PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Similarities and differences in the associations between patient safety culture dimensions and self-reported outcomes in two different culture settings: A national cross-sectional study in Palestinian and Belgian hospitals
AUTHORS	Najjar, Shahenaz; Baillien, Elfi; Vanhaecht, Kris; Hamdan, Motasem; Euwema, Martin; Vleugels, Arthur; Sermeus, Walter; Schrooten, Ward; Hellings, J; Vlayen, A

VERSION 1 – REVIEW

REVIEWER	Zenewton André da Silva Gama Professor in the Department of Collective Health at the Federal University of Rio Grande do Norte, Brazil.
REVIEW RETURNED	30-Jan-2018

GENERAL COMMENTS	General comments: The manuscript is about a relevant subject (patient safety culture) and the research question is original and interesting for improving the quality of care and patient safety in hospitals. Objectives are achievable with the proposed study design, despite limitations. The method was well planned and described but requires additional information to provide a better interpretation and reproducibility of the study, especially with respect to the description of variables and data analysis. The results are consistent with the purpose of the study, but some relevant results in the tables were not highlighted and discussed. Finally, it is important to discuss the validity of the chosen outcome measures as well as some contradictory results. The conclusion should be reviewed because it brings affirmations not supported by the study. Specific concerns: - Abstract and all the manuscript: (1) Self-reported outcomes are used in a generic form, giving the impression that you measured other clinical outcomes or adverse events. It is necessary puts clear that you measured "Patient Safety Culture self-related outcomes". (2) Do not use the term "safety culture" or "safety dimensions", but always "Patient Safety Culture (PSC)", "PSC dimensions" or "PSC self-related outcomes". The definition and instrument that you choose is specific and limited for healthcare and patients. - Background: (1) p.5, lines 46-51: You affirm that "it has been widely translated and validated in several languages and countries, including Belgium, England, Norway, Scotland, the Netherlands and Palestine [17-21]", but I did not see any Belgium validity study in these references. Please, inform the validity study in Belgium hospitals. (2) p.6, lines 2-5: "Frequency of event reporting" and "Overall perceptions of safety" are considered to be two of the 12 original

	HSOPSC dimensions, but you defined dimensions only as "self-related outcome". It is true that these dimensions are related to results of PSC, but you should make it clear that they are PSC dimensions in the original HSOPSC and in all validity studies, as you confirm in your previous study (Najjar et al. 2013 - reference 21) . - Method: (1) p.8, lines 43-48: It is unclear whether HSOPSC negative questions were reversed as a step prior to data analysis. For some items, "strongly agree" is a negative aspect of PSC, but for others it is positive. If they have not reversed, it is important to repeat the analyzes with this correction. If they were correctly reversed, it is important to specify in the text. (2) p.9, lines 2-14: 2.1. The data types of the dependent and control variables were not presented. Some appear to be categorical (eg, staff position), so they would not be appropriate for Pearson's correlation analysis of table 4. Please specify the type of data and range of each of them (eg, patient safety grade is continuous and ranges 0-10?). 2.2. In order to have the best comprehension and comparison with other studies, use the measure "percentage of positive responses" to analyze the two composite outcomes that have multiple items, as recommended in the original version of the questionnaire (Sorra et al. 2012, reference 14). Observe in your table 2 that the average score has little variation, most with average in the 3 points. 2.3. Justify why these variable controls were chosen. For example, working hours are not related to any variable dependent (table 4). It is interesting to use a bivariate analysis as a criterion for selecting the control variables or to select aggregating variables from the smallest to the largest, for example: profession, hospital size and country. (3) p.10, lines 25-28: Please add to the text which two of the 12 original HSOPSC dimensions were not distinguished by factorial analysis and were excluded from the analysis. (4) p.9, lines 51-54: It may not be necessary to use standardized beta values (B). If all independent variables are of the same type, non-standardized Betas are easier to understand because they are presented in the same unit of measure of the dependent variable and not the amount of standard deviations. However, if you prefer to use standardized Beta, explain in the text how to interpret this value. - Results: (1) p.10, lines 23-33: In the text on table 3, explain the significant differences found. (2) p.10, lines 35-55: 2.1. We cannot include categorical variable in the Pearson correlation (staff position). 2.2. Add comment about the force of the correction, because in many cases it is weak. (3) in table 3, specify the statistical test used. - Discussion: (1) p.13, lines 27-32 and lines 41-45. The sentence "The study results demonstrate that at least two ... in each country" is redundant with "every outcome measure ... two HSOPSC dimensions in the Palestine and Bengian samples". (2) You included three control variables in the hierarchical regression analysis, however, although they were significant in some cases, they were not explained or interpreted in the text. (3) "Overall perception of safety" and "Overall grade on patient safety" are two possibility to measure similar things. Why do they have different results in the regression analysis? Likewise, "Frequency of events" and "Number of events reported" that measure the same thing. In addition, you have conflicting results in Palestine, when the Staff improves, increases the "FR" (B = 0.09) and decreases the "NER" (B = -0.20) in the table 5. Why? You need

