of ACS mutually exclusive? The analysis in the table do not suggest this, since there are separate p-values for each category. This could be explained in the text, I could not find this specified in the methods. The same applies to the life-threatening events.<br><br>The changes to the manuscript should again be checked for English language. |

| | |
|---|---|
| **REVIEWER** | Boehnke, Jan Rasmus<br>University of Dundee, School of Health Sciences |
| **REVIEW RETURNED** | 08-Jan-2021 |

| | |
|---|---|
| **GENERAL COMMENTS** | Thank you for letting me review the revision of the manuscript. The authors responded to some of the issues raised by the reviewers and the manuscript is more readable. Nevertheless, the reporting still needs further improvement and several of the points raised by the reviewers need to be addressed more coherently or transparently. Also, many of the requests by several reviewers were aimed at providing more detail in the manuscript, but the points were only (and sometimes only partly) addressed in the responses to reviewers.<br><br>MAJOR<br>The two major points that remain from the reviewer comments across several of the reviews submitted are the following.<br><br>1) Firstly, the team decided not to implement changes addressing reviewer#2's and my own points regarding statistical comparisons across gender. Such statistical comparisons, whether performed in a more complex multivariate model or extending the presented analyses by an interaction with gender in each case would answer whether clinicians or telephone triage nurses should actually use different symptom lists for phone calls by males and females. Only if this interaction analysis is statistically significant (and potentially even: if the analysis shows a relevantly sized effect), a recommendation could be made that a symptom should be considered for some callers, but not for others. This has a number of consequences for the manuscript.<br><br>1a) The authors responded to this point: " As pointed out in the introduction and discussion, it is crucial for clinician and telephone triage nurse to know what symptoms are associated with the diagnosis ACS, and whether this is different in men than women." As stated above, the analysis does not provide any evidence whether the predictors actually differ across gender of the caller. |

| | |
|---|---|
| | 1b) It is notable that a number of the studies on which the authors build their analyses and the evidence they use actually performed such tests to support their claims regarding gender differences (e.g., REFs 11, 12 (on aggregate), 13, 16 (derivable from LR+ reporting)). And REF17 makes no claims regarding the usefulness of gender differences in practice, the authors do not claim that the descriptive differences shown in the paper should be used for triage, quite to the contrary, "Discriminating ACS in patients with chest discomfort who contacted primary care OHS is difficult in both women and men." This should be considered when placing the current study in the context of previous research and justifying its current approach.<br><br>1c) REF23 does not provide evidence that states "retrosternal chest pain was discriminative for diagnosing ACS in both sexes was also reported in a study among 2,475 patients with acute chest pain in a multicentre ED-study". The study evaluated only a differential effect across ethnicity and does not make any reference to gender ("sex" in that study only treated as a covariate).<br><br>1d) The conclusion of the paper reads now much more in line with the evidence, after changes to the terminology (diagnostic, association etc) were implemented. Nevertheless, the conclusion is still not supported by the presented analyses: "There were more similarities than differences in symptoms associated with the diagnosis ACS for women and men. Important exceptions were severity, type, and radiation of pain, and in women a pale face, and in men sweating." This statement even calls for a more complicated set of analyses as the "similarities" could only be shown with statistical tests of equality and the statistical tests to support the conclusions about "important exceptions" are not presented as we do not know whether their association is actually statistically different across genders (applies also to the paragraph on the bottom of page 11). As these are the words of the team as to how their results should be interpreted, I respectfully disagree that the "by-gender" analysis would actually require them to write a different paper: it is instead an essential part to support the Conclusion they present.<br><br>2) The second point relates more generally to the fact that the authors chose not to apply more robust validation procedures of the associations (see also reviewer#3's comments), the impact of missing data was not properly accounted for, and that no evaluation of the strength of the relationships (i.e. their (relative) predictive power and the number and type of potential diagnostic errors has been undertaken) is presented. The paper falls therefore short of a number of good practice recommendations for the development of diagnostic algorithms (e.g., TRIPOD for an overview of which points should be considered and |

