Author's response to reviews

Title: Functional capacity, physical activity and muscle strength assessment of individuals with non-small cell lung cancer: A systematic review of instruments and their measurement properties.

Authors:

Catherine L Granger (catherine.granger@mh.org.au)
Christine F McDonald (christine.mcdonald@austin.org.au)
Selina M Parry (Selina.parry@austin.org.au)
Cristino C Oliveria (cc.oliveira@student.unimelb.edu.au)
Linda Denehy (l.denehy@unimelb.edu.au)

Version: 2 Date: 22 February 2013

Author's response to reviews: see over
Dear Dr Chap

Thank you for reviewing our manuscript titled “Functional capacity, physical activity and muscle strength assessment of individuals with non-small cell lung cancer: A systematic review of instruments and their measurement properties”. Please find below our response and actions to the reviewers comments.

<table>
<thead>
<tr>
<th>Reviewer 1 comments</th>
<th>Authors response</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Major revisions</strong></td>
<td>Thank you for reviewing our article. We acknowledge that your point is well justified and therefore have altered the manuscript to better synthesize the results from studies as a group (instead of providing results from only individual studies) and by doing this we hope the new revised manuscript allows readers to compare the 13 tests more readily. We have addressed this point by adding another table to the paper (based on your recommendations in a subsequent comment below). The new table (Table 4) synthesizes the evidence on measurement properties from studies as a whole and provides a comparison of each test to allow the reader to more easily compare the 13 tests. We feel this markedly improves the paper. This table is referred to in the text and we have also added the following line to the start of the ‘Study results’ section: “Study results are summarised in Error! Reference source not found. and the sections below. The stair-climbing test, six-minute walking test (6MWT) and incremental-shuttle walk test (ISWT)”</td>
</tr>
<tr>
<td>The review of the outcome measures (OM) by the authors presents a substantial effort to “identify, evaluate, synthesis and compare” the measurement properties in participants with NSCLC”. As presented the manuscript succeeds to “identify and evaluate” the literature to date with a clear set of search strategies presented. On the aims of synthesizing and comparing the OMs, however, it is felt that in its current state the manuscript falls short of these goals. Effort is required to improve the manuscript in these aspects in order to improve the usefulness of the review in the literature.</td>
<td></td>
</tr>
</tbody>
</table>
performed the best out of the 13 tests reviewed, primarily due to lack of studies investigating measurement properties of the other 10 tests (Table 4).”

The old table 4 (Inter-rater reliability, intra-rater reliability and measurement error associated with outcome measures) has been moved to the appendix as Appendix table number 4. This table only provided data from two studies and we feel this would be better suited to the appendix.

Minor revisions

Methods: Please justify why only the lowest score on the COSMIN was used. Does this affect its applicability to the components under investigation?

The COSMIN scoring system developed by Terwee and colleagues states that the overall score for each item should be obtained by using the lowest score (excellent, good, fair or poor) recorded for any question within the item [1]. For study consistency and methodology transparency we did not deviate from these recommendations. In the manuscript this has been altered to explain this and now reads “The overall score for each item was obtained by using the lowest score (excellent, good, fair or poor) recorded for any question within the item, as recommended by the COSMIN scoring system [1].”

Detailed information about the COSMIN checklist (and a copy of the checklist) is available at http://www.cosmin.nl/the-cosmin-checklist_8_5.html. If the editors agree, we can provide a link/address to this website within the manuscript.

Methods: Please elaborate on the choice to exclude longitudinal studies providing indirect evidence of responsiveness.

Elaborate on what the indirect evidence was and also why to exclude these studies considering this seems to be the element most lacking in the evaluations included.

Methods: Please also elaborate why if a relevant sub-component of a battery measure included an OM of interest, why it was excluded?

These types of studies were excluded from the review based on the guidelines for conducting systematic reviews of measurement properties by de Vet and colleagues [2]. According to these recommendations:

1. Longitudinal studies in which indirect evidence for responsiveness of the outcome measures may be found; and

2. Studies which validated an alternative outcome measure against one of the outcome measures of interest in which indirect evidence for validity may be found were excluded. de Vet and colleagues [2] state
that these types of articles should be excluded from systematic reviews of measurement properties because “it is difficult to interpret the evidence for validity or responsiveness provided in these studies, because no hypotheses about these measurement properties have been formulated or tested in them” [2]. For study consistency and methodology transparency we did not deviate from these recommendations. In addition,

3. Studies which investigated or evaluated a battery measure which included one of the desired outcome measures as a sub-component were excluded because battery measures are designed to be used in their entirety and provide clinicians with different information to the outcome measure of interest when used in isolation.

