PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Are there gender disparities in symptom presentation or triage of patients with chest discomfort at primary care out-of-hours services? An observational study
AUTHORS	van der Meer, Manon; Appelman, Yolande; Rutten, Karlijn; van der Graaf, Yolanda; Nathoe, Hendrik; Doevendans, Pieter; Smit, Michelle; Verheij, Emmy; Botermans, Anne; Rutten, Frans

VERSION 1 – REVIEW

REVIEWER	Larisa Burke University of Illinois at Chicago, USA
REVIEW RETURNED	05-Jun-2019

GENERAL COMMENTS	Manuscript Review Larisa Burke BMJ Open 6/5/2019 Summary Generally a well written manuscript which reports on differences in symptoms, urgency allocation and ACS diagnosis for patients seeking treatment for chest discomfort in the Netherlands. However, the way the statistical analyses and results are presented could be improved. Abstract Regarding these values: “5.22 versus 7.26 minutes, p-value=0.003, and 6.27 versus 7.22 minutes, p=0.087, respectively.” What are these values (means?). Can you add a statistics reflecting how much the values vary (eg. SD). What statistical test was used? Regarding these values: (95.7% versus 88.2%, p-value=0.331). Can you add the n’s? What statistical test was used? “In 37.7% of cases we did not enough receive information from the patients’ GP to determine a diagnosis. This did not bias our results as determinants were similar between participants and non-participants.” This statement wasn’t clear when first reading. Suggest: 37.7% of cases could not be included as participants… this did not bias our results because characteristics were similar between participants and non-participants.

	Data Analysis By what standard were covariates selected for the multivariable modeling? For logistic regression models: -Was ACS diagnosis the outcome or was urgency allocation the outcome? -Currently the logistic regression model presented in table 3 doesn't test whether "Women with chest discomfort were not under-triaged compared to men with chest discomfort". Using a gender interaction term with the logistic model instead of stratifying by gender would allow for gender comparisons. The interaction term would test that they ACS/urgency allocation OR for men (OR=13) is/isn't different from the ACS/urgency allocation OR for women (OR=5). If stratifying by gender no statistical comparisons between genders is possible. The analyses were repeated after adding all potential life-threatening diagnoses to ACS, including pulmonary embolism, pneumothorax, aortic dissection and acute heart failure since a high urgency would be appropriate in all such cases." Are these analyses reported on? What is the result? Why was this done? Results In the results write-up, it is unclear what standard is being used to present when there is a true difference between groups versus when there is not. For instance, it is stated that there is a group difference in stabbing pain between women with and without ACS based on statistically non-significant 3% difference (pain was less frequent in women and men with ACS than in those without ACS; in women 15.8% vs. 18.8%, p-value=0.073). But a 7% non-statistically significant difference between women and men getting high-urgency allocation was noted as supporting no difference between the groups (women with an ACS got at least as often a high urgency allocation as men with an ACS (95.7% vs. 88.2%, p-value=0.331) There needs to be a clear standard for what is/is not reported as a group difference in the findings of this manuscript. If reporting non-statistically significant results, researchers should note that the results may be due to chance/may not be actually represent the true effect in the population. "women with an ACS got at least as often a high urgency allocation as men with an ACS (95.7% vs. 88.2%, p-value=0.331); see Table 3" Is this a separate test that was done unrelated to the ORs that are presented in table 3? This is confusing because the details of the test and the p-value of 0.331 don't appear in table 3. "The chance of receiving a high urgency allocation with ACS was not affected by age" Can authors explain how this was determined? Is this for both men and women? It doesn't seem to support adjusting for age in the logistic regression model. Why is the finding with the logistic regression that women/men with ACS have a 13/5 times greater odds of a high urgency allocation not reported in the results when discussing table 3? It seems to be a main focus of table 3 and the results. It should be noted that the OR for women has a quite large confidence interval, meaning authors can't be confident of the true value.

	Discussion Discussion of group differences and statistical significance is much clearer/well presented in discussion. “Women were not under-triaged, and those with ACS received at least as many high urgency allocations as men”. Unclear if this is referring to the logistic regression results in table 3. If so, it does not seem to be a correct way to summarize (correct = both men and women with ACS had a higher odds of having urgent allocation versus non-urgent). No gender comparison possible due to stratification/no interaction term. “Fifth, the low number of ACS cases did not allow for multivariable logistic regression analysis in men and women separately.” You did do a logistic regression of men and women separately in table 3. I think authors are referring specially to assessing symptoms but this not stated. Conclusion “Women with chest discomfort were not under-triaged compared to men with chest discomfort in the primary care OHS.” Don’t know what statistical test supports this conclusion. The reviewer provided a marked copy with additional comments. Please contact the publisher for full details.