| | |
|---|---|
| | reported, BMJ 2015; 350:g7594. PMID: 25569120). And although the authors claim that they are not aiming to develop one and the research question and conclusion have been adjusted accordingly, the paper is still riddled with references to the question what clinicians and triage nurses should be alert to in patient communication.<br>I would therefore suggest that the authors provide clarification on this point in the limitations section: "As the intention of our analysis was to describe whether symptoms were different in patients with ACS from patients without ACS in women and men separately, none of our results can be used to inform the development of interview schedules." This could be extended by "For this purpose, an evaluation of the differential relation with gender as well as a robust evaluation of the predictive power of the symptoms would have needed to take place which are beyond the scope of the current paper". But I feel the shorter sentence would be enough to be explicit for readers what the extent of the use of the results is.<br>Alternatively, the language around the use of symptoms in diagnostic interviews in multiple places of the paper could be toned down, but I agree that this information is useful for context.<br><br><br>OTHER POINTS<br>3) The team did not offer any clarification in the manuscript regarding this original point:<br>"2) The introduction and discussion could benefit from a more differentiated take on what the authors mean when they use the term gender in this case? The introduction should cover this as this is one of the key variables and it is crucial for understanding about what exactly it is the team wanted to learn and about what we are learning as readers. It seems from the nature of the data covered it would be self-identified/-signified (how many of the callers did not fall into the male/female categories?), but this is only what I can infer from the presented information at the moment."<br>While a response to reviewers is appreciated, the clarification is required for readers of the manuscript.<br><br><br>4) The authors did not address the point regarding missing data sufficiently. The team chose not to provide any imputed results, despite substantial missing data. This is clear and evident from their report. But in that case, the sample size for each statistic needs to be reported as this would be different for most statistical tests presented in the paper.<br><br><br>5) Figure 1 was not available in the word document or the uploaded appendices. The points raised by reviewers regarding patient flow and selection remain therefore |

| | |
|---|---|
| | unaddressed at the moment. This refers to the following point from the previous review and several parts of this process are still not sufficiently described in the manuscript.<br>OLD-3d) The sample selection is also not described in sufficient detail (e.g., STROBE guidelines). How many calls were available? How many of these had the relevant codes and key words? How large was the random sample? How many of the selected were below the age of 18, didn't live in the area, or were calls of low quality? (The results section reads to me as if the 1795 are only the last number of this process, but this is not clear either, it is simply the number of calls analysed.)<br><br><br><br>6) The point regarding a justification of the sample size, raised by two reviewers, was not addressed. Multiple methods for the sample size calculation in naturalistic and particularly diagnostic studies exist (e.g., https://doi.org/10.1016/j.jbi.2014.02.013, DOI: 10.1111/j.1553-2712.1996.tb03538.x) and could have been applied. The justification presented by the authors referring to a "convenient number of patients" is also incomplete, as it does not define why this number is convenient and sufficient for the purposes of the current study.<br><br><br>7) These two points have not affected any changes in the manuscript<br>OLD-3) A bit more detail on the process of selecting the interviews could be useful. [could be addressed once the flow diagram is included, but from the text alone this remains unclear. This addresses STROBE items 6-7 and reporting is required.]<br>OLD-3c) How was the random sample drawn with Excel? Were all calls listed and then random numbers assigned? [The selection procedure is still not described in the manuscript; this STROBE item no 6 and reporting required.]<br><br><br><br>MINOR<br>8) "Another ED-study among 1,334 patients with ACS showed that regardless of ethnics status the most common presenting symptom was retrosternal pain/discomfort of any intensity." It is unclear why this data is presented as the current paper deals with gender, not with ethnicity.<br><br>9) Finally, the revision has many typos, i.e. more thorough copy-editing is necessary. |

