Author’s response to reviews

Title: Evaluating patient experiences in decentralised acute care using the Picker Patient Experience Questionnaire; methodological and clinical findings

Authors:

Ann-Chattrin Leonardsen (dleo@online.no)
Vigdis Grøndahl (vigdis.a.grondahl@hiof.no)
Waleed Ghanima (waleed.ghanima@so-hf.no)
Espen Storeheier (espen.storeheier@iomk.no)
Anders Schönbeck (anders.schonbeck@vgregion.se)
Nina Mikkelsen (nina-carine.mikkelsen@sarpsborg.com)
Guro Letting (gurs@fredrikstad.kommune.no)
Thor Løken (thor.asbjorn.loken@moss.kommune.no)
Réné Holst (rholst@health.sdu.dk)
Lars-Petter Jelsness-Jørgensen (lars.p.jelsness-jorgensen@hiof.no)

Version: 1 Date: 01 Aug 2017

Author’s response to reviews:

BMC Health Services Research

Editor

Torleif Ruud

Re-Submission of manuscript: BHSR-D-16-01527 Evaluating patient experiences in decentralised acute care using the Picker Patient Experience Questionnaire; methodological and clinical findings
Dear Associate Editor and Reviewers

We would like to thank the associate editor and reviewers for their comments on our paper. Below, You will find the issues raised by the associate editor and reviewers, followed by our response to these remarks. Changes have also been highlighted in the revised manuscript (manuscript with track changes). We have done major revisions to increase the reader-friendliness, as well as to create a “red-line” throughout the manuscript. For Your information, tables under 1 page have been included in text (according to author guidelines), while tables over 1 page have been presented on separate pages at the end of the manuscript.

We hope You find our efforts to improve the quality of the submitted paper satisfactory, and that You once more will consider our paper for publication in the BMC Health Services Research.

Associate editor

Comment 1: The reference to statisticjcal method is to an article in 1972 (45 years ago) and should be replaced by newer references to statistical methods used.

Response: We thank the Associate editor for this input, and have consequently added a more recent reference to the original 1972 reference (Nelder and Wedderburn introduced and described this method).

Comment 2: The tables are generally in the form of reporting descriptive data or correlation coefficients for each of the five sites and the whole material, but without statistical tests (except the correlations).

Response: We thank You for this comment. With reference to the reviewers’ comments as well as the associate editors’ comment 4, we have chosen to report results as a whole, not separated to each of the MAWs, as we did not aim to compare results across MAWs.

Comment 3: There are no tables showing analyses of linear mixed models, even if a few results of such are mentioned under results.

Response: This is a valid point. We have now included information about factors included in the linear mixed models in the revised version of the manuscript (see table 5).
Comment 4: The aims of the articles do not include analyses of differences between the sites (MAWs), still all tables use a lot of space to show detailed data on MAWs and individual items. A resubmitted paper after major revisions is expected to use statistical methods relevant to the aims and research questions, and present these analyses in tables replacing some of the present tables.

Response: We completely agree with the associate editor on this point. Consequently, we have chosen to report results as a whole (see comment 2), and not separated down to each MAW, since the aim was not to compare. Ref response to comment 3, we have presented a table for the linear mixed model analysis, which we used to explore whether different factors (CCIS, EQ5D3L, gender, age, and so on) were associated with the patients’ experiences.

Reviewer 1, Suzanne Fredericks

Comment 1: The study is one which is interesting; however what is the rationale for examining patients experiences in MAW? The reason for this study has not been clearly explicated.

Response: We thank the reviewer for this important remark, and have included a clarification of the rationale for the study (as well as information about the qualitative study, ref comment 2 and 3), adding information that reads: “Information about patient experiences in primary healthcare, particularly in decentralised acute healthcare, are either limited or lacking. Only two qualitative papers describing patients’ perspectives on MAWs have been published to date, finding for example that patients view MAW’s as “almost a hospital”, but at the same time different with regard to person centeredness and diagnostic opportunities. Furthermore, the relative influence of socio-demographic variables, length of stay, self-rated health and comorbidity on these experiences remains unknown. Such knowledge could be useful when planning and improving alternatives to hospital treatment, such as MAWs”.

Comment 2: Also, why examine experiences using a quantitative design. Won't you get more information about the phenomenon of interest by doing a qualitative design. What is the impact of this study? What are the benefits of this type of study? This needs to be clearly explained.

