Author's response to reviews

Title: Validity, reliability and responsiveness to change of the Italian Palliative care Outcome Scale. A multicentre study of advanced cancer patients.

Authors:

- Massimo Costantini (costantini.massimo@asmn.re.it)
- Elisa Rabitti (elisa.rabitti@ausl.re.it)
- Monica Beccaro (monica.beccaro@asmepa.org)
- Flavio Fusco (flavio.fusco@asl3.liguria.it)
- Carlo Peruselli (carlo.peruselli@gmail.com)
- Pietro La Ciura (laciura.pietro@alice.it)
- Alessandro Valle (alessandro.valle@tin.it)
- Cinzia Suriani (cinziasuri@alice.it)
- Maria A Berardi (alejandra.berardi@irst.emr.it)
- Danila Valenti (d.valenti@ausl.bo.it)
- Felicita Mosso (fmosso@libero.it)
- Piero Morino (piero.morino@asf.toscana.it)
- Giovanni Zaninetta (zaninetta@gmail.com)
- Giorgio Tubere (g.tubere@asl1.liguria.it)
- Massimo Piazza (mpiazza52@gmail.com)
- Michele Sofia (msofia@aogarbagnate.lombardia.it)
- Silvia Di Leo (silvia.dileo@asmn.re.it)
- Irene J Higginson (irene.higginson@kcl.ac.uk)

Version: 4
Date: 26 January 2016

Author's response to reviews: see over
Dear Editor,

please find enclosed our revised research article (MS: 9326686861673504), that we are submitting for publication in ‘BMC Palliative Care’. Following the suggestion of reviewer # 2 we have changed the title from “Outcomes for patients with advanced medical illness: a multicentre study of validity and predictors of response to change” to “Validity, reliability and responsiveness to change of the Italian Palliative care Outcome Scale. A multicentre study of advanced cancer patients”.

We have amended the manuscript according to reviewers’ suggestions. In the revised uploaded version the revised parts are highlighted in red. More specifically:

Reviewer # 1
Major Compulsory Revisions

• **Lines 202-203:** “The comments were transcribed verbatim, grouped according to question and common themes were identified. The POS was then modified accordingly for formal testing.” The authors should better describe the qualitative analysis. How was it done? Content analysis according to Bardin, Thematic analysis, etc.

The point raised by the reviewer is correct. Unfortunately, we did not perform a formal qualitative analysis and, as reported – now more extensively – in the text (lines 205-208) we just analysed the transcripts of the interviews.

• **Authors used two widely used instruments for the convergent and divergent analyses (EORTC QLQ-C15Pal and Facit-sp); however, their psychometric properties were not previously evaluated in Italy. It is a potential bias for the study. The authors could discuss the limitation of using non-validated instruments for the convergent/divergent analyses.**

The two questionnaires, as pointed out by the reviewer are widely used. Their psychometric properties were never evaluated in Italy, but an official Italian version was obtained by EORTC and FACIT organisations. When the study was designed, other questionnaires with the same construct and validated in Italian samples were
not available. This can be seen as a limitation of this study. We comment this point at the end of the discussion section. (lines 440-443)

- **... and also the inclusion of patients to the retest without confirmation of clinical stability.**
  At the end of the discussion section (lines 443-449), we have included a specific comment related to the limitation of the reliability analysis that was performed without confirmation of clinical stability before retesting the patients.

- **The authors did not discuss adequately several problems during the validation. It should be a better contribution to the literature if they show clearly the limitations of the tool and further discuss them.** For example: *The Cronbach’s alpha was only marginally adequate (~0.70), and should be presented with the 95% confident intervals.*
  The tool, as all instruments aimed at assessing QoL dimensions in palliative care patients, shows some limitation. In this new revised version we discuss more in deep the limitations of the POS, specifically of some items. (line 378-381 and 428-438). Cronbach’s alphas have now been reported with their 95% CIs, in text and tables.