REVIEWER	Sofia Sederholm Lawesson Department of Cardiology and Department of Medicine and Health Linköping University Hospital SE-58185 Linköping Sweden
REVIEW RETURNED	25-Jul-2019

GENERAL COMMENTS	The present study aimed to compare the genders in symptom presentation and in triaging of chest pain patients contacting primary care by phone. In total, 276 women and 242 men in the Netherlands in 2013 and 2014 were included. Differences in symptoms between patients with and without ACS were in general small, for both genders. Women and men with ACS received equally often a high urgency allocation (95.7% versus 88.2%, p-value=0.331). 1. The objectives of the study ought to be focusing on new research areas. It is already wellknown that, 1) in an unselected chest pain population men will have ACS to a higher extent than women and, 2) in an ACS population women have less chest pain and more atypical symptoms than men. Anyhow, the question whether there is a gender difference in triaging chest pain patients based on phone interviewing is interesting. 2. In addition, it would be interesting to evaluate the NTS system, i.e. whether this system correctly identify the most critically ill patients including ACS patients and if it works equally well in both genders. It would also be interesting to identify factors associated with the NTS score. 3. Strangely, in the abstract, the question whether there is a gender difference in triaging chest pain patients based on phone interviewing is never answered, i.e. whether men were more often

	triaged than women as having an emergency condition, as ASC. The percentage of ACS patients (sex stratified) triaged as emercency/urgent is presented, but not whether this differ from chest pain patients not recieving an ACS diagnosis. Instead, the difference in lenght of phone calls between ACS and non-ACS patients, stratified by sex, is presented, which is not among the objectives of the study. 4. The main problem with the present study is a lack of power comparing triage between the genders, escpecially using multivariable adjustment (only adjusting for age was done). According to the STROBE protocol a study size consideration was done, which I cannot find (power calculation). Only 23 women and 34 men with ACS were included. The logistic regression analysis adds little without the possibility to adjust for possible confounders (i.e. age, co-morbidities, socioeconomic data etc) - now only age-adjustment is done. The very wide CIs point out the lack of power/too small study cohort (too few ACS cases). A simple Chi-square test could be done (but does not fix the main problem with too few ACS patients included). 4. Another problem is the risk of selection bias. In more than one third of the cases the authors did not get access to the final discharge diagnosis of the patient. 5. There is a limitation of data gathered about the patients, not only on symptom presentation in more detail, but on co-morbidities and other important information that could be gender biased such as socioeconomic data, living condition, educational level etc. In addition, the information seems to have been gathered in an unstructured manner. 6. The authors should be more careful on how they describe their results from the statistics. A p-value of less than .05 is most often considered as identifying a statistically significant difference between groups. In the result section, several ”differences” between groups are highlighted in spite of a p-value far higher than .05, e.g. ”Women and men with ACS were on average older (women 66.8 vs. 62.8 years, p-value=0.184, and men 68.1 vs. 58.8 years, p-value=0.224)”. 7. A minor limitation is the lack of information on type of ACS in a fairly large percentage of the men.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1 Larisa Burke, University of Illinois at Chicago, USA

Generally a well written manuscript which reports on differences in symptoms, urgency allocation and ACS diagnosis for patients seeking treatment for chest discomfort in the Netherlands. However, the way the statistical analyses and results are presented could be improved.

1. Abstract. Regarding these values: “5.22 versus 7.26 minutes, p-value=0.003, and 6.27 versus 7.22 minutes, p=0.087, respectively.” What are these values (means?). Can you add a statistics reflecting how much the values vary (eg. SD). What statistical test was used?

We thank the reviewer. These values were means. In the revised manuscript and in Table 1 we added the standard deviation (SD). As mentioned on page 6 under statistical analyses, we used the student's t-test or Mann-Whitney U test (if not normally distributed) for categorical variables.

We revised this sentence to: "In women with ACS compared to women without ACS, mean duration of telephone calls was discriminative; 5.22 (SD 2.53) versus 7.26 (SD 3.11) minutes, p-value=0.003.''