Response: The reviewer raises an interesting discussion. The questionnaire study was part of a PhD project that also included interviews with patients, allowing a mixed method approach to exploring patient experiences with MAWs in the thesis. We used qualitative studies to gain in-depth information about patients’ experiences related to quality and safety when staying in MAWs. Results from the qualitative study have been published in The Scandinavian Journal of Primary Health Care, as well as the European Journal for Patient Centered Health Care. These
studies are now referred in the introduction section (see manuscript with track changes). Yet, information from qualitative studies lack the opportunity for statistical generalisation, as well as questions about transferability. The quantitative study adds another dimension of patient experiences, as well as the opportunity to investigate potential predictors to these experiences. No other present studies on patients’ experiences from MAWs have been identified. See also comment 3.

Comment 3: The present study doesn't seem to add anything new or innovative to the existing literature. Collecting data on patient experience using a questionnaire doesn't seem to be the best approach.

Response: We are sorry that the reviewer did not find that the submitted study added anything new or innovative. Of course we agree that the reviewer may have a valid point, even so we would like to argue that assessing patient experience and its potential associations in a new healthcare service is of vital importance. Particularly in order to improve quality of care. In our view, even though patient experiences have been investigated in several settings, these results may not be transferable due to potential differences across e.g., healthcare services and patient populations.

Reviewer 3, Andrew Garratt

Comment 1: Translation and testing of the PPE-15 should be made clear in the abstract.

Response: We thank the reviewer for this important remark, and have added this in the abstract as well as in the introduction: ”The aims of this study were: a) to translate and validate the Picker Patient Experience Questionnaire (PPE-15) in Norwegian, and b) assess patient experiences in decentralised acute care, and potential factors associated with these experiences”.

Comment 2: Background, first sentence. Patient experiences often include evaluations of structure and even outcomes.

Response: We thank the reviewer for making us aware of this. We have included this important element in the sentence, which now reads: ”The patient experience refers to how the patients, their families and other persons who participate in their care feel about the process and structure of care, as well as the outcomes of care “.
Comment 3: The authors include several references to the Norwegian questionnaires developed by the Norwegian Knowledge Centre for the Health Services. These questionnaires have been developed and tested for measurement properties in Norwegian patients. Why did they choose to use a questionnaire that was not developed in Norway, had not been translated and validated with Norwegian patients? Was there not an appropriate questionnaire available that had been developed with input from and testing with Norwegian patients?

Response: We thank the reviewer for this comment. Certainly, there are several Norwegian tested questionnaires. The final choice of questionnaire was a result of several discussions between the researchers and collaborating physicians in the community (through the ADMS/Klinisk utvalg KAD) over a one-year period. The translated, validated instruments that existed were considered too extensive. Short forms consisting of only six to eight questions (such as the NORPEQ) were considered to be too short and not to cover all desired areas. There were also discussions about whether patients were able to answer a questionnaire at all. Finally, all of the participants in the project planning process (as well as the co-authors) agreed on a questionnaire: the Picker Patient Experience Questionnaire (PPE-15). This information is now included in the revised manuscript (see manuscript with track changes).

Comment 4: Data collection. I am not familiar with the Brislin reference from 1970. International standards now exist for the translation of patient-reported questionnaires (Wild, Value in Health 2009; COSMIN checklist; CAHPS have their own guidelines). How many translators were involved each way, did they work independently, how was disagreement dealt with (for example, was there a meeting)?