- **There was clearly a floor effect regarding “Wasted time” and “Personal affairs” and also “Information” and “Share feelings”.** It should be discussed.
  There was a problem in the test-retest of “share feelings” with regards to the patients and to a lesser extent in relation to the staff. Why? They should discuss it better.
  These points are now discussed in the Discussion section (line 428-438)

**Minor Essential Revisions:**

- **Line 215 - To assess test-retest reliability of the POS, we asked inpatient hospices an additional POS assessment 24-48 hours after T1 assessment (T2).** The authors need to explain why 24-48h was chosen for the retest. In general it is a very short time to retest patients. The reference (or some similar reference) could be used to explain why to retest patients so fast (PMID: 24447633).
We estimated test-retest reliability with an interval of 24-48 hours. We thought it was an acceptable compromise between the need to avoid recall bias and the need to retest patients in a stable clinical condition. A recent systematic review (Paiva CE, 2014) recommended an interval of 24-48 hours for retesting patients with advanced cancer in palliative care when evaluating symptoms and 2-7 days more for other QoL dimensions. As suggested by the reviewer, we included this reference in support of our choice, (now ref # 22) and we slightly modify the methods section. (line 223)

• **Why performing the Preliminary field-testing and assessment of feasibility, content and face validity of the version 1? The author could go directly to the translation process using adequate translation methodology (version 2)**

We used version 1 in the preliminary assessment of the POS because when we started the first phase of the study, the POS version 1 translated by Franco Toscani (ref # 21) was the only Italian version available.

• **An important thing to be considered in test-retest is clinical stability. Authors assumed “that the quality of life of the patients should be rather stable”. How did they assume it?**

We cannot be sure, without any external variables measuring clinical stability, that the QoL of the patients are stable. We have assumed that because the probability of a rapid clinical deterioration, possible in palliative patients, should be not so frequent. This limit has been now commented in the limitation of the study at the end of the Discussion section (line 443-449).

• **Line 246 – “A Cronbach’s alpha of around 0.70 indicates good internal consistency without homogeneity”. The authors should correct the phrase, since “around” 0.70 is not correct. It should be higher than 0.70! Or between 0.70 and 0.95. Please use the recommendations of Terwee et al. (ref 21).**

In our opinion the intervals of Cronbach’s alpha indicating a good internal consistency (for example those proposed by Terwee et al) should not be considered as a formal threshold (good is ≥0.70, bad is below), but range of values that should help the researcher in interpreting the results. For the POS self-assessed by the patients, we found a Cronbach’s alpha of 0.67 at T0 and of 0.72 at T1. We have
changed the sentence taking into account the suggestion of the reviewer. Now the text is “According to the criteria proposed by Terwee CB, et al (23), Cronbach’s alpha ≥0.70 indicates good internal consistency without homogeneity.” (line 257)

- **Lines 249-250 - Responsiveness to change – Why not use an anchor-based method asking patients if they were better, the same or worse? It would be easier to conclude responsiveness to change in this case. ES could be calculated for each group of response.**

  The anchor-based method requires an external criterion (clinical or from the patient) to inform as to changes scores that are meaningful. It is used to interpret the magnitude of the observed changes and to estimate minimal important differences for PROs. Unfortunately, we did not plan in the protocol this type of analyses because we did not want to overload the PCUs participating to this spontaneous study. Nevertheless, the magnitude of the effect sizes can be useful indicators for interpreting the magnitude of the observed changes, as recommended by Norman GR, et al (QoL Research 2007) We have briefly commented this point in the limitation of the study at the end of the discussion section.

- **Line 256 – “A regression analysis was flitted to the data...” Please cite the type of regression used. Linear regression?**

  A linear regression analysis was used. The text was modified accordingly. (line 268)

- **Lines 261-262 – Please cite the independent variables.**

  We have cited in the method section all the independent variables included in the model.

- **Lines 308-309 There was little missing data for POS assessments; the highest was for the items ‘information’ and ‘personal affairs’ (3.3% each). The authors should include in the Methods section an assessment about the acceptable percentage of missing items. Is 3.3% acceptable?**

  The acceptable percentage of missing values for the items does not exist, as it depends by the aims of the specific study where the POS is used. We think that 5% is an acceptable proportion taking into account that the POS was assessed in a
sample of patients in a palliative care setting. We have included this point in the methods section (line 230).

- **Lines 313-314** – “Cronbach’s alpha for the 10 POS items at admission was 0.67”. *It is quite a low score. It should be presented in a table (supplementary at least) together with the Cronbach’s alpha if items deleted. Are there any item that excluded improves score?*
  
  We have modified Additional file 4, now including two new columns reporting, for each POS item, the corrected item-total correlations and the Cronbach’s Alpha if item deleted. We have briefly reported these figures in the results section. We have also reported the Cronbach’s alpha of the POS scale with its 95% confidence intervals.

