PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Satisfaction of 30,402 Callers to a Medical Helpline of the Emergency Medical Services Copenhagen: A Retrospective Cohort Study
AUTHORS	Zinger, Nienke Doreen; Blomberg, Stig Nikolaj; Lippert, Freddy; Collatz Christensen, Helle

VERSION 1 – REVIEW

REVIEWER	JOSE JOAQUIN MIRA UNIVERSIDAD MIGUEL HERNANDEZ, SPAIN
REVIEW RETURNED	20-Mar-2019

GENERAL COMMENTS	This is a good study. Direct, simple and relevant aim. Methods are described in a right way. Results, methods and conclusion are related to this aim. Limitations are included and they are reasonable. In this study authors take advantage of the systematic evaluation of the non-emergency medical helpline in the Capital Region of Denmark. So, they can describe what people think about this new service. They present results considering certain variables that are coded in each phone call. So, they can identify factors related to satisfaction. Statistics used are right and they have been applied in the right way. However, the sample (30,402) should let them segment the sample to provide additional outcome. For example, what happen with the subjects when they repeat call in a short time or during the time of this study. In fact, they do it when calls concerning 0-4 years old have been analyzed. This study is based on two questions. I have some doubts if the term questionnaire is right for this study. I think authors must justify why they include to some extent or to a moderate extent in the pool of satisfied subjects. The response rate was 23%. Authors could explain why they ruled out making a study asking people who refuse replied when their call ends. For example, this issue should be explained in the limitation section. Ethics issues are not included.

REVIEWER	kira Leeb Victorian Agency for Health Information Melbourne, Australia
REVIEW RETURNED	24-Mar-2019

GENERAL COMMENTS	While the study provides a point in time understanding of how patients are using a triage call line when in need of non-urgent

	medical help, it is difficult to understand the key messages of the study given the analytical approach. The sample included in the study is sufficiently robust such that many of the multiple comparisons between respondents and non-respondents would be significant. However it is difficult to discern if any of these differences are clinically meaningful. Understanding the meaningfulness of any statistical differences is further hampered when all positive response categories are included in the "satisfaction" group and then compared to the "dissatisfied" group. The authors have chosen to label as satisfied those that were highly satisfied ("to a great extent") and those that were less so (either some or moderately satisfied). The analysis may be stronger if the authors entertain an analysis that only considers "top box" respondents against those that were either limited or not at all satisfied. By doing the analysis in this way may point to places where telephone line support would absolutely be useful or where follow up surveys could be administered to understand the true differences in patient experience with these types of initiatives. It is clear that all health care systems are suffering from similar pressures of increased demand, aging populations and increased chronic disease while at the same time dealing with fiscal constraints. Successful initiatives that will effectively keep patients out of ED or gp care are important. However, a healthcare system that doesn't distinguish between providing high quality care and mediocre care is likely not one that will be efficient or sustainable in the long run. Knowing which patients are most likely to benefit from telephone triage lines and targeting these patients for optimal care may be more likely to add value than simply dividing patients between two outcomes and comparing responses between them. In short, I think the authors would serve the audience better if they presented some hypotheses about which populations they believe might be very well served by a telephone triage line and then set about testing those hypotheses rather than analyzing all comparisons to see what might be statistically significantly different and reporting on those differences. More specific comments: table 1 - likely easier to review if the table was simplified to show respondents and non respondents rather than breaking down the respondents by satisfied/dissatisfied. Also not all figures in the text match the table (e.g. call forwarded to physician). Figure 3 has an apparent dip both in respondents and satisfaction (May 17). Can the authors suggest any reasons? Discussion would be strengthened by distilling a few key messages from the analysis to allow the audience to understand the take home messages from the study.

REVIEWER	Matthieu Heidet Assistance Publique - Hôpitaux de Paris (AP-HP) SAMU 94 Hôpitaux universitaires Henri Mondor Créteil France

REVIEW RETURNED	26-Mar-2019

GENERAL COMMENTS	The article is overall clear and well-written. The settings, objectives, methods and results are globally clear, and the subject is of strong importance for EMS systems. Nevertheless, some questions remain, and should be clarified for publication, which i would support afterwards: INTRODUCTION 1) "Such an OOH telephone line has been established [...]" : i don't understand what is a "comprehensive EMS system" ? I am unable to assess the relevance of the reference (7), apparently written in Danish. 2) "Furthermore, a subgroup analysis [...]" : is the "frequent use" of callers aged 0-4 known, or suspected ? And is there a reference to support this asumption ? MATERIALS AND METHODS 1) Study design and settings a) Can the authors precise if callers can be forwarded from 112 to 1813, and what was the proportion of such forwards in their study sample ? It could be interesting to add this characteristic to the model, because more worried patients initially calling 112 could be more dissatisfied with the final response mode. b) Can the authors detail the process of randomization ? I miss details to understand it. 2) Data collection a) can the authors justify the choice to classify callers as "satisfied" when the latters responded being satisfied "to a moderate extent" ? It may have been more discriminant to consider only the most satisfied callers (to a large + to a great extent). This question would be partially answered by giving détails on the size of each class. 3) Statistical analyses a) Authors should consider the use of the terms "univariable" and "multivariable" instead of uni/multivariate (single outcome: satisfaction). b) Can the authors explain and/or justify the choice for a full fitted model ? Did they try other strategies to select the variables of their model (stepwise, other) ? c) Why did the authors perfom subgroup analysis on children only ? Can they justify the reason for not considering another subgroup in other age classes (5-100 yo seems wide and non-homogeneous : why not analysing older patients as well ?). Indeed, satisfaction rates vary in classes 60-79 and > 80 yo (table 2). RESULTS 1) Characteristics of study subjects a) 1.731.556 calls were "eligible" (not "included" ?)

	b) Response rate 23.0 % : i cannot acknowledge this proportion. 23% of which sample ? c) In the paragraph "Multivariate logistic regression", before referring to table 2 for exhaustive detailed results, authors should consider adding important results in text (somatic injury : OR, IC95), telephone consultation, waiting time, call-taker type. d) Table 2, OR for triage response : the authors should consider inverting the presentation of their results (reference : telephone consultation) so that results are similar to those presented in the abstract (face-to-face consultation : OR 2.18 [1.96-2.43] ; telephone consultation : 1). e) Table 2: authors should consider displaying p-values for adjusted OR DISCUSSION 1) Sample selection a) The process of randomization may have biased the sample selection (cf. data collection) b) Responders may have been only the most satisfied/dissatisfied patients, as authors state ("self-selection of people who responded"). Yet, i could not find a discussion on Figure 3, which illustrates that potential bias, through an apparent trending relationship between satisfaction and numbers of respondents (cf. may 17) 2) Evaluation of satisfaction a) the discussion lacks considering qualitative evaluation of the conversation between the caller and the call-taker (why would physicians provide less satisfying reponses ?) b) the reasons for dissatisfaction are not discussed herein. It seems that callers request a measurable response, not only an advice (especially callers aged 18-59 ?). The authors should consider discussing the expectations of the callers, based on references if applicable. I understand that patients request a measurable response for measurable affections (injuries, GP encounter). REFERENCES Reference n.7 is written in Danish: i am unfortunately unable to assess its appropriateness. Other references are appropriate and appear up-to-date, to the best of my knowledge.