	to explain this in discussion. - Conclusions: (1) I believe that the authors should not conclude on what dimensions should be improved in each country but on the study's ability to actually identify PSC dimensions related to PSC outcomes. The priorities for improvement are seen in table 2 and the conclusions based only on the factors associated with the result dimensions can give a false impression of improvement priority. I also believe that the authors should be more critical about the validity of these outcome measures, because although they try to measure similar aspects, they may have a contradictory or different association with the dimensions and between different countries. (2) "Improve staffing", among other conclusions, are not supported by results. Note in table 5 that staffing improvement is related to worse results in 4 of the 8 outcome indicators, being significant for NER in Palestine and FER in Belgium.

REVIEWER	Hana Brborović, MD, PhD University of Zagreb, School of Medicine, Andrija Stampar School of Public Health, Croatia
REVIEW RETURNED	28-Feb-2018

GENERAL COMMENTS	Thank you for the opportunity to review this paper. It is very well written, and it addresses the very important topic in healthcare. The authors have invested a lot of knowledge and attention to the matter. This results will contribute to the growing knowledge of patient safety culture. There are just a few issues I would like to encourage the authors to address: Introduction – it would be helpful for the international readers to briefly explain the reasons you decided to compared PSC in Belgium and Palestine and to briefly describe the healthcare system: How big is the healthcare workforce? Are there similarities in the healthcare system between the two countries? Are there any important differences that may influence PSC? Methods 1. please briefly describe how the confidentiality was ensured regarding the returning of the filled out questionnaires? 2. Also, it would be helpful to add the references to the validation studies in the Methods section, wherever you mention the validation 3. For the readers who would be interested in matching the sample, please explain the procedure (is it done manually or using a software etc.) 4. It would be useful for the audience unfamiliar with the hierarchical regression model to explain what is this analysis used for, how you decided to perform it and not, for example, multiple regression models, is there any literature background to support your choice? Discussion Regarding PSC dimension Staffing and the question “We use more agency/temporary staff than is best for patient care“. Do the hospitals use agency/temporary staff? This is in some countries, not the case. How are the healthcare systems addressing the lack of healthcare workers?

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

Reviewer Name: Zenewton André da Silva Gama Institution and Country: Professor in the Department of Collective Health at the Federal University of Rio Grande do Norte, Brazil.			
Section	**Comment**	**Response and modifications**	**Page and line**
All manuscript	Do not use the term "safety culture" or "safety dimensions", but always "Patient Safety Culture (PSC)", "PSC dimensions" or "PSC self-related outcomes". The definition and instrument that you choose is specific and limited for healthcare and patients.	Thank you for pointing at this issue. We have now modified this matter throughout the entire manuscript, as you indicated by the highlighted changes.	Entire manuscript
Abstract	Self-reported outcomes are used in a generic form, giving the impression that you measured other clinical outcomes or adverse events. It is necessary puts clear that you measured "Patient Safety Culture self-related outcomes".	We agree with this comment and have now modified this issue in the abstract and revised manuscript	Abstract and entire manuscript
Background	p.5, lines 46-51: You affirm that "it has been widely translated and validated in several languages and countries, including Belgium, England, Norway, Scotland, the Netherlands and Palestine [17-21]", but I did not see any Belgium validity study in these references. **Please, inform the validity study in Belgium hospitals.**	As it seems, we had indeed forgotten to include this reference. It has now been added to the reference list no. 25 and 26	Page 5 and Reference section. Page 24, reference no. 25 & 26
Background	p.6, lines 2-5: "Frequency of event reporting" and "Overall perceptions of safety" are considered to be two of the 12 original HSOPSC dimensions, but you defined dimensions only as “self-related outcome”. It is true that these dimensions are related to results of PSC, but you **should make it clear that they are PSC dimensions in the original HSOPSC and in all validity studies, as you confirm in your previous study (Najjar et al. 2013 - reference 21) .**	We thank the reviewer for pointing at this, and have revised accordingly.	Page 6, paragraph 1, line 1-4.
Method	p.8, lines 43-48: It is unclear whether **HSOPSC negative**	Thank you for noticing this missing information in the	Page 10, under

