

## 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

|                            |   |
|----------------------------|---|
| <b>TITLE (PROVISIONAL)</b> | Help-seeking behaviour outside office hours in Denmark, the Netherlands, and Switzerland: a questionnaire study exploring responses to hypothetical cases |
| <b>AUTHORS</b>             | Huibers, L; Keizer , E; Carlsen, Anders; Moth, Grete; Smits, Marleen; Senn, Oliver; Christensen, Morten   |

### VERSION 1 – REVIEW

|                        |   |
|------------------------|---|
| <b>REVIEWER</b>        | Dr Sarah Neill<br>University of Northampton, UK |
| <b>REVIEW RETURNED</b> | 12-Oct-2017                                     |

|                         |  |
|-------------------------|--|
| <b>GENERAL COMMENTS</b> | <p>You have conducted a very rigorous piece of work but it is inherently flawed by the choice of method - use of hypothetical scenarios. As you note this introduced a potential social acceptability bias. This isn't the only problem as scenarios are answered hypothetically when respondents are not affected by the emotions of a real situation, consequently they respond differently. This makes your results, and that of any other studies using this approach, meaningless in my opinion. In addition the development of your research instrument (questionnaire and scenarios) appears to have been derived from professional perspectives, yet you wanted to find out about patient's and parent's decision making. The two things are not aligned and risk producing results which have little or no value in real life situations where patients/parents are making decisions about help seeking without professional input.</p> <p>I also note that your review of the literature is focused on other similar quantitative studies. None of these studies is able to generate data which furthers the understanding of why people seek help the way they do. For that you need theory building qualitative research. You did not review the qualitative research around parents/patients help seeking.</p> <p>Your research has taken so much time and effort on the part of all involved, including your participants and has yielded results which do not contribute to health care. I am of the opinion that this is bordering on unethical when an alternative design, underpinned by a more comprehensive literature review might have generated so much more.</p> |
|-------------------------|--|

|                        |   |
|------------------------|---|
| <b>REVIEWER</b>        | Corinne Chmiel<br>Institute of Primary Care, University of Zurich Switzerland |
| <b>REVIEW RETURNED</b> | 25-Oct-2017   |

|                         |  |
|-------------------------|--|
| <b>GENERAL COMMENTS</b> | This is a thoroughly planned and methodological found publication. |
|-------------------------|--|

|  |   |
|--|---|
|  | <p>Some minor issues should be taken care of, then nothing should stand in the way of publishing the manuscript:</p> <ul style="list-style-type: none"> <li>- Some revisions of the English language</li> <li>- Figure 1 is not clearly visible in the current form.</li> <li>- Concerning the Discussion, Conclusion and Implications I would like to add some personal considerations, which might help to improve the relevance of the manuscript:</li> </ul> <p>The astonishing finding of the study is not only the difference in health care seeking of pediatric emergencies, but the lacking difference in health care seeking among the adult population of all three countries. This is especially astonishing, considering the very different health care systems and out of hours care services which are being compared with each other in this study. Especially astonishing is that gate keeping systems as well as non-gate keeping systems show almost the same distribution of out of hours usage. What are possible reasons here for? Is European mentality to similar in these countries to expect differences? If this should be the case, why does the difference only appear in the pediatric setting? Why does a managed care system not show an impact on out of hours care usage? Should managed care not result in reduced unnecessary usage of resources? Please discuss these issues in the discussion section of the manuscript and adapt the conclusion and implications section accordingly.</p> <p>The other suggestions for improvements can be appreciated in the attached PDF as comments.</p> <p>I wish the Authors all the best for the future research activity.</p> <ul style="list-style-type: none"> <li>- The reviewer also provided a marked copy with additional comments. Please contact the publisher for full details.</li> </ul> |
|--|---|

|                        |  |
|------------------------|--|
| <b>REVIEWER</b>        | Catherine Pope<br>University of Southampton UK |
| <b>REVIEW RETURNED</b> | 16-Nov-2017                                    |

|                         |  |
|-------------------------|--|
| <b>GENERAL COMMENTS</b> | <p>This paper and one based on the same study came to me for review from separate journals. There is very significant overlap in the description of the methods and other aspects of the reporting in both papers and therefore some of my comments may be repeated in both reviews. The editors should ask the authors for the parallel paper in each case to inform their decisions.</p> <p>The analysis is descriptive and the findings confirm what is known such and it is difficult to see what this paper really adds, despite drawing comparative data from 3 countries. When discussing the existing literature on p13 the authors note the importance of health system differences that seem a more interesting avenue for research but which are not the topic of this paper.</p> <p>The design of this study is not clear. The authors refer to a cross sectional observational study in the abstract and methods section p4. This is a little misleading as the survey investigates stated preferences based on hypothetical scenarios, and the paper needs to be much clearer that the data describe preferences not actual behaviour. For example on p4 the aim is stated as ‘ to study how individuals from different age groups in 3 countries...react to acute health problems occurring outside office hours’ – it may be worth inserting ‘hypothetical scenarios about’ before the word acute. Another example is on p12 line 30 ‘more often contacted OOH care’</p> |
|-------------------------|--|

|  |  |
|--|--|
|  | <p>should be rephrased as 'expressed preference for contacting OOH care'.</p> <p>The 6 hypothetical scenarios used vary in the depiction of urgency (and thus in 'appropriate' disposition) yet the analysis assumes that all the preferences are equivalent and score 1 (or zero) - on p9 line 43 "we calculated a score between 0 and 6 for the number of cases for which OOH contact had been chosen" suggests that some scores of 3 would have a very different meaning (in terms of 'appropriateness' of disposition) than other scores of 3. I was confused about this. It did not seem that there was a pre-specified analytical plan which I might expect for such an analysis. It may be that this just needs better explanation and justification but I would urge that the paper has an expert statistical review.</p> <p>The authors note that the use of invented cases may introduce social desirability bias and may not represent actual behaviour –this seems to be a significant weakness.</p> <p>A power calculation is mentioned on p8 but not justified and then on p13 the paper concedes that the study lacked power. The paper goes on to suggest that other factors (not studied) may cause the variation p14 rather undermining the analysis presented.</p> <p>The Netherlands and Swiss samples are drawn from consumer panels which the authors concede is problematic p9 line 32. There is variation in the response rates reported and concerns about the representativeness of the eventual samples – see p10.</p> <p>The paper would benefit from a better description and justification of the analytical approach.</p> |
|--|--|

|                        |   |
|------------------------|---|
| <b>REVIEWER</b>        | Matthew Booker<br>University of Bristol, UK |
| <b>REVIEW RETURNED</b> | 22-Nov-2017                                 |

|                         |   |
|-------------------------|---|
| <b>GENERAL COMMENTS</b> | <p>Thank you for the opportunity to review this manuscript, which discusses the important area of 'out of hours' care access. It presents a component of research (which I believe from the authors' references in the document to be part of a larger project) exploring how randomly-selected questionnaire respondents indicate they would be likely to respond to hypothetical urgent care case scenarios. The sample is drawn from Denmark, Netherlands and Sweden. The emphasis is upon the difference of indicated behaviours between countries, with attempts made to correct for a limited number of potentially confounding variables.</p> <p>The topic of decision-making in urgent care is undoubtedly very important, and the work presented does look at this from a different and interesting angle. In particular, the construction of hypothetical cases is an interesting method, and there is much merit to the significant work the authors have undertaken. Indeed, the process of the development of these cases has clearly been somewhat of a strength and should be commended.</p> |
|-------------------------|---|

However, I do have a number of comments about the manuscript in its present format, that I feel would need to be addressed before I could recommend it for publication in BMJ Open. I have summarised these below.

