PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Optimisation of telephone triage of callers with symptoms suggestive of acute cardiovascular disease in out of hours primary care: observational design of the Safety First study
AUTHORS	Erkelens, Daphne; Wouters, Loes; Zwart, D.L; Damoiseaux, Roger; De Groot, Esther; Hoes, Arno; Rutten, Frans

VERSION 1 – REVIEW

REVIEWER	Amir Mirhaghi Nursing and Midwifery Care Research Center, School of Nursing and Midwifery, Mashhad University of Medical Sciences, Mashhad, Iran
REVIEW RETURNED	14-Nov-2018

GENERAL COMMENTS	The necessity of study is well-developed. The authors indicated that telephone recordings with complaints that lead to suspicion of ACS or TIA/stroke will be included. Purposeful sampling may lead to the selection bias. Select cases on both ICPC codes and keywords in free text may end up to the underestimation of the error. It is probable that several patients with nonspecific complaints such as jaw pain or urinary incontinence which can be related to the myocardial infarction or stroke may have been missed. Therefore, random sampling is recommended. The study does not assess the reliability of NTS. The result of validity studies is more meaningful when readers are aware of reliability coefficients. The authors may be interested in adding a reliability assessment phase to the protocol. It is suggested that if the study is designed to mix the quantitative and qualitative phases, authors describe which kinds of mixed methods need to be used. However, it is appropriate to use diagnostic validity studies, readers may also be interested in over-triage and under-triage rates which is not mentioned in the protocol. Kind Regards,

REVIEWER	Joanne Turnbull University of Southampton, UK
REVIEW RETURNED	16-Nov-2018

GENERAL COMMENTS	A clear protocol that describes a very interesting sounding study. A couple of minor points. 1, If it is a requirement of the published protocol to clearly detail all ethical issues, then the section on ethics is brief (e.g. no details about recruitment/consent of interview participants. However, it's clear that the study has received ethical approval.

	2. Study limitations are covered breifly in the abstract but perhaps warrant a fuller discussion in the body of the paper?

REVIEWER	Linda Huibers Senior researcher Research Unit for General Practice, Aarhus, Denmark
REVIEW RETURNED	07-Dec-2018

GENERAL COMMENTS	**Optimisation of telephone triage of callers with symptoms suggestive of acute cardiovascular disease in out of hours primary care: design of the safety first study** This protocol refers to an ongoing study. The protocol mostly describes the cross-sectional study, of which data seems to be collected (at least partly), but assessment of recordings and data collection of the substudies probably is ongoing. The dates of the study are mentioned in the manuscript (i.e. ethics and dissemination paragraph). **General:** I feel that the manuscript will benefit from a language revision. Also, the abbreviation OHS-PC is a bit unusual; perhaps the authors could consider using GPC or OOH-PC? Abbreviations are not used consequently throughout the manuscript. I have one main question: The protocol mainly describes one observational study, whereas the research questions seem to be answered by this observational study together with three substudies. I am not sure why these are called substudies as they include new data and have different aims. So it is not clear to me why only the observational study is described in detail, whereas the substudies have a brief general description. All studies seem relevant to answer the research questions as mentioned also in the introduction. **Abstract:** I assume that the objective of the project is: “describing, understanding and improving the diagnostic process and urgency location in callers with symptoms of acute cardiovascular disease, in order to improve both efficiency and safety of telephone triage in this domain”. These objectives seem to be answered using four studies (i.e. a cross-sectional study and three substudies), but only the first one is described. In the abstract, the objective is quite unspecific. **Key words:** Check MesH term ‘after hours’, which has several entry terms as a substitute for ‘out of hours services in primary care’. **Strengths and limitations:** Perhaps use bullets for strengths too, as for limitations.

	Could one also have selection bias due to the algorithm to include cases?
	
	Introduction:
	
	The introduction overall covers all relevant information to understand the aims for this project, and see the relevance. Yet, the authors should consider restructuring the order of paragraphs/topics, as this is not always clear. In particular the information on the exact process of triage with the NTS, including identification of the main symptom and question ranking, could be a paragraph in the methods section.
	
	Also, there are several statements that could use a reference.
	
	Looking at the third research question, I miss a bit of background information in the introduction. Articles on the use of clinical decision support systems, its use by nurses, and history taking in telephone triage are present.
	
