Author's response to reviews

Title: The Italian version of the Physical Therapy Patient Satisfaction Questionnaire (PTPSQ-I(15). Psychometric properties in a sample of inpatients

Authors:

- Carla Vanti (cavanti@tin.it)
- Paolo Pillastrini (paolo.pillastrini@unibo.it)
- Marco Monticone (marco.monticone@fsm.it)
- Daniele Ceron (daniele.ceron@gmail.com)
- Francesca Bonetti (fra.bonetti@alice.it)
- Raffaella Piccarreta (piccarreta@unibocconi.it)
- Andrew Guccione (andrewguccione@cox.net)
- Francesco S Violante (francesco.violante@unibo.it)

Version: 2 Date: 5 February 2014

Author's response to reviews: see over
Dear Editor,
	hank you for your comments regarding our paper.

We are pleased to submit our point-by-point response to the changes requested on our paper, “The Italian version of the Physical Therapy Patient Satisfaction Questionnaire (PTPSQ-I(15). Psychometric properties in a sample of inpatients”.

We hope that we have made all the changes requested prior to publication and our manuscript is now acceptable. In any case, we waiting for your critique and/or further suggestions. Changes have been made in bold in the text to facilitate your review.

We look forward to hearing your response and thank you for your consideration.

Sincerely yours,

Paolo Pillastrini

EDITOR

1. Please include context information in the background section of the abstract.

We included the requested information in the background section.

REVIEWER: PETER MICHAEL KENT

2a. All sections: You state that the aim this study was to explore the psychometric properties of PTPSQ-I(15) but you also describe the relationships between satisfaction and patient characteristics. This additional aspect would be better signaled if you add this second aim (taken from your outpatient study): (2) to investigate the relationships between the characteristics of the patients and physical therapists and the indicators of satisfaction.

We added this second aim, both in the Abstract and in the Background section of the text.

2b. As an extension of this, the discussion section would be strengthened if you compared and contrasted these findings with those of the outpatient study. For example, the finding in the current study that satisfaction was higher when treated by a male physiotherapist, whereas in the outpatient study, satisfaction was higher when treated by a female physiotherapist – what might those results mean?

We expanded the Discussion section comparing and contrasting these results with our previous ones on outpatients. Unfortunately, regarding some differences (e.g. the impact of the physical therapist gender on the level of satisfaction), we have only been able to propose hypotheses in the absence of Italian literature on this topic.

3. Methods section: More detail is needed so that readers of this paper need not refer to the other paper for basic details, including: who administered the questionnaires and whether patients self completed the answers. If, as in the your previous study, patients were asked the questions, is it your intention to recommend that this is the usual mode of administration - with all the attendant issues of social desirability affecting the item responses? What was the mechanism by which participants could have reported problems of item comprehension?

Following this advice, we added the following sentences in the Methods section: “Questionnaires were presented by research assistants to each participant, who was assured that his or her physical therapist was blinded to the results. Items were presented to each participant in written form. Participants answered each question verbally, and
research assistants filled in the answers. Research assistants could repeat questions but could not change wording. If a participant altered his or her response, the assistant noted the change on the form; if the participant did not choose any answer, the assistant did not mark any box. The levels of patient understanding and the time needed to answer were recorded for each item by the research assistant.”

4. Methods section: What determined your decisions about whether the sample size was adequate?

We agree with this suggestion and decided to specify in the Methods section how the final sample size was determined as follows: The final sample size was based on the “rule of 10” patients per item, giving a final expected sample of 150 subjects (Terwee et al).

We added this reference: Terwee CB, Bot S, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. JCE 2007;60:34-42.

5. Methods section: State the study period in which participants were recruited.

We added the study period.

6. Methods section: More explanation is required for the use of Pearson correlation coefficients (a parametric statistic), when you also state that you have used nonparametric statistics in other parts of the analysis because “the item response numerical and group means were not normally distributed”.

From a descriptive point of view, the Pearson correlation coefficient can be defined and calculated even if the distribution is not normal. As the reviewer points out, its significance, instead, can be evaluated only under the assumption of normal distribution. For our data, the sample size is big enough, and asymptotic results hold. Nonetheless, for the sake of completeness, we now also report the Spearman correlation coefficient and its p-value, even if in all cases the results are aligned with those provided by the Pearson correlation coefficient.

7. Methods section: In your previous paper you indicate that, as there are no Italian language instruments available that would have allowed an examination of convergent validity, concurrent validity was examined using the visual analog scale (VAS) and a 5-point scale of global perceived effect (GPE) as criterion standards. You also indicate that in this context, these criterion standards are best conceptualized as indicators of divergent validity as they measure different constructs than patient satisfaction. However, it seems that as you are focused on divergent validity, then the low and non-significant correlation found with the VAS is an indication of strong divergent validity. You report “Divergent validity was moderate for the GPE and not significant for the VAS”, but I believe this would be more accurately stated as “Divergent validity was moderate for the GPE and strong for the VAS”.