**VERSION 2 – AUTHOR RESPONSE**

| Comments reviewer 4 | Response from authors |
|---|---|
| Firstly, the team decided not to implement changes addressing reviewer#2's and my own points regarding statistical comparisons across gender. Such statistical comparisons, whether performed in a more complex multivariate model or extending the presented analyses by an interaction with gender in each case would answer whether clinicians or telephone triage nurses should actually use different symptom lists for phone calls by males and females. Only if this interaction analysis is statistically significant (and potentially even: if the analysis shows a relevantly sized effect), a recommendation could be made that a symptom should be considered for some callers, but not for others. This has a number of consequences for the manuscript.<br><br>1a) The authors responded to this point: " As pointed out in the introduction and discussion, it is crucial for clinician and telephone triage nurse to know what symptoms are associated with the diagnosis ACS, and whether this is different in men than women." As stated above, the analysis does not provide any evidence whether the predictors actually differ across gender of the caller.<br><br>1b) It is notable that a number of the studies on which the authors build their analyses and the evidence they use actually performed such tests to support their claims regarding gender differences (e.g., REFs 11, 12 (on aggregate), 13, 16 (derivable from LR+ reporting)). And REF17 makes no claims regarding the usefulness of gender differences in practice, the authors do not claim that the descriptive differences shown in the paper should be used for triage, quite to the contrary, "Discriminating ACS in patients with chest discomfort who contacted primary care OHS is difficult in both women and men." This should be considered when placing the current study in the context of previous research and justifying its current approach.<br><br>1c) REF23 does not provide evidence that states "retrosternal chest pain was discriminative for diagnosing ACS in both sexes was also reported in a study among 2,475 patients with acute chest pain in a multicentre ED-study". The study evaluated only a differential effect across ethnicity and does not make any reference to gender ("sex" in that study only treated as a covariate).<br><br>1d) The conclusion of the paper reads now much more in line with the evidence, after changes to the terminology (diagnostic, association etc) were implemented. Nevertheless, the conclusion is still not supported by the presented analyses: "There were more similarities than differences in symptoms associated with the diagnosis ACS for women and men. Important exceptions were severity, type, and radiation of pain, and in women a pale face, and in men sweating." This statement even calls for a more complicated set of analyses as the "similarities" could only | For this rebuttal we have consulted dr. M. van Smeden (statistician).<br>We have adjusted methodological terms to clarify the aim and nature of this study, and have performed -as requested- new analyses including interaction analyses and have added mean differences with 95% confidence intervals to table 1.<br><br>We have now also further clarified the nature of this study. The study should be considered a *diagnostic factor study*, analogous to a prognostic factor study (Riley et al. 2013). It was explicitly not our aim to develop a diagnostic algorithm or model. Instead, our study aims to identify the relevant diagnostic factors associated with ACS in people suspected of ACS.<br><br>Abstract and introduction of the revised manuscript:<br>*"**Objective**s: To identify clinical variables that are associated with the diagnosis acute coronary syndrome (ACS) in women and men with chest discomfort who contact out-of-hours primary care (OHS-PC) by telephone, and to explore whether there are indications these variables differ among women and men."*<br><br>Ref: Riley RD, Hayden JA, Steyerberg EW, Moons KG, Abrams K, Kyzas PA, Malats N, Briggs A, Schroter S, Altman DG, Hemingway H and Group P. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS medicine*. 2013;10:e1001380<br><br>1a) We thank the reviewer for suggesting interaction analysis across gender, and we have performed such analysis accordingly (table 1). The results of the two-way interaction analysis (gender x covariate) indicate there are indeed some significant differences (at $\alpha$ 0.05) in symptoms related to ACS among men and women. We now conclude in the revised manuscript that these give an indications of gender differences, but that the impact on predicting ACS still needs to be further investigated.<br><br>1b) We have rephrased the sentence in the introduction of the revised manuscript |