REVIEWER	Patricia Wilkie President National Association for Patient Participation, UK
REVIEW RETURNED	26-Mar-2019

GENERAL COMMENTS	1. Is abstract accurate, balanced and complete. The objectives state that more information is needed on factors that influence caller satisfaction in order to increase the quality of such services. Sadly the information sought is very limited with the authors focusing entirely on demographic characteristics of the caller and 2 questions about satisfaction. I appreciate the limitations of the

	study. However triage needs to consider patient choice and patient autonomy. 2. Research Ethics, consent and information for participants. If there are no ethical considerations that should be stated. 3. I do not know the rules about consent for research in Denmark. There was an intervention a text message comprising 2 questions following a call to the medical helpline. The results were then linked to the patient registration. Should the caller be informed about this? 4. Are outcomes clearly defined 5. Statistics Please see comments below under 9. 6. There is a long list of references. Please check doi references 7. Results The sample of those receiving question texts is meant to be randomly selected. To confirm this one could compare the distribution of demographic and other independent variables between the Receivers and the Non-receivers. Because the demographic data is not available for the Non-respondents among the Receivers , such a comparison is not possible, and in any case would just confirm the randomness. However, it is worth comparing the demographic details between Responders and Non-responders, to see whether any particular category is more or less likely to respond. This cannot be done directly, but it can be done in a better way than the authors have done. For simplicity I ignore the minor exclusions, and assume that there are 102,473 + 30,402 = 132,875 Receivers. Consider Sex. Assume that the distribution among the Receivers is the same as among the non-receivers. If the mailing is random it should be, apart from sampling errors, which will here be a small. Among the Non-receivers there are said (in Table 1) to be 901247 Females and 742677 Males, which add up to 1,643,824. This is 57,230 short of the stated total of 1,701,154, and it is not clear where these have gone. The percentages given are 53.0% and 43.7% which add up to 96.7%, a shortfall of 3.3%. The proportions of Females and Males based on the correct sum are 54.823% and 45.177%. There are similar shortfalls among the Satisfied and Dissatisfied, and these all need to be explained or corrected. The totals of Females plus Males among Satisfied plus Dissatisfied is 29,824 (not 30,402). Adding the given 102,743 Non-respondents gives 132,567. Now assume that these are split among the Sexes in the same way as the Non-receivers. This gives, in whole numbers, 72,677 Females and 59,890 Males. Now deduct the Respondents, 16,650 Females and 13,174 Males. This gives for the Non-respondents an estimated 56,027 Females and 46,715 Males, 54.5% and 45.5%. This means that the third sentence of the paragraph at the top of page 8 should start: "Respondents were more often Female (55.8% vs 54.5%) ... ". These are small differences, and I have not done the same calculations for the other categories, but there might well be bigger differences than shown.

	In any case the shortfalls need to be explained. If there is a category "Sex not known" then it should be included explicitly; and the same with the other characteristics" " I would like the authors to explain why they have put answers with Category 3 in with 4 and 5, instead of with 1 and 2. One can see from Figure 2 that Categories 4 and 5 add up to about 65% to 70% anyway. One can argue that being satisfied "to a moderate extent" sounds like being not fully satisfied. It is also not clear (but this may be a matter of English versus Danish) what is the difference at the foot of Figure 2 between "to a great extent" and "to a large extent", but the latter is given on page 5 as "to some extent". 8. Are results presented clearly 9. Discussions and conclusions justified by results? It is not clear whether the gender of the caller for a child of 4 and under was included. The caller for this young child may not be the parent and records may not be known about them. They are not really justified because authors should really not have put the middling group with the satisfied but wot the less satisfied. 10. Are study limitations discussed fully? It is indeed a short questionnaire. However the first question was “are you overall satisfied with the contact you had with the medical helpline?” If respondents answered yes to that question the second question “ were your questions answered during the contact with the medical helpline 1813?” is really irrelevant. So it is not surprising that most respndents answered yes to both questions. Perhaps the questions were put in the wrong order? Discussion about patient satisfaction is very limited. Authors could mention the expectations of callers to the helpline in terms of the expected qualification of the staff of the helpline, of whether the caller got better. The description of satisfaction is very limited 11. Supplementary reporting 12. Plagiarism etc. 13. Standard of written English

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

Methods:

Comment: I think authors must justify why they include to some extent or to a moderate extent in the pool of satisfied subjects.
Reply: Thank you for asking this very important question. We indeed also seriously considered to only take the distinctly satisfied callers and compare their characteristics with the rest of the respondents. However, since this medical helpline was introduced with the aim to be more patient friendly than the old (more dispersed) system, we decided that it would be of most interest to discern the very dissatisfied (e.g. ‘to a limited extent’ and ‘not at all’ answer categories), from those who are at least ‘to a moderate extent’ content with the service they received. Our aim was to try to identify certain characteristics that make callers very likely to become dissatisfied. Only after this identification and improvement procedures for this subpopulation, we would like to focus on the callers who are only moderately satisfied. Yet, we decided to write this article with the focus on characteristics that could influence satisfaction as opposed to dissatisfaction, because we envision that policy makers’ and

healthcare professionals should focus on system improvement rather than on making satisfaction outcomes less worse.

Discussion:

Comment: Authors could explain why they ruled out making a study asking people who refuse replied when their call ends.
Reply: Thank you for your suggestion. We agree that it would be very interesting to know why callers who received a text message did not respond to the survey. However, because of ethical reasons, it is not allowed to approach the non-respondents of this study.

Reviewer 2

Methods:

Comment: The authors have chosen to label as satisfied those that were highly satisfied ("to a great extent") and those that were less so (either some or moderately satisfied). The analysis may be stronger if the authors entertain an analysis that only considers "top box" respondents against those that were either limited or not at all satisfied. By doing the analysis in this way may point to places where telephone line support would absolutely be useful or where follow up surveys could be administered to understand the true differences in patient experience with these types of initiatives
Reply: Thank you for asking this very important question. We indeed also seriously considered to only take the distinctly satisfied callers and compare their characteristics with the rest of the respondents. However, since this medical helpline was introduced with the aim to be more patient friendly than the old (more dispersed) system, we decided that it would be of most interest to discern the very dissatisfied (e.g. 'to a limited extent' and 'not at all' answer categories), from those who are at least 'to a moderate extent' content with the service they received. Our aim was to try to identify certain characteristics that make callers very likely to become dissatisfied. Only after this identification and improvement procedures for this subpopulation, we would like to focus on the callers who are only moderately satisfied. Yet, we decided to write this article with the focus on characteristics that could influence satisfaction as opposed to dissatisfaction, because we envision that policy makers' and healthcare professionals should focus on system improvement rather than on making satisfaction outcomes less worse.