Title and abstract – I have concerns that the title in its present form does not accurately reflect the methodology. Calling the work a 'cross-sectional observational study of help-seeking' to me implies that the participants' actual consulting behaviour/choices were analysed (and to me, implies the use of routine data in some way). The manuscript actually presents an analysis of the response to randomly invited questionnaires about hypothetical actions in artificially constructed case-situations. Whilst I feel that this a very reasonable way to explore an aspect of help-seeking behaviour, it is a study about how people say they might act in a hypothetical situation, and as such I think the title needs to be reworded to reflect that very clearly "e.g. a questionnaire study exploring responses to hypothetical cases... etc etc" to avoid confusion. Similarly the abstract needs rewording a little to avoid that confusion.

Methods - I think the use of hypothetical case vignettes (personally, I would prefer this terminology over 'invented' cases) is a very reasonable way of exploring what people might say they would do in certain circumstances. I think there needs to be much more reference to studies that have used hypothetical case vignettes successfully (including a more detailed critical discussion of the pros and cons of this method) to give more weight to why this is an appropriate method. I also think it would be useful to reference other studies that have used hypothetical cases as a way of exploring decision making in other settings (i.e. non urgent care) - there are a vast array of these in the literature.

Age groups – the authors have selected to send questionnaires to 3 age groups (0-4, 30-39 and 50-59 years). I note that the authors have explained their reasons for this to give 'adequate power' and/or to be able to author hypothetical cases the situations these groups may experience most closely. However, I do have some significant concerns about the relatively narrow age selections that are therefore included in this study, and how representative as a whole conclusions might be. For example, exclusion of the elderly and retired is a potentially significant flaw, as is exclusion of young working adults. I note that the authors cite discovering high variation in one of their previous studies in 0-4 year olds and 25-35 year olds as a justification, but then the age band of 30-39 chosen for this study would not fit with this and therefore seems a bit confusing? I think more detail about the justification for these selections, and a much more critical discussion of the implications of excluding other age groups is needed.

Questionnaires – the authors note that 'questions on factors related to help seeking were part of a larger study and will be described in further detail in another scientific article'. Whilst I understand the need to avoid duplicate publication, I think as a minimum a brief overview of the significance of these questions should be included here. I have struggled a little to understand how this work helps inform the collective knowledge base about decision making without this, as its omission makes this work quite narrow in focus and of more restricted interest. If analysis has not been undertaken yet, a couple of lines about the hypothesis would help understanding how these questions fit with the data presented in this paper, and how

|  |   |
|--|---|
|  | <p>the two analyses might complement the wider decision-making research base. The case vignettes are included in the appendix, but is it possible also to include a version of the full questionnaire (ideally in English, but if not possible then in the relevant languages?) as a supplementary file for transparency and clarity?</p> <p>Case development – the process described here seems very robust and detailed with good piloting, psychometric evaluation and consensus. Was a patient or lay representative included in the development of the cases at any stage? I wasn't quite clear if so.</p> <p>Data collection – Do the authors believe there is likely to be any bias associated with the different modality of questionnaire distribution and completion (i.e. paper versus online?). Has any sub analysis been done to determine if there is a difference between this – I am guessing it may be associated with confounding due to education level or economic status?</p> <p>Results and analysis – The crude and adjusted rates of consultation are presented in a table, but this does not break down by hypothetical condition/case – is this because there is no statistical difference between any of the vignettes as individuals, or is this a limitation of the study that it is not powered adequately to do this? Is there anyway that a graphical or diagrammatic representation of how the groups responded to each of the hypothetical cases could be included? For example, it would be very interesting to see visually the spread of “consult OOH” –vs- “don't consult OOH” for the different cases. I do think some visual representation of this is needed. I also had real difficulty reading and understanding figure 1 with the colourings as greyscale – this may be my peer review copy but I think it does need to be reproduced in colour and some attention given to how well that figure stands alone without detailed textual description.</p> <p>Discussion – the authors have explained that the intention of the study was to focus on the differences between countries, correcting for potential confounding of socio-demographic variables, and indeed it seems the study was set up as such. I am not sure I completely understand the full rationale for analysing the data this way around – what is the importance of understanding the difference between countries more so than the effect of socio-economic status, deprivation, education, finance etc? Clearly the authors must have a rationale for looking at the differences between countries, but at the moment as a reader this does not come through entirely clearly at the moment – perhaps a few lines in a box or bullet section at the outset would help explain this? As such it is at times a little difficult to understand the practical significance of some of the analysis as it is presented – for example, why is it of any real interest or practical significance that a fairly narrow age band of individuals (30-39 year olds) would hypothetically not consult any differently if they are from the Netherlands vs Denmark? I do wonder if some of the interest in this might be around the questions about help-seeking (see my above comment) which makes their inclusion here even more important.</p> <p>Strengths and limitations – The youngest age group (0-4) seems – from the authors reference to their previous work and from the data presented – to be of interest. I think parental roles and family constitution will be hugely influential here (single parent versus 2-parent families, availability of grandparents, extended family, other</p> |
|--|---|

|  |   |
|--|---|
|  | <p>children in the family and therefore previous parental experience versus first child etc etc) – there are a really significant number of shaping influences here that – whilst extremely interesting – can not be explained by this data alone. I think it is necessary to reflect on how this fits in with this understanding – see above.</p> <p>In summary, I think there is much merit in the contents of this work as it does provide evidence that looks at an important issue. However, prior to recommending publication I think it needs to be more accurately titled and described, it needs a much clearer explanation of why the question of comparing countries is important and relevant, why the groups selected were appropriate, and some additional detail as mentioned above in the results, conclusions and discussion sections to adequately explore the limitations.</p> |
|--|---|

|                        |   |
|------------------------|---|
| <b>REVIEWER</b>        | Professor Daniel Lasserson<br>University of Birmingham, UK. |
| <b>REVIEW RETURNED</b> | 28-Nov-2017   |

|                         |   |
|-------------------------|---|
| <b>GENERAL COMMENTS</b> | <p>Huibers and colleagues have undertaken a very interesting and important study, comparing help seeking behaviour in the out of office hours time period, across three European countries.</p> <p>I have several questions and comments for the research team to consider. Overall, my comments are focussed on how much of the differences between countries could be due to the different patient selection mechanisms, and that this should be considered more deeply in the discussion.</p> <p>Comments by section:</p> <p>Methods:</p> <ol style="list-style-type: none"> <li>1. Dutch and Danish Ooh data were used to identify frequent presenting problems and not data from Swiss OOH primary care. Were there differences between Dutch and Danish data? Excluding Swiss data may introduce bias into the results if different prevalence conditions are used in each setting.</li> <li>2. Were cases discussed with researchers and GPs in each country? Given that this study was undertaken in different countries, more clarity is needed on the representation of each country in the case study development.</li> <li>3. Why wasn't the readability of the Dutch and German versions piloted for readability, like the Danish?</li> <li>4. More detail is needed about the power calculation (based on other empirical data, or assumptions?)</li> <li>5. More detail is needed about recruitment methods – it is clear that the invitation is different (post vs email), were there differences in patient selection methods? The Analysis section seems to suggest that recruitment methods were very different between settings.</li> <li>6. The lack of children's data from Switzerland needs to be explained. This is a weakness of the study in terms of a multi-country comparison.</li> </ol> <p>Results:</p> <ol style="list-style-type: none"> <li>7. Different demographic data are presented in each country, this should be made clearer in the methods. This makes comparisons across datasets more difficult. I think the research team should include how this may have introduced bias later in the discussion.</li> <li>9. Figure 1. The legibility needs to be improved.</li> </ol> <p>Discussion:</p> |
|-------------------------|---|