We thank the reviewer for this suggestion and we agree with his proposal. As a consequence, we modified the statement as suggested.

8. Methods section: State how the participants in the retest-subsample were selected – what was the randomization procedure and how was it administered?

We clarified the randomization procedure.
9. Results section: You state that “No questions were left unanswered and yet Table 2 shows that there were missing responses of varying frequency to nine of the 15 questions. Is it that you intended to indicate that no individual item was completely unanswered?

We thank the reviewer for this suggestion and we modified the text accordingly.

10. Results section: I believe that further justification is required as to why in the Factor Analysis, you favoured a one-factor model over the two factor model. The two factor model showed moderate to high loadings on factor one for all items except questions 11, 14 and 15, while questions 11 and 14 loaded highly on factor two. It would be ideal if this justification was consistent with the way in which you reached conclusions about favoring a multi-factor model in your published study of outpatients.

Thank you for this comment. Our decision concerning the number of factors was based upon the following considerations.
1) Moving from the model with one factor F1, to a model with two-factors F21 and F22, the loadings of the first three variables in Table 5 (Q09, Q19, and Q12) with F21 are higher compared to those with F1. Nonetheless, for remaining variables (Q23, Q08, Q21, Q10, Q07, Q20, Q22, Q13, Q26) their relation with F21 was weaker to that with F1, and we observed that for some variables, also the loadings with F22 are moderate (from 0.3 to 0.5 or some variables). In other words, moving from one to two factors lead to a relatively weaker first factor, and to a second factor which surely is related to Q14 and Q11 (only partially unexplained by F1), but also, to a lesser extent, to those which were previously clearly connected to F1. This can make the interpretation of the factors weak.
2) Analyzing carefully the correlations between the variables left partially unexplained by the single factor F1 (Q14, Q11 and Q15) and all the others, we noted that they actually show a moderate relation with many of the others, which is perfectly described by their moderate loading with F1 itself. Actually, this also explains why moving to a two-factor model gives a second factor partially explaining some of the variables explained by the first. Therefore, we concluded that the one-factor model coherently and more neatly describes the interrelations among the involved variables. We have added additional text in the revised version of the paper to explain our reasoning.

11. Results section: In my view, Table 2 would be more interpretable for the reader if:
(i) The missingness for each question was expressed as a proportion of the 149 patients that completed a questionnaire, and
(ii) The responses (1 to 5) were expressed as a proportion of the patient who did answer that question.

Thank you for the advice. We modified the table as suggested.

12. Results section: Please also add a caption to Table 5 explaining what the bold and italised font indicates, such as ‘The numbers in bold indicate high correlations (.30>r<.60) and in italics indicate medium correlations (r>.60)’. Similar captions are required for Tables 3 and 4.

We added the requested explanations to the captions.

13. Discussion section: Limitations component. The approach that you have taken in determining psychometric properties in these two studies is based on Classical Test
Theory. For completeness, it would be useful to also mention that there are more modern approaches based on Item Response Theory, such as Rasch Analysis, that are now available to evaluate questionnaire psychometrics with greater precision than the classical approach.

We mentioned the approaches based on Item Response Theory in Limitations section of the discussion.

14. Results section: Test-retest stability component. This sentence “We also analyzed the consistency between the scores assigned to each item in the two administrations” Might be more clear if reworded to be “We also analyzed the consistency between the scores assigned to each item in the two administrations OF THE QUESTIONNAIRE”.

We made this correction in the text.

15. Discussion section: Please consider these grammatical suggestions: “This may be a reflection of cultural attitudes specific to Italians COMMA as Issa et al. found that co-payERs did not affect patient satisfaction with post-operative physical therapy following total hip arthroplasty”.

Following the suggestion n#41 of this point-by-point file, we modified this sentence to better explain that our results are very different from those by Issa et al.

16. Acknowledgments Section: “The Authors wish to thank...” in UK English, authors does not need to be capitalized.

We made this correction in the Acknowledgments Section.

REVIEWER: SUSAN ROUSH

17. More information needs to be provided on how the surveys were administrated. When, and by whom, were subjects approached? At what point during their admission were subjects surveyed, e.g. upon admission or in their first few days as an impatient, in the middle of their inpatient stay, or at discharge? If the subject data were consistently collected at a particular point in subjects’ inpatient stay, it should be noted. If this was not consistent, it needs to be described as a limitation.

We added this information on Methods section.