2. Regarding these values: (95.7% versus 88.2%, p-value=0.331). Can you add the n's? What statistical test was used?

We added the n's in the revised abstract (22/23 in women and 30/34 in men with ACS). We used the Chi square test to calculate this p-value.

The revised abstract reads as follows: "Women with ACS received a high urgency allocation in 22/23 (95.7%) and men with ACS in 30/34 (88.2%), p-value=0.331."

3. "In 37.7% of cases we did not receive information from the patients' GP to determine a diagnosis. This did not bias our results as determinants were similar between participants and non-participants."

This statement wasn't clear when first reading. Suggest: 37.7% of cases could not be included as participants… This did not bias our results because characteristics were similar between participants and non-participants.

We thank the reviewer for the valuable suggestion. We incorporated it in the revised manuscript.

"37.7% of cases could not be included as participants, since we did not receive information from the patients' GP to make a diagnosis. This did not bias our results because patient characteristics were similar between participants and non-participants."

4. Data Analysis. By what standard were covariates selected for the multivariable modeling?

We are pleased to elucidate this further. First of all, given the low number of patients with ACS, we decided not to perform a multivariable regression analysis with ACS as the outcome. The more because on some symptom variables we had missing values as is unfortunately rather common when using routine care data. We also decided beforehand not to go for a full multivariable analysis with urgency allocation (high vs. low) as the outcome. This, because we were merely interested in disparities in urgency allocation among women and men with chest discomfort, also after correction

for ACS, and age. Thus, we did not select covariates based on p-values with univariate analysis.

5. For logistic regression models: Was ACS diagnosis the outcome or was urgency allocation the outcome?

Please see our answer to the previous question. Urgency allocation was the outcome.

We changed the text in the methods section as follows: ''We used multivariable logistic regression analysis with urgency allocation (high vs. low) as the outcome to assess differences between men and women with chest discomfort, also after adjustment for ACS and age.''

6. Currently the logistic regression model presented in Table 3 doesn't test whether "Women with chest discomfort were not under-triaged compared to men with chest discomfort". Using a gender interaction term with the logistic model instead of stratifying by gender would allow for gender comparisons. The interaction term would test that they ACS/urgency allocation OR for men (OR=13) is/isn't different from the ACS/urgency allocation OR for women (OR=5). If stratifying by gender no statistical comparisons between genders is possible.

Thank you, agreed. We performed the analyses as recommended and added the results in the text. Moreover we revised Table 3 and added an extra Table (Table 2).

7. The analyses were repeated after adding all potential life-threatening diagnoses to ACS, including pulmonary embolism, pneumothorax, aortic dissection and acute heart failure since a high urgency would be appropriate in all such cases."

Are these analyses reported on? What is the result? Why was this done?

We indeed did not reported on these results. It concerned in total nine patients (two women and one man with pulmonary embolism, one women and one man with pneumothorax, a woman with aortic dissection, three women with acute HF). These patients also need a high urgency allocation. That was the reason why we originally repeated our multivariable analysis with urgency allocation (high vs. low) as the outcome and considering as covariate patients with either ACS or LTD (57+9 patients).

We decided to remove this sentence because it may confuse readers and it does not really provide added value.

8. Results. In the results write-up, it is unclear what standard is being used to present when there is a true difference between groups versus when there is not. For instance, it is stated that there is a group difference in stabbing pain between women with and without ACS based on statistically non-significant 3% difference (pain was less frequent in women and men with ACS than in those without ACS; in women 15.8% vs. 18.8%, p-value=0.073). But a 7% non-statistically significant difference

between women and men getting high-urgency allocation was noted as supporting no difference between the groups (women with an ACS got at least as often a high urgency allocation as men with an ACS (95.7% vs. 88.2%, p-value=0.331) There needs to be a clear standard for what is/is not reported as a group difference in the findings of this manuscript. If reporting non-statistically significant results, researchers should note that the results may be due to chance/may not be actually represent the true effect in the population.

We agree with the reviewer and changed our results paragraph by only mentioning statistically significant differences.

9. "Women with an ACS got at least as often a high urgency allocation as men with an ACS (95.7% vs. 88.2%, p-value=0.331); see Table 3"

Is this a separate test that was done unrelated to the ORs that are presented in Table 3? This is confusing because the details of the test and the p-value of 0.331 don't appear in Table 3.