|  |   |
|--|---|
|  | 10. I don't think it is fair to put the differences in results entirely down to the differences between countries. The methods of recruitment of participants differed as well and this could have introduced a bias which may explain the results. This should appear as a separate section in the discussion. |
|--|---|

**VERSION 1 – AUTHOR RESPONSE**

**Point-by point response**

|   |  |
|---|--|
| <b>Editorial Requirements:</b>  |  |
| <p>Reviewer 3 raises concerns about overlap between this manuscript and another paper submitted to a different journal. Please provide the parallel paper and a justification to explain the similarity between the two.</p>  | <p>The parallel paper is titled “Factors related to out-of-hours help-seeking for acute health problems: a survey study using case scenarios”.</p> <p>We decided to write two papers on this study, to be able to focus on the specific help seeking intentions (wait&amp;see, self-care, own GP next day, out-of-hours primary care, emergency department, ambulance care) and on factors related to help-seeking outside office hours.</p> <p>The first paper, as submitted to BMJ Open, we focus on clinical help seeking behavior of citizens of different countries, describing the actual intended help seeking, which could be related to health care systems.</p> <p>The second paper focuses on factors related to intended help seeking outside office hour (as a dichotomous outcome variable), thus enabling us to study multiple relevant factors related to help seeking, based on extensive literature. Here we also included the variable ‘country’, to investigate whether a difference remained after correcting for other known factors.</p> <p>Combining these aims in one article would result in too high information density.</p> |
| <p><b>Reviewer: 1</b><br/> <b>Reviewer Name: Dr Sarah Neill</b><br/> <b>Institution and Country: University of Northampton, UK Please state any competing interests: None declared.</b></p>   |  |
| <p>You have conducted a very rigorous piece of work but it is inherently flawed by the choice of method - use of hypothetical scenarios. As you note this introduced a potential social acceptability bias. This isn't the only problem as scenarios are answered hypothetically when respondents are not affected by the emotions of a real situation, consequently they respond differently. This makes your results, and that of any other studies using this approach, meaningless in my opinion. In addition the development of your research instrument (questionnaire and scenarios) appears to have been derived from professional perspectives, yet you wanted to find out about patient's and parent's decision making. The two things are not aligned and risk producing results which have little or no value in real life situations where</p> | <p>A relevant point of discussion that we have discussed in our research team as well. We concluded that the chosen method was the best way to study help-seeking, taking pragmatic issues into consideration. Indeed, citizens may have different help-seeking behaviour in emotional situations; however, it is difficult to ask patients about their choice in emotional situations directly after their decision. Not all patients and caregivers will be able or willing to talk to an interviewer or complete a questionnaire at the moment of contact with health care. Also, we were interested in patients who decided to wait and see, and this group is hard to include in real life. Furthermore, we aimed to compare help-seeking behaviour in different countries, for which constructed and identical case scenarios were desirable.</p>  |

|  |  |
|--|--|
| <p>patients/parents are making decisions about help seeking without professional input.</p>  | <p>We have added the following in the strengths and limitations section of our manuscript: <i>“The chosen design of using invented cases to measure intended help-seeking behaviour had several strengths and limitations. Strengths were that the respondents received the same cases, making comparisons more straightforward, and that persons who do not use OOH care or healthcare at all were also included. A limitation was the risk of introducing social desirability bias, with the response not representing actual behaviour. Additionally, the absence of emotional reactions that occur in real-life situations could have influenced the response. However, according to the theory of planned behaviour, behaviour is mainly determined by behavioural intentions. A review of literature on theory of planned behaviour concluded that behavioural intentions do predict behavior, while Nagai (2015) found that help-seeking intentions are an important predictor of help-seeking behavior. Several studies used hypothetical case scenarios in out-of-hours care and other settings. Thus, we found that the chosen design was the most feasible and appropriate in relation to our aim.”</i></p> <p>We do not agree with the comment on perspectives. Our goal was to select cases that were representative for the included countries, as occurring in out-of-hours care. Here we both had professionals and citizens giving feedback on our cases. The expert panel was used to include different ranking of cases, indeed based on a professional perspective, as this is the main starting point in discussion on irrelevant use of out-of-hours care. Indeed, there is a difference between professionals and citizens on the assessment of relevance, but that is another discussion. Medically irrelevant does not necessarily mean that a contact was irrelevant/inappropriate from a social/personal perspective.</p> |
| <p>I also note that your review of the literature is focused on other similar quantitative studies. None of these studies is able to generate data which furthers the understanding of why people seek help the way they do. For that you need theory building qualitative research. You did not review the qualitative research around parents/patients help seeking.</p> | <p>In our study, we focus on factors of help-seeking rather than motives. Qualitative studies known to us focus mostly on motives, using interviewing or questionnaires. Several studies showed that worry and anxiety are important drivers for contacting OOH services. Parents with young children contact OOH services often, especially for non-urgent problems, with worry, fear, and lack of control as main motives (Hugenhotlz, 2009; Kallestrup 2003). We did not include this in our manuscript as this was beyond the scope of our article.</p>  |
| <p>Your research has taken so much time and effort on the part of all involved, including your</p>   | <p>We feel that this statement of the reviewer is rather strong (using the word unethical), and we</p>   |



|  |   |
|--|---|
| <p>participants and has yielded results which do not contribute to health care. I am of the opinion that this is bordering on unethical when an alternative design, underpinned by a more comprehensive literature review might have generated so much more.</p> | <p>do feel that we add knowledge to our field of research.</p> <p>Overcrowding of OOH care is a well-known problem in many western countries, with negative consequences (see our introduction). To handle the situation we need to understand help-seeking behaviour. One way is to examine differences between countries, so that countries can learn from each other. To our knowledge there are no studies that have compared out-of-hours help-seeking, so we think a literature review cannot answer our question.</p> <p>We also feel that the chosen design was appropriate, taking pros and cons into account, as added to the discussion section, paragraph strengths and limitations (see also above): <i>“The chosen design of using invented cases to measure intended help-seeking behaviour had several strengths and limitations. Strengths were that the respondents received the same cases, making comparisons more straightforward, and that persons who do not use OOH care or healthcare at all were also included. A limitation was the risk of introducing social desirability bias, with the response not representing actual behaviour. Additionally, the absence of emotional reactions that occur in real-life situations could have influenced the response. However, according to the theory of planned behaviour, behaviour is mainly determined by behavioural intentions. A review of literature on theory of planned behaviour concluded that behavioural intentions do predict behavior, while Nagai (2015) found that help-seeking intentions are an important predictor of help-seeking behavior. Several studies used hypothetical case scenarios in out-of-hours care and other settings. Thus, we found that the chosen design was the most feasible and appropriate in relation to our aim.”</i></p> |
| <p><b>Reviewer: 2</b><br/> <b>Reviewer Name: Corinne Chmiel</b><br/> <b>Institution and Country: Institute of Primary Care, University of Zurich, Switzerland Please state any competing interests: None declared</b></p>  |   |
| <p>This is a thoroughly planned and methodological found publication.<br/> Some minor issues should be taken care of, then nothing should stand in the way of publishing the manuscript:</p>   |   |
| <p>Some revisions of the English language</p>  | <p>Our manuscript was revised by a translator before initial submission. Based on the reviewer's suggestions we made some adjustments and a translator briefly checked the revisions.</p> <p>However, we did not perform language editing in the appendix, as the cases in Danish, Dutch, and</p>   |