| | |
|---|---|
| be shown with statistical tests of equality and the statistical tests to support the conclusions about "important exceptions" are not presented as we do not know whether their association is actually statistically different across genders (applies also to the paragraph on the bottom of page11). As these are the words of the team as to how their results should be interpreted, I respectfully disagree that the "by-gender" analysis would actually require them to write a different paper: it is instead an essential part to support the Conclusion they present. | into*: "In a recent Dutch OHS-PC study among 23 women and 34 men with ACS, and 253 women and 208 men without ACS, symptoms associated with ACS in women and men seemed quite similar, and the authors conclude that discriminating ACS in patients with chest discomfort who contacted primary care OHS is difficult in both women and men."*<br><br>1c) In table 2 and figure 1 from reference 23 (Rubini et al, 2014 JAMA) 'mid chest' pain (by us rephrased into retrosternal chest pain), and radiation to the arm were associated with the diagnosis ACS in both sexes.<br>We have adjusted the sentence to:<br>*"Our finding that retrosternal ('mid') chest pain, and radiation of pain to the arm was associated with the diagnosis ACS in both sexes was also reported in a study among 2,475 patients with acute chest pain in a multicentre ED-study."*<br><br>1d) We agree with the reviewer that strong conclusions regarding similarity would require alternative methods of analyses, such as equivalence testing by setting equivalence bounds, which we believe to be out of scope of this article. We have therefore formulated our conclusions more carefully, following the suggestion of the reviewers, as we have detailed in our answers 1a) above. |
| 2. The second point relates more generally to the fact that the authors chose not to apply more robust validation procedures of the associations (see also reviewer#3's comments), the impact of missing data was not properly accounted for, and that no evaluation of the strength of the relationships (i.e. their (relative) predictive power and the number and type of potential diagnostic errors has been undertaken) is presented. The paper falls therefore short of a number of good practice recommendations for the development of diagnostic algorithms (e.g., TRIPOD for an overview of which points should be considered and reported, BMJ 2015; 350:g7594. PMID: 25569120). And although the authors claim that they are not aiming to develop one and the research question and conclusion have been adjusted accordingly, the paper is still riddled with references to the question what clinicians and triage nurses should be alert to in patient communication.<br>I would therefore suggest that the authors provide clarification on this point in the limitations section: "As the intention of our analysis was to describe whether symptoms were different in patients with ACS from patients without ACS in women and men separately, none of our results can be used to inform the development of interview schedules." This could be extended by "For this purpose, an evaluation of the differential relation with gender as well as a robust | We are very familiar with the TRIPOD-criteria, as they were partly developed at our research center. However, as we have not developed a prediction model, the TRIPOD guideline does not apply to our study.<br><br>We have now clarified this in the discussion of the revised manuscript,<br>including the reviewer's suggested sentence:<br>*"As the intention of our analysis was to describe whether symptoms were different in patients with ACS from patients without ACS in women and men separately, none of our results can be used to adjust interview questions for the triage nurses. For this purpose, prediction rule development with multivariable analyses is necessary. Also, with such multivariable analysis it would be truly investigated whether the potential differences in sex are clinically relevant in prediction of ACS. "*<br><br>For diagnostic factor studies there are no formal reporting guidelines. Therefore, we used the STROBE guidelines for observational |

| | |
|---|---|
| evaluation of the predictive power of the symptoms would have needed to take place which are beyond the scope of the current paper". But I feel the shorter sentence would be enough to be explicit for readers what the extent of the use of the results is.<br>Alternatively, the language around the use of symptoms in diagnostic interviews in multiple places of the paper could be toned down, but I agree that this information is useful for context. | research. We have carefully re-examined the STROBE guidelines and added some additional information in the revised manuscript, in particular to figure 1 we now provide detailed information about the selection process.<br><br>Missing data is indeed a limitation of the study, we have mentioned this in the discussion section:<br>*"Another limitation is missing values on some clinical variables, a phenomenon common when using routine care data, and therefore, the results should be interpreted with caution*." |
| 3. The team did not offer any clarification in the manuscript regarding this original point:<br>"2) The introduction and discussion could benefit from a more differentiated take on what the authors mean when they use the term gender in this case? The introduction should cover this as this is one of the key variables and it is crucial for understanding about what exactly it is the team wanted to learn and about what we are learning as readers. It seems from the nature of the data covered it would be self-identified/-signified (how many of the callers did not fall into the male/female categories?), but this is only what I can infer from the presented information at the moment."<br>While a response to reviewers is appreciated, the clarification is required for readers of the manuscript. | We have added now 'self-identified' gender to the method section of the revised manuscript.<br><br>*"Gender considered the self-identified gender of the patient."* |
| 4. The authors did not address the point regarding missing data sufficiently. The team chose not to provide any imputed results, despite substantial missing data. This is clear and evident from their report. But in that case, the sample size for each statistic needs to be reported as this would be different for most statistical tests presented in the paper. | Yes, we agree this is a limitation of the study. Nevertheless, we have decided not to perform multiple imputation for missing data for this particular study because our interest was in the univariable associations and the interaction with gender, thus, to identify potential diagnostic factors. We agree that the missing data may introduce some bias (in particular, if missing data is not completely at random), and thus we acknowledge this as a limitation in the discussion of the revised manuscript. Arguably the effect of this bias may be less severe than when, for instance, a full prediction model would have been developed, in which the combination of missingness on several variables might contribute to a much larger reduction in the available sample in a complete case analysis. |
| 5. Figure 1 was not available in the word document or the uploaded appendices. The points raised by reviewers regarding patient flow and selection remain therefore unaddressed at the moment. This refers to the following point from the previous review and several parts of this process are still not sufficiently described in the manuscript.<br>OLD-3d) The sample selection is also not described in sufficient detail (e.g., STROBE guidelines). How many calls | We apologize to the reviewer he could not found Figure 1. Hopefully, it is now assessable in the revised manuscript. |