	questions were reversed as a step prior to data analysis. For some items, "strongly agree" is a negative aspect of PSC, but for others it is positive. If they have not reversed, it is important to repeat the analyzes with this correction. If they were correctly reversed, it is important to **specify in the text**.	original manuscript, which is pivotal to share with the readers. In fact, we had reversed the negatively worded questions of the instrument. Therefore, we have now added "For the purpose of our analysis, negatively worded items were reversed coded ……"	statistical analysis section, first paragraph, line 1-3.
Method	p.9, lines 2-14: 2.1. The **data types** of the **dependent and control variables were** not presented. Some appear to be categorical (eg, staff position), so they would not be appropriate for Pearson's correlation analysis of table 4. Please specify the type of data and range of each of them (eg, patient safety grade is continuous and ranges 0-10?).	We agree it is important to add this in the manuscript, and adopted accordingly.	Page 9 & 10, Paragraphs under dependent variables and control variables sections, the last two paragraphs and the first three lines in page 10.
Method	In order to have the best comprehension and comparison with other studies, use the **measure "percentage of positive responses** " to analyze the two composite outcomes that have multiple items, as recommended in the original version of the questionnaire (Sorra et al. 2012, reference 14). Observe in your table 2 that the average score has little variation, most with average in the 3 points.	We align with this suggestion and changed the table based on your advice.	Page 29, table 2.
Method	Justify why **these variable controls** were chosen. For example, working hours are not related to any variable dependent (table 4). It is interesting to use a bivariate analysis as a criterion for selecting the control variables or to select aggregating variables from the smallest to the largest, for	The control variables were selected based on previous research and/or their significant correlations with the outcomes. We included that to the text as requested to be clearer.	Page 10, the first three lines.

	example: profession, hospital size and country.		
Method	p.10, lines 25-28: Please add to the text **which two of the 12 original HSOPSC dimensions were not distinguished by factorial analysis and were excluded from the analysis.**	Thank you for mentioning this matter. We have now modified the text to be clearer, in line with your comment.	Page 10, first paragraph, line 6-10.
Method	p.9, lines 51-54: It may not be necessary to use standardized beta values (B). If all independent variables are of the same type, non-standardized Betas are easier to understand because they are presented in the same unit of measure of the dependent variable and not the amount of standard deviations. However, **if you prefer to use standardized Beta, explain in the text how to interpret this value.**	We have now taken care of this matter in accordance with the suggestion.	Page 11, first paragraph, line 7 - 11.
Results	p.10, lines 23-33: In the text on **table 3, explain the significant differences found**.	We have now deleted these significant values from the table. A reason is that the values were significant due to the large sample size (2836 participants). As we used the matching technique, the two samples got very similar characteristics and numbers regarding the ones mentioned on table 3. Thanks for noticing that important point.	Page 30, Table 3.
Results	p.10, lines 35-55: 2.1. We **cannot include categorical variable in the Pearson correlation (staff position).**	Yes, we removed staff position from the table and used Spearman's correlation test as the variables	Page 12. We modified the "correlation between PSC dimensions and outcomes of HSOPSC section." line 2. Also, page 31,

			table 4.
Results	Add comment about **the force of the correction**, because in many cases it is weak.	Could the Reviewer further explain this question, as it is not clear to us what is meant? Thank you very much.	
Results	in **table 3, specify the statistical test used**.	Adhering to an earlier comment about this table (and our response there), we have now omitted this information from the table.	Page 30, Table 3.
Discussion	p.13, lines 27-32 and lines 41-45. The sentence "The study results demonstrate that at least two ... in each country" is **redundant with "every outcome measure ... two HSOPSC dimensions in the Palestine and Bengian samples".**	Thanks for pointing this out. We have now deleted this from the text.	Page 14-15. Last and first paragraph respectively.
Discussion	"Overall perception of safety" and "Overall grade on patient safety" are two possibility to measure similar things. **Why do they have different results in the regression analysis? Likewise, "Frequency of events" and "Number of events reported" that measure the same thing. In addition, you have conflicting results in Palestine, when the Staff improves, increases the "FR" (B = 0.09) and decreases the "NER" (B = -0.20) in the table 5. Why? You need to explain this in discussion.**	It worth to note that "Frequency of events reported" is subscale of the PSC which assess the perceptions of participants on how <u>often</u> the events/ mistakes of all kinds (harmed or not) are being reported in their work unit/ hospital by the staff (themselves) and other workers. In comparison, the "number of events reported" is an outcome measure that measures the <u>number</u> of events that were reported by the staff in the last 12 months. While both are related in the way they explore reports on adverse events, they actually are distinct as they look at two different issues (frequency versus total number). That is, frequency does not necessarily mean that the reports are dealing with a	Based on my response and as they are not similar things. I did not explain it in the discussion. Any suggestion?

