Author's response to reviews

Title: A critical analysis of test-retest reliability in instrument validation studies of cancer patients under palliative care: a systematic review

Authors:

Carlos E Paiva (caredupai@gmail.com)
Eliane M Barroso (embarroso@uol.com.br)
Estela C Carneseca (estatistica@hcancerbarretos.com.br)
Cristiano P Souza (crispeixoto10@hotmail.com)
Felipe T dos Santos (fthome10@hotmail.com)
Rossana VM López (rossana.nap@hcancerbarretos.com.br)
Bianca SR Paiva (bsrpaiva@gmail.com)

Version: 3
Date: 5 January 2014

Author's response to reviews:

January, 5th 2014

To
Editor-in-Chief
BMC Medical Research Methodology

Dear Editor:

Ref.: Ms. No. 8359593381097364
Title: “A critical analysis of test-retest reliability in instrument validation studies of cancer patients under palliative care: a systematic review”.

First of all we would like to thank the careful revision of our manuscript. We believe that the changes made were able to improve the scientific content of our manuscript. Below we provide you a list of changes made with point-by-point discussions. The changes have been highlighted in gray within the manuscript.

Comments made by the Reviewer 1 (Philipa Davies):

1. Abstract, methods: It would be helpful to include the dates of the searches
Response: Abstract, methods: “A systematic search of PubMed, EMBASE, PsychInfo, CINAHL, and SCIELO was performed”. New phrase: “A systematic search of PubMed (1966 to June 2013), EMBASE (1980 to June 2013), PsychInfo (1806 to June 2013), CINAHL (1980 to June 2013), and SCIELO (1998 to June 2013) was performed”.

2. Abstract, results: “Multi-symptom instruments were retested over a shortened interval when compared to the HRQoL instruments (24 hours vs. 168 hours, respectively; p=0.001)”. State that these results are medians.
Response: Abstract, results: “Multi-symptom instruments were retested over a shortened interval when compared to the HRQoL instruments (24 hours vs. 168 hours, respectively; p = 0.001)”. New text: “Multi-symptom instruments were retested over a shortened interval when compared to the HRQoL instruments (median values 24 hours and 168 hours, respectively; p = 0.001)”.

3. Methods section: Some of the information in this section could be re-ordered to be more consistent with PRISMA.

- Data sources: “Figure 1 depicts the search strategy”. Figure one is the PRISMA flow diagram showing the search results and would fit better within the results section of the paper.

Response: As suggested, the text “Figure 1 depicts the search strategy” was deleted from the methods (data sources). The new phrase “Figure 1 summarizes the identification and selection of studies.” was inserted in the results section (first paragraph).

- Search strategy: First paragraph. This is not the search strategy as such. It would fit better further on, after the inclusion/exclusion criteria and prior to data extraction.

The first paragraph was included after inclusion/exclusion criteria.

Response: The following text was moved from the section ‘search strategy’ to the first paragraph of “data extraction”. Methods, data extraction, new text: “Initial searches (titles and abstracts) were conducted independently by CPS and FT. The studies with full text available were further reviewed, the data were independently extracted by two other reviewers (CEP, CPS), and the data were verified by a third reviewer (BSRP). A standardized data collection form was used. The data collected included study demographics…”

Inclusion and exclusion criteria were grouped in the topic “Eligibility criteria”. We moved it into the beginning of the methods section, after study design and before data sources.

4. Results section: There is some inconsistency in the referencing in this section. On some occasions, the studies referred to are referenced and in other cases not. I think that it would be helpful if each results statement references the studies to which it refers as this information is not always easy to deduce from the results tables or is not given within them.

As suggested, the references were completely revised in this section.