We agree it is confusing. 95.7% vs. 88.2%, p-value=0.331 (Chi square test) is based on comparing all women with chest discomfort (276) to men (242). Although, it can be calculated from Table 3, the corresponding p-value is not mentioned there.

In the revised manuscript (methods and results) we now better clarify which comparison and data was assessed univariately with the chi square test and what data with multivariable analysis.

10. "The chance of receiving a high urgency allocation with ACS was not affected by age" Can authors explain how this was determined? Is this for both men and women? It doesn't seem to support adjusting for age in the logistic regression model.

Why is the finding with the logistic regression that women/men with ACS have a 13/5 times greater odds of a high urgency allocation not reported in the results when discussing Table 3? It seems to be a main focus of Table 3 and the results. It should be noted that the OR for women has a quite large confidence interval, meaning authors can't be confident of the true value.

We thank the reviewer we can clarify this now. As already mentioned before (answers to question 4-6), we adapted our multivariable analysis to the wish of the reviewer and made a new Table 2 and adapted Table 3. The results of the multivariable analysis with urgency allocation (high vs. low) as the outcome are reported and discussed in the revised manuscript. Indeed this is an important outcome. We were interested in disparities in symptoms, but also in urgency allocation between women and men with chest discomfort.

We now start our multivariable analysis with women (n=276) vs. men (n=242) with chest discomfort, and than adjusted for ACS (and age), while originally the starting point was ACS (n=57) and than adjusting for sex, with as a result assessing small numbers comparing women with ACS (23) to men with ACS (34) going along with broad confidence intervals.

11. Discussion. Discussion of group differences and statistical significance is much clearer/well presented in discussion.

“Women were not under-triaged, and those with ACS received at least as many high urgency allocations as men”. Unclear if this is referring to the logistic regression results in Table 3. If so, it does not seem to be a correct way to summarize (correct = both men and women with ACS had a higher odds of having urgent allocation versus non-urgent). No gender comparison possible due to stratification/no interaction term.

We thank the reviewer and as already mention before, we adapted our multivariable analysis. In addition we added Table 2 to better clarify the relation, also univariately. The new results are presented in the results section, Tables 2 and 3, and are discussed in the revised discussion paragraph.

12. “Fifth, the low number of ACS cases did not allow for multivariable logistic regression analysis in men and women separately.” You did do a logistic regression of men and women separately in Table 3. I think authors are referring specially to assessing symptoms but this not stated.

The reviewer is correct, we were not clear in this. We indeed meant multivariable analysis of symptoms with urgency allocation as the outcome.

In the revised manuscript we clarified this as follows: ‘’Fifth, missing values on symptoms prevented us from full multivariable analysis with urgency allocation (high vs. low) as the outcome, and the low number of ACS cases let us decide to refrain from multivariable logistic regression analysis considering symptoms and with ACS (yes/no) as the outcome ACS.

13. Conclusion. “Women with chest discomfort were not under-triaged compared to men with chest discomfort in the primary care OHS.” Don’t know what statistical test supports this conclusion.

This was based on 95.7% vs. 88.2%, p-value=0.331 (Chi square test). We understand that we were possibly not always clear about the domain of investigation. In this case, the whole study population of women and men with chest discomfort.

Reviewer: 2 Sofia Sederholm Lawesson, Department of Cardiology and Department of Medicine and Health, Linköping University Hospital, Linköping Sweden

The present study aimed to compare the genders in symptom presentation and in triaging of chest pain patients contacting primary care by phone. In total, 276 women and 242 men in the Netherlands in 2013 and 2014 were included. Differences in symptoms between patients with and without ACS were in general small, for both genders. Women and men with ACS received equally often a high urgency allocation (95.7% versus 88.2%, p-value=0.331).

1. The objectives of the study ought to be focusing on new research areas. It is already well known that, 1) in an unselected chest pain population men will have ACS to a higher extent than women and, 2) in an ACS population women have less chest pain and more atypical symptoms than men. Anyhow, the question whether there is a gender difference in triaging chest pain patients based on phone interviewing is interesting.

Our research question was not focused on either aspect, but on gender disparities in symptom presentation or triage of patients with chest discomfort at primary care out-of-hours services. In the domain chest discomfort we could only find one study that analyzed the results similar to us by comparing symptoms of women with ACS to those without and men with ACS to men without in the ED.[1] We did not find any study executed in the OHS-PC, or evaluating disparities in urgency allocation. To answer questions related to these two issues relevant for clinical practice, much more research is needed.