18. Why were items that didn’t relate to inpatients included in the survey: e.g. #13 (‘It was easy to schedule visits after my first appointment’) or Q10 (‘The clinic scheduled appointments at convenient times.’)? It is difficult to determine what effect these items had on the statistical analysis, but since there were only 15 items total, the inclusion of 2 or 3 accounted for a non-negligible percentage of the survey. More explanation of the basis for the inclusion of these items is needed.

We agree with this concern and provided more explanations on these two items in the Methods section (Acceptability paragraph), as follows: “On the first day of in-hospital admission, the staff promptly informed patients of the possibility of scheduling follow-up visits at convenient times in order to avoid misunderstandings on Q10 and Q13 and to help them in giving the appropriate answers.”
19. Concerning the non-parametric tests of significance: no alpha level was defined, but I assume alpha = .05. Given that, too many tests were conducted without apparent modification of the alpha level to accommodate the accumulating Type I error rate. Also, as noted on Page 10, multiple comparisons were conducted “between all the pairs of means,” again raising questions about accumulating Type I error.

It has to be specified that a one-way ANOVA was considered for each variable. This has been better clarified in the introduction of the subsection describing the procedure (Dependency of satisfaction on external variables). The majority of variables have 2 levels, and therefore rejecting the null ANOVA hypothesis that the means of the ranks are equal implies that the two considered levels present means of the ranks significantly different. As for the variables with more than 2 levels, we only have 3 variables for which the null hypothesis that the means of the ranks are all equal is rejected. For those cases, we now consider non-parametric post-hoc tests (as suggested in Siegel and Castellan (1988) Nonparametric statistics for the behavioral sciences; MacGraw Hill Int., NewYork. pp 213-214), which were not mentioned in the previous version of the manuscript since they cannot be applied using standard statistical software (such as SPSS). The tests were applied using the statistical software R (“pgirmess” package). This has been explicitly mentioned in the paper. Table 7 has been modified so as to make as clear as possible which pairs of means (of ranks) are significantly different one from another and the alpha level is now explicitly declared.

20a. Since subjects saw a “variety of health care professionals, various technical and assistive personnel” (pg. 13), how do the authors know the subjects responded to the questionnaire specifically as it related to their physical therapy experience? More detail in this regard would be very helpful.

In our opinion, the Introduction section of the PTPSQ (“You recently received physical therapy services at our facility. Because we strive to deliver the best possible physical therapy services, we are interested in learning from patients how we might improve or enhance our services …”) is useful to clarify that a subject is requested to answer specifically relating to his/her physical therapy experience. Following the suggestion of the reviewer, we added this detail in the reviewed text.

20b As the authors note, inpatient satisfaction with physical therapy has never/rarely been reported. One of the reasons for this (at least in the US) is that the acuity level of inpatients typically is so high it is difficult to have confidence in gathered data. Of particular concern is whether patients are able to accurately identify who among the professionals they see are physical therapists, and what treatment they receive is physical therapy. (The additional detail requested in Item 1, above, may address this concern, or more detail on the typical inpatient experience in Italian hospitals may also clarify this issue.)

We thank the reviewer for this suggestion. In Italian hospital setting, an inpatient can easily identify which type of professional he/she sees, due to the different uniforms and the different identification badges, showing the name, the picture and the qualification of each professional. Moreover, in Italy physical therapy is delivered only by physical therapists, because other professionals as the “physical therapists assistants” do not exist.

We hope these clarifications we added to the text are sufficient to explain typical characteristics of Italian hospital setting and experience.
21. Diagnosis information on the subjects is missing. This deficiency becomes important given the last sentence in the Discussion: “...sustained across different health conditions (e.g., neurologic, musculoskeletal, etc.).” If the current subjects didn't have neurological or musculoskeletal diagnoses, what diagnoses were represented among the subjects? If this information is not available, this is a limitation.

We thank the reviewer for this suggestion. We added to the Results section that 107 subjects had neurological diagnoses and 41 had musculoskeletal ones. We modified also the Discussion sentence, to avoid any misunderstanding.

22. Considering the analysis related to Divergent Validity: In addition to reporting the significance of the correlation coefficients, please add the coefficient of determination, indicating the amount of shared variance.

In the case when pair-wise correlations are considered (association between two variables), the determination coefficient coincides with the squared correlation coefficient. For this reason, we think that inserting also the determination coefficient might make the discussion a bit heavier without providing any relevant information and, also, might induce in the reader the wrong idea that multiple regression is used.

23. I do not understand this statement from Page 10: “Instead of being based upon the means of the response, this test is based upon the means of the ranked values, the rank being the position of a value in the ordered sequence of values.” Please clarify.

In the new version of the paper the notion of ranks and the main intuition underlying the Kruskal-Wallis and the Wilcoxon tests are hopefully better explained.