|  |  |
|--|--|
|  | German were based literally on these cases in English. Thus the wording cannot be altered but should state the exact wording used in the project. A translator has done an idiomatic check of the English cases before translation.  |
| - Figure 1 is not clearly visible in the current form.   | We added a new version of figure 1.  |
| <p>- Concerning the Discussion, Conclusion and Implications I would like to add some personal considerations, which might help to improve the relevance of the manuscript:</p> <p>The astonishing finding of the study is not only the difference in health care seeking of pediatric emergencies, but the lacking difference in health care seeking among the adult population of all three countries. This is especially astonishing, considering the very different health care systems and out of hours care services which are being compared with each other in this study. Especially astonishing is that gate keeping systems as well as non-gate keeping systems show almost the same distribution of out of hours usage. What are possible reasons here for? Is European mentality to similar in these countries to expect differences? If this should be the case, why does the difference only appear in the pediatric setting? Why does a managed care system not show an impact on out of hours care usage? Should managed care not result in reduced unnecessary usage of resources? Please discuss these issues in the discussion section of the manuscript and adapt the conclusion and implications section accordingly.</p> | <p>Indeed, if we look at the overall results (OOH or not), we do not find a difference for adults aged 30-39 year, but we found a small but significant difference for adults aged 50-59 years. A different pattern was found when looking at help-seeking actions, with more ED contacts for Swiss adults and more OOH primary care contacts for Danish and Dutch citizens.</p> <p>We had described this in the discussion, but in line with the comments of the reviewer, we have emphasized this more: “... <i>Yet, we found a difference for Swiss adults aged 50-59 years who more often chose to contact OOH care than Danish and Dutch adults. Swiss adults more often answered ‘wait and see’, but they also more often chose ‘ED’. The difference in healthcare systems (with or without gate-keeping) seems to influence the intended help-seeking behaviour. ... A healthcare system based on gate-keeping may thus lead to less (unnecessary) use of the ED, but not necessarily to lower use of OOH care in general.</i>”</p> |
| The other suggestions for improvements can be appreciated in the attached PDF as comments.   | We have made adjustments accordingly.  |
| Title: rather a questionnaire based interview than an observational study?   | We have changed our title into: “ <i>Help-seeking behaviour outside office hours in Denmark, the Netherlands, and Switzerland: a questionnaire study exploring responses to hypothetical cases</i> ”   |
| Abstract, conclusion: delete lower the workload  | We agree with the suggestion and adjusted accordingly.   |
| Introduction, line 3: work pressure: workload for the involved OOH personel  | We prefer work pressure in this sentence, adjusting into: “work pressure <i>for OOH staff</i> ”.   |
| <p>Methods, cases: “Item selection was done using Rasch analysis to ensure that all the items included in the test were sufficiently unidimensional and to maximize the test information across the interested continuum of the latent constructs. This resulted in the selection of six</p>   | We have rewritten this sentence to improve clarity: “ <i>The cases were treated as items in a Rasch analysis. This was done to eliminate redundant cases with respect to estimating the latent variable for intention to seek help. Cases were reduced, and we selected six cases for children and six for adults.</i> ”   |

|  |  |
|--|--|
| cases for children and six for adults.” → Sentence unclear (across the interested continuum....)   |  |
| Discussion, strengths and limitations: Necessary numbers according to the power calculation achieved?  | We added the following: <i>“Finally, to obtain an eight percent difference between groups, we needed 600 respondents; this was not achieved for all age groups”.</i>   |
| Discussion, strengths and limitations: Highlight as strength international collaboration to a subject which all countries have to deal with likewise. Out of Hours care is a strong and complex issue in all the countries.  | We made some adjustments to highlight this: <i>“We were able to include citizens from three countries for our study by using a consumer panel in two countries. OOH care is a complex issue, which currently faces challenges in many European countries”.</i> |
| Table 1, title: delete here mean, %  | We deleted this, as we can see that this information is presented in the rows further down.  |
| Table 1: explain here why for CH no data on 0-4 years is shown   | We added a footnote to clarify this: <i>“Switzerland had no age group 0-4 years, due to restrictions of the consumer panels”.</i>  |
| Table 2: delete contents of the (.....)  | Adjusted.  |
| Figure 1: Figure not well readable in this black and white format. Bars/colours should be labeled and figure legend should be added  | We adjusted the figure.  |
| Appendix, box: The title "Cases for children" appears twice  | Adjusted.  |
| Appendix, table 1/2: Also here % mentioned twice but mean in the left column only once. Consistent usage of labelling throughout the manuscript  | Adjusted.  |
| Appendix, table 2: This footnote is incomprehensible   | Adjusted: <i>“Information was only available on children for the general population, whereas information on the respondents was on parent/care-giver, who was the decision maker and answered the questionnaire”.</i>  |
| Appendix, table 3: This reference belongs in the methods section and not in a table footnote   | We referred to the consumer panels in the methods section, paragraph design and population, reference 14. We added the second reference.   |
| <b>Reviewer: 3</b><br><b>Reviewer Name: Catherine Pope</b><br><b>Institution and Country: University of Southampton UK Please state any competing interests: Currently a co-investigator on a NIHR funded project examining sense making in relation to urgent care.</b> |  |
| This paper and one based on the same study came to me for review from separate journals. There is very significant overlap in the description of the methods and other aspects of the reporting in both papers and therefore some of my                                  | <p>The methods are indeed similar, as both articles present different results of the same data collection.</p> <p>As written above: We decided to write two papers</p>   |