Results:

Comment: likely easier to review if the table was simplified to show respondents and non respondents rather than breaking down the respondents by satisfied/dissatisfied. Also not all figures in the text match the table (e.g. call forwarded to physician).
Reply: Thank you for your comment. We agree that table 1 and the accompanying text can be a bit confusing when put next to each other. Therefore, we changed the sentence order in the manuscript, hoping that it would not confuse our readers anymore.
Previous text: "Table 1 shows the characteristics of the respondents, divided into satisfied and dissatisfied respondents, and the non-receivers. On all tested characteristics, the respondents differed from the non-receivers ($p<0.0001$). Respondents were more often female (54.8% vs 53.0%), were younger (median age 28 vs 29), and called more often for a somatic illness (47.7% vs 45.5%) or somatic injury (24.5% vs 19.1%). They were also more often offered a face-to-face consultation (53.3% vs 45.4%) and received less often a telephone consultation (36.4% vs 41.5%). Furthermore, respondents called more often during weekdays (14.9% vs 12.8%), had more often a nurse as the

first call-taker (78.6% vs 74.4%) and their calls were less often forwarded to a physician (10.7% vs 10.8%).”
New text: “Respondents were more often female (54.8% vs 53.0%), were younger (median age 28 vs 29), and called more often for a somatic illness (47.7% vs 45.5%) or somatic injury (24.5% vs 19.1%). They were also more often offered a face-to-face consultation (53.3% vs 45.4%) and received less often a telephone consultation (36.4% vs 41.5%). Furthermore, respondents called more often during weekdays (14.9% vs 12.8%), had more often a nurse as the first call-taker (78.6% vs 74.4%) and their calls were less often forwarded to a physician (10.7% vs 10.8%). Table 1 shows the characteristics of the respondents, divided into satisfied and dissatisfied respondents, and the non-receivers. On all tested characteristics, the respondents differed from the non-receivers ($p<0.0001$).”

Comment: Figure 3 has an apparent dip both in respondents and satisfaction (May 17). Can the authors suggest any reasons?
Reply: Thank you for pointing this out. We noted this dip (actually, a dip in both May’s) as well, but unfortunately we cannot find an any explanation for this.

Discussion:

Comment: Discussion would be strengthened by distilling a few key messages from the analysis to allow the audience to understand the take home messages from the study.
Reply: Thank you for your comment. We agree that the take home messages should be clear in order to make such a paper relevant for other settings. We hope we have phrased the conclusion paragraph of this study specific enough to be clear and relevant.

Reviewer 3

Introduction:

Comment: "Such an OOH telephone line has been established [...]" : i don't understand what is a "comprehensive EMS system" ? I am unable to assess the relevance of the reference (7), apparently written in Danish.
Reply: Thank you for your comment. Upon revision, we agree that this phrase might cause confusion and that it does not attribute to the message of the paper. Therefore, this phrase is now removed from the manuscript.

Comment: Is the ‘frequent use’ of callers aged 0-4 known, or suspected? And is there a reference to support this assumption?
Reply: When we compared the distribution of the population age of the citizens that are covered by this EMS system with the frequency of the calls split by age, we could see that calls for 0-4 year olds are relatively frequently made.

Methods:

Comment: Can the authors precise if callers can be forwarded from 112 to 1813, and what was the proportion of such forwards in their study sample? It could be interesting to add this characteristic to the model, because more worried patients initially calling 112 could be more dissatisfied with the final response mode.
Reply: Thank you for this nice suggestion, we totally agree that this would be a very interesting characteristic and study topic. Unfortunately in our study, patients who were forwarded to 1813 after calling 112 did not receive the satisfaction survey.

Comment: Can the authors detail the process of randomization? I miss details to understand it.
Reply: We are sorry to hear that such an important part of the methods section was not clear. Every day 200 people who called the previous day to the medical helpline were picked out to receive the text message for the survey. This implies that approximately 10% of the callers of the previous day were retrospectively asked to rate their satisfaction regarding the medical helpline.
Previous text: “Every day, a random sample of 200 callers were sent a text message to the phone number they called the medical helpline with.”
New text: “Every day, a random sample of 200 callers of the previous day were sent a text message to the phone number they called the medical helpline with.”

Comment: Can the authors justify the choice to classify callers as "satisfied" when the latters responded being satisfied "to a moderate extent"? It may have been more discriminant to consider only the most satisfied callers (to a large + to a great extent). This question would be partially answered by giving détails on the size of each class.
Reply: Thank you for asking this very important question. We indeed also seriously considered to only take the distinctly satisfied callers and compare their characteristics with the rest of the respondents. However, since this medical helpline was introduced with the aim to be more patient friendly than the old (more dispersed) system, we decided that it would be of most interest to discern the very dissatisfied (e.g. ‘to a limited extent’ and ‘not at all’ answer categories), from those who are at least ‘to a moderate extent’ content with the service they received. Our aim was to try to identify certain characteristics that make callers very likely to become dissatisfied. Only after this identification and improvement procedures for this subpopulation, we would like to focus on the callers who are only moderately satisfied. Yet, we decided to write this article with the focus on characteristics that could influence satisfaction as opposed to dissatisfaction, because we envision that policy makers’ and healthcare professionals should focus on system improvement rather than on making satisfaction outcomes less worse.

Comment: Authors should consider the use of the terms "univariable" and "multivariable" instead of uni/multivariate (single outcome: satisfaction).
Reply: Thank you for your remark. We followed your suggestion and changed the terms in the manuscript.

Comment: Can the authors explain and/or justify the choice for a full fitted model ? Did they try other strategies to select the variables of their model (stepwise, other) ?
Reply: Again thank you for asking this very important question. We also executed a stepwise multivariable analysis (with both backward and forward selection), and these analyses showed that the same variables were put in the model. Since we were not fully able to base hypotheses about potential relevant variables on previously published scientific literature, we decided to present this full fitted model in the paper.

Comment: Why did the authors perfom subgroup analysis on children only ? Can they justify the reason for not considering another subgroup in other age classes (5-100 yo seems wide and non-homogeneous : why not analysing older patients as well ?). Indeed, satisfaction rates vary in classes 60-79 and > 80 yo (table 2).
Reply: Thank you for your question. We agree that it would be very interesting to analyze other age groups in detail as well. We decided to analyze the characteristics of the callers for the 0-4 year old because this subgroup was a relatively large part of the callers to the medical helpline. In addition, we had the feeling that the general public opinion was that parents/guardians of 0-4 year old children who called the medical helpline for their child were in general less satisfied with the service. After our initial raw analysis, we became curious why we did not see this pattern in our results.
In the manuscript, we changed the phrase “A subgroup analysis was performed to analyze the characteristics of the satisfied callers for 0-4 year old children with the variables that were found to be

statistically significant in the multivariable analysis." to highlight that we focused on the 0-4 year old instead of the heterogeneous subgroup of the remaining respondents.
Previous text: "A subgroup analysis was performed to compare the satisfied callers for 0-4 year old children with those being 5-100 years old for the variables that were found to be statistically significant in the multivariate analysis."
New text: "A subgroup analysis was performed to analyze the characteristics of the satisfied callers for 0-4 year old children with the variables that were found to be statistically significant in the multivariable analysis."