- **Lines 381-385** – *This paragraph should be referenced and improved.*
  
  We have changed and referenced the paragraph (line 396).

**Discretionary Revisions**

- **Lines 118, 119**: there is an extra comma “(10 inpatient hospices, 10 home-based PCTs,)”
  
  We deleted the extra comma.

- **Lines 152-153** – *The following phrase could be deleted…” The European Organisation for Research and Treatment of Cancer quality of life core 15 palliative questionnaire (EORTC QLQ-C15-PAL).”*
  
  According to reviewer’s suggestion, we delete the first phrase.

- **Line 211** - *...the functional status through the Eastern Cooperative Oncology Group (ECOG) scale. COG. The authors should correct the name of the PS scale.*
  
  We deleted the word “COG” from the text.

- **Lines 407-408** – “Five percent of patients declined to take part in the study”. *In fact, there was fifty percent of refusal…and it was not a low percentage. I believe it was a typo.*
There was a mistake in the presentation of the results in Figure 1 and in the text. We have modified both the Figure and the text, by separating non-eligible patients by eligible-patients who did not give the consent to the study.

- **The authors could discuss the limitation of using non-validated instruments for the convergent/divergent analyses and also the inclusion of patients to the retest without confirmation of clinical stability.**
  We have discussed this point at the end of the Discussion section.

- **Additional file 7. There is a lack of point in the number 028 (last column to the right).**
  We modified the number, now 0.28.

**Reviewer # 2**
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**

- **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.**
  In the revised version of the paper, we have modified the title, now more focused on the key words, as suggested by the reviewer.

- **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.**
According to the reviewer’s suggestion we have integrated the introduction with a specific sentence providing the rationale for the second part of the aims (line 98).

- **Study aim is not clearly stated (may be also due to some typos in them 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. We agree with this comment. According to the reviewer’s suggestion, we have revised the aims and corrected some typos in the paragraph.

- **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.

We agree that the two terms could be misleading. In the statistical section we replaced “Internal consistency (reliability) of the scales” with “internal consistency”. We also changed the abstract and the discussion section.

- **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.**

We have modified the previous additional file 5 (test-retest for patients scores), additional file 6 (test-retest for staff), and additional file 7 (agreement between the
two POS versions) by including a new column with the proportion agreement within 1 score. We have also estimated the ICC for the total POS score and we have reported that in both the text and in the tables. Accordingly, we have included a brief statistical section about ICC, and we have modified the results section. Additional files 5 and 6 have been transformed in tables (now table 3 and 4)

- **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.

We agree with the observation of the reviewer. This point is now discussed in the Discussion section of the article. (line 396)

- **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.
The regression analysis shows that patients admitted to home care had an improvement in POS scores six times greater than patients admitted to inpatient hospices. We agree with the reviewer that the interpretation of these findings are rather complicated. The association between precocity of provision of palliative care and improvement is difficult to demonstrate in this study. Following the suggestion of the reviewer, we have completely changed the paragraph, avoiding any conclusion from these findings. (line 404)

- **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.**

In the new version we have included this point as a limitation of the study.

**Minor essential revisions**

- **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, it is also important to clarify which are the main differences between tools (POS_P vs POS_S) and between versions (briefly in the box).**

According to the reviewer’s suggestion we have included a paragraph in the Methods section with some explanation of the differences between POS version 1 and 2. (line 146). We did not have summarised the information in a specific box, but we have referred the reader (with a specific sentence and the reference # 14) to the POS web page.

- **Page 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.**
According to reviewer’s suggestion, we deleted the sentence that might be confusing for the reader. For the same reason, we also removed the definition “POS system” from the text.

- **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.

We agree with this point that was noticed also by Reviewer 1. A mistake occurred in the presentation of the results in the flow chart of Figure 1 and in the text. We have modified both the Figure and the text, by separating non-eligible patients by eligible-patients who did not give the consent to the study.

Discretionary revisions

- **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).**

We agree with this suggestion and we removed the word consent.

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

We agree, the correct figure is 115/295 and it also include early death. We have modified accordingly the text.
All authors agreed to the content of the revised article and to its re-submission to ‘BMC Palliative Care’, and declared, as reported in a specific section of the text, they have no competing interests.

With best regards
Yours sincerely

Massimo Costantini, MD
Palliative Care Unit,
IRCCS Arcispedale S. Maria Nuova, Reggio Emilia (Italy)
Email: costantini.massimo@asmn.re.it