	Methods and analysis:
	
	- *Design:*
	  - Minor details: abstract says '3,000', while here is written 'over 3,000'. Are caller characteristics and patient-specific items the same thing (as it says 'these')? Do you collect information on the caller and/or the patient?
	  - The authors mention that they collect information on history taking and caller characteristics discussed during the conversation, for the observational study. I could not find more detailed information on this: how is this done, what information is collected, and which characteristics are collected from the conversation (in addition to information from the registration notes)?
	  - These 'registered notes', are they available in the electronic records of the service?
	  - What is an electronic case record form?
	  - Why do you ask the patient's own GP to provide the final diagnosis? It probably is because this information is not available elsewhere, but perhaps this can be mentioned? Several countries have national registries to gain this information, and could question this method.
	- *Setting:*
	  - Perhaps the authors could clarify the meaning of 'collaboration' a bit? Six locations in the region, but also six telephone triage centers?
	  - The organisation of the out-of-hours primary care service is now described in the introduction, but could be moved to this paragraph to improve overview.
	  - Is 'electronic patient record' the same as 'callmanager'?

	○ The text on weighted sample is unclear to me. Does this account for all codes in table 2? And what do you do exactly? Also, it is not 'setting' but more 'data collection'. - *Inclusion and exclusion criteria:* ○ Do the authors mean interrater variability due to different person coding? Perhaps they could add a reference? ○ Can the authors describe how they defined the key word selection? And do they plan to validate/test their algorithm for inclusion of calls? ○ Could (some of) the exclusion criteria result in selection bias, for example GPs who refuse to provide information? - *Data analysis:* ○ I do not understand the sensitivity analyses, reading the a) and b) information. Could the authors clarify their plans? ○ 'history items' and 'caller characteristics' are mentioned here, but I miss a clear description of this in the previous text. The authors also write "variable selection will be based on literature review … and on univariable analysis". Does this account for these history items and caller characteristics, or for other variables? It could help to have a specific paragraph on this, as most information now is written at "design". - *Power calculation:* ○ Prior pilot study: was the same algorithm for inclusion of contacts used? ○ The last sentence is unclear; I suggest to delete or move "both with follow up data". ○ Is it correct that you used 10 cases per variable as rule of thumb. If so, 23 should be 22 variables (2000 divided by 9 divided by 10). ○ Why did you decide to use two different numbers of variables possible in the analyses (i.e. 23 and 50)? It would help to clarify this, as it seems a bit random. **Substudies:** - (i) ○ The authors write "… often points at flaws in the triage process". Could they provide a reference to support this? This article seems to show a lower role of telephone triage as cause of safety incidents (Smits M et al. Patient safety in out-of-hours primary care: a review of patient records. BMC Health Serv Res 2010;10:335). ○ The comparison seems to focus on call and patient characteristics. I would expect that history taking is also important here. It says "… will evaluate the triage conversations", but it is not explained how this is done.

	○ I am not sure what diagnostic accuracy will be calculated here? - (ii) ○ If I understand correctly, the authors will collect new data, rather than use the data from the observational study. ○ One could also think of including suggestive questioning. **Discussion:** The authors start with a summary of their studies, followed by a more detailed motivation for several aspects. I am aware of the different nature of a discussion for a protocol article, but I feel that there is relatively much overlap with the introduction. - Paragraph "Efficiency and patient safety": the last sentence is not so clear; could the authors explain this in more detail? - Paragraph "Users of the NTS": the authors bring up very interesting points. I wonder if some of the information could be addressed earlier in the manuscript, for example in the methods section (for example the process of using NTS and the information collected in the substudies). **Conclusion:** No statement is referring to the highly relevant part about "Users of the NTS".

VERSION 1 – AUTHOR RESPONSE

Overview major revisions BMJ Open

Comments reviewers	Response from Author
1. Formatting amendments	
1.1 Required amendments will be listed here; please include these changes in your revised version: Please provide better qualities figures, ensuring the figures are not pixelated when zoomed in on. Figures can be supplied in TIFF, JPG or PDF format (figures in DOCUMENT, EXCEL or POWERPOINT format will not be accepted), we also request that they have a resolution of at least 300 dpi and 90mm x 90mm of width. *figure uploaded only 96 dpi, should be at least 300 dpi.	We adjusted the figure (Figure 1. Flowchart of the Safety First study) from JPG to PDF, ensuring a resolution of at least 300 dpi.
2. Editor Comments to Author	
2.1 Please revise your title so that it includes your study design. This is the preferred format for the journal	The Safety First study is in fact an observational study project of which the main study has a cross-sectional design. The other additional three studies have a case-control and interview study design respectively.