We hope that we made clear enough in our discussion paragraph that a clinician is interested in how he/she possibly could discriminate those with ACS from those without, if information from electrocardiography and troponin levels is not (yet) available.

For answering that question adequately, it is necessary to compare women with ACS to women without ACS, and men with ACS to men without ACS. This is different from comparing women to men with ACS. We agree with the reviewer that there are many studies, including systematic reviews specifically addressing this latter issue.

2. In addition, it would be interesting to evaluate the NTS system, i.e. whether this system correctly identify the most critically ill patients including ACS patients and if it works equally well in both genders. It would also be interesting to identify factors associated with the NTS score.

We evaluated the eventual urgency allocation (after digitomizing urgency in high and low) which is a result of the recommended urgency by the NTS system and the triage nurse who can overrule the NTS system. We preferred to use the 'eventual urgency allocation' in our analyses and in the revised Table 3 compared men and women with chest discomfort, also after considering ACS and age.

As also mentioned in our answers to the first reviewer, we pay more attention to these results of urgency allocation in the revised manuscript.

3. Strangely, in the abstract, the question whether there is a gender difference in triaging chest pain patients based on phone interviewing is never answered, i.e. whether men were more often triaged than women as having an emergency condition, as ACS. The percentage of ACS patients (sex stratified) triaged as emergency/urgent is presented, but not whether this differ from chest pain patients not receiving an ACS diagnosis. Instead, the difference in length of phone calls between ACS and non-ACS patients, stratified by sex, is presented, which is not among the objectives of the study.

We agree we did not adequately mention our findings on triage in the abstract. We are pleased to adjust this in the revised abstract.

Regarding phone calling; this is key in the triage, and the duration of it is in a way a surrogate outcome for clarity of the communication. Our finding that (i) calls in women with chest discomfort did not last longer than in men, and (ii) the calls of women with ACS significantly lasted shorter than in women without ACS, both point in the direction that triage nurses do not have more problems of discriminating ACS in women than in men calling the OHS-PC.

The revised abstract reads as follows:

"Women and men with chest discomfort received similar high urgency allocation (crude and adjusted odds ratio after correction for ACS and age; 1.03 (95%CI 0.72-1.48) and 1.04 (95%CI 0.72-1.52), respectively). Women with ACS received a high urgency allocation in 22/23 (95.7%) and men with ACS in 30/34 (88.2%), p-value=0.331."

4. The main problem with the present study is a lack of power comparing triage between the genders, especially using multivariable adjustment (only adjusting for age was done). According to the STROBE protocol a study size consideration was done, which I cannot find (power calculation). Only 23 women and 34 men with ACS were included. The logistic regression analysis adds little without the possibility to adjust for possible confounders (i.e. age, co-morbidities, socioeconomic data etc.) - now only age-adjustment is done.

The very wide CIs point out the lack of power/too small study cohort (too few ACS cases). A simple Chi-square test could be done (but does not fix the main problem with too few ACS patients included).

For our multivariable regression analysis, the outcome was urgency allocation (high vs. low) and not ACS. In the original paper we did indeed assess men and women with ACS separately which resulted in broad confidence intervals. As recommended by reviewer 1, we now repeated the multivariable analysis with sex as an interaction term. This allows for more firm conclusions the more because we assess all patients (276 women and 242 men) with chest discomfort.

We agree that the number of patients with ACS (in total 57) was small. We therefore only compared ACS vs. no ACS univariably, applying Chi square testing for categorical variable and Student's t-test or Mann-Whitney U test for continuous variables, thus, being well aware of the shortcomings of low numbers.

We followed the STROBE protocol, but in an observational study comparing characteristics of two groups, without a single hypothesis or intervention, a power calculation is not really possible.

For multivariable logistic regression analysis the 'rule of thumb' of Harrell could be applied.[2] With urgency allocation as the outcome, as we did, 9 to 10 variables could be analyzed (lowest category 'low urgency' 95/10=9.5). We preferred not to do this, but only assessed sex, ACS and age. This we did also because we had missing values on symptoms (as is common with routine care data), and symptoms together with 'gestalt' of the clinician is already incorporated in the variable ACS.

The calculations of the reviewer would be correct if we would have performed MVA with ACS (yes vs.

no) as the outcome. We refrained from doing so because of the low numbers of ACS.

In the revised manuscript we now more clearly stress the risk of change finding due to small numbers when it comes to the comparison of ACS patients vs. no ACS patients. We deleted the sentence about power calculation.