Response: We thank the reviewer for this comment. We have consequently referred to Wild, 2005 Value in Health, which fit our translation procedure, as well as to the COSMIN checklist, and added this information: ”The PPE-15 has not previously been translated into Norwegian, and forwards and backwards translation were consequently performed according to recommendations in the literature: Two professional bilingual translators with Norwegian as their mother tongue performed two independent translations into Norwegian. After comparing the translations and synthesizing these into one, the questionnaire subsequently underwent a backward translation to English by a translator with English as her mother tongue. Finally, three independent individuals evaluated the questionnaire by comparing the English and Norwegian versions with regard to semantic, idiomatic, experiential, and conceptual equivalence. Following this procedure and prior to statistical testing, the Norwegian PPE-15 underwent testing of face validity. This was done by distributing the questionnaire to 10 patients prior to the study period in order to assess the adequacy, appropriateness and understandability of the questionnaire, including language and scoring instructions. Patient feedback did not reveal any problematic issues in any of these aspects. Following these procedures, a final version of the PPE-15 was approved and tested”. 
Comment 5: EQ-5D. Most Norwegian EQ-5D-3L studies use the UK Measurement and Valuation of Health value set (Dolan, Medical Care, 1997) with EQ-5D scores that range from -0.59 to 1. This value set is based on the time trade-off which is the most widely used and accepted method for the EQ-5D. Reference 24 for the "Europe VAS value set" does not include the development of the value set but rather a clinical study relating to construct validity. King et al (ref 24) also assessed Cronbach’s alpha which is inappropriate for a preference weighted instrument such as the EQ-5D. King et al also state that the scores range from 0 to 1 which is not correct. What about states worse than dead? Justification for the choice of weightings is needed in a single sentence together with an appropriate reference.

Response: Thank You very much for making us aware of this. The reference has of course been replaced, supporting this information: “ The EQ-5D-3L consists of the EQ-5D descriptive system that measures health-related quality of life on five dimensions: mobility, self-care, usual activities, pain/discomfort, anxiety/depression, and the EQ visual analogue scale (EQ VAS). Responses are scored according to three levels: 0 (no problem), 1 (some problems) to 2 (severe problems). The EQ-5D-3L score was used as an overall EQ-5D-3L index score by assigning weights to each level of each dimension according to the Europe VAS value set “ Hence, we did not use the TTO calculation.

Comment 6: Patient ratings of experiences are affected by data collection methods including where the data are collected. Better ratings of experiences have been found when questionnaires are administered where care is received. The patients completed questionnaires at home but where were they returned to - the hospital? What are the potential implications of this? This point might be taken up in the discussion.

Response: The questionnaires were returned to the first authors’ office address. We have added the following information under 'limitations': “Increasing time since discharge seems to result in poorer patient experiences scores, and patients who self-report their experiences at home following discharge may consequently have an increased risk of recall bias. While collecting data before discharge from MAWs could have increased the number of respondents and reduced the risk of recall bias, the responses could also have been influenced by potential interruptions and influence of the personnel “.

Comment 7: Was it the study nurses who telephoned the patients? Please state.

Response: The patients were telephoned by the first author, and this is now stated in the revised manuscript.
Comment 8: How were patients selected for inclusion in the study? The results state that 1235 patients out of 2182 admitted patients received a questionnaire. Were the included patients representative of the 2182 patients? There was considerable variation in the proportions receiving questionnaires across the five MAWS and yet in the methods it is stated that a "standardised inclusion procedure" was used. The different proportions receiving a questionnaire at each of the MAWS more than suggests that this was not the case across the MAWS. This issue may relate to inclusion or recruitment but either way it makes comparisons of the five MAWS highly problematic.

Response: Thank You for this comment. The objective was to include all patients that were admitted to the MAWs and fulfilled the inclusion criteria, in a one-year period. However, due to various circumstances this methodological approach was not always followed. This is of course quite frustrating and also a clear drawback of our study. Despite these drawbacks however, we view that it is of importance to publish our findings. We have emphasised this information under 'limitations': "Sadly, not all patients who were discharged from the five MAWs were invited to participate, as planned. This could indicate a potential selection bias. Based on input from the various MAWs and study follow-up, there did not seem to be any pattern regarding who was invited and who was not. The reasons seemed to be lack of time due to many tasks, and forgetfulness. Another drawback is that we did not collect information on those who did not receive a questionnaire, which of course makes it impossible to assess any differences between those who were invited and those who were not. Moreover, the response rate is relatively low. It is therefore difficult to know if those not responding had worse experiences, which has been reported in the literature. Comparison of non-responders and responders furthermore revealed significant differences in gender and age”. In one of the wards, an administrative employee took the responsibility for inviting patients, and in another ward one of the administrative nurses did the same. In their absence a decrease in the proportion of patients being included was observed. In retrospect, it would have been beneficial to engage a study nurse in each of the participating wards.

Comment 9: 68 patients responded to the retest questionnaire but how many were selected to randomly receive a questionnaire? Were reminders used? What were the data collection procedures for the test-retest questionnaire?