|  |   |
|--|---|
| <p>comments may be repeated in both reviews. The editors should ask the authors for the parallel paper in each case to inform their decisions.</p>   | <p>on this study, to be able to focus on the specific help seeking intentions (wait&amp;see, self-care, own GP next day, out-of-hours primary care, emergency department, ambulance care) and on factors related to help-seeking outside office hours.</p> <p>The first paper, as submitted to BMJ Open, we focus on clinical help seeking behavior of citizens of different countries, describing the actual intended help seeking, which could be related to health care systems.</p> <p>The second paper focuses on factors related to intended help seeking outside office hour (as a dichotomous outcome variable), thus enabling us to study multiple relevant factors related to help seeking, based on extensive literature. Here we also included the variable 'country', to investigate whether a difference remained after correcting for other known factors.</p> |
| <p>The analysis is descriptive and the findings confirm what is known such and it is difficult to see what this paper really adds, despite drawing comparative data from 3 countries. When discussing the existing literature on p13 the authors note the importance of health system differences that seem a more interesting avenue for research but which are not the topic of this paper.</p>  | <p>This article describes the overall help seeking of populations of three countries in our study. Our results confirm other studies in this field, but in addition show that with similar design for adults only small differences exist when taking all help seeking actions into account, related to health care system issues, and some differences exist for children, which provides room for discussion.</p> <p>The added value of our study is that we were able to use one design in three countries, which enabled us to compare intended help seeking behaviour in detail.</p>   |
| <p>The design of this study is not clear. The authors refer to a cross sectional observational study in the abstract and methods section p4. This is a little misleading as the survey investigates stated preferences based on hypothetical scenarios, and the paper needs to be much clearer that the data describe preferences not actual behaviour. For example on p4 the aim is stated as ' to study how individuals from different age groups in 3 countries...react to acute health problems occurring outside office hours' – it may be worth inserting 'hypothetical scenarios about' before the word acute. Another example is on p12 line 30 'more often contacted OOH care' should be rephrased as 'expressed preference for contacting OOH care'.</p> | <p>We have adjusted the text, to avoid misunderstanding. We had included the phrase "<i>intended</i>" to clarify that our study investigated help seeking based on hypothetical scenarios, but we chose not to repeat this for every statement in the results section, to improve readability. By no means, we want to risk being misleading, and thus we went through our manuscript again to add "<i>intended</i>" at essential sentences. In particular, we replaced the word "<i>contacted</i>".</p>  |
| <p>The 6 hypothetical scenarios used vary in the depiction of urgency (and thus in 'appropriate' disposition) yet the analysis assumes that all the preferences are equivalent and score 1 (or zero) - on p9 line 43 "we calculated a score between 0 and 6 for the number of cases for which OOH contact had been chosen" suggests that some scores of 3 would have a very different meaning (in terms of 'appropriateness' of disposition) than other scores of 3. I was confused about this. It</p>   | <p>It is correct that a score of 3 can have different meanings if you take appropriateness into account. The assessment of urgency levels was used to realise a distribution of cases with different urgency and expected health care needed in order to include a variation of help seeking. In the current paper, we are only interested in the extend of OOH help seeking, without taking appropriateness into account. In other words, we aimed to measure the degree of</p>  |

|  |   |
|--|---|
| <p>did not seem that there was a pre-specified analytical plan which I might expect for such an analysis. It may be that this just needs better explanation and justification but I would urge that the paper has an expert statistical review.</p>                  | <p>help seeking outside office hours regardless of its appropriateness.</p> <p>We have decided upon this, as appropriateness is a different topic, with varying perspective between professionals and citizens (as a reviewer above already mentioned). As this was not clear for the reviewer, we have made some small adjustments in our manuscript (methods sections, paragraph development of questionnaires and paragraph cases).</p> <p>A statistician was involved in the development and analyses of our study.</p>   |
| <p>The authors note that the use of invented cases may introduce social desirability bias and may not represent actual behaviour –this seems to be a significant weakness.</p>   | <p>In the discussion, we added some lines about the pros and cons of using case scenarios (see also above): <i>“The chosen design of using invented cases to measure intended help-seeking behaviour had several strengths and limitations. Strengths were that the respondents received the same cases, making comparisons more straightforward, and that persons who do not use OOH care or healthcare at all were also included. A limitation was the risk of introducing social desirability bias, with the response not representing actual behaviour. Additionally, the absence of emotional reactions that occur in real-life situations could have influenced the response. However, according to the theory of planned behaviour, behaviour is mainly determined by behavioural intentions. A review of literature on theory of planned behaviour concluded that behavioural intentions do predict behavior, while Nagai (2015) found that help-seeking intentions are an important predictor of help-seeking behavior. Several studies used hypothetical case scenarios in out-of-hours care and other settings. Thus, we found that the chosen design was the most feasible and appropriate in relation to our aim.”</i></p> |
| <p>A power calculation is mentioned on p8 but not justified and then on p13 the paper concedes that the study lacked power. The paper goes on to suggest that other factors (not studied) may cause the variation p14 rather undermining the analysis presented.</p> | <p>For the power calculation, we made assumptions on the expected differences in the use of OOH care. We assumed a difference of 8% to be a relevant difference. We have adjusted our manuscript: <i>“A power calculation showed that we needed 600 returned questionnaires per age group to be able to find 8% difference between countries, which we considered a clinical relevant difference”</i>.</p> <p>We removed our remark in the discussion section on p13 on lack of power.</p>  |
| <p>The Netherlands and Swiss samples are drawn from consumer panels which the authors concede is problematic p9 line 32. There is variation in the response rates reported and</p>   | <p>A general concern with consumer panels is the potential lack of representativeness; to make sure that our selection was representative (which used several key variables to make the selection, as</p>   |

|   |   |
|---|---|
| <p>concerns about the representativeness of the eventual samples – see p10.</p>   | <p>described in the manuscript) we did an extra comparison. We made the following adjustment to clarify this: “We also performed descriptive analyses to compare respondents with the general population in the Netherlands and Switzerland. <i>This was done because we wanted to check the representativeness of the consumer panels that we used in these two countries</i>”.</p> <p>We made an adjustment in the discussion section, paragraph strengths and limitations: “Our Danish sample was representative for the general population, and our Dutch and Swiss panels were also able to select quite representative samples for a range of background characteristics <i>although some small statistically significant differences existed</i>.”</p> |
| <p>The paper would benefit from a better description and justification of the analytical approach.</p>  | <p>We reread our paragraphs power calculation and analysis and made adjustments to clarify and justify.</p>   |
| <p><b>Reviewer: 4</b><br/> <b>Reviewer Name: Matthew Booker</b><br/> <b>Institution and Country: University of Bristol, UK Please state any competing interests: None declared</b></p>  |   |
| <p>Thank you for the opportunity to review this manuscript, which discusses the important area of ‘out of hours’ care access. It presents a component of research (which I believe from the authors’ references in the document to be part of a larger project) exploring how randomly-selected questionnaire respondents indicate they would be likely to respond to hypothetical urgent care case scenarios. The sample is drawn from Denmark, Netherlands and Sweden. The emphasis is upon the difference of indicated behaviours between countries, with attempts made to correct for a limited number of potentially confounding variables.</p> <p>The topic of decision-making in urgent care is undoubtedly very important, and the work presented does look at this from a different and interesting angle. In particular, the construction of hypothetical cases is an interesting method, and there is much merit to the significant work the authors have undertaken. Indeed, the process of the development of these cases has clearly been somewhat of a strength and should be commended.</p> <p>However, I do have a number of comments about the manuscript in its present format, that I feel would need to be addressed before I could recommend it for publication in BMJ Open. I have summarised these below.</p> |   |
| <p>Title and abstract – I have concerns that the title in its present form does not accurately reflect the methodology. Calling the work a ‘cross-sectional observational study of help-seeking’ to me</p>  | <p>We can see that our wording was not accurately enough, and we have adjusted our manuscript in line with the suggestions of the reviewer. We changed our title into: “<i>Help-seeking</i>”</p>  |