Results:

Comment: 1.731.556 calls were "eligible" (not "included" ?)
Reply: Thank you for pointing this out, we have changed it according to your suggestion in the manuscript.
Previous text: "Of the 1,843,094 calls during the study period, 1,731,556 calls were included (Figure 1)."
New text: "Of the 1,843,094 calls during the study period, 1,731,556 calls were eligible (Figure 1)."

Comment: Response rate 23.0 % : i cannot acknowledge this proportion. 23% of which sample ?
Reply: Thank you for your comment. We acknowledge that the term 'included' in the phrase "(…) 1,731,556 calls were included" is not the most appropriate one. Therefore, we changed this word in the manuscript into 'eligible'.
Of those in first instance eligible 1,731,556 callers, 133,055 received a text message in the study period. Of them, 30,582 responded. 30,582/133,055x100% = 22.98%. However, 180 of these respondents did have either an invalid age, made a pocked call, or answered 'don't know' or 'not applicable'. This left 30,402 respondents eligible for the final analysis.

Comment: In the paragraph "Multivariate logistic regression", before referring to table 2 for exhaustive detailed results, authors should consider adding important results in text (somatic injury : OR, IC95), telephone consultation, waiting time, call-taker type.
Reply: Thank you for your comment. Following your advice, we have added those odds-ratios and confidence intervals to the text in the manuscript. Furthermore, we moved the reference to table 2 to a place further down the paragraph.
Previous text: "Table 2 shows the associations between patient characteristics and satisfaction. Calling for a somatic injury was statistically significantly associated with satisfaction. People who received a telephone consultation were less likely to be satisfied. People were also less likely to be satisfied when they called during GP office hours and when they had a waiting time of more than 10 minutes. No statistically significant association was seen between consultation time and satisfaction. In the univariate analysis, the profession of the first call-taker was associated with satisfaction. Adding the variable to the multivariate model did not have an effect. Yet, people who were forwarded to a physician were less likely to be satisfied."
New text: "Calling for a somatic injury was statistically significantly associated with satisfaction (OR: 1.87, 95% CI: 1.64-2.13). People who received a telephone consultation were less likely to be satisfied (OR: 0.46, 95% CI 0.41-0.51). People were also less likely to be satisfied when they had a waiting time of more than 10 minutes (OR: 0.61, 95% CI: 0.53-0.70) and especially a waiting time more than 20 minutes (OR: 0.34, 95% CI: 0.28-0.40). No statistically significant association was seen between consultation time and satisfaction. In the univariable analysis, the profession of the first call-taker was associated with satisfaction. Adding the variable to the multivariable model did not have an effect. Yet, people who were forwarded to a physician were less likely to be satisfied (OR: 0.74, 95% CI: 0.64-0.85) (Table 2)."

Comment: Table 2: authors should consider displaying p-values for adjusted OR.
Reply: Thank you for your comment. To highlight which variables were shown to be statistically significant associated with satisfaction, we added asterisks (*) in table 2.

Comment: Table 2, OR for triage response : the authors should consider inverting the presentation of their results (reference : telephone consultation) so that results are similar to those presented in the abstract (face-to-face consultation : OR 2.18 [1.96-2.43] ; telephone consultation : 1).
Reply: Thank you for looking so closely at our results. We have seriously considered your suggestion, but we hope we can explain why we would like to keep the face-to-face consultation as the reference value in table 2. The medical helpline was initially established to provide citizens who would like to access an emergency department of a hospital outside GP office hours with a time slot. Therefore, the function of giving telephone advice is was initially not the primary aim of this medical helpline, although not less important in practice. We hope that the readers of this paper note that the OR of 2.18 in the abstract corresponds with the 0.46 from table 2 (as 1/0.46=2.18).

Discussion:

Comment: The reasons for dissatisfaction are not discussed herein. It seems that callers request a measurable response, not only an advice (especially callers aged 18-59?). The authors should consider discussing the expectations of the callers, based on references if applicable. I understand that patients request a measurable response for measurable affections (injuries, GP encounter).
Reply: Thank you for this very nice and important suggestion. We certainly agree that this could be an interesting and important topic. Yet, we think that such an analysis is outside the scope of this particular paper. However, we will take it in considering to examine it in another paper.

Comment: the discussion lacks considering qualitative evaluation of the conversation between the caller and the call-taker (why would physicians provide less satisfying reponses ?)
Reply: Thank you for this very interesting suggestion. We agree that this is an important topic indeed. Although we considered this outside the scope of this study, we will certainly take such a study into consideration for the future.

Comment: The process of randomization may have biased the sample selection (cf. data collection).
Reply: Thank you for your comment. Initially, this was our concern as well. However, when we compared the characteristics of the respondents with the characteristics of the non-invited (i.e. callers who did not receive a text message), we only did see very small differences between these two groups. Therefore, we think we can carefully presume that the respondents resemble all callers.

Comment: Responders may have been only the most satisfied/dissatisfied patients, as authors state ("self-selection of people who responded"). Yet, i could not find a discussion on Figure 3, which illustrates that potential bias, through an apparent trending relationship between satisfaction and numbers of respondents (cf. may 17)
Reply: Thank you for pointing this out. We noted this dip (actually, a dip in both May's) as well, but unfortunately we cannot find an any explanation for this.

Reviewer 4

Results:

Comment: I would like the authors to explain why they have put answers with Category 3 in with 4 and 5, instead of with 1 and 2. One can see from Figure 2 that Categories 4 and 5 add up to about 65% to 70% anyway. One can argue that being satisfied "to a moderate extent" sounds like being not fully satisfied.

It is also not clear (but this may be a matter of English versus Danish) what is the difference at the foot of Figure 2 between "to a great extent" and "to a large extent", but the latter is given on page 5 as "to some extent".
Reply: Thank you for asking this very important question. We indeed also seriously considered to only take the distinctly satisfied callers and compare their characteristics with the rest of the respondents. However, since this medical helpline was introduced with the aim to be more patient friendly than the old (more dispersed) system, we decided that it would be of most interest to discern the very dissatisfied (e.g. 'to a limited extent' and 'not at all' answer categories), from those who are at least 'to a moderate extent' content with the service they received. Our aim was to try to identify certain characteristics that make callers very likely to become dissatisfied. Only after this identification and improvement procedures for this subpopulation, we would like to focus on the callers who are only moderately satisfied. Yet, we decided to write this article with the focus on characteristics that could influence satisfaction as opposed to dissatisfaction, because we envision that policy makers' and healthcare professionals should focus on system improvement rather than on making satisfaction outcomes less worse.
Also thank you for pointing out to the discrepancy in the translation of the answer categories between Figure 2 and the text of the manuscript. The answer categories were chosen on the basis of five even steps between the items on the answer scale (i.e. in Danish the difference between 'to a great extent' and 'to a large extent' could be interpreted as a same difference between for example 'to a large extent' and 'to a moderate extent').