4. Another problem is the risk of selection bias. In more than one third of the cases the authors did not get access to the final discharge diagnosis of the patient.

In our study a selection of patients participated, a common finding in any study when informed consent of patients or the willingness of physicians is needed to provide medical information. Selection, however, does not necessarily mean bias; that participation is confounder for the outcome. In our study, it is unlikely that the willingness of the GP to provide the eventual diagnosis is related to the triage or outcome of ACS weeks/months earlier.

We were in the fortunate circumstances that we could compare non-participants with participants, and important patient characteristics did not differ significantly (page 8 at the top of Results section 'There were no differences in sex, age, duration of the telephone calls, and urgency allocation between participants and patients in whom the medical diagnosis could not be retrieved.") Retrieval of outcomes was hampered in one third because the general practitioner did not cooperate in those situations.

5. There is a limitation of data gathered about the patients, not only on symptom presentation in more detail, but on co-morbidities and other important information that could be gender biased such as socioeconomic data, living condition, educational level etc. In addition, the information seems to have been gathered in an unstructured manner.

We used routine care data, and many of the mentioned items are not part of routine care at the PC-OHS. Moreover, as is common with routine care data, there were missing data on certain variables. We clearly mentioned this shortcoming in the limitations paragraph. Importantly, however, we could assess in total seven symptoms, three patient characteristics (sex, age and history of CVD) and the duration of the calls.

6. The authors should be more careful on how they describe their results from the statistics. A p-value of less than .05 is most often considered as identifying a statistically significant difference between groups. In the result section, several "differences" between groups are highlighted in spite of a p-value far higher than .05, e.g. "Women and men with ACS were on average older (women 66.8 vs. 62.8 years, p-value=0.184, and men 68.1 vs. 58.8 years, p-value=0.224)".

We agree with the reviewer and in the revised manuscript we only mention in the text significant differences.

7. A minor limitation is the lack of information on type of ACS in a fairly large percentage of the men.

We agree, but unfortunately, that is everyday clinical practice in the Netherlands, where cardiologists not always report whether a patient with MI has had a NSTEMI or STEMI.

References

1. DeVon HA, Rosenfeld A, Steffen AD, Daya M. Sensitivity, Specificity, and Sex Differences in Symptoms Reported on the 13-Item Acute Coronary Syndrome Checklist. Am Heart Assoc. 2014. 3(2):e000586.
2. Riley RD, Snell KIE, Ensor J, Burke DL, Harrell FE, Moons KGM, Collins GS. Minimum sample size for developing a multivariable prediction model: PART II – binary and time-to-event outcomes. Stat Med. 2019. 30;38(7):1276-1296

VERSION 2 – REVIEW

REVIEWER	Sofia Sederholm Lawesson Department of Medicine and Care, Division of Cardiovascular Medicine Linköping University Sweden
REVIEW RETURNED	15-Sep-2019

GENERAL COMMENTS	Reviewer comments: The manuscript is now better with a more concise and relevant research question and the results from the statistics are described more properly. As there were few cases with ACS in the study cohort, it is good that the main analyses are focused on whether there are gender differences in triaging of chest pain patients (log reg with urgent allocation as the dependent variable, comparing the genders, adjusting for age and ACS). The comparison between patients where a final diagnosis could be retrieved and those where this was not possible showing no important differences implies that there is probably not a selection bias present. Most reviewer comments are answered in a satisfactory manner. Anyhow I still have some comments and questions regarding this paper: I find the question whether there is a gender difference in telephone triaging of chest pain patients both new and relevant, and the most important part of this paper. I agree with the authors that a strength of this study is the capturing of the patients first time symptom description before encountering health care staff. Anyhow, the question whether there are gender differences in symptoms may be problematic to answer from this study, as no formal questionnaire was used, including detailed information on symptom characteristics, location, duration etc. I thus ask for clarification if questions regarding severity of pain, radiation of chest pain, shortness of breath and symptoms related to autonomous nervous system activation were asked to all patients? I.e. if a patient answers that the pain severity is 7/10 he/she will be