To explain this more clearly we have altered the manuscript slightly and referenced our justification. This now reads “Studies validating an alternative test against an outcome measure of interest (which provide indirect evidence for validity) and longitudinal studies (which provide indirect evidence for responsiveness) were excluded because such studies have not specifically formulated or tested hypotheses about the measurement properties [2]. Studies evaluating a battery measure including a relevant sub-component were also excluded as they are designed to be used in their entirety.”

<table>
<thead>
<tr>
<th>Results: Tables – Please provide reference numbers for manuscripts in Tables if possible. Allows for easier reference</th>
<th>Changed as recommended. All tables now include reference numbers.</th>
</tr>
</thead>
</table>

| Tables – suggests in footnotes that sorted by time since publication, however, that does not seem to be the case provide the amount | Changed as recommended. Articles in the tables are now sorted initially by type of test and then according to time since publication. |

| The results as currently displayed present much detail concerning the individual studies included which is helpful as supplementary material. What is lacking is a sense of summarization of the findings. If the utility of this review in the literature will be to provide future researchers | Thank you for this suggestion. As noted above, we have changed Table 4 to now synthesize results and compare the 13 different tests. This table now gives readers information on what is known and what is not known about the |
through “synthesis” with a sense of whether current tools are useful and based on how much data, this message is not clear from the data as presented. Would it be possible to summarize into one table the quantitative outcomes across the studies with a measure of the performance of the instruments across studies.

Ie:

<table>
<thead>
<tr>
<th>N studies N cases measures of performance (range etc.)</th>
<th>Rel. Val. Respons.</th>
</tr>
</thead>
<tbody>
<tr>
<td>FC</td>
<td>PA</td>
</tr>
<tr>
<td>MS</td>
<td></td>
</tr>
</tbody>
</table>

Discussion: A large proportion of the information contained in the Discussion section should be included in the results section.

The following sentence has been moved from discussion to the results section: “With every 50m improvement in 6MWT, survival improved by 13% [3] and patients walking ≥400m pre-chemotherapy had greater survival time [4].”

Response to reviewer 2:

<table>
<thead>
<tr>
<th>Reviewer comment</th>
<th>Author response</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Major revisions</strong></td>
<td>Thank you for your response. We agree that this is an important area of research and addresses a significant gap in the literature.</td>
</tr>
</tbody>
</table>
higher risk for adverse operative outcomes continues to be a priority.

I think the paper would benefit from a stronger, more detailed introduction, so that readers have a better understanding of how this paper fits within the broader literature on NSCLC and exercise science.

The authors appear to be making some major assumptions about the average reader’s understanding of the measures assessed, their use in other populations (for example, walk tests and CPET in lung transplantation and COPD populations), and how they relate to patient health status, HRQL and even survival (to which the authors refer later in the paper). BMC Cancer is not a lung cancer or thoracic surgery journal and a few comments about exercise capacity, current gold standards, and how NSCLC is known to affect exercise physiology seem important to ‘set the stage’ for the reader.

Some readers may not know what terms such as ‘functional capacity’ actually mean. Moreover, some justification to link why measures of muscle strength such as hand-held dynamometry might be related to measures like the 6MWT is also advisable. As a physiotherapist, I am familiar with this literature, but I suspect many readers of BMC Cancer will not bring this background to their reading of the paper. I am not recommending lengthy explanations, but I would recommend re-working the introduction to include additional well-crafted background information.

The introduction has been re-worked and we have elaborated on these points. We have added information regarding how and why NSCLC affects exercise physiology and included definitions of functional capacity, physical activity and muscle strength for the readers. We have also referred to the gold standard tests and the reason why there is a wide variety of different tests used.

The first paragraph now reads: “Non-small cell lung cancer (NSCLC) is associated with significant disease burden, impaired physical status and diminished physical activity [5, 6]. Due to the disease and treatment (surgery, chemotherapy and or radiotherapy) adverse physiological and psychological effects are prevalent in NSCLC, particularly exercise intolerance, weakness and impaired gas exchange and commonly a cycle of functional decline ensues [5]. Increasingly exercise interventions targeted at preventing the functional decline associated with NSCLC or improving the physical status prior to or after cancer treatment are the focus of research trials [7]. Three commonly used endpoints are functional capacity “the maximal capacity of an individual to perform aerobic work or maximal oxygen consumption” [8]; physical activity “any bodily movement produced by skeletal muscles that results in energy expenditure” [9]; and muscle strength “the maximum voluntary force or torque brought to bear on the environment under a given set of test conditions” [10]. The gold standard instruments (outcome measures) to assess these outcomes are laboratory based, which are not always feasible for use in research or clinical practice [11]. Therefore, a wide variety of instruments have been used to assess changes in these outcomes in the NSCLC literature.”
In terms of methodology, the search strategy and rating criteria are reasonably well-described.