Comment: There are similar shortfalls among the Satisfied and Dissatisfied, and these all need to be explained or corrected. The totals of Females plus Males among Satisfied plus Dissatisfied is 29,824 (not 30,402). Adding the given 102,743 Non-respondents gives 132,567. Now assume that these are split among the Sexes in the same way as the Non-receivers. This gives, in whole numbers, 72,677 Females and 59,890 Males. Now deduct the Respondents, 16,650 Females and 13,174 Males. This gives for the Non-respondents an estimated 56,027 Females and 46,715 Males, 54.5% and 45.5%. This means that the third sentence of the paragraph at the top of page 8 should start: "Respondents were more often Female (55.8% vs 54.5%) ... ". These are small differences, and I have not done the same calculations for the other categories, but there might well be bigger differences than shown. In any case the shortfalls need to be explained. If there is a category "Sex not known" then it should be included explicitly; and the same with the other characteristics" "
Reply: Thank you very much for pointing this out and for looking closely at our results. Unfortunately variables such as gender, reason for encounter, triage response and first call-taker could not be obtained from all callers. To make this more clear, we added the category 'missing' for these variables in table 1 of the manuscript.

Discussion:

Comment: It is not clear whether the gender of the caller for a child of 4 and under was included. The caller for this young child may not be the parent and records may not be known about them.
Reply: Thank you for your comment, we are sorry to hear that it was unclear from our manuscript. The demographic characteristics that were obtained, are from the patient about whom the medical helpline is called for. Unfortunately we indeed do not have information available about the actual caller (e.g. a parent/guardian) of the 0-4 year old patients.

Comment: Discussion about patient satisfaction is very limited. Authors could mention the expectations of callers to the helpline in terms of the expected qualification of the staff of the helpline, of whether the caller got better. The description of satisfaction is very limited.
Reply: Thank you for this interesting suggestion. We agree that this could have been an influencing factor and therefore we added this topic to the manuscript.

Previous text: “The multivariate analysis also showed that people whose call was forwarded to a physician were less likely to be satisfied. This might be induced by the reason why the call was forwarded in the first place, which were probably the more complex calls.”
New text: “The multivariable analysis also showed that people whose call was forwarded to a physician were less likely to be satisfied. This might have been induced by the reason why the call was forwarded in the first place, which were probably the more complex calls. Besides, it could have been influenced by a difference in expectation callers had about their call-taker.“

VERSION 2 – REVIEW

REVIEWER	JOSE JOAQUIN MIRA Universidad Miguel Hernandez, Spain
REVIEW RETURNED	10-May-2019

GENERAL COMMENTS	Authors have done a good work. All suggestions and comments have been replied point-by-point. New details have been included and many of them represent limitations of this study. However, this section (limitations) has not been modified. In my opinion, authors may consider this new information and include new limitations according their explanations. For example: authors did not take in account the number of callers receiving SMS and who did not reply to the survey; patients calling 1813 after 112 were excluded; the actual caller of the 0-4 year old patients. This is an interesting study, the topic is relevant and is not usually studied. I think author could make additional review of their manuscript including the responses to reviewers into the text.

REVIEWER	Kira Leeb Victorian Agency for Health Information Melbourne, Australia
REVIEW RETURNED	19-May-2019

GENERAL COMMENTS	The study provides interesting insight into patient satisfaction with health care delivery through the use of a medical helpline. While the results are worthy of reflection, the authors' choice to look at greater than top-box results from the Likert 5-point scale - top 3 responses as opposed to top 2 responses - suggests the results might be unduly weighted positively. This too when the response to the second question of did you have your question answered was identified as "90% responded with at least to a moderate extent". While it is important to know whether patients are satisfied with a service, the more important aspect of patient surveys is to understand how to improve services. The fact that the authors do not reflect more substantially on their low response rate of 23% - a response rate that is markedly below some much longer emailed or mailed surveys - suggests that this method of trying to understand patient satisfaction with this service delivery model is of limited value. This and the lack of speculation about what the results could be used for (is it sufficient to know that X% were "to a moderate extent" satisfied) makes me question the utility of the study and its results. In the introduction the authors state: "Analyzing patient satisfaction scores can provide information

	about whether interventions result in better outcomes from the perspective of the patient, and consequently improve the quality of patient-centered healthcare systems." While the study indicates that this type of survey yields results (sending a 2 question text message that is linkable to patient demographics), there appears to be little applicability to the study objective of determining whether the intervention improves patient-centred care. This might be achieved if the authors reflected more on how the results could be used and if the authors provide at least some breakdown of top-box versus top 3 responses analyses.

REVIEWER	Matthieu Heidet SAMU 94 et Urgences Hôpitaux universitaires Henri Mondor Assistance Publique-Hôpitaux de Paris
REVIEW RETURNED	19-May-2019

GENERAL COMMENTS	The authors provided quick, detailed and argued responses to every question that was asked by the reviewers. Their manuscript is clear and, overall, of importance. A few points still need to be clarified to me : INTRODUCTION 1) Previous Comment: Is the 'frequent use' of callers aged 0-4 known, or suspected? And is there a reference to support this assumption? Previous Reply: When we compared the distribution of the population age of the citizens that are covered by this EMS system with the frequency of the calls split by age, we could see that calls for 0-4 year olds are relatively frequently made. Can the authors state that this assumption relies on internal data ? METHODS 1) l.31 : would the authors please explain the randomization process of the 2000 patients ? E.g., were the latters the 2000 first patients of the day before ? The last 2000 ? Was the randomization rate 1:1, 1:2, else ? What is the rationale for the randomization ? Authors could use the following link to provide more détails : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4938277/ 2) previous Reply : Since we were not fully able to base hypotheses about potential relevant variables on previously published scientific literature, we decided to present this full fitted model in the paper. The authors should state this clearly in their methods section. DISCUSSION 1) Previous Comment: The process of randomization may have biased the sample selection (cf. data collection). Reply: Thank you for your comment. Initially, this was our concern as well. However, when we compared the characteristics of the respondents with the characteristics of the non-invited (i.e. callers who did not receive a text message), we only did see very small differences between these two groups. Therefore, we think we can carefully presume that the respondents resemble all callers.

	Can the authors state this point in the discussion section ?