	allocated the highest urgency level without asking further questions. Is then the questions about radiation and associated symptom still asked? In addition, I miss questions about chest pain characteristics in the decision tree. The authors state that in men a stabbing chest pain was more common in patients without ACS then among those with ACS. Is this based on what patients had said spontaneously during the call or were questions about pain characteristics asked in a structured manner? If all the questions described in the decision tree, adding pain characteristics, is asked to all patients a comparison between the genders is satisfactory, otherwise probably not. An important limitation of the study is that there is no data on IHD risk factors and co-morbidities. This should be clearer from the manuscript. Taking such information into account is something every doctor do when determining the probability of ACS when meeting patients with chest pain. Results, triage: Please describe the percentage of women and men that received a high urgency allocation among patients not having ACS. Results, medical diagnosis: The authors state that women relatively more often had NSTEMI than men. This is probably true, and concurrent with previous studies, but I notice that there is much missing data among men on type of ACS, which is not commented anywhere in the text, only in Table 4. Discussion, minor comment: Strengths and limitations are most often placed in the end of the discussion part, and not in the middle of it. I, personally, would prefer if the results of the study where discussed in relation to previous research before the strength and limitation part. Discussion, minor comment: “the low number of ACS cases let us decide to refrain from multivariable logistic regression analysis considering symptoms and with ACS (yes/no) as the outcome ACS” should probably be rephrased to “the low number of ACS cases let us decide to refrain from multivariable logistic regression analysis considering symptoms with ACS (yes/no) as the outcome” Discussion: “Multiple previous studies compared symptoms of women and men with ACS, and only one single study compared symptoms similarly as we did; comparing women with and without ACS, and men with and without ACS” – this comparison is although problematic due to few ACS cases, thus no firm conclusions can be made. Table 2. Please explain the odds ratio! I guess this ORs comes from a logistic regression analysis where urgency allocation is the dependent variable and sex and age are the co-variates. This is neither described in the Table footnote, nor in the statistical part of the method section. Table 3: The finding the ACS patients were 6.36 times higher chance being allocated high urgency could be commented. It would anyhow be very disappointing if this was not the case. I think the ORs and CI for ACS should also be presented from the full model, i.e. with sex, age and ACS as co-variates.

VERSION 2 – AUTHOR RESPONSE

Comments to Reviewer 2: Sofia Sederholm Lawesson, Department of Cardiology and Department of Medicine and Health, Linköping University Hospital, Linköping Sweden

1. The manuscript is now better with a more concise and relevant research question and the results from the statistics are described more properly. Most reviewer comments are answered in a satisfactory manner.

We thank the reviewer for these positive remarks.

2. Could you clarify if questions on severity of pain, radiation of chest pain, shortness of breath and symptoms related to autonomous nervous system activation were asked in all patients? I.e. if a patient answered with pain severity 7/10, he/she will be allocated to the highest urgency level. Are then still these other questions asked for?

As mentioned on page 5 (mehods) all these four questions are part of the decision tree of the Netherlands Triage Standard (NTS). Indeed, if a patient has acute severe chest pain (≥7 on a scale from zero to 10) the NTS directly recommends an ambulance (U1). Nevertheless, in everyday practice, in the large majority, also the other three questions are asked in such a situation.

3. The question whether there are gender differences in symptoms may be problematic to answer from this study, as no formal questionnaire was used, including detailed information on symptom characteristics, location, duration etc. In addition, I miss questions about chest pain characteristics in the decision tree.

If all the questions described in the decision tree, adding pain characteristics, is asked to all patients a comparison between the genders is satisfactory, otherwise probably not.

We agree with the reviewer that we can only compare men and women with chest discomfort on the symptoms reported. We disagree with the suggestion that all questions in the decision tree need to be answered by all participants. This is unrealistic; even if a questionnaire is applied, missing values occur,[1] but of course more often with routine care data. Importantly, however, it does not bias the results on the comparison between men and women as long as missing values are not related to sex category.

4. The authors state that in men a stabbing chest pain was more common in patients without ACS then among those with ACS. Is this based on what patients had said spontaneously during the call or were questions about pain characteristics asked in a structured manner?

In the majority of cases this was not spontaneously reported, but after a question of the triagist.

Pain characteristics were asked for in a structured manner and divided into pressing and stabbing. It was reported in 349 (67.4%) patients (249 pressing and 90 stabbing).

In the revised manuscript we mention in the Methods paragraph:

"…. Such as symptoms, pain characteristics, medical history,…."

5. An important limitation of the study is that there is no data on IHD risk factors and co-morbidities. This should be clearer from the manuscript. Taking such information into account is something every doctor does when determining the probability of ACS when meeting patients with chest pain.