	We changed the title so that it includes the overall design (observational study).
2.2 Re the strengths and limitations section: please include each limitation as a main bullet point	In the revised document we adjusted the strengths and limitations section accordingly.
3. Reviewer 1 Comments to Author	
3.1 The authors indicated that telephone recordings with complaints that lead to suspicion of ACS or TIA/stroke will be included. Purposeful sampling may lead to selection bias. Select cases on both ICPC codes and keywords in free text may end up to the underestimation of the error. It is probable that several patients with nonspecific complaints such as jaw pain or urinary incontinence which can be related to the myocardial infarction or stroke may have been missed. Therefore, random sampling is recommended.	We disagree with the reviewer on this point. Indeed, as in any study a selection occurs, but that does not necessarily lead to selection bias. Our aim was to study patients with symptoms (more or less) suggestive of ACS or TIA/stroke. For this reason we included ICPC codes that somehow may resemble patients from these domains (e.g. ICPC codes included such as K75, K89, K90, but also K02, N17, N18, N19 etc.). Our broad sampling increases the generalizability of our results. Our sampling is not the same as purposeful sampling.
3.2 The study does not assess the reliability of NTS. The result of validity studies is more meaningful when readers are aware of reliability coefficients. The authors may be interested in adding a reliability assessment phase to the protocol.	Reliability coefficients (i.e. internal consistency expressed by Cronbach's alpha or test-retest reliability) would be important measures if the main goal was to evaluate the reliability of the NTS. However, our main aim was to describe the diagnostic accuracy of NTS with clinical outcomes as the reference. For that reason, sensitivity, specificity, positive and negative predictive values are best options.
3.3 It is suggested that if the study is designed to mix the quantitative and qualitative phases, authors describe which kinds of mixed methods need to be used.	It would be better to use the term sequential explanatory design; we first collect quantitative data (cross-sectional study and case-control study) followed by gathering qualitative data (interview study and conversation analysis). We adjusted the text accordingly in the revised document.
3.4 However, it is appropriate to use diagnostic validity studies, readers may also be interested in over-triage and under-triage rates which is not mentioned in the protocol.	We will also indirectly evaluate and discuss 'over-triage' and 'under-triage' rates, realising that identifying false-positives (those without ACS or TIA/stroke, who got a high urgency) and false-negatives (those with ACS or TIA/stroke, who got a low urgency) not simply means incorrect triage. For example, a 70 years old man who receives a high urgency for acute chest pain lasting for half an hour, with heavy transpiration, nausea and shows not to have an ACS after investigations at the ED is a false-positive case, but in our opinion not really a case of 'over-triage'. We can only calculate false-positives and negatives. Over-triage and under-triage are different concepts than false-positives or -negatives, that are at least partly defined by the risk we are willing to take as physicians and as society.
4. Reviewer 2 Comments to Author	
4.1 If it is a requirement of the published protocol to clearly detail all ethical issues, then the section on ethics is brief (e.g. no details about recruitment/consent of interview participants). However, it's clear that the study has received ethical approval.	The recruitment and informed consent procedure of the interviewees and expert panel members in the additional studies is in accordance with 'good clinical (research) practice'. We consider it more worthwhile to provide detailed information on the ethics in the