| | |
|---|---|
| were available? How many of these had the relevant codes and key words? How large was the random sample? How many of the selected were below the age of 18, didn't live in the area, or were calls of low quality? | |
| 6. The point regarding a justification of the sample size, raised by two reviewers, was not addressed. Multiple methods for the sample size calculation in naturalistic and particularly diagnostic studies exist (e.g., https://doi.org/10.1016/j.jbi.2014.02.013, DOI: 10.1111/j.1553-2712.1996.tb03538.x) and could have been applied. The justification presented by the authors referring to a "convenient number of patients" is also incomplete, as it does not define why this number is convenient and sufficient for the purposes of the current study. | We thank the reviewer for the reference with sample size suggestion, however this reference includes methods for diagnostic accuracy studies (sens/spec), likelihood ratio's and AUC and therefore does not apply to our study. Unfortunately there seem to be no methods for formal sample size calculation of a diagnostic factor study (unlike diagnostic modeling studies), which is why we choose a convenient sample.<br><br>*"Adequate methods for sample calculation of a diagnostic factor study is yet lacking. We therefore included a convenient number of patients, that was, at least 80 patients with ACS in each sex category. This number was chosen primarily based for practical ad feasibility reasons."* |
| 7a. A bit more detail on the process of selecting the interviews could be useful. [could be addressed once the flow diagram is included, but from the text alone this remains unclear. This addresses STROBE items 6-7 and reporting is required. | We kindly refer to our answer to question 5 above. |
| 7b. How was the random sample drawn with Excel? Were all calls listed and then random numbers assigned? The selection procedure is still not described in the manuscript; this STROBE item no 6 and reporting required. | We have indeed applied the randomization approach as described by the reviewer. We have adjusted the sentence in the method section of the revised manuscript:<br>*"We listed all available calls of these patients and assigned random numbers with the Random Number Generator (RAND) function in Microsoft Excel to retrieve a random sample."* |
| 8. "Another ED-study among 1,334 patients with ACS showed that regardless of ethnics status the most common presenting symptom was retrosternal pain/discomfort of any intensity." It is unclear why this data is presented as the current paper deals with gender, not with ethnicity | We have added a reference about ethnicity on the request of another reviewer in the first rebuttal, to add information about other possible factors that might affect how patients describe their symptoms. |
| 9. Finally, the revision has many typos, i.e. more thorough copy-editing is necessary. | Thank you. We have double checked our English language throughout the paper. |
| **Comments reviewer 3** | |
| The authors have addressed most of my comments adequately.<br>However, I do not agree with the reasons why a sample size calculation was not performed. The fact that this is an observational study does not imply that sample size calculation is not possible. There are different ways to assess in advance whether the sample size will be sufficient to provide meaningful answers to the research questions. One approach is to calculate the width of 95% confidence for proportions derived from the study in advance, assuming a desirable level of precision. The sample size required to estimate a proportion with a desired precision of 5% based | Please also see our answer to question 6 of reviewer four.<br>We agree with the reviewer that in retrospect a sample size calculation could have been performed based on the power or desired precision of confidence intervals for the univariable associations. However, for interaction this would have been slightly more difficult. Therefore, we stick to the convenient sample size was chosen primarily based on practical and feasibility reasons. |