REVIEWER	DR Patricia Wilkie National Association for Patient Participation(N.A.P.P.) UK
REVIEW RETURNED	23-May-2019

GENERAL COMMENTS	The authors do need to take on board the statistical comments of all the reviewers that they have not yet done. If they are not prepared to do so I would reject The actual numbers in the various degrees of satisfaction seem to appear only approximately in Figure 2. But from that we can see that, if we put Categories 3, 4 and 5 as satisfied, versus 1 and 2 as not, the split is about 90% to 10% (in fact 89.5% v 10.5%), whereas if we split as 4+5 versus 1+2+3 it is about 70% to 30%. Comparing a very large versus a very small group is less likely to show up differences than comparing two groups rather closer in size. An alternative, which makes it much more complicated, is to compare three categories 1+2, 3 and 4+5. To the extent that there are differences in responses between those with different characteristics, as shown in Table 2, these might well be more significant if category 3 is even partially more like 1+2 than it is like 4+5. This should at least be investigated. In Table 1 the authors still compare Respondents versus Non-Receivers, rather than versus Non-Respondents. II think I previously suggested how they could estimate the counts for Non-Respondents by assuming the same proportions among Receivers as among Non-Receivers, and then estimating numbers of Non-Respondents by subtraction. The authors seem to have ignored this suggestion and have commented only on the numerical discrepancies pointed out, because of missing values. I am also now concerned that the authors seem concerned to ensure that their results will be acceptable to policy makers. This is not good research methodology

VERSION 2 – AUTHOR RESPONSE

Reviewer 1:

Comment: Authors have done a good work. All suggestions and comments have been replied pointby-point. New details have been included and many of them represent limitations of this study.However, this section (limitations) has not been modified. In my opinion, authors may consider this new information and include new limitations according their explanations. For example: authors did not take in account the number of callers receiving SMS and who did not reply to the survey; patients calling 1813 after 112 were excluded; the actual caller of the 0-4 year old patients. **Reply:** We were sorry to hear that you still felt that the discussion section of this study was not strong enough. We hope that you agree with the following changes and addition to the discussion:

People who did not reply:

To get a better impression of the representativeness of the respondents for the total population, we were suggested to estimate the characteristics of the non-respondents by assuming that the receivers

had the same proportions as the non-receivers of the questionnaire. Please find in the last column of Table 1 the difference in proportions between the respondents and the estimated nonrespondents. Based on this estimation, the following phrases was added to the discussion section of this manuscript:

Old text: “The low response rate may have induced a selection bias by self-selection of people who responded to the questionnaire, which was also indicated by the differences in characteristics between the respondents and the non-receivers in this study.”
New text: “The low response rate and the fact that the questionnaire could not be sent to analog telephones may have induced a selection bias by self-selection of people who responded to the questionnaire. With estimating the characteristics of the non-respondents based on the nonreceivers, it seemed that respondents were less often older than 80 years, called more often for a somatic injury and received a face-to-face consultation.”

Patients calling 1813 after 112:
We agree that this a very interesting subpopulation to explore. However, our system that distributed the satisfaction questionnaire, could not include these people since it has no access to the telephone numbers of those callers. We hope we can solve this technical limitation for future studies. To make it more clear, we have added the following sentence to the ‘data collection and processing’ paragraph of the methods section:

New text: “Patients who were referred to the medical helpline after calling 112 were excluded for selection, because from them there were no telephone numbers available in the system.”

The actual caller of the 0-4 year old:
We totally agree that it should be kept in mind that the demographic characteristics of the 0-4 year old children (presented in this manuscript) do not equal the characteristics of the people who called on behalf of them. However, we also decided to study this subpopulation because they are not often taken into account in other OOH satisfaction questionnaires. We have added the following sentence to the discussion:

New text: “The way the questionnaire was distributed also limited the study because the respondent might not have been the patient to whom the answers were linked. That means that the caller could have other demographic characteristics than was assumed in this study. This is especially a relevant limitation for the analysis of the callers for the 0-4 year old patients.”

Reviewer 2:
Comment: The study provides interesting insight into patient satisfaction with health care delivery through the use of a medical helpline. While the results are worthy of reflection, the authors' choice to look at greater than top-box results from the Likert 5-point scale - top 3 responses as opposed to top 2 responses - suggests the results might be unduly weighted positively. This too when the response to the second question of did you have your question answered was identified as "90% responded with at least to a moderate extent". While it is important to know whether patients are satisfied with a service, the more important aspect of patient surveys is to understand how to improve services. The fact that the authors do not reflect more substantially on their low response rate of 23% - a response rate that is markedly below some much longer emailed or mailed surveys - suggests that this method of trying to understand patient satisfaction with this service delivery model is of limited value. This and the lack of speculation about what the results could be used for (is it sufficient to know that X% were "to a moderate extent" satisfied) makes me question the utility of the study and its results. In the introduction the authors state: "Analyzing patient satisfaction scores can provide information about whether interventions result in better outcomes from the perspective of the patient, and consequently improve the quality of patient-centered healthcare systems." While the study indicates that this type of survey yields results (sending a 2 question text message that is linkable to patient demographics), there appears to be little applicability to the study objective of determining whether the intervention improves patient-centred care. This might be achieved if the authors reflected more on how the

results could be used and if the authors provide at least some breakdown of top-box versus top 3 responses analyses.

Reply: Thank you for sharing your thoughts about this study in general and the methodological choices that we have made. Since we heard more concerns about our initial decision to also categorize the patients responding 'to a moderate extent' as being satisfied, we decided to run our analysis again and then classify only the 'to a great extent' and 'to a large extent' answer categories as being satisfied. Whereas this difference in classification lead to some minor changes in the results, the conclusion of this study remained the same. For a more detailed analysis, we kindly refer you to the results section of this paper.

We agree that the response rate of this study is a limitation that should be taken into account when interpreting the results. To get a better view on the possibility of a selection bias, we were suggested to make an estimation of the characteristics of the non-respondents by assuming that the receivers of the questionnaire had the same proportions as the non-respondents. We added the following sentences to the discussion section of the manuscript:

New text: "The low response rate and the fact that the questionnaire could not be sent to analog telephones may have induced a selection bias by self-selection of people who responded to the questionnaire. With estimating the characteristics of the non-respondents based on the nonreceivers, it seemed that respondents were less often older than 80 years, called more often for a somatic injury and received a face-to-face consultation."
Whereas we certainly agree with you that satisfaction surveys can provide valuable information about understanding how to improve healthcare services, we would like to explain that this was not the objective of setting up this small survey. The aim of this –relatively easy to arrange – quality measurement method was to monitor satisfaction over time, instead of scrutinizing and evaluating satisfaction. With this paper, we hoped to give an overview of the satisfaction of the callers in general, and could help to further scrutinize it with a longer questionnaire if needed.

Reviewer 3:
Comment: INTRODUCTION

1) Previous Comment: Is the 'frequent use' of callers aged 0-4 known, or suspected? And is there a reference to support this assumption?

Previous Reply: When we compared the distribution of the population age of the citizens that are covered by this EMS system with the frequency of the calls split by age, we could see that calls for 0-

4 year olds are relatively frequently made.

Can the authors state that this assumption relies on internal data ?

Reply: When looking at the proportion of 0-4 year old children in the region around Copenhagen registered by the Danish statistics registry (unpublished data from Statistics Denmark), we saw that around 5.7% of the total population were aged 0-4 (first quartile of 2017). However, when we look at the descriptives of this study, we found 16.4% of non-respondents and 18.5% of respondents who called for a 0-4 year old child. Here, we based our statement on. Is this explanation in your opinion enough, or would you like us to reference this reasoning including the unpublished Statistics Denmark source in the manuscript?