We agree that this is a limitation. Indeed, most physicians ask for previous IHDs and co morbidities. Interestingly, however, these item are not often investigated in the setting of suspected ACS (and therefore seldom reported in previous studies), and are not part of the 13-items validated symptom questionnaire often used in the USA.[1]

In the revised manuscript we added to the limitations paragraph of the Discussion "Since this is not part of the NTS, risk factors for ischaemic heart disease and co morbidities could not be evaluated."

6. Results, triage: Please describe the percentage of women and men that received a high urgency allocation among patients not having ACS.

We now present (i) the data of the whole population of suspected ACS and (ii) of those with ACS separately, which is common practice because results of those without ACS may easily be distracted from these.

"Both women and men with chest discomfort received most often a high urgency allocation (U1, U2) (women 65.6% vs. men 64.9%). Also in those with an ACS, women and men received as often a high urgency allocation (95.7% vs. 88.2%, p-value=0.331). See Table 2."

If the editor considers it worthwhile to also mention the results of those without an ACS separately, then we are very willing to do so.

7. Results, medical diagnosis: The authors state that women relatively more often had NSTEMI than men. This is probably true, and concurrent with previous studies, but I notice that there is much missing data among men on type of ACS, which is not commented anywhere in the text, only in Table 4.

We agree that the higher incidence of 'non-classified myocardial infarction' in men hampers the interpretation of the comparison between men and women regarding NSTEMI and STEMI.

In the revised Results section we deleted that sentence and mention the following:

"The distribution of unstable angina, NSTEMI, STEMI and 'non-classified myocardial infarction' are presented in Table 4. Men had more often 'non-classified myocardial infarction' (9/34 (26.5%) vs. 2/23 (8.7%))."

8. Discussion, minor comment: Strengths and limitations are most often placed in the end of the discussion part, and not in the middle of it. I, personally, would prefer if the results of the study where discussed in relation to previous research before the strength and limitation part.

We thank the reviewer and changed the revised Discussion accordingly.

9. Discussion, minor comment: "the low number of ACS cases let us decide to refrain from multivariable logistic regression analysis considering symptoms and with ACS (yes/no) as the outcome ACS" should probably be rephrased to "the low number of ACS cases let us decide to refrain from multivariable logistic regression analysis considering symptoms with ACS (yes/no) as the outcome."

We indeed changed the sentence in the revised test to:
"The low number of ACS cases let us decide to refrain from multivariable logistic regression analysis comparing symptoms with ACS (yes/no) as the outcome."

11. Discussion: "Multiple previous studies compared symptoms of women and men with ACS, and only one single study compared symptoms similarly as we did; comparing women with and without ACS, and men with and without ACS" – this comparison is although problematic due to few ACS cases, thus no firm conclusions can be made.

The fact that we had a low number of cases of ACS in our study has nothing to do with our remark about previous studies, and therefore, this sentence still holds true.

12. Table 2. Please explain the odds ratio! I guess this ORs comes from a logistic regression analysis where urgency allocation is the dependent variable and sex and age are the co-variates. This is neither described in the Table footnote, nor in the statistical part of the method section.

In Table 2 we only compare sex category to urgency allocation (high vs. low) and this is done by univariable logistic analysis. In Table 3 we performed multivariable analysis with correction for age and ACS.

In the revised manuscript we changed the sentence (methods) on page 7 as follows:

"We used both univariable and multivariable logistic regression analysis with urgency allocation (high vs. low) as the outcome to assess differences between women and men with chest discomfort. For multivariable analysis, after adjustment for the diagnosis ACS and age."

13. Table 3: The finding the ACS patients were 6.36 times higher chance being allocated high urgency could be commented. It would anyhow be very disappointing if this was not the case. I think the ORs and CI for ACS should also be presented from the full model, i.e. with sex, age and ACS as co-variates.

Do we correctly understand that the reviewer wants us to mention these results also in the Results paragraph?

In the revised manuscript we added to the results paragraph:
"Men and women with ACS received much more often a high urgency allocation than those who showed not to have an ACS (crude OR 6.36 (95%CI 2.49-16.24)."

References

1. DeVon HA, Rosenfeld A, Steffen AD, Daya M. Sensitivity, Specificity, and Sex Differences in Symptoms Reported on the 13-Item Acute Coronary Syndrome Checklist. Am Heart Assoc. 2014. 3(2):e000586.