Response: In order to measure test-retest reliability, a subgroup of responders was invited to fill out the PPE-15 a second time, approximately 3 weeks after the first completion. Invitations to participate in retests were sent out consecutively as completed questionnaires were received, until a number of 154 participants was reached. The literature suggests that only patients in a stable condition should be included in the re-test. Hence, a question was added on the retest,
asking whether the patients’ condition was unchanged (stable, had deteriorated or had improved, and the study only included the patients reporting to be stable/unchanged. This information has been included in the paper (see manuscript with track changes), as well as the following: “In retrospect, we could have invited a larger number of patients in the retest, yet the number of patients needed in these test have been subject of debate. For instance some have advocated that a sample size of 50 could be sufficient or a starting point, while others have highlighted the need for larger sample sizes and more robust test-retest data. Indeed, a systematic review found that the median number included in retest analysis was 60”. Reminders were not used in the retest study.

Comment 10: Approximately 3 weeks but what was the SD and range in days. Did the authors include a question to assess whether patients had received further health care in the period between test and retest, which might have influenced their responses to the second questionnaire?

Response: Unfortunately, exact time from first completion of questionnaire to retest was not calculated. Unfortunately, we did not ask whether they had received further health care in this period, just if their condition had been stable or not. Additional information about the retest procedure has been added to the paper (see manuscript with track changes).

Comment 11: Methods, statistical analysis. Please be more precise about the nature of testing for face validity (see COSMIN checklist) which should come before the Statistical methods.

Response: We are grateful for this comment. The nature of testing for face validity has been placed together with information about the translation process (and in-line with the COSMIN checklist), before ’Statistical methods’. We have also done changes in both the methods section and results section (See manuscript with track changes).

Comment 12: Content validity is a qualitative judgement about the degree to which an instrument adequately reflects the construct being measured. This is a widely accepted definition (See COSMIN checklist).

Response: We have added information in line with the COSMIN checklist, together with, information about the translation process, and also included information regarding face validity testing (see manuscript with track changes, as well as response 11).
Comment 13: Did the authors consider assessing structural validity by means of factor analysis or better still, confirmatory factor analysis? The PPE-15 has a unidimensional structure based on earlier work and hence it is important that this is demonstrated in Norwegian patients. After content validity, this is arguably the most important measurement property because it underpins dimensionality and hence scoring procedures and further testing including Cronbach’s alpha. The level of Cronbach’s alpha is related to the number of items in a scale and hence for a lengthy scale such as the PPE-15, a high level of alpha is expected. This highlights the importance of testing for structural validity, the results of which show whether it is appropriate to sum the 15 items.

Response: We are grateful for this remark and of course agree that factor analysis is important to investigate the structure of a questionnaire. However, in the original development of the PPE-15 and later studies using this questionnaire, no form of factor analysis has been performed. The PPE-15 dimensions/domains were developed based on input from patients during testing of face validity, and based on what items patients viewed as semantically associated. In the current study, the same procedure was followed – and information has now been added in the revised manuscript. We have however not been able to perform a confirmatory factor analysis (which could have been methodologically interesting), and have also added further information on this under limitations. In any way we believe that this would not have altered how results are presented – since there are no dimensional scores – merely as dichotomised item scores. We have added the following (under the discussion section, PPE-15 validity and reliability): “The PPE-15 has primarily been tested in hospitalised patients. The dimensional structure is based on input from patients during its development and not on any statistical tests, such as e.g., factor analysis. The latter might be because no dimensional score exist. Consequently, the dimensions reported only reflect what items are perceived by patients to be semantically related. Similar testing have been performed in Sweden. As a consequence of the procedures used in prior studies and scoring instructions, a factor analysis was not used in the current study. However, during the assessment of face validity, patients did not report any specific issues related to neither the content of PPE-15 nor the items’ corresponding dimensions”.

Comment 14: The test-retest reliability of individual questions should also be considered by means of kappa/weighted kappa as appropriate.

Response: We thank the reviewer for this input. Consequently, kappa analysis for analysis of test-retest reliability have been added in the revised manuscript (see manuscript with track changes).

Comment 15: Table 1. EQ-5D/CCI - give the score ranges and meaning of lowest and highest scores as footnotes.
Response: This has been done.

Comment 16: Table 2. A short summary of item content alongside the item number is recommended rather than the footnote. Eg NORPEQ "1 Understanding doctors".
Response: This has been revised as suggested (see manuscript with track changes).

