Reviewer's report

Title: Outcomes for patients with advanced medical illness: a multicentre study of validity and predictors of response to change

Version: 3 Date: 26 November 2015

Reviewer: Cinzia Brunelli

Reviewer's report:

This paper reports on the cultural adaptation and psychometric evaluation of the Italian version of the Palliative care Outcome Scale (POS), patient and staff version, in a cancer patient palliative care setting. A systematic review has shown that the tool is widely used internationally but lacks formal validation in a number of languages, among which Italian. This paper reports useful significant evidence to overcome barriers to widespread use of common outcome assessment tools in PC.

I have a number of observations which may improve the presentation of this report.

Major compulsory revisions

1. I suggest to clarify the focus of the paper already from the title. Key words present in the title should be: POS, Italian, cancer, palliative care. The same is true for the abstract.

2. The background section well explains the need of “appropriate and validated translation in different cultures and languages of existing tools”, but lacks to provide a rational for the second part of the aim (also mentioned in the title) i.e: to examine predictors of changes after transition to PC.

3. Study aim is not clearly stated (may be also due to some typos in the punctuation). As for the title I suggest to mention POS, Italian, cancer, palliative care. Regarding the second part of the aim - “....... to determine if the POS could detect changes in the transition to palliative care”- actually this is the same as evaluating responsiveness, already mentioned as one psychometric property. I suggest to drop this and mention only the aim to examine predictors of changes on POS scale scores, after transition to PC.

4. Page 8 line 243: I suggest not to use “reliability” as a synonym of “internal consistency”. Although this terminology is common in the literature, it can be quite misleading; actually internal consistency may be also defined as “intra-scale reliability” which, however, is different from “test-retest reliability” -often abbreviated as “reliability”. Consequently the words “was reliable”, page 11 line 360 in the discussion and page 2 line 60, in the abstract, should be substituted with “proved acceptable internal consistency” or some similar expression.

5. The test-retest results on the patient version (additional file 5, appendix) are
quite low (range of Kappa 0.04 to 0.55) mainly for some items. I think this deserves more attention. I suggest to move the table in appendix to the main paper, maybe integrating it also with data in additional file 4; also I suggest to add to the table the “proportion agreement within 1 score”, which can help understanding the magnitude of the problem. ICC on the total POS score should also be added. Consider to move also staff reported reliability results.

6. Reliability data also deserve extensive comment in the discussion section comparing it to those reported in the original validation paper (reference 11), and maybe in other from other language validation. For example it is likely that the low Kappa for item “personal affairs” is due to the skewness of the item distribution (high number of 0 values) as the % agreement is acceptable (72.7%) ; different is the interpretation of the same value of Kappa (0.12) for the item “other symptom” : here the % agreement is also very low and this is probably due to the content of the item which makes the test-retest design probably inadequate and this could be confirmed by the higher patient-staff agreement on this item (% agreement is 47% , Kappa 0.44). Other very low Kappa values should be discussed as well .

7. The conclusions drawn in the first two paragraphs in page 12 (lines 375-385) are quite difficult to understand . The higher improvement in home care setting is not necessarily due to longer survival; (i.e couldn’t it be that lower improvements in hospice setting are due to the fact that some of patient were already in home care before, so that their condition was better at baseline which in turn left less room for improvement?); if the hypothesis is that improvement after transition to PC is associated to time to death, then the association to “time to death” and not to “setting” should be examined. Ultimately, it is not clear to me how data collected in PC setting (home and hospice) can support “earlier provision of palliative care, in other settings, such as in hospitals and home care (?) , where patients spend most of their time”. (same reasoning for last sentence of the conclusion page 13 line 420 and for the association with educational level/socio-economical status, lines 381-385) I suggest to clarify or to drop these statements.

8. Among the limitations of the study I suggest to add that the two tools used to assess construct validity of POS ( namely FACIT-SP and QLQC15-PAL) were adequately translated but not formally validated into Italian language.

Minor essential revisions

9. In the methods I suggest to describe a bit more clearly which of different POS measures (patient reported, staff reported) and of different versions available (V1 and V2) have been used in this study. A box could be used to schematize it and different acronyms, to be used also throughout the paper, could help the reader (i.e. POS_P_V1 for “patient reported” vs POS_S_V2 for staff reported version 2) . As different versions are used in the different phases of the study, bit is also important to clarify which are the main differences between tools (POS_P vs POS_S) and between versions (briefly in the box).

10. Pag 5 line 147-149 the sentence mentioning the “Patient “ care Outcome Scale may confuse the reader further : it is one more version , not clear how
different from others, and not relevant in this study where only PC centres were involved. I suggest to drop it from here and from the aim statement.

11. Results. From Figure 1, 295 consecutive patients were “screened for eligibility” (not “proposed to enter the study”, as stated in the text) and 150 “were enrolled “. Exclusion for refusal to participate happened in 25 patients (8.5% of screened); written as it is now in the text it seems that 145 (49%) patients (295-150) “did not provide IC” which would wrongly indicate a very low acceptability of the tool. I suggest to revise the first paragraph of the results.

Discretionary revisions

12. If my interpretation in comment 7 is correct, I suggest to drop the word “consent” from line 358 page 11 (it is not a problem of consent, it is a problem of eligibility due to poor physical/cognitive conditions).

13. IS 116/295 correct? isn’t it 115 (92+22+1)?

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

'I declare that I have no competing interests"