	individual papers to follow of the Safety First study.
4.2 Study limitations are covered briefly in the abstract but perhaps warrant a fuller discussion in the body of the paper?	We tried to write a concise study protocol paper covering our main cross-sectional study (focussing on the context, design and methods), but also the additional studies. Unfortunately word limitation does not allow for details on all parts. Again, a detailed limitation section will be published in the individual papers to follow.
5. Reviewer 3 Comments to Author	
5.1 **General:** I feel that the manuscript will benefit from a language revision. Also, the abbreviation OHS-PC is a bit unusual; perhaps the authors could consider using GPC or OOH-PC? Abbreviations are not used consequently throughout the manuscript.	In literature multiple abbreviations circulate to describe out of hours primary care (e.g. OHS, OOH, GPC). We prefer OHS-PC, but when the editor considers another abbreviation a better fit to the journal, we will be happy to change it throughout the document. In addition, we thoroughly checked our manuscript and adjusted it as much as possible. If the editor believes a language revision is needed, we will ask a native English speaker to check our manuscript.
5.2 **General:** I have one main question: The protocol mainly describes one observational study, whereas the research questions seem to be answered by this observational study together with three substudies. I am not sure why these are called substudies as they include new data and have different aims. So it is not clear to me why only the observational study is described in detail, whereas the substudies have a brief general description. All studies seem relevant to answer the research questions as mentioned also in the introduction.	We agree that 'additional studies' is better and changed the revised text accordingly. We describe the observational cross-sectional study, in more detail because it is our largest study within Safety First. By combining multiple data sources we can provide a more holistic view on the triage of patients suspected of an ACS or TIA/stroke.
5.3 **Abstract:** I assume that the objective of the project is: "describing, understanding and improving the diagnostic process and urgency allocation in callers with symptoms of acute cardiovascular disease, in order to improve both efficiency and safety of telephone triage in this domain". These objectives seem to be answered using four studies (i.e. a cross-sectional study and three substudies), but only the first one is described. In the abstract, the objective is quite unspecific.	Indeed, the objectives will be answered with data from four studies (main cross-sectional study and additional studies). Apart from the main study, the additional studies are also mentioned in the abstract. Unfortunately, the word limit doesn't allow for a more detailed explanation of all four studies. The overall objective is summarized in our research questions in the introduction.
5.4 **Key words:** Check MesH term 'after hours', which has several entry terms as a substitute for 'out of hours services in primary care'.	We thank the reviewer and we added the MeSH term 'after hours care' to our keyword section.
5.5 **Strengths and limitations:** Perhaps use bullets for strengths too, as for limitations.	In the revised document we use bullets now.

5.6 **Strengths and limitations:** Could one also have selection bias due to the algorithm to include cases?	By definition there is selection in a non-experimental (observational) study without randomisation. However, the risk of selection bias was limited by using a clinically relevant broad domain (see explanation 3.1).
5.7 **Introduction:** The introduction overall covers all relevant information to understand the aims for this project, and see the relevance. Yet, the authors should consider restructuring the order of paragraphs/topics, as this is not always clear. In particular the information on the exact process of triage with the NTS, including identification of the main symptom and question ranking, could be a paragraph in the methods section.	We think that this part should stay in the Introduction for contextual reasons and that it will help the readers understand how the NTS works in the triage process.
5.8 **Introduction:** Also, there are several statements that could use a reference.	We changed the text and added a reference accordingly.
5.9 **Introduction:** Looking at the third research question, I miss a bit of background information in the introduction. Articles on the use of clinical decision support systems, its use by nurses, and history taking in telephone triage are present.	We added some background information on the third research question (i.e. lack of knowledge on users of the NTS and relation between users and performance of the system) in the revised introduction.
5.10 **Design:** abstract says '3,000', while here is written 'over 3,000'.	We adjusted the text accordingly.
5.11 **Design:** Are caller characteristics and patient-specific items the same thing (as it says 'these')? Do you collect information on the caller and/or the patient? The authors mention that they collect information on history taking and caller characteristics discussed during the conversation, for the observational study. I could not find more detailed information on this: how is this done, what information is collected, and which characteristics are collected from the conversation (in addition to information from the registration notes)?	The main source of information are the backed up triage conversations that were re-listened. Detailed information collected about the patient concerned symptoms, signs, medical history, and if mentioned, drug use. Also registered were onset and duration of the main symptom and time and duration of the call. The urgency allocation was extracted from the registered notes. In our revised manuscript we added more detailed information.
5.12 **Design:** These 'registered notes', are they available in the electronic records of the service?	Yes. We mention this now in the revised manuscript.
5.13 **Design:** What is an electronic case record form?	This is a generally acknowledged notification for an electronic data collection tool.
5.14 **Design:** Why do you ask the patient's own GP to provide the final diagnosis? It probably is because this information is not available elsewhere, but perhaps this can be mentioned? Several countries have national registries to gain this information, and could question this method.	We thank the reviewer for this valuable addition. We added this information in our revised manuscript.