|   |  |
|---|--|
| <p>implies that the participants' actual consulting behaviour/choices were analysed (and to me, implies the use of routine data in some way). The manuscript actually presents an analysis of the response to randomly invited questionnaires about hypothetical actions in artificially constructed case-situations. Whilst I feel that this a very reasonable way to explore an aspect of help-seeking behaviour, it is a study about how people say they might act in a hypothetical situation, and as such I think the title needs to be reworded to reflect that very clearly “e.g. a questionnaire study exploring responses to hypothetical cases... etc etc” to avoid confusion. Similarly the abstract needs rewording a little to avoid that confusion.</p> | <p><i>behaviour outside office hours in Denmark, the Netherlands, and Switzerland: a questionnaire study exploring responses to hypothetical cases”.</i></p> <p>The description of the design was changed into: <i>“A questionnaire study exploring responses to six hypothetical cases ...”.</i></p> <p>In addition, we went through the manuscript again, adding ‘<i>intended</i>’ to help-seeking as well as ‘<i>hypothetical</i>’ to case scenarios to clarify our study design and avoid the risk of misunderstanding.</p>  |
| <p>Methods - I think the use of hypothetical case vignettes (personally, I would prefer this terminology over ‘invented’ cases) is a very reasonable way of exploring what people might say they would do in certain circumstances. I think there needs to be much more reference to studies that have used hypothetical case vignettes successfully (including a more detailed critical discussion of the pros and cons of this method) to give more weight to why this is an appropriate method. I also think it would be useful to reference other studies that have used hypothetical cases as a way of exploring decision making in other settings (i.e. non urgent care) - there are a vast array of these in the literature.</p>                               | <p>We have replaced “predefined cases” by “hypothetical cases” throughout the manuscript.</p> <p>Also, we added a statement on pros and cons of using case scenarios in the discussion section (as mentioned above): <i>“The chosen design of using invented cases to measure intended help-seeking behaviour had several strengths and limitations. Strengths were that the respondents received the same cases, making comparisons more straightforward, and that persons who do not use OOH care or healthcare at all were also included. A limitation was the risk of introducing social desirability bias, with the response not representing actual behaviour. Additionally, the absence of emotional reactions that occur in real-life situations could have influenced the response. However, according to the theory of planned behaviour, behaviour is mainly determined by behavioural intentions. A review of literature on theory of planned behaviour concluded that behavioural intentions do predict behavior, while Nagai (2015) found that help-seeking intentions are an important predictor of help-seeking behavior. Several studies used hypothetical case scenarios in out-of-hours care and other settings. Thus, we found that the chosen design was the most feasible and appropriate in relation to our aim.”</i></p> <p>Furthermore, we have added some references to studies that have used hypothetical case scenarios in out-of-hours settings as well as other settings.</p> |
| <p>Age groups – the authors have selected to send questionnaires to 3 age groups (0-4, 30-39 and 50-59 years). I note that the authors have explained their reasons for this to give 'adequate power' and/or to be able to author hypothetical cases the situations these groups may experience most closely. However, I do have</p>  | <p>Including citizens from all age groups would have been most optimal. However, this was not feasible: cases are different for several age groups, in particular for children and adults, and health care needed is different for cases depending on the age group (we have plans to also investigate appropriateness of help-seeking).</p>   |

|   |  |
|---|--|
| <p>some significant concerns about the relatively narrow age selections that are therefore included in this study, and how representative as a whole conclusions might be. For example, exclusion of the elderly and retired is a potentially significant flaw, as is exclusion of young working adults. I note that the authors cite discovering high variation in one of their previous studies in 0-4 year olds and 25-35 year olds as a justification, but then the age band of 30-39 chosen for this study would not fit with this and therefore seems a bit confusing? I think more detail about the justification for these selections, and a much more critical discussion of the implications of excluding other age groups is needed.</p> | <p>If including citizens from all age groups, analyses are still relevant for different age bands; including enough citizens per age band to achieve sufficient power would make the study too extensive and expensive to conduct.</p> <p>Therefore, we have chosen to focus on a few age groups that have shown some differences in a previous study. Extrapolating our findings to the entire population would be imprecise, in particular since our findings vary for the different age groups. Therefore, we have been careful to describe our findings per age group, without making statements on a population level. We have carefully reread our manuscript and made adjustments accordingly.</p> <p>Indeed 30-39 is not completely in line with the findings of our previous study. We defined our age groups in a way that groups were as homogenous as possible. Younger adults vary a lot, from students to parents.</p> <p>We agree with the reviewer that including elderly would be interesting too. We had to define a limited number of groups, as stated above. Parents of young children and young adults were the most interesting groups, as they more often contact out-of-hours care for non-urgent health problems that could be seen during office hours later. We found it most relevant to focus on age groups that could have room for adjusting behavior. Furthermore, consumer panels were the best method to get representative selections of citizens in the Netherlands and Switzerland. In these panels elderly are underrepresented and not representative, thus introducing selection bias.</p> <p>We have made some adjustments in our methods section to justify our selection: <i>“We composed the age group of individuals aged 30-39 years, as we expected more homogeneity in this group than in the group of individuals aged 25-35 years.”</i> Furthermore, we have added a statement on the implication of choosing specific age groups in the discussion: <i>“The use of three age groups with varying results limited the generalisability of our results to the entire population of the included countries. The results could be rather different for other groups, such as the elderly.”</i></p> |
| <p>Questionnaires – the authors note that ‘questions on factors related to help seeking were part of a larger study and will be described in further detail in another scientific article’. Whilst I understand the need to avoid duplicate publication, I think as a minimum a brief overview of the significance of these questions should be included here. I have struggled a little to understand how this work</p>  | <p>We can see the potential confusing with regard to our two papers. We have given an explanation about the two papers above, clarifying the relevance of the current manuscript:</p> <p>We decided to write two papers on this study, to be able to focus on the specific help seeking intentions (wait&amp;see, self-care, own GP next day, out-of-hours primary care, emergency</p>   |