Comment 17: Abstract "scarce".
Response: Revised

Comment 18: Abstract. "investigate" is rather vague. Please replace with "assess" throughout.
Response: Revised as suggested.

Comment 19: Abstract. Replace "negatively influenced" with "negatively associated" which is more appropriate given the limitations of the study design.
Response: Revised as suggested.

Reviewer 4, Øyvind Bjertnes

Comment 1: Consider changing the title: 479 is the response n; is it an evaluation of patient experiences or MAWs?
Response: Thank You for this comment. We have changed the title according to the suggestions of the reviewer. We hope that the new title better reflects the content of the paper, it now reads "Evaluating patient experiences in decentralised acute care using the Picker Patient Experience Questionnaire; methodological and clinical findings".

Comment 2: There's no mention on the MAW literature in the Introduction, neither general literature nor patient satisfaction/experiences. A paragraph on this would be useful, in addition to a little bit more information about this particular health service.
Response: Thank You for making us aware of this important fact. We have added this information about MAWs: "Throughout Norway, MAWs are organised differently, some being located in nursing homes, some in “houses of health, in local medical centres in relation to a casualty or a hospital. Some MAWs have employed their own doctors dedicated to the service on a 24-hour basis, while other places have employed doctors only during daytime. All of the MAWs have daily doctors’ visits on weekdays. Moreover, the MAWs differ in terms of number of beds and services offered”, as well as on patient experiences, that reads: “The patient experience refers to how the patients, their families and other persons who participate in their care feel about the process and structure of care, as well as the outcomes of care. There is a growing recognition that patients’ perspectives are essential in achieving high quality care. Integrating patients’ perspectives into the evaluation of healthcare delivery is important because they indicate ways to improve care, enhance strategic decision making, meet patients’ expectations, effectively manage and monitor healthcare performance, and document benchmarks for healthcare organisations. Due to the lack of a common definition, the measurement of patient experiences remains challenging. Despite its different meanings, patient experience is often used interchangeably with terms such as patient satisfaction, perceptions or preferences. However, patient experiences are usually considered less subjective than patient satisfaction because patients may be satisfied with healthcare even though they have negative experiences and vice versa”. After submitting this paper, two of our qualitative papers have been published, and we have consequently referred to these in the background section (see manuscript with track changes).

Comment 3: Introduction: I wonder if psychometric testing is one of the goals of the article? There’s a range of tests, but no mention of this as one of the aims of the study.

Response: We thank the reviewer for this important remark, and have added this as an aim of the study, which now reads “The aims of this study were: a) to translate and validate the Picker Patient Experience Questionnaire (PPE-15) in Norwegian, and b) to assess patient experiences in decentralised acute care, and potential factors associated with these experiences”. We have added information on testing of the PPE-15 both under statistical analysis and the results (see manuscript with track changes).

Comment 4: Methods: Please describe the reasons for choosing the PPE-15.

Response: This information is now included in the revised manuscript: “The final choice of questionnaire was a result of several discussions between the researchers and collaborating physicians in the community over a one-year period. Even though other translated and validated instruments existed, they were viewed either as too extensive or too short (not covering all areas of interest)” (and see manuscript with track changes).
Comment 5: Methods: please present more information about the translation process and results.

Response: Information about the translation process and testing of the PPE-15 has been added (see manuscript with track changes).

Comment 6: Methods: why did you not test the PPE-15 with cognitive interviews? How can you document that this instrument consists of important aspects for MAW patients? The questionnaire was developed and validated in a hospital setting in another country.

Response: The PPE-15 was developed in collaboration with healthcare personnel and other experts. Questionnaire development included literature reviews, in-depth qualitative interviews with patients and their families, cognitive interviews, pilot tests and evaluation of reliability and validity. In Sweden (comparable healthcare system to Norway) the PPE-15 has been used since 1998, but tested in 2012. Hölund claimed that the strength of the PPE-15 was that the items related to what patients defined to be real problems in care situations, that the language was easy to understand, and that it had few evaluating questions. The MAWs are meant as an alternative to hospitalisation, and has been evaluated as ”almost a hospital”. Since the PPE-15 had been tested with patient interviews in a Nordic setting, we did not consider cognitive interviews in our study. Nevertheless, testing of face validity revealed that patients viewed the questions as relevant for expressing their experiences with staying at a MAW.