5.15 **Setting:** Perhaps the authors could clarify the meaning of 'collaboration' a bit? Six locations in the region, but also six telephone triage centers?	In the revised text we explain that 'Primair Huisartsenposten' is a collaboration of six OHS-PC locations with each their own OHS-PC location and telephone triage centre.
5.16 **Setting:** The organisation of the out-of-hours primary care service is now described in the introduction, but could be moved to this paragraph to improve overview.	We would like to refer to our explanation on 5.7; we think for contextual reasons it is better to leave it in the introduction.
5.17 **Setting:** Is 'electronic patient record' the same as 'callmanager'?	Yes, the electronic patient records of the OHS-PC are the same as 'Callmanager'. We added this in the revised manuscript.
5.18 **Setting:** The text on weighted sample is unclear to me. Does this account for all codes in table 2? And what do you do exactly? Also, it is not 'setting' but more 'data collection'.	We now put this part under the heading 'data collection.' We realized that the term 'weighted sample' might be confusing. Instead, we describe the distribution of sampling of ICPC codes within our study. This distribution was based on the actual distribution of ICPC codes, to reflect daily practice.
5.19 **In-/exclusion criteria:** Do the authors mean inter-rater variability due to different person coding? Perhaps they could add a reference?	Yes. We added two references.
5.20 **In-/exclusion criteria:** Can the authors describe how they defined the key word selection? And do they plan to validate/test their algorithm for inclusion of calls?	We defined the keyword selection as an additional filter following the ICPC code selection to prevent inclusion of misclassified telephone triage conversations. E.g., a triage conversation about a nosebleed which was incorrectly labelled as K02.
5.21 **In-/exclusion criteria:** Could (some of) the exclusion criteria result in selection bias, for example GPs who refuse to provide information?	This could indeed result in selection but very unlikely to selection bias. It is very unlikely that the patient characteristics and relation to the outcomes (ACS, TIA/stroke) we used are related to this specific GP characteristic.
5.22 **Data analysis:** I do not understand the sensitivity analyses, reading the a) and b) information. Could the authors clarify their plans?	The sensitivity analyses will be applied in a composite of diagnoses that need urgent medical management. E.g. the composite outcome of ACS plus pulmonary embolism, acute heart failure and thoracic aortic dissection. In the case of TIA/stroke, subarachnoid haemorrhage, and epilepsy. The sensitivity analyses will provide additional information on the diagnostic accuracy of the NTS for all emergency diagnoses in our study domains (suspected ACS or suspected TIA/stroke).
5.23 **Data analysis:** 'history items' and 'caller characteristics' are mentioned here, but I miss a clear description of this in the previous text. The authors also write "variable selection will be based on literature review … and on univariable analysis". Does this account for these history items and caller characteristics, or for other variables? It could help to have a specific	Our variable selection indeed accounts for history items and caller characteristics. It concerns patient-specific items, which can be used in everyday practice during telephone triage. The information on data collection of history items and caller characteristics is described earlier in the protocol paper (see also 5.11).

paragraph on this, as most information now is written at “design“.	
5.24 **Power calculation:** Prior pilot study: was the same algorithm for inclusion of contacts used?	No, not exactly. In the pilot study a smaller domain was used: only patients with ‘chest pain’ and ‘neurologic deficit’ were included. This is an advantage when we validate externally; it may help generalise our results.
5.25 **Power calculation:** The last sentence is unclear; I suggest to delete or move “both with follow up data“.	We deleted this part of the last sentence accordingly in the revised text.
5.26 **Power calculation:** Is it correct that you used 10 cases per variable as rule of thumb. If so, 23 should be 22 variables (2000 divided by 9 divided by 10).	Indeed 2,000/9/10=22.2 variables may be analysed. We changed 23 to 22 as suggested by the reviewer.
5.27 **Power calculation:** Why did you decide to use two different numbers of variables possible in the analyses (i.e. 23 and 50)? It would help to clarify this, as it seems a bit random.	We speculated that possibly more variables needed to be evaluated in the domain suspected TIA/stroke than in the domain ACS. Patients within the ‘neurological deficit’ domain may present themselves with a larger variety of symptoms, therefore we included more variables.
5.28 **Substudies: (i)** The authors write “… often points at flaws in the triage process“. Could they provide a reference to support this? This article seems to show a lower role of telephone triage as cause of safety incidents (Smits M et al. Patient safety in out-of-hours primary care: a review of patient records. BMC Health Serv Res 2010;10:335).	In the revised manuscript we added a reference here.
5.29 **Substudies: (i)** The comparison seems to focus on call and patient characteristics. I would expect that history taking is also important here. It says “… will evaluate the triage conversations“, but it is not explained how this is done.	We consider history taking with exploration of the symptoms and signs as part of patient characteristics. We more explicitly mention this in the revised text.
5.30 **Substudies: (i)** I am not sure what diagnostic accuracy will be calculated here?	The diagnostic accuracy for the determinant “safe triage handling according to the expert yes/no” against the final outcome “calamity or no calamity”. We clarify this in the revised text.
5.31 **Substudies: (ii)** If I understand correctly, the authors will collect new data, rather than use the data from the observational study.	This is correct. We changed our text accordingly to make this clear.
5.32 **Substudies: (ii)** One could also think of including suggestive questioning.	We thank the reviewer and added this item.
5.33 **Discussion:** The authors start with a summary of their studies, followed by a more detailed motivation for several aspects. I am aware of the different nature of a discussion for	To the best of our knowledge, we focussed on the context of the clinical problem in the introduction, and on possible implications in our discussion. Only if items from the context were