|   |   |
|---|---|
| <p>helps inform the collective knowledge base about decision making without this, as its omission makes this work quite narrow in focus and of more restricted interest. If analysis has not been undertaken yet, a couple of lines about the hypothesis would help understanding how these questions fit with the data presented in this paper, and how the two analyses might complement the wider decision-making research base. The case vignettes are included in the appendix, but is it possible also to include a version of the full questionnaire (ideally in English, but if not possible then in the relevant languages?) as a supplementary file for transparency and clarity?</p> | <p>department, ambulance care) and on factors related to help-seeking outside office hours. The first paper, as submitted to BMJ Open, we focus on clinical help seeking behavior of citizens of different countries, describing the actual intended help seeking, which could be related to health care systems. The second paper focuses on factors related to intended help seeking outside office hour (as a dichotomous outcome variable), thus enabling us to study multiple relevant factors related to help seeking, based on extensive literature. Here we also included the variable 'country', to investigate whether a difference remained after correcting for other known factors.</p> <p>We made an adjustment in the methods section, paragraph development of questionnaires: "The questions on factors related to help-seeking were part of a larger study and will be described in further detail in another scientific article <i>focusing on factors related to intended help-seeking outside office hours.</i>"</p> <p>We did not want to overload the reader with information that is not relevant for the current manuscript. Yet, access to the entire questionnaire can indeed be helpful and give information for the interested readers. Thus, we have added the versions for adults and children in English as a supplementary file. As these contain the written case scenarios, we have deleted those boxes from the manuscript.</p> |
| <p>Case development – the process described here seems very robust and detailed with good piloting, psychometric evaluation and consensus. Was a patient or lay representative included in the development of the cases at any stage? I wasn't quite clear if so.</p>   | <p>After developing our cases we had <i>"two email feedback rounds with eight individuals"</i>. As this might remain unclear, we have changed <i>"individuals"</i> into <i>"lay persons"</i>.</p>   |
| <p>Data collection – Do the authors believe there is likely to be any bias associated with the different modality of questionnaire distribution and completion (i.e. paper versus online?). Has any sub analysis been done to determine if there is a difference between this – I am guessing it may be associated with confounding due to education level or economic status?</p>  | <p>For the Netherlands and Switzerland, only one modality of data collection was included, namely internet questionnaires, making a comparison of different modalities impossible. For Denmark, paper questionnaires were sent, with the option to answer online. About 20% of response was done online. A previous study from Ebert et al (2018), showed that some differences exist between respondents using paper versus online modalities, when offered both. We also found some differences, doing sub analyses for Denmark.</p> <p>The question is whether this affected our study. In general, it could be that citizens who are part of a consumer panel, and thus interested in online questionnaires, are different from the general population. Therefore, we have done a comparison with the general population (see</p>   |

|  |   |
|--|---|
|  | <p>results section paragraph 1). Also, we have correct for background variables in our final analyses.</p> <p>We have added the following in our discussion, paragraph strengths and limitations: <i>“We found some difference in intended help-seeking between the three countries after correcting for differences in several background variables. Yet, different recruitment methods may have introduced some bias, although the effect on differences between the countries and differences between populations and culture remains unclear.”</i></p>  |
| <p>Results and analysis – The crude and adjusted rates of consultation are presented in a table, but this does not break down by hypothetical condition/case – is this because there is no statistical difference between any of the vignettes as individuals, or is this a limitation of the study that it is not powered adequately to do this? Is there anyway that a graphical or diagrammatic representation of how the groups responded to each of the hypothetical cases could be included? For example, it would be very interesting to see visually the spread of “consult OOH” –vs- “don’t consult OOH” for the different cases. I do think some visual representation of this is needed. I also had real difficulty reading and understanding figure 1 with the colourings as greyscale – this may be my peer review copy but I think it does need to be reproduced in colour and some attention given to how well that figure stands alone without detailed textual description.</p>   | <p>The crude and adjusted rates were calculated using the sumscore on the dichotomous variable OOH/no OOH.</p> <p>We have not shown differences on case scenario level, as we were hesitant to present this much information. Considering the comments of the reviewers and the aim of our article, we have adjusted figure 1, adding information on tests of difference on case level. The power is sufficient, equal to the power for the sumscore.</p> <p>We adapted figure 1, in response to comments of several reviewers. Apparently, the range of intended help-seeking plus OOH care was not clear from the previous version.</p>   |
| <p>Discussion – the authors have explained that the intention of the study was to focus on the differences between countries, correcting for potential confounding of socio-demographic variables, and indeed it seems the study was set up as such. I am not sure I completely understand the full rationale for analysing the data this way around – what is the importance of understanding the difference between countries more so than the effect of socio-economic status, deprivation, education, finance etc? Clearly the authors must have a rationale for looking at the differences between countries, but at the moment as a reader this does not come through entirely clearly at the moment – perhaps a few lines in a box or bullet section at the outset would help explain this? As such it is at times a little difficult to understand the practical significance of some of the analysis as it is presented – for example, why is it of any real interest or practical significance that a fairly narrow age band of individuals (30-39 year olds) would hypothetically not consult any differently if they are from the Netherlands vs Denmark? I do wonder if some of</p> | <p>We believe that it is both interesting to focus on differences between countries as on factors influencing help-seeking behavior. See the comment to the editor.</p> <p>We have addressed the second aim in another study on the same data. Describing both aims would be too extensive for one article, and would result in losing interesting and relevant information.</p> <p>High demands in out-of-hours care is a challenge for many European countries, and we study this area to see if there is room for improvement. We first investigated if there is a difference in help-seeking between countries, to see if one could learn from each other.</p> <p>Added in the discussion: <i>“OOH care is a complex issue, which currently faces challenges in many European countries.”</i></p> |

|   |   |
|---|---|
| <p>the interest in this might be around the questions about help-seeking (see my above comment) which makes their inclusion here even more important.</p>   |   |
| <p>Strengths and limitations – The youngest age group (0-4) seems – from the authors reference to their previous work and from the data presented – to be of interest. I think parental roles and family constitution will be hugely influential here (single parent versus 2-parent families, availability of grandparents, extended family, other children in the family and therefore previous parental experience versus first child etc etc) – there are a really significant number of shaping influences here that – whilst extremely interesting – can not be explained by this data alone. I think it is necessary to reflect on how this fits in with this understanding – see above.</p> | <p>We agree with the reviewer that this difference is interesting. As we refer to above, we test the influence of social support and the amount of children on help-seeking in a second article that aimed to look at factors related to help-seeking outside office hours.</p> <p>A previous study, comparing Danish and Dutch help-seeking in out-of-hours care using actual register data already found a difference for this age group. This study was the basis for selection of this particular age group to check whether the difference was consistent. Several hypotheses could be related to this difference, as are also suggested by the reviewer, which will be addressed in the second article.</p> |
| <p>In summary, I think there is much merit in the contents of this work as it does provide evidence that looks at an important issue. However, prior to recommending publication I think it needs to be more accurately titled and described, it needs a much clearer explanation of why the question of comparing countries is important and relevant, why the groups selected were appropriate, and some additional detail as mentioned above in the results, conclusions and discussion sections to adequately explore the limitations.</p>  |   |
| <p><b>Reviewer: 5</b><br/> <b>Reviewer Name: Professor Daniel Lasserson Institution and Country: University of Birmingham, UK.</b><br/> <b>Please state any competing interests: None declared.</b></p>   |   |
| <p>Huibers and colleagues have undertaken a very interesting and important study, comparing help seeking behaviour in the out of office hours time period, across three European countries.</p> <p>I have several questions and comments for the research team to consider. Overall, my comments are focussed on how much of the differences between countries could be due to the different patient selection mechanisms, and that this should be considered more deeply in the discussion.</p>  | <p>We thank the reviewer for acknowledging the relevance of our study.</p>  |
| <p>Comments by section:</p>   |   |
| <p>Methods:<br/> 1. Dutch and Danish Ooh data were used to identify frequent presenting problems and not data from Swiss OOH primary care. Were there differences between Dutch and Danish data?</p>  | <p>We looked at the top 30 of registered complaints, using ICPC coding from OOH primary care services of one Danish region (data from a previous study and one Dutch data (with routine ICPC coding), both for all contacts as per contact</p>  |