Comment 7: Methods: Neutral answers=non-problem - please explain why this is not handled as missing.

Response: This choice was based on the methods used in prior studies using the PPE-15, and after mail-communication with researchers at the Picker Institute, who developed the questionnaire (jfr Steve Sizmur PhD Chief Statistician Picker Institute Europe Buxton Court 3 West Way Oxford OX2 0JB Tel. +44 (0) 1865 208123 Fax. +44 (0) 1865 208101 Email: Steve.Sizmur@PickerEurope.ac.uk Web: www.pickereurope.org ). This information has been included in the manuscript, that reads: “Because no method for calculating missing items exists for the PPE-15, and based on recommendations from statistical expertise at the Picker Institute (personal communication – available from the first author upon request), missing items were not included in the analysis”.

Comment 8: Methods: Please elaborate on the face validity testing with 10 patients

Response: This has been elaborated on, as requested (see manuscript with track changes).
Comment 9: Methods: please include total n and response n for the test-retest, and clarify what variables was tested in the test-retest (only total score?).

Response: This has consequently been included. In the retest, the scores of each of the variables/items of the PPE-15 was tested using the unweighted Kappa score between each of the variables of the PPE-15 (see table 3).

Comment 10: Results: what was the reasons for only giving 1235 patients (of 2182) the questionnaire?

Response: Thank You for this comment. The objective was to include all patients that were admitted to the MAWs and fulfilled the inclusion criteria, in a one-year period. However, due to various circumstances this methodological approach was not always followed. This is of course quite frustrating, and also a clear drawback of our study. Despite these drawbacks however, we view that it is of importance to publish our findings. We have emphasised this information under 'limitations': "Sadly, not all patients who were discharged from the five MAWs were invited to participate, as planned. This could indicate a potential selection bias. Based on input from the various MAWs and study follow-up, there did not seem to be any pattern regarding who was invited and who was not. The reasons seemed to be lack of time due to many tasks, and forgetfulness. Another drawback is that we did not collect information on those who did not receive a questionnaire, which of course makes it impossible to assess any differences between those who were invited and those who were not. Moreover, the response rate is relatively low. It is therefore difficult to know if those not responding had worse experiences, which has been reported in the literature". In one of the wards, an administrative employee took the responsibility for inviting patients, and in another ward one of the administrative nurses did the same. In their absence a decrease in the proportion of patients being included was observed. In retrospect, it would have been beneficial to engage a study nurse in each of the participating wards.

Comment 11: Results: Why is the ICC between wards negligible? The results in table 3 shows quite large differences on single items. Did you test the ICC-wards for the total score?

Response: Some of the variation in the patients’ scores can be attributed to individual experiences as well as to aspects of the different locations (e.g., the staffing situation and the type of services they offer). The care wards and patients (identity) were consequently included as random effects in the mixed method analysis, to account for the inhomogeneity among patients and inhomogeneity among wards. Insignificant variables were removed from the model one at a
time until only significant effects remained. The ICC was calculated to explore the proportion of random variation. The random effects analysis revealed a negligible variation among wards (ICC<0.001), whereas the random variation between patients within wards contributed to 21% of the total random variation. This indicates that the differences in results shown in table 4, could be caused by random effects due to inhomogeneity among patients, not among wards.

Comment 12: Results: Please explain "face validity revealed no problematic issues".
Response: This sentence has been removed, and information about testing of face validity has been placed together with information about the translation process, and in line with the COSMIN checklist (see manuscript with track changes).

Comment 13: Results: Test-retest ICC: total score? Response rate?
Response: Thank You for making us aware of this unclearity. A total of 154 patients responded to the retest, of which 68 of them reported their condition as stable/unchanged. This information is now included in the revised manuscript. We have done an unweighted Kappa score instead of the ICC scoring for test-retest, in-line with comments from reviewer 3, Andrew Garrat (see methods section, as well as results, table 3).

Comment 14: Discussion: Needs more discussion on the validity of the PPE-15 for this service/patient group;
Response: We thank You for this comment, and have included more in the discussion on the validity for the PPE-15 in this healthcare service (see manuscript with track changes).