New text: "Thereafter, a subgroup analysis was performed to analyze the characteristics of the satisfied callers for 0-4 year old children, who were relatively frequent callers based on the distribution of the population by age in the Copenhagen region."

Comment: METHODS

1) l.31 : would the authors please explain the randomization process of the 2000 patients ? E.g., were the latters the 2000 first patients of the day before ? The last 2000 ? Was the randomization rate 1:1,

1:2, else ? What is the rationale for the randomization ? Authors could use the following link to provide more détails :

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4938277/

Reply: Thank you for notifying us to this article to clarify our randomization method. We added our way of sampling (simple random sampling method) to our methods section and included the reference.

Old text: “Every day, a random sample of 200 callers of the previous day were sent a text message to the phone number they called the medical helpline with.”
New text: “Every day, 200 callers of the previous day were selected by a simple random sampling method (23) and sent a text message to the phone number they called the medical helpline with.”

Comment: 2) previous Reply : Since we were not fully able to base hypotheses about potential relevant variables on previously published scientific literature, we decided to present this full fitted model in the paper.

The authors should state this clearly in their methods section.

Reply: We now did, we hope you agree with the new sentence.

Old text: “A full fitted model without a selection was created with: gender, age, reason for encounter, triage response, time of the call, waiting time, consultation time, profession of first calltaker and being forwarded to a physician.”
New text: “For the multivariable analysis, a full fitted model without a selection was created, since there was no solid evidence available in previously published scientific literature about potential relevant variables. Variables that were entered to the model were: gender, age, reason for encounter, triage response, time of the call, waiting time, consultation time, profession of first calltaker and being forwarded to a physician.“

Comment: DISCUSSION

1) Previous Comment: The process of randomization may have biased the sample selection (cf. data collection).

Reply: Thank you for your comment. Initially, this was our concern as well. However, when we compared the characteristics of the respondents with the characteristics of the non-invited (i.e. callers who did not receive a text message), we only did see very small differences between these two groups. Therefore, we think we can carefully presume that the respondents resemble all callers.

Can the authors state this point in the discussion section ?

Reply: With the help of the instructions of one of the other reviewers of this paper, we made an estimation of the characteristics of the non-respondents by assuming that the characteristics of the receivers were divided with the same proportions as the non-receivers. When we then compared the characteristics of the respondents with the estimated non-respondents, we found that the respondents were less often older than 80 years (2.4% vs 7.9%), called more often for a somatic injury (24.4% vs 17.5%) and received more often a face-to-face consultation (53.3% vs 43.0%).

We have added these results to the discussion section of the manuscript:

Old text: “However, the study was limited by the low response rate, the way the questionnaire was distributed and the form of the questionnaire. The low response rate may have induced a selection bias by self-selection of people who responded to the questionnaire, which was also indicated by the differences in characteristics between the respondents and the non-receivers in this study. Yet, the relevance of these small differences may be doubted. A study from the Netherlands that interviewed non-respondents of an OOH GP cooperative questionnaire found that most non-respondents gave reasons for not responding that were not directly related to their contact with the GP cooperative (16).”
New text: “However, the study was limited by the low response rate, the way the questionnaire was distributed and the form of the questionnaire. The low response rate and the fact that the questionnaire could not be sent to analog telephones may have induced a selection bias by selfselection of people who responded to the questionnaire. When estimating the characteristics of the non-respondents based on the non-receivers, it seemed that respondents were less often older than 80 years, called more often for a somatic injury and received more often a face-to-face consultation. Yet, the relevance of these estimated differences may be doubted. A study from the Netherlands that interviewed non-respondents of an OOH GP cooperative questionnaire found that most nonrespondents gave reasons for not responding that were not directly related to their contact with the GP cooperative (16).”

Reviewer 4:

Comment: The actual numbers in the various degrees of satisfaction seem to appear only approximately in Figure 2. But from that we can see that, if we put Categories 3, 4 and 5 as satisfied, versus 1 and 2 as not, the split is about 90% to 10% (in fact 89.5% v 10.5%), whereas if we split as 4+5 versus 1+2+3 it is about 70% to 30%. Comparing a very large versus a very small group is less likely to show up differences than comparing two groups rather closer in size. An alternative, which makes it much more complicated, is to compare three categories 1+2, 3 and 4+5.

Reply: Thank you very much for going over those numbers so thoroughly. Following your suggestion, we have made the analysis by classifying the satisfied respondents as being satisfied when they answered “to a great extent” and “to a large extent” and incorporated it in the manuscript. For the results of this analysis, we kindly refer to the methods and results section of our manuscript. Whereas this new classification resulted in some minor changes in the results section, the conclusion of this study did not change.

Comment: To the extent that there are differences in responses between those with different characteristics, as shown in Table 2, these might well be more significant if category 3 is even partially more like 1+2 than it is like 4+5. This should at least be investigated. **Reply:** Please find our answer in the reply above.

Comment: In Table 1 the authors still compare Respondents versus Non-Receivers, rather than versus Non-Respondents. II think I previously suggested how they could estimate the counts for Non-Respondents by assuming the same proportions among Receivers as among Non-Receivers, and then estimating numbers of Non-Respondents by subtraction. The authors seem to have ignored this suggestion and have commented only on the numerical discrepancies pointed out, because of missing values.

Reply: Thank you very much for explaining us a more methodological appropriate method to estimate the counts of the non-respondents. Excuse us for misunderstanding this suggestion you made in the first revision round. Following your addition in the second revision, we now hope we understood you correctly. Please find below the table that we made to estimate the nonrespondents. We have incorporated the last column of this table in Table 1 of our manuscript. We think this method gave some interesting insights, so thank you for this suggestion.

	Non-receivers (n=1701154)	Estimation receivers (n=132875)	Respondents (n=30402)	Estimation nonrespondents (n=102473)	Difference % respondents vs % estimation non-respondents
Sex					
Female	901247 (53.0%)	70395	16650 (54.8%)	53745 (52.3%)	2.5%
Male	742677 (43.7%)	58010	13174 (43.3%)	44836 (43.6%)	-0.3%
Missing	57230 (3.4%)	4470	578 (1.9%)	3892 (3.8%)	-1.9%
Age (years)					
0-4	278601 (16.4%)	21761	5625 (18.5%)	16136 (15.7%)	2.8%
5-17	230482 (13.5%)	18003	5421 (17.8%)	12582 (12.2%)	5.6%
18-39	518393 (30.5%)	40491	8022 (26.4%)	32469 (31.6%)	-5.2%
40-59	294642 (17.3%)	23014	6555 (21.6%)	16459 (16.0%)	5.5%
60-79	208682 (12.3%)	16300	3658 (12.0%)	12642 (12.3)	-0.3%
≥80	113127 (6.7%)	8836	723 (2.4%)	8113 (7.9%)	-5.5%
Reason for encounter					
Somatic illness	773868 (45.5%)	60446	14506 (47.7%)	45940 (44.7%)	3.0%
Somatic injury	324253 (19.1%)	25327	7432 (24.4%)	17895 (17.4%)	7.0%
Psychiatric illness	10842 (0.6%)	847	129 (0.4%)	718 (0.7%)	-0.3%
Other	592191 (34.8%)	46255	8335 (27.4%)	37920 (36.9%)	-9.5%
Triage response					
Face-to-face consultation	772583 (45.4%)	60345	16194 (53.3%)	44151 (43.0%)	10.3%
Telephone consultation	706467 (41.5%)	55181	11077 (36.4%)	44104 (42.9%)	-6.5%
Ambulance	54071 (3.2%)	4223	1160 (3.8%)	3063 (3.0%)	0.8%
Other	168033 (9.9%)	13125	1971 (6.5%)	11154 (10.9%)	-4.4%
Time of the call					
Daytime weekday	216978 (12.8%)	16948	4515 (14.9%)	12433 (12.1%)	2.8%
Daytime OOH	409131 (24.1%)	31957	5075 (16.7%)	26882 (26.2%)	-9.5%
Evening/night OOH	1075045 (63.2%)	83970	20812 (68.5%)	63158 (61.5%)	7.0%
Waiting time (minutes)					
0-3	860874 (50.6%)	67242	15561 (51.2%)	51681 (50.3%)	0.9%