|   |   |
|---|---|
| <p>Excluding Swiss data may introduce bias into the results if different prevalence conditions are used in each setting.</p>  | <p>type. We made an overall assessment of types of complaints and organ groups. Some differences existed between Danish and Dutch data, which could also be the result of differences in ICPC coding. A previous study has found that reasons for encounter were quite similar (Huibers LAMJ, Moth G, Bondevik G, Kersnik J, Huber CA, Christensen MB, Leutgeb R, Remmen R, Wensing M. Diagnostic scope in out-of-hours primary care services in 8 European countries: an observational study. <i>BMC Fam Pract</i> 2011;12:30. doi: 10.1186/1471-2296-12-30).</p> <p>The reviewer suggests that excluding Swiss data of OOH primary care may have introduced bias. The Swiss health care system is different from the Dutch and Danish system, as described in our methods section. This may result in some differences in prevalence of health problems in different healthcare services. Yet, the cases have been checked by two Swiss colleagues, who found them relevant for Switzerland. Furthermore, the six cases for adults are all in the top 17 of frequent reasons for encounter, as studied by Huber et al (<i>Out-of-hours demand in primary care: frequency, mode of contact and reasons for encounter in Switzerland. J Eval Clin Pract</i> 2011;17(1):174-9).</p> <p>As the health problems described in the cases were explicitly checked to match frequent reasons for encounter in the Swiss healthcare system, we added this to our method section <i>“The relevance of the health problems described was checked and found relevant for the Swiss healthcare system.”</i></p> |
| <p>2. Were cases discussed with researchers and GPs in each country? Given that this study was undertaken in different countries, more clarity is needed on the representation of each country in the case study development.</p> | <p>Yes, the cases were discussed with researchers and GPs from each country, though the majority came from Denmark and the Netherlands. From Swiss two GP-researchers contributed to the case study development.</p>  |
| <p>3. Why wasn't the readability of the Dutch and German versions piloted for readability, like the Danish?</p>   | <p>We decided to only pilot our questionnaire in Denmark, due to pragmatic considerations. We mostly expected possible readability issues in the general phrasing of questions. This was taken into account when translating the questionnaire into Dutch and German, using forward-backward translation. Therefore, we expected the risk of readability problems in Dutch and German limited, and few comments were reported by the responders, although we cannot exclude readability issues.</p> <p>Methods section, paragraph pilot testing: <i>“Due to pragmatic consideration, we only performed a pilot test in Denmark”.</i></p>  |

|   |   |
|---|---|
| <p>4. More detail is needed about the power calculation (based on other empirical data, or assumptions?)</p>  | <p>For the power calculation we made assumptions on the expected differences in the use of OOH care. We added: “A power calculation showed that we needed 600 returned questionnaires per age group to be able to find an 8% difference between countries, <i>which we considered a clinical relevant difference.</i>”</p>  |
| <p>5. More detail is needed about recruitment methods – it is clear that the invitation is different (post vs email), were there differences in patient selection methods? The Analysis section seems to suggest that recruitment methods were very different between settings.</p> | <p>Information on the recruitment methods is written in the methods paragraph on design and population: “We used the Danish Civil Registration System to randomly select representative individuals among the five Danish regions. We excluded individuals living in institutions and individuals with address protection. The Dutch and Swiss samples were selected using consumer panels (The Netherlands: TNS Nipo; Switzerland: Respondi and Bilendi). The Dutch sample represented the population on age, gender, and region (0-4 years), and age, gender, region, education, and ethnicity (both adult age groups). For Switzerland, it was only possible to include adults selected on age by using two panels to reach 600 respondents, <i>as information about children of panel members was not available.</i>”</p> |
| <p>6. The lack of children’s data from Switzerland needs to be explained. This is a weakness of the study in terms of a multi-country comparison.</p>   | <p>The Swiss consumer panel companies were unfortunately unable to select parents of children 0-4 years of age. We added information to our methods section: “For Switzerland, it was only possible to include adults selected on age by using two panels to reach 600 respondents, <i>as information about children of panel members was not available.</i>”</p>   |
| <p>Results:</p>   |   |
| <p>7. Different demographic data are presented in each country, this should be made clearer in the methods. This makes comparisons across datasets more difficult. I think the research team should include how this may have introduced bias later in the discussion.</p>          | <p>We are unsure to what data the reviewer refers, but we assume that he refers to the comparison between respondents and non-respondents as presented in the appendix (table 1, 2, and 3). As described in the methods section, paragraph analysis, we have compared the respondents and non-respondents. Due to restrictions on data availability of the general population, we did not have access to all demographic variables for each country.</p> <p>And, as shown in table 1, the demographic data vary in the different countries. Therefore, we have adjusted for age, gender, education, and ethnicity in table 2.</p> <p>Some small differences existed, and we have added information to the discussion section, paragraph strengths and limitations: “... and our</p>   |

|   |  |
|---|--|
|   | Dutch and Swiss panels were also able to select quite representative samples for a range of background characteristics <i>although some small statistically significant differences existed.</i> "   |
| 9. Figure 1. The legibility needs to be improved.   | We have adjusted figure 1 after comments of the reviewers.   |
| Discussion:   |  |
| 10. I don't think it is fair to put the differences in results entirely down to the differences between countries. The methods of recruitment of participants differed as well and this could have introduced a bias which may explain the results. This should appear as a separate section in the discussion. | We added the following to the limitation section in the discussion: " <i>We found some difference in the intended help-seeking between the three countries, after correcting for differences in several background variables. Yet, different recruitment methods may have introduced some bias, although the effect on differences between the countries and differences between populations and cultures remains unclear.</i> " |

#### VERSION 2 – REVIEW

|                        |  |
|------------------------|--|
| <b>REVIEWER</b>        | Chmiel Corinne<br>University Hospital Zürich, Institute of Primary Care, University of Zurich, Pestalozzistrasse 24, CH-8091 Zürich, Switzerland, Email: corinne.chmiel@usz.ch, Tel: 0041 44 255 98 55, Fax: 0041 44 255 90 97, www.hausarztmedizin.uzh.ch |
| <b>REVIEW RETURNED</b> | 07-Mar-2018  |

|                         |   |
|-------------------------|---|
| <b>GENERAL COMMENTS</b> | The revision was performed thoroughly and all concerns of the reviewers were addressed. The manuscript can be accepted for publication. |
|-------------------------|---|

|                        |   |
|------------------------|---|
| <b>REVIEWER</b>        | Matthew Booker<br>University of Bristol, UK |
| <b>REVIEW RETURNED</b> | 09-Mar-2018                                 |

|                         |   |
|-------------------------|---|
| <b>GENERAL COMMENTS</b> | Thank you for the opportunity to review a revised version of this manuscript. Whilst I note that other peer reviewers' comments covered different areas to my own initial points, I feel that the issues I highlighted previously have been mostly addressed to a satisfactory extent to support publication. |
|-------------------------|---|

|                        |   |
|------------------------|---|
| <b>REVIEWER</b>        | Daniel Lasserson<br>Institute of Applied Health Research, University of Birmingham, UK. |
| <b>REVIEW RETURNED</b> | 09-Mar-2018   |

|                         |   |
|-------------------------|---|
| <b>GENERAL COMMENTS</b> | I thank the authors for responding to my comments. The current manuscript reads well and makes clear and valuable findings which will be helpful for designers and providers of OOH primary care. |
|-------------------------|---|