REVIEWER: SYLVIA VON MACKENSEN

24. Background: "Divergent validity was investigated..." In my opinion the correlation with the Global Perceived Effect Questionnaire is convergent validity and not divergent validity. To assess construct validity, you should first establish convergent validity, before testing for divergent validity.

We thank the reviewer for this suggestion, but also as we noted in our previous paper on the same questionnaire, VAS and GPE, in the context of satisfaction, are best conceptualized as indicators of divergent validity, because they measure different constructs than patient satisfaction. Using the VAS a patient rates the pain amount; using the GPE he/she judges the effectiveness of the treatment in relation to his/her needs (“How would you rate the overall effectiveness of the treatment received in relation to your needs?”). As a consequence, in our opinion GPE could also be considered as an indicator of divergent validity. Moreover, following the suggestion of another reviewer (see question n#7 in this point-by-point file), we modified our text, but we maintained the same line of reasoning.

25. Psychometric Characteristics, Acceptability: "No questions were left unanswered...", but in table 2 there are 9 items with missing values, please explain.

Following the suggestion of another reviewer (see question n#9 in this point-by-point file), we modified this sentence to explain this apparent contradiction. The new version is: “no individual item was completely unanswered.”
26. Background: "The original version of the questionnaire consists of 26 items, of which 20 explore ..." This sentence is a bit confusing, when deleting 5 items from 26 a version of 21 items should result. Since in the original paper already only 20 items were taken, it would be better to reformulate.

We added “American” to better explain this concept.

27. Background: "Divergent validity was investigated..." the sentence in order to avoid confusion. In the original paper the following is stated: ‘From the instruments in the Compendium and other sources, 20 items were generated for the instrument that we used in our study.’

We added “of the Italian version” to better explain this concept.

28. Background: "...and indicated high internal validity..." instead of validity it should read consistency.

Corrected in the text.

29. Background: "...and to adapt the original..." instead of and it should read in order to.

Corrected in the text.

30. Background: Reference 3 is not fitting in my opinion, in the text it is reported that the Italian version of the PTPSQ-I was administered to 315 Italian outpatients, while reference 3 reports about the Oswestry index for low back pain in Arab population.

Reference #3 was chosen because it refers to a similar modification of another questionnaire (the Oswestry Disability Questionnaire) which requested to delete some items, due to the different cultural context (Arab population) and the high number of unanswered items. This situation was very similar to ours.

31. Background: please provide some information on the answer categories of the PTPSQ-I(15)

We added the suggested information.

32. Background: "... 5 items were excluded from the PTPSQ-I..." please explain why these items were excluded.

Reply
We explained the reasons for these exclusions.

33. Background: “Both the US version and the Italian version...”

Corrected in the text.

34. Methods: "The administration of questionnaires followed the same methods employed in our previous study on PTPSQ.” The procedure should be explained here shortly as well,
since the reader has not always the possibility to access the author article, but would need this information in order to understand the study procedure.

We added this information in the Methods section.

35. Methods: use the direct article 'the' for the questionnaires, since there is only this one existing "...the PTPSQ-I(15), ... the 5-point Likert-type scale evaluating the Global Perceived Effect..."

Corrected in the text.


37. Statistical analysis: ICC (3,1) should be explained to the reader. Please provide reference such as Shrout and Fleiss, Intraclass correlations: uses in assessing rater reliability, Psychological Bulletin 1979 86:420-428 or equivalent reference

We added this reference to explain ICC (3,1).

38. Results, Subjects: "...our sample consists of 148 inpatients..."

Corrected in the text.

39. Psychometric Characteristics, Reliability: You should refer in the text that Q15 (r=0.261) has a item-total correlation below the critical value of r=.0.30, by contrat Q11 has a r=0.526 which is quite good.

Corrected in the text.

40. Discussion: "Similar reliability coefficients...were obtained in an Italian validation study on an instrument for patient evaluation of general practice care." I do not really understand what this comparison means, since you refer to another instrument where different or similar values have no meaning for the reliability values of the PTPSQ-I(15).

We deleted this reference.

41. "This may be a reflection of cultural attitudes specific to Italians as Issa et al found..." this reference confirms the opposite finding compared to your paper. Issa et al. found that co-pay amount did not significantly affect patient satisfaction, while in your study you found a sig. relationship between payment and patient satisfaction with patients on co-payment
showed the lowest satisfaction in the PTPSQ-I(15). Please explain your statement in the discussion or reformulate it that it is clear that you found the contrary results.

Following this suggestion, we modified this sentence to better explain that our results are very different from those by Issa et al.

42. In table 1 and 7 I would say 'working' instead of worker, which can be easily misunderstood as blue or white collar workers, by contrast you want to distinguish here patients who are actually working or not as far as I understand.

Corrected in the tables.