| | |
|---|---|
| on 95% confidence interval can be calculated. | |
| A second approach is to determine the relevant difference in proportions between men and women with and without ACS that should be detectable. In the current manuscript, all statistically significant differences are implicitly assumed to be relevant and conversely all non-significant differences are dismissed as irrelevant. For instance, in line 37-40 on page 9, the following is written:<br>"Chest pain was the most common complaint, both in those with and without an ACS; in women with and without ACS 98.8% and 93.1% (p=0.055), and in men 92.4% and 94.5%, respectively (p=0.364). "<br>The difference in proportions for women is bigger than for men (5.7% vs 2.1%), but neither are significant. However, could the difference for women be clinically relevant?~~.~~ There are numerous other examples of the same approach, for instance in lines 8-10 on page 10. Here the authors write:br />"Recognition of symptoms being similar to a previous cardiac event was associated with ACS in men (52.9% vs. 32.1%, p=0.004), but not clearly for ACS in women (32.5% vs. 21.4%, p=0.108). "<br>Admittedly, the difference in proportions for men is clearly bigger than for women (20.8 vs 11.1%) but could the difference be relevant in women?<br>The minimum difference in proportions that would be detectable with sufficient power could have been calculated before the study for a range of proportions to provide more robustness to the results.<br>Strictly speaking, all analyses in a study that is not supported by a sample size calculation are exploratory, as the authors admit in their reply. However, no mention of this is made anywhere in the manuscript, most obviously not among the weaknesses of the study. Meanwhile, p-values are ubiquitous in the results section. This is at odds with the descriptive nature of the study and also with the description "an observational study with descriptive univariable | We thank the reviewer for this valuable suggestion to add more detailed information about the results. We have now added mean differences with 95% confidence intervals for continuous variables and differences of proportions to table 1.<br><br>The reviewer raises the important point of knowing whether differences are clinically relevant. With our exploratory analysis it is not possible to adequately answer that question. For that purpose multivariable regression analysis would be necessary, to assess the interrelationship of symptoms.<br><br>We have added information to the revised discussion:<br>*"As the intention of our analysis was to describe whether symptoms were different in patients with ACS from patients without ACS in women and men separately, none of our results can be used to adjust interview questions for the triage nurses. For that purpose prediction rule development with multivariable analyses is necessary. Only with multivariable analysis it can be truly investigated whether the potential differences are clinically relevant in prediction of ACS."* |
| I am also not convinced about the reasoning behind the lack of confounding for age. The goal of the study is to look at gender differences in presentation of complaints of acute chest discomfort. If the type of complaints of acute chest discomfort depends on age, then finding the true association between complaints and presence of ACS requires adjustment for confounding by age (if age is significantly different between men and women irrespective of ACS or not, this is not specified). | Confounding by age would be an issue if the goal of the analyses was one where the causal relation between certain determinants and the presence of ACS was of interest. The interest in this article is explicitly not to make causal inferences but to identify the relevant diagnostic factors associated with ACS in people suspected of ACS. |
| Finally, I am a little confused about some results in the table 2. The first part of the table is about ACS yes or no and type of ACS. Are these types of ACS mutually exclusive? The analysis in the table do not suggest this, since there are separate p-values for each category. This could be explained in the text, I could not find this specified in the methods. The same applies to the life-threatening events. | Thank you for giving us the opportunity to clarify information of table 2.<br>The types of ACS are indeed mutually exclusive. We have now added to the introduction (according to the European Society of Cardiology (ESC) guidelines): "*For the diagnosis of ACS an abnormal electrocardiogram (ST and/or T wave ) and/or elevated blood levels of troponin I or T are needed. ACS may than be further subdivided in ST-elevation myocardial infarction (STEMI) and* |

| | |
|---|---|
| | *non-ST-elevation myocardial infarction (NSTEMI) if the troponin levels are elevated [1]. If troponin levels are not elevated (or increase over time), it is unstable angina pectoris (UAP). [1] "*<br><br>Ref: Roffi et al. 2015 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: Task Force for the Management of Acute Coronary Syndromes in Patients Presenting without Persistent ST-Segment Elevation of the European Society of Cardiology (ESC). *European heart journal*. 2016;37:267-315.<br><br>The types of LTEs are also mutually exclusive. |