a protocol article, but I feel that there is relatively much overlap with the introduction.	needed for explanation of the implications of our studies, we mention these in the discussion
5.34 **Discussion:** Paragraph "Efficiency and patient safety": the last sentence is not so clear; could the authors explain this in more detail?	We revised our text to: this knowledge might put the general view on (i) analysing calamities, and (ii) the weight that is assigned to improvement measures in a different perspective.
5.35 **Discussion:** Paragraph "Users of the NTS": the authors bring up very interesting points. I wonder if some of the information could be addressed earlier in the manuscript, for example in the methods section (for example the process of using NTS and the information collected in the substudies).	We already added information on the users of the NTS in the introduction. To prevent overlap we did not add this information again to the methods section.
5.36 **Conclusion:** No statement is referring to the highly relevant part about "Users of the NTS".	We added a statement on the "users of the NTS" part in the conclusion of the revised manuscript.

VERSION 2 – REVIEW

REVIEWER	Amir Mirhaghi Nursing and Midwifery Care Research Center, Mashhad University of Medical Sciences, Mashhad, Iran
REVIEW RETURNED	22-Jan-2019

GENERAL COMMENTS	I enjoyed reading manuscipt. I thank you for responses. Regards

REVIEWER	Joanne Turnbull University of Southampton, UK
REVIEW RETURNED	05-Feb-2019

GENERAL COMMENTS	I have reviewed this revised manuscript and the responses from the authors to the reviewers comments. I feel that the authors have satisfactorily addressed the reviewers comments

REVIEWER	Linda Huibers Research Unit for General Practice, Aarhus, Denmark
REVIEW RETURNED	30-Jan-2019

GENERAL COMMENTS	Dear authors, Thank you for your response on my comments, which in most cases was satisfactory. I have some remaining comments/questions: • (5.2) I agree that describing four studies in detail in one protocol article is hardly possible, and therefore I understand the choice of describing the observational cross-sectional study in detail. As a reader cannot assess the methodology of

	the additional studies to the same degree, I would suggest making this choice explicit in the study protocol. • (5.8) One can argue whether some statements are 'general knowledge' or not. The following statements could benefit from a reference: o Yet, comparable studies on validity of telephone triage systems in primary care settings are limited. *(the word limited suggest that there are references)* o There are no other studies available that assessed the validity of the NTS for telephone triage in the OHS-PC setting or validated the NTS to the final diagnosis instead of to surrogate markers. (5, 8) *(not sure whether reference 8 addresses the NTS)* o Over the last decade, out of hours primary care in the Netherlands has been reorganised from small practices into larger OHS-PC. *(a reference on an article describing out-of-hours care in the Netherlands would be helpful)* o On the other hand, overestimation of urgency results in unnecessary high workload, high referral rates, high costs, and potentially in iatrogenic damage of the caller. • (5.9 & 5.33) The authors have added more information as well as reference 11. They have addressed this issue (i.e. information (i.e. clinical decision support systems, use and history taking in telephone triage) in more detail in the discussion. One could also check O'Cathain et al. I conclude that the authors have chosen to focus primarily on the problem with the NTS in the introduction and to refer to results of international studies mainly in the discussion. • (5.11) Thank you for the clarification. So is it correct to conclude that if information is not discussed in the call, the information is not asked and data is lacking? • (5.27) Could the authors add this in the manuscript, to clarify their choice not just for me, but also for the readers of the article? • (5.29) I am unable to find this clarification in the text. • (5.31) Also here, I have difficulties finding the adjustments in the revised manuscript. I can see that the authors use a sample of the case-control study. Kind regards