Comment 15: more discussion of possible causes to low ICC-wards (low n for instance)
Response: We thank the reviewer for this comment, and have added this: “The results from our study show that the random effect attributed to the ward to which patients were admitted had very little influence on patient experiences. This may be due to a relatively small sample size, or that participants were included from one county where the MAW routines were quite homogenous, due to joint collaborative efforts. Interestingly, prior studies have argued that for most quality aspects, including the ward into analyses of patient experiences is important for a number of quality indicators”, as well as under the limitation section (see manuscript with track changes).
Comment 16: more methodological discussions on the reasons for not finding any expected predictors of patient experiences (age, self-perceived health), which is in contrast to almost all literature in this field

Response: We thank the reviewer for this request, and have added the following in the revised manuscript: “It might be speculated if our findings occurred because the PPE-15 uses a dichotomised score”. In addition we have added more in the discussion- as well as under the limitations section regarding e.g. sample size and selection process.

Comment 17: clarify recall bias under limitations;

Response: Clarification has been included, as well as a reference (see manuscript with track changes).

Comment 18: what does the literature say about differences between responders and non-responders in this field?

Response: We have not been able to identify any studies on patient experiences with MAWs as newly established acute care units. It could be discussed, of course, whether non-reponders are too old or to sick to answer, or wheter only those who have either positive or negative experiences answer. This would only be speculations, but we have added this information in the manuscript, under limitations: ”Sadly, not all patients who were discharged from the five MAWs were invited to participate, as planned. This could indicate a potential selection bias. Based on input from the various MAWs and study follow-up, there did not seem to be any pattern regarding who was invited and who was not. The reasons seemed to be lack of time due to many tasks, and forgetfulness. Another drawback is that we did not collect information on those who did not receive a questionnaire, which of course makes it impossible to assess any differences between those who were invited and those who were not. Moreover, the response rate is relatively low. It is therefore difficult to know if those not responding had worse experiences, which has been reported in the literature”.

Comment 19: how many patients would you have liked to include in the test-retest?

Response: Theory on test-retest methodology does not give an accurate answer to this question. This information has been included in the revised manuscript: “In retrospect, we could have invited a larger number of patients in the retest, yet the number of patients needed in these test have been subject of debate. For instance some have advocated that a sample size of 50 could be
sufficient or a starting point, while others have highlighted the need for larger sample sizes and more robust test-retest data. Indeed, a systematic review found that the median number included in retest analysis was 60”.

Comment 20: Tables: the tables occurred two times, which was a little bit confusing.
Response: We are sorry for this and have consequently revised this in our re-submission. According to guidelines for authors, tables under one page shall be included in the text, while larger tables shall be on separate sheets. We have now followed these instructions.

Comment 21: All in all, an important topic, but many questions and comments to address. The aims are related to MAW experiences and predictors of these experiences, but the manuscript is full of psychometric results and comparisons of MAW units, making it hard to find a "red line" throughout the manuscript (in a way it looks like four aims, not two).
Response: Thank You for finding our topic important. The psychometric results are now gathered and referred under the statistical analysis section as well as under the results section (PPE-15, validity and reliability), which we hope make the paper more reader-friendly. The aim “to translate and validate the PPE-15 in Norwegian” has been added. We have also tried to create the ”red line” throughout the manuscript, focusing on patient experiences with MAWs, and factors associated with these experiences.

Comment 22: Personally I would have liked a separate validation paper on the PPE-15 in this setting, then additional analysis in other papers. Has the instrument been validated in a separate paper or is this the validation?
Response: We certainly considered publishing the validation in a separate paper. There may be both pros and cons related to both alternatives, which we realise. In general however, it is quite difficult to publish a separate paper on the methodological evaluation of a translated questionnaire that has been existing for quite some time. The case would have been different if we indeed had been developing a new questionnaire. Moreover, due to remarks from international reviewers on our studies, that they seemed of a national/nordic interest, we decided to include psychometrical testing and clinical data in the same paper.
Concluding remarks

The comments presented in this letter are meant as complementary to the submitted revised manuscript with “Track Changes”, as well as a clean copy. We hope our revisions are deemed sufficient by the associate editor and the reviewers, and that the paper will be accepted for publication in the BMC Health Services Research. If further revisions are requested, we will still be at Your disposal. Thank You for considering our revised manuscript for publication.

Yours Sincerely

Ann-Chatrin Linqvist Leonardsen