## VERSION 3 – REVIEW

| | |
|---|---|
| **REVIEWER** | Groen, Henk<br>Rijksuniversiteit Groningen, Epidemiology |
| **REVIEW RETURNED** | 15-Apr-2021 |

| | |
|---|---|
| **GENERAL COMMENTS** | The authors have addressed my questions clearly and elaborately. I am satisfied with the answers regarding sample size and complaints specific for men and women which was addressed by regression with interaction terms. However, if a signifcant interaction is found, the direction of the interaction should be mentioned.<br>The fact that the study design is now described as a 'diagnostic factor study' adds to the clarity regarding the aim of the study.<br>The ages of men and women with and without ACS are presented in table 1 and the absence of a significant interaction indicates that age does not have a differential effect on the risk of a diagnosis of ACS between men and women. |

| | |
|---|---|
| **REVIEWER** | Boehnke, Jan Rasmus<br>University of Dundee, School of Health Sciences |
| **REVIEW RETURNED** | 20-Apr-2021 |

| | |
|---|---|
| **GENERAL COMMENTS** | Thank you for letting me review the revision of the manuscript. The original interpretations of the results have either been adjusted or supported by appropriate analyses. The much clearer presentation of the research question and more careful interpretation of the results also makes several points raised by me in the past obsolete (as pointed out by the team).<br><br>MINOR<br>1) There are some row-alignment issues in pdf version of table 1 (e.g., rows "Mean patient's introduction duration in min (SD)" and below; not visible in the HTML version). |

| | |
|---|---|
| | They could be due to format conversion, but if not, this should be corrected. |

## VERSION 3 – AUTHOR RESPONSE

| Comment reviewer 3 | Response from authors |
|---|---|
| The authors have addressed my questions clearly and elaborately. I am satisfied with the answers regarding sample size and complaints specific for men and women which was addressed by regression with interaction terms. However, if a significant interaction is found, the direction of the interaction should be mentioned. | We thank the reviewer for the suggestion and have now mentioned the direction of the interaction for the symptoms in which a significant sex interaction was found.<br><br><u>Results:</u><br>*"Only in women radiation to the jaw had an association with ACS (women 50.0% vs. 22.9%, p=0.007, men 23.6% vs. 30.4%, p=0.312, gender interaction p=0.015) and severe pain (8 or more on a Numeric Rating Scale 0-10) (65.4% vs. 38.1%, p=0.006, men 2.6% vs. 11.3%, p=0.098, gender interaction p=0.007), which had a differential effect towards the risk of ACS in women. Only in men, stabbing pain was very rare in those with ACS (8.4% vs. 26.5%, p<0.001,) however this had not have a differential effect on the diagnosis of ACS between men and women (gender interaction p=0.141). Of the autonomous nervous system (ANS)-related symptoms, nausea/vomiting and dizziness/near fainting were not associated with ACS in either sex. A pale or ashen face was associated with ACS in women (55.6% vs. 35.5%, p=0.019, gender interaction p=0.545), and sweating in men (52.4% vs. 38.1%, p=0.015, gender interaction p=0.418), however without a differential effect on the risk of diagnosis of ACS between women and men."* |
| **Comment reviewer 4** | **Response from authors** |
| There are some row-alignment issues in pdf version of table 1 (e.g., rows "Mean patient's introduction duration in min (SD)" and below; not visible in the HTML version). They could be due to format conversion, but if not, this should be corrected. | We apologize to the reviewer about the row-alignments issues of Table 1. We have resolved the issues in this version and hopefully it is better readable now. To be sure, we have also added a PDF version of table 1 to the submission. |