VERSION 2 – AUTHOR RESPONSE

Overview minor revisions BMJ Open

Additional comments reviewer	Response from Author

(5.2) I agree that describing four studies in detail in one protocol article is hardly possible, and therefore I understand the choice of describing the observational cross-sectional study in detail. As a reader cannot assess the methodology of the additional studies to the same degree, I would suggest making this choice explicit in the study protocol.	We agree with this suggestion of the reviewer, and we added a sentence to explicit our choice to only describe the observational cross-sectional study in more detail, and not the other three studies. "We explicitly chose to describe the observational cross-sectional study in more detail and not the other three studies because it is the largest study within the Safety First project, and for readability." (See page 12).
(5.8) One can argue whether some statements are 'general knowledge' or not. The following statements could benefit from a reference: a) Yet, comparable studies on validity of telephone triage systems in primary care settings are limited. *(the word limited suggest that there are references)* b) There are no other studies available that assessed the validity of the NTS for telephone triage in the OHS-PC setting or validated the NTS to the final diagnosis instead of to surrogate markers. (5, 8) *(not sure whether reference 8 addresses the NTS)* c) Over the last decade, out of hours primary care in the Netherlands has been reorganised from small practices into larger OHS-PC. *(a reference on an article describing out-of-hours care in the* *Netherlands would be helpful)* d) On the other hand, overestimation of urgency results in unnecessary high workload, high referral rates, high costs, and potentially in iatrogenic damage of the caller.	a) We thank the reviewer and added a reference accordingly (page 4, Van Ierland et al. 2011). b) Indeed, reference 8 does not address the NTS system directly. However, it shows there are no other systematic reviews that describe the validity of the NTS telephone triage against the final diagnosis, at least until 2017. We therefore considered this study relevant as a reference. c) We thank the reviewer and added a reference of a study on the history of out-of-hours primary care in the Netherlands. Pease see on page 5; Smits et al. 2017). d) We added a reference to this sentence; page 7, Coster et al 2017.
(5.9 & 5.33) The authors have added more information as well as reference 11. They have addressed this issue (i.e. information (i.e. clinical decision support systems, use and history taking in telephone triage) in more detail in the discussion. One could also check O'Cathain et al. I conclude that the authors have chosen to focus primarily on the problem with the NTS in the introduction and to refer to results of international studies mainly in the discussion.	We thank the reviewer and added two references of O'Cathain et al. in the revised text of the introduction to the sentence: "However, previous studies on OHS-PC telephone triage in the United Kingdom described that the clinical background of triage nurses, the range of their experience, their gender and their attitudes to risk did not affect the triage decisions made." (See page 6).
(5.11) Thank you for the clarification. So is it correct to conclude that if information is not	That is indeed correct.

discussed in the call, the information is not asked and data is lacking?	
(5.27) Could the authors add this in the manuscript, to clarify their choice not just for me, but also for the readers of the article?	We now added this information in the revised text on power calculation. “We speculated that more variables would be needed to evaluate suspected TIA/stroke cases than in patients in the domain ‘suspicion of ACS’ because the former patients may present themselves with a larger variety of symptoms.” (See page 11).
(5.29) I am unable to find this clarification in the text.	We agree with the reviewer that this was not sufficiently clear and adjusted the revised text accordingly. “A researcher blinded to the outcome will extract information regarding call characteristics, caller and symptom characteristics (including medical history) and urgency allocation, by re-listening the archived triage calls and by using the registered information. All data will be inserted in a database.” (See page 12).
(5.31) Also here, I have difficulties finding the adjustments in the revised manuscript. I can see that the authors use a sample of the case-control study.	We will use a sample from the case-control triage conversations and transcribe these conversations. These transcripts (new data) are the subject of research. “We will transcribe a sample of the conversations from the case-control study following the Jefferson conventions. These transcripts that can be considered as new data will then be qualitatively analysed with established conversation analysis techniques.“ (See page 13).

VERSION 3 – REVIEW

REVIEWER	Linda Huibers Research Unit for General Practice, Denmark
REVIEW RETURNED	01-May-2019

GENERAL COMMENTS	I thank the authors for their careful reply to my comments, to clarify these issues.