3-6	286752(16.9%)	22398	5234 (17.2%)	17164 (16.7%)	0.5%
6-10	235531 (13.8%)	18397	4244 (14.0%)	14153 (13.8%)	0.2%
11-20	240072 (14.1%)	18752	4138 (13.6%)	14614 (14.2%)	-0.6%
≥20	77914 (4.6%)	6086	1225(4.0%)	4861 (4.7%)	-0.7%
Consultation time (minutes)					
0-3	641846 (37.7%)	50134	11083 (36.5%)	39051 (38.0%)	-1.6%
3-6	740206 (43.5%)	57817	13702 (45.1%)	44115 (42.9%)	2.1%
6-10	264892 (15.6%)	20690	4764 (15.7%)	15926 (15.5)	0.2%
≥10	54210 (3.2%)	4234	853 (2.8%)	3381 (3.3%)	-0.5%
First call-taker					
Nurse	1265043 (74.4%)	98811	23898 (78.6%)	74913 (72.9%)	5.7%
Physician	388509 (22.8%)	30346	5683 (18.7%)	24663 (24.0%)	-5.3%
Priority physician	20527 (1.2%)	1603	192 (0.6%)	1411 (1.4%)	-0.7%
112	12 (0.0%)	1	0 (0.0%)	1 (0.0%)	0.0%
Missing	27063 (1.6%)	2114	629 (2.1%)	1485 (1.4%)	0.6%
Call forwarded to a physician					
Yes	184250 (22.3%)	14392	3237 (22.6%)	11155 (22.2%)	0.4%
No	641846 (77.7%)	50134	11083 (77.4%)	39051 (77.8%)	-0.4%

We incorporated these results in the manuscript as well. We added the last column of the table above as the last column in table 1 of the manuscript. Furthermore, we described the results the following:

New text: “Assuming that the receivers of the questionnaire have the same proportions of characteristics as the non-receivers, the respondents were less often older than 80 years (2.4% vs 7.9%), called more often for a somatic injury (24.4% vs 17.4%) and received more often a face-to-face consultation (53.3% vs 43.0%).”

VERSION 3 – REVIEW

REVIEWER	Kira Leeb Victorian Agency for Health Information
REVIEW RETURNED	25-Jul-2019

GENERAL COMMENTS	I am comfortable with the extent to which the authors have responded to the previous rounds of feedback, save one. I'm unclear why the authors seem unwilling to respond to any comments about response rate. On the one hand, some will say generically that a 23% response rate is reasonable for a patient satisfaction survey. On the other hand, surely they expected a greater response rate given the mode of delivery and only 2 questions to answer. Anecdotally evidence suggests that patients are not all that willing to respond to surveys that are sent via text. It would have been good to see any speculation on whether this might have impacted the study at hand.

REVIEWER	Dr Patricia Wilkie President National Association for Patient :Participation in Primary Care Surrey, England, UK
REVIEW RETURNED	05-Aug-2019

GENERAL COMMENTS	Please ensure that you are very clear e.g.(line 55 , page 4) who the calls referred to. by adding on child's behalf and check throughout the paper. I am not adding any additional major corrections at this stage in the process. You have addressed the major statistical flaws. But I must add that it is extremely disappointing that there were no patient representatives involved in the design of the study, particularly a study with a major focus on patient satisfaction. This is a real weakness. Finally I would have liked to have seen more up to date references on patient satisfaction.There are much better and more recent ones than 2004!

VERSION 3 – AUTHOR RESPONSE

Reviewer 2:

Comment: I am comfortable with the extent to which the authors have responded to the previous rounds of feedback, save one. I'm unclear why the authors seem unwilling to respond to any comments about response rate. On the one hand, some will say generically that a 23% response rate is reasonable for a patient satisfaction survey. On the other hand, surely they expected a greater response rate given the mode of delivery and only 2 questions to answer. Anecdotally evidence suggests that patients are not all that willing to respond to surveys that are sent via text. It would have been good to see any speculation on whether this might have impacted the study at hand.

Reply: Thank you for going over the paper again and considering our suggested changes. We hope that our remarks about the response rate in the discussion section will be sufficient.

"However, the study was limited by the low response rate (…). A study from the Netherlands that interviewed non-respondents of an OOH GP cooperative questionnaire found that most non-respondents gave reasons for not responding that were not directly related to their contact with the GP cooperative (16)."

Reviewer 4:

Comment: Please ensure that you are very clear e.g.(line 55 , page 4) who the calls referred to. by adding on child's behalf and check throughout the paper.

I am not adding any additional major corrections at this stage in the process. You have addressed the major statistical flaws. But I must add that it is extremely disappointing that there were no patient representatives involved in the design of the study, particularly a study with a major focus on patient satisfaction. This is a real weakness.

Finally I would have liked to have seen more up to date references on patient satisfaction.There are much better and more recent ones than 2004!

Reply: Thank you for going over the paper again and considering our suggested changes. We agreed that line 55 on page 4 could have been phrased more clearly and changed it from:

"Furthermore, a subgroup analysis was performed of calls concerning 0-4 year old children, because of their frequent use of the medical helpline."

into the following:

"Furthermore, a subgroup analysis was performed of calls concerning 0-4 year old children, because of the frequent use of the medical helpline for this group."

We critically reviewed the rest of the text in the paper as well to check whether it is clear that it refers to people calling on a child's behalf, and we think there are no uncertainties left on this aspect.

We totally agree with you that it would have been a real asset for the results of this study if we could have more in debt interviews with our respondents. Unfortunately, this was unfeasible for this study. We hope it could be done in the near future.
We were sorry to hear your remark about the references. Unfortunately, we did not find any newer references that are related to the subject of our paper and hope that the current ones are, despite their publication dates, sufficient for the message of the manuscript.