		different event; it could be so that various people are reporting about the same event. Consequently, this reflects more or less how 'open' the individuals are in reporting cases. In contrast, the number of events more or less aims to reflect how many cases there were. The same applies to "Overall perception of safety" and "Overall grade on patient safety". These two dimensions are two different outcomes. "Overall perception of safety" measures whether the procedures and systems are perceived as good at preventing errors as related to patient safety problems. "Overall grade on patient safety" measures the participants' estimation of the overall grade on patient safety for their work area/unit. The overall Patient Safety Grade percent positive response is calculated by combining the percentage of respondents that answered "Excellent" and "Very Good", and dividing by the total number of respondents that answered that variable. Please have a look on table 1, page 28.	
Conclusions	I believe that the authors should not conclude on what dimensions should be improved in each country but **on the study's ability to actually identify PSC dimensions related to PSC outcomes.** The priorities for	Thanks for pointing this out. Yes we modified the conclusion.	Page 19, conclusion section.

	improvement are seen in table 2 and the conclusions based only on the factors associated with the result dimensions can give a false impression of improvement priority. **I also believe that the authors should be more critical about the validity of these outcome measures, because although they try to measure similar aspects, they may have a contradictory or different association with the dimensions and between different countries.**		
Conclusions	"Improve staffing", among other conclusions, are not supported by results. Note in table 5 that staffing improvement is related to worse results in 4 of the 8 outcome indicators, being significant for NER in Palestine and FER in Belgium.	Thanks for pointing this out. Yes we modified the conclusion.	Page 19, conclusion section.

Reviewer 2

Reviewer Name: Hana Brborović, MD, PhD

Institution and Country: University of Zagreb, School of Medicine, Andrija Stampar School of Public Health, Croatia

Section	**Comment**	**Response and modifications**	**Page and line**
Methods:	It would be helpful for the international readers to briefly explain the reasons you decided to compare PSC in Belgium and Palestine and to briefly describe the healthcare system: How big is the healthcare workforce? Are there similarities in the healthcare system between the two countries? Are there any important differences that may influence PSC?	Thank you highlighting this point. Please find the modifications on the last paragraph.	Page 6, paragraph two, and line 12-17.
Methods:	Please briefly describe how the confidentiality was ensured regarding the returning of the filled	Indeed, it is important to include this information in the manuscript. We have now added two lines on	Page 8, first paragraph.

	out questionnaires?	that point.	Last three lines of the paragraph.
Methods:	Also, it would be helpful to add the references to the validation studies in the Methods section, wherever you mention the validation	Done with thanks.	Page 7, paragraph one and two.
Methods:	For the readers who would be interested in matching the sample, please explain the procedure (is it done manually or using a software etc.)	We thank the Reviewer for pointing this out and have now added this information in the revised manuscript.	Page 8. The last two lines of paragraph two.
Methods:	It would be useful for the audience unfamiliar with the hierarchical regression model to explain what is this analysis used for, how you decided to perform it and not, for example, multiple regression models, is there any literature background to support your choice?	While Hierarchical Regression Analyses are widely applied in a range of disciplines particularly investigating variances in outcomes (in contrast to, for example, odds ratios), we agree our readers might not always be familiar with this technique. Therefore, we have now briefly explained the benefit of this analysis for our current research question, and added a reference for readers interested in this material.	Page 10, last paragraph highlighted in red.
Discussion:	Regarding PSC dimension Staffing and the question "We use more agency/temporary staff than is best for patient care". Do the hospitals use agency/temporary staff? This is in some countries, not the case. How are the healthcare systems addressing the lack of healthcare workers?	In the studied hospitals, especially Ministry of health (MoH) hospitals in Palestine, temporary contracts are not a common issue/ practice. Staffs are recruited on full-time bases by a central HR unit at the MoH level. However in Belgium they address the lack of healthcare workers by using temporary and part time contracts.	