5. Discussion, pg. 14, 3rd paragraph: “Three studies [21-23] compared 2 different time frame intervals for the retest. Two of them [21-22] measured global HRQoL 3 hours and 7 days after the first evaluation; the test-retest results were 0.84-0.93 and 0.63 at 3 hours and 7 days, respectively. The other validation study [23] evaluated cancer symptoms using the Edmonton Symptom Assessment System (ESAS) scale. It found higher test-retest values at shorten time intervals, with the exception of the symptom fatigue”. This feels more like results than discussion and therefore I would suggest moving it onto the results
The following text was excluded from the discussion section and moved to results section (scores of retest, 3rd paragraph): “Three studies [13, 19, 37] compared 2 different time frame intervals for the retest. Two of them [13, 19] measured global HRQoL 3 hours and 7 days after the first evaluation; the test-retest results were 0.84-0.93 and 0.63 at 3 hours and 7 days, respectively. The other validation study [37] evaluated cancer symptoms using the Edmonton Symptom Assessment System (ESAS) scale. That study found higher test-retest values for shorter time intervals, with the exception of the symptom of fatigue (Table 5).” In addition, the text “Three studies [13, 19, 37] evaluated the retest reliability at 2 different time points (< 24 hours and 1 week after the first evaluation); in general, a lower time interval was associated with a better retest analysis result.” was included in the discussion (discussion, 7th paragraph).

6. Minor essential revisions. Methods, exclusion criteria: Missing number (3) in the reasons for exclusion.

Response: The exclusion criteria were completely revised (as suggested by the reviewer 2). The new text: “Studies were excluded for any of the following reasons: (1) the study was not published as a full article (i.e., conference proceedings were excluded); (2) the study contained pediatric data; or (3) the publication was a duplicate publication.”

Comments from the Reviewer 2 (Camilla Zimmerman):

1. Under search strategy, on page 6, it is not clear which discrepancies were rectified by referring to the original articles. Are you referring to differences in opinion regarding which articles should be reviewed?

Response: The phrase “Discrepancies were rectified by referring to the original articles” was not correct. Thus, it was excluded from the manuscript.

2. Under exclusion criteria, page 7, “The study was a validation of a needs instruments”- incorrect grammar and not clear what a needs instrument is. Please clarify.

Response: We agree with reviewer. This phrase was excluded from the eligibility criteria. The exclusion criteria were completely revised (please see below).

3. As well, it is redundant to list all the various types of studies that were excluded, which did not fall into the inclusion criteria. In other words, if inclusion criteria are “validation study of a multidimensional quality of life instrument or a multidimensional symptom assessment instrument”, then it is obvious that studies were excluded if that were not validation studies, focused on only one symptom, or evaluated a quality of care instrument. Likewise, if the inclusion is a population composed mainly of advanced cancer patients, then there does not need to be an exclusion criterion for “not composed mainly of cancer patients”. Therefore please edit the “exclusion criteria” section accordingly.

Response: The exclusion criteria were edited accordingly.
Previous text: “Studies were excluded for any of the following reasons: (1) the study was not published as a full article (i.e., conference proceedings were excluded); (2) the study contained pediatric data; (3) the population studied was not composed mainly of advanced cancer patients; (4) the publication was a duplicate publication; (5) the study was not a validation study; (6) the study was a validation of a quality of care instrument; (7) the study was a validation of a needs instruments; (8) the study was a validation of symptom instruments focusing on only one symptom (for example, pain and fatigue multi-item instruments); (9) the study was a validation of instruments that only included psychological symptoms; or (10) the study examined quality of life instruments with #2 evaluated domains.”

New text:” Studies were excluded for any of the following reasons: (1) the study was not published as a full article (i.e., conference proceedings were excluded); (2) the study contained pediatric data; or (3) the publication was a duplicate publication”.

4. Under scores of retest (page 11), there should not should be an explanation in the results for why the trend for retest results favoring shorter time intervals were nonsignificant (ie due to small numbers and lack of power). This is an interpretation, and should be mentioned in the discussion section (as it already is). Should just be mentioned in the results that these were nonsignificant trends, without offering an explanation here. As well, in the discussion it should be mentioned that another reason contributing to the nonsignificant results might be the large interquartile range for some of the domains.

Response: In Results, page 11, Scores of retest.