A bit more explanation about the COSMIN checklist and its domains/items seems warranted. Specifically, the authors should indicate which four items were dropped from the checklist. This would contribute to the paper’s methodological transparency. Figure 1 describing the inclusion of studies and the authors’ decision-making was very helpful.

The Results section is clearly written and this section is definitely enhanced by filled in details missing from the manuscript text. I’m not sure that these should be included as supplementary material as these seem integral to readers' understanding the paper and I would encourage their inclusion with the main paper at publication.

With respect to the Discussion, the study’s limitations are reasonably well-addressed. I think it is important that they note the inherent problems of including studies with patients with mixed types of cancers (not just NSCLC) since lung neoplasms are likely to affect exercise capacity and gas exchange in significantly different ways from cancers at other sites. It was not sufficiently clear why they chose to include these studies in the first place.

Further consideration in the Discussion concerning the large SDs seen in some of the
measures reported in Appendix 3 would be welcomed. The age ranges for patients with NSCLC cited in the tables would suggest that some consideration of comorbidities and their contribution to this variability in functional capacity/exercise tolerance should be mentioned. Many of these patients are smokers and may have COPD or heart disease that may influence their exercise performance, and still others will have arthritic conditions. Some studies included patients receiving chemotherapy – some chemotherapeutic agents have cardiotoxic side effects which can in turn affect functional exercise capacity. The authors would do well to consider that there is considerable heterogeneity amongst patients with NSCLC, and specifically in terms of ‘performance status’.

In addition to the limitations section of the Discussion however, the authors do a good job of contextualizing their results within the larger body of literature. The authors are thoughtful/reflective and the claims they have made are grounded in the data presented.

**Minor revisions**

1) In both Methods and Results, I think the terms Search one and Search two should be reformatted as Search 1 and Search 2 as it was sometimes confusing to have the ‘one’ and ‘two’ lower case in the text. This will enhance readability.

   - Changed as recommended. Now reads “Search 1” and “Search 2”

2) Methods, pages 5-6: The methods for Search 2 are overly ‘broken up’ into headings with one sentence underneath each one (see lines 130 – 142 for example). I think some re-writing so that the ideas have a more narrative flow would help the paper.

   - We have re-worked the methods section to reduce the number of sub-headings and help the flow of the paper. The following sub-heading have been combined into one section: Part 1: Information sources, search and study selection

     Part 2: information sources, search, study collection, data collection process and data extraction

3) Discussion (1st paragraph): lines 283 – 286. I take issue with the statement “Research into the

   - Thank you for this comment. We now understand that the second line of the discussion
benefits of exercise for individuals with NSCLC is relatively recent and this review has demonstrated that research into the suitability of instruments for use in exercise trials is embryonic” for two reasons. First, the importance of exercise/early mobilization following thoracic surgical procedures has long been recognized. This is standard clinical practice. Moreover the role of exercise in a wide array of post-surgical and oncologic populations outside of NSCLC has been researched previously and I would suggest that some of this literature is certainly informative to the NSCLC population (the authors would do well to consult the work of K. Courneya for example). I would suggest rewording this sweeping statement or at least making their argument clearer.

The second reason I find this statement problematic is that the authors are introducing a very different concept at the outset of this discussion from the focus and results that have been presented up to that point. The paper focuses on evaluating and validating an array of clinical assessment measures used in the NSCLC population. The results presented are NOT about the benefits of exercise, but rather about how these tests perform as assessment tools. While these tests may indeed ultimately be used to evaluate the benefits of exercise in NSCLC, to bring this up at the beginning of the discussion is confusing for readers.

This sentence now reads “There is the risk of publication bias, where studies which have found poor measurement properties have not been published. Given that registration of studies evaluating measurement properties is not standard practice, the extent of this is unknown [2].

In closing, once the authors make these changes, I would then recommend that this paper be published in BMC Cancer. Thank you for the opportunity to review this very interesting manuscript.

Thank you for your review of our paper.
Thank you for the chance to publish in BMC Cancer. We look forward to your feedback and we are willing to alter other aspects of our manuscript as you see fit.

We hope you consider publishing our work.

Yours sincerely,

[Signature]

Catherine Granger