VERSION 2 – REVIEW

REVIEWER	Hana Brborović, MD, PhD University of Zagreb, School of Medicine, Andrija Stampar School of Public Health, Croatia
REVIEW RETURNED	27-Apr-2018

GENERAL COMMENTS	Thank you for the opportunity to review the revised version of the manuscript. The revised version is clearer and more comprehensive than the previous one. The authors have put effort to answer reviewers' comments. There are just a few minor issues left to address: 1. You have briefly described the healthcare system in the West Bank. I would suggest to add a brief description of the Belgian healthcare system as well. It will be interesting to briefly explain to your worldwide audience just in few general points the similarities and differences in the healthcare system between these two countries. Also, it will be very useful to describe why you decided to compare the two countries. Was it a part of a project or international exchange program etc.? 2. The second part of this sentence is unclear. Link between what and what? "Earlier research has demonstrated a link between organizational culture, outcomes and adverse event rates [6-8], financial performance [9] and patient satisfaction [10]." 3. Please clarify. What do you mean by favorable impact on the two composite outcome dimensions (which ones)? "The 10 different PSC dimensions are expected to have a favorable impact on the two composite outcome dimensions (multiple items) and the two single-item outcomes" 4. You mention few studies but only cite one. "Despite efforts to measure PSC in terms of dimensions and outcomes to improve patient safety, few studies have examined the specific predictive value of the dimensions in terms of HSOPSC outcomes [27]." 5. Abbreviations should be used evenly throughout the whole manuscript, not just in some parts. For example, the abbreviations of dimensions. 6. There seems to be a gap between these two sentences, or is the first one redundant in this part of the discussion? "Respondents were asked to provide an OGPS in their work area/unit and to indicate the number of events they reported over the past 12 months. The overall grade for patient safety (OGPS) was found to be particularly significant in building a constructive learning system based on previous mistakes in Palestine and in improving staffing levels in Belgium."

VERSION 2 – AUTHOR RESPONSE

Comment	Response
Editorial Request:	
Before submitting your revision please check one more time for typos/ grammatical errors e.g. page 10: "A total of 2,836 healthcare staff was participated in the study" should be "A total of 2,836 healthcare staff participated in the study"	Done
Reviewer: 2 **Reviewer Name: Hana Brborović, MD, PhD** **Institution and Country: University of Zagreb, School of Medicine, Andrija Stampar School of Public Health, Croatia**	
1. You have briefly described the healthcare system in the West Bank. I would suggest to add a brief description of the Belgian healthcare system	Thank you for highlighting this point. Please find our modifications in page 6, last paragraph with

as well. It will be interesting to briefly explain to your worldwide audience just in few general points the similarities and differences in the healthcare system between these two countries. Also, it will be very useful to describe why you decided to compare the two countries. Was it a part of a project or international exchange program etc.?	track changes.
2. The second part of this sentence is unclear. Link between what and what? “Earlier research has demonstrated a link between organizational culture, outcomes and adverse event rates [6-8], financial performance [9] and patient satisfaction [10].”	Totally agree with you. We added and that made our sentence clearer. Page 5, first paragraph.
3. Please clarify. What do you mean by favorable impact on the two composite outcome dimensions (which ones)? “The 10 different PSC dimensions are expected to have a favorable impact on the two composite outcome dimensions (multiple items) and the two single-item outcomes”	Done. Please find our modification in page 6, paragraph 1.
4. You mention few studies but only cite one. “Despite efforts to measure PSC in terms of dimensions and outcomes to improve patient safety, few studies have examined the specific predictive value of the dimensions in terms of HSOPSC outcomes [27].”	Indeed. We added the missed references (17 & 20). Page 6, paragraph 2.
5. Abbreviations should be used evenly throughout the whole manuscript, not just in some parts. For example, the abbreviations of dimensions.	Done. But we kept some full words to be clearer. Abbreviations were used mainly in the tables.
6. There seems to be a gap between these two sentences, or is the first one redundant in this part of the discussion? “Respondents were asked to provide an OGPS in their work area/unit and to indicate the number of events they reported over the past 12 months. The overall grade for patient safety (OGPS) was found to be particularly significant in building a constructive learning system based on previous mistakes in Palestine and in improving staffing levels in Belgium.”	Yes, the first one was redundant. We deleted.