Previous paragraph: “Although trends favoring shorter time intervals for the retest in the studies with satisfactory results for the retest statistical analysis (value # 0.70) were observed, the small number of studies reviewed did not provide sufficient statistical power to demonstrate a statistically significant difference”.

New paragraph: “There was a non-significant trend favoring shorter time intervals for the retest in the studies with adequate results for the retest statistical analysis (value # 0.70) in comparison with those with non-satisfactory results (value <0.70) (Table 4).

Discussion section, 7th paragraph: Previous text: “There was a trend of shorter time periods in the adequate test-retest reliability results when compared with the scores with inadequate results (less than 0.7). One reason contributing to the non-significant results might be the large interquartile range for some of the domains; since few studies were analyzed, there was insufficient statistical power for further conclusions.”

5. Minor. Under exclusion criteria on page 7, there seems to be a number missing (number 3, before “the population studied was not composed mainly of advanced cancer patients…)

It was corrected. Please see the above revision of the exclusion criteria.
6. It is mentioned on page 14 that “we believe that symptoms are more prone to fluctuations than overall HRQoL scores”. I agree with this statement, but it would carry more weight if it was rephrased as less tentative and followed by a reference. At very least, a reason for this belief should be given.

Previous text: “In fact, because of concerns about reassessing an unstable patient, some authors (n = 7) reapplied the questionnaires at very short intervals (i.e., less than 24 hours). We believe that symptoms are more prone to fluctuations than overall HRQoL scores. Consequently, we observed that multi-symptom instruments are generally retested within a shorter time frame compared with HRQoL instruments”.

New text: “In fact, because of concerns about reassessing an unstable patient, some authors (n = 7) reapplied the questionnaires at very short intervals (i.e., less than 24 hours). Jim et al. [48] investigated daily and intraday changes in the fatigue, depression, sleep, and activity scores in a cohort of cancer patients undergoing chemotherapy. Significant changes were observed over time. Additionally,Dimsdale et al. [49] investigated cancer-related fatigue every hour for 72 consecutive hours and observed a diurnal variation in fatigue. HRQoL, on the other hand, is a multidimensional construct that encompass physical, psychological, social, and spiritual domains. In general, instruments that measure HRQoL use recall periods of 7 days. Although HRQoL is not commonly assessed on a daily basis, it is expected to behave stably over a few days, especially the social, existential, and global domains. Consequently, we observed that multi-symptom instruments are generally retested within a shorter time frame than HRQoL instruments.”

7. Is there a time interval that you would suggest for test-retest analysis for symptom and quality of life measures, respectively, for palliative populations? This should be useful for future research.

Response: The following text were added in the discussion section (7th paragraph): “Considering the median time interval used in the studies with adequate test-retest results, in addition to the findings from studies that used two different time intervals for the retest, we can recommend that patients under palliative care for advanced cancer should be retested somewhere around 24 to 48 hours later when evaluating cancer symptoms and 2 to 7 days later when assessing HRQoL. However, we believe that the most important factor is not the time itself but rather confirmation of clinical stability before retesting patients.”

8. Table 4. Suggest renaming the table so that it better describes the results. Also, suggest moving the heading “median (Q1-Q3)” to right underneath “Time interval (hours), because the median

Response: As suggested, we corrected the Table 4. The heading “median (Q1-Q3)” was moved to underneath “time interval (hours)”. In addition, we renamed the table. New title: “Median values of time intervals of studies with non-adequate (<0.70) and adequate (#0.70) test-retest values.”
• We also included a copy of PRISMA guideline, as suggested by the Editor, as a supplementary file. The phrase “(see the PRISMA checklist in the Supplementary file)” was inserted in the last paragraph of methods section.

• The manuscript was revised by a professional copyediting service (American Journal of Experts) regarding the English language.

We would like to thank the opportunity to have our work under consideration for publication in the BMC Medical Research Methodology.

With kindly regards,

Carlos Eduardo Paiva, M.D., Ph.D. (corresponding author)
Department of Clinical Oncology
Division of Breast and Gynecology
Barretos Cancer Hospital
Barretos, SP, Brazil