OARSI recommended performance-based tests to assess physical function in osteoarthritis of the hip or knee: Authors, reply.

Fiona Dobson, Kim Bennell, Rana Hinman, Ewa Roos, Haxby Abbott, Paul Stratford, Aileen Davis, Rachelle Buchbinder, Lynn Snyder-Mackler, Paul Hansen, Julian Thumboo, Yves Henrotin

PII: S1063-4584(13)00897-2
DOI: 10.1016/j.joca.2013.07.011
Reference: YJOCA 2950

To appear in: Osteoarthritis and Cartilage

Received Date: 11 July 2013
Accepted Date: 13 July 2013


This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Letter to the Editor

OARSI recommended performance-based tests to assess physical function in osteoarthritis of the hip or knee: Authors’ reply.

Sir,

We thank Professor Dekker for his letter (1) and interest in our paper (2). We welcome Professor Dekker’s challenges as part of the healthy peer review process, and offer the following responses.

Professor Dekker firstly questions the feasibility of the OARSI recommendations questioning both the choice of several tests (instead of a single test) and the feasibility of the individual tests, on the basis that they require a substantial investment of time and resources. Instead, he proposes a single test (the Timed Up and Go test, TUG) as a better alternative. We disagree on both counts. Firstly, the rationale to include a set of tests was based on the desire to represent several activity themes (or physical performance domains) relevant to people with hip or knee osteoarthritis. The reasons for incorporating multiple activity themes in the recommended set is outlined in the article (2), and principally includes: content validity, measurement theory, the need to capture change across different interventions and requirements of clinic practice. Further, we considered evidence that the tests from different activity themes have different change trajectories that are important for detecting change at different time points. For example, following total joint replacement, the use of a single test, such as the TUG, may be useful for early detection of change, when this test is shown to be responsive, however it may not be as useful over longer time intervals, where the test has been found to plateau 8-9 weeks post-surgery with ceiling effects evident at 9-10 weeks post-surgery (3). Inclusion of multiple activity themes, such as stair climbing, would enable detection of change at later points (3). Additionally, the TUG has been shown to be the least responsive measure from a suite of performance-based measures to detect change following different interventions (4, 5). The ceiling effect associated with this measure may limit its ability to detect change in people with hip or knee osteoarthritis who are functioning at higher levels. We certainly agree with the statement made by Professor Dekker that “three separate tests, as suggested in the present recommendation, provide more information” (1), but disagree that the inclusion of these three simple tests is at a “level of sophistication only required in trials of specific exercise modalities” (1), for the above reasons. Furthermore, we agree that it will be “a great step forward if a single and simple test” (1) could be recommended for “all future trials... on pharmacological interventions, surgical interventions, and non-pharmacological interventions” (1), but have concluded that, based on the available evidence, such a recommendation cannot currently be supported.

Regarding the feasibility of the individual tests recommended, we would like to highlight that all candidate tests considered for selection were required to be feasible, that is, they did not require specialist equipment and could mostly be performed within the field (i.e. hospital/clinic/rooms setting). This specific attribute was established in the methodology.

The second concern raised was that there is insufficient evidence to support the recommendations of the specific tests and that personal preferences, instead of empirical evidence, “had a strong influence on the recommendation” (1). As highlighted in our previous systematic review, we agree
that measurement property evidence is not complete for any performance-based tests used to
assess people with hip or knee osteoarthritis (6), and hence selection on measurement evidence
alone simply was not possible. Indeed, our concluding recommendation was that “future research
priorities should be directed towards expanding the measurement-property evidence of the
recommended tests” (2). However, in face of the limitations, the tests that were recommended
were those with the best available measurement property evidence, as this factor was given priority
and was most strongly weighed in the selection process. Other important factors included
representation of the identified key activity themes relevant to the population, feasibility and
statistical properties of the scoring method used.

Professor Dekker additionally states that “the authors’ preference for tests set to time” rather than
distance or target number “is entirely based on reasoning, not on empirical data”. Reasoning does
lead us to conclude that a time, or number of repetitions for a timed interval, would result in fewer
missing values owing to persons not being able to complete the test. In addition, however, we did
examine empirical evidence from the group’s databases to support this preference. For example,
two popular walk tests are the 6-minute walk test and an alternative 400 m walk test. To obtain a
score for the latter test, a person must be able to walk 400 m. In Table 1, the 6-minute walk test
results obtained for 83 people prior to total knee joint replacement are provided. Had the 400 m test
been performed, only 54% (45/83) of patients would have completed the test resulting in substantial
missing data. Similar results were evident favouring the 30-second chair-stand test over the five-
repetition chair-stand test.

Table 1 Results for all six-minute walk tests completed compared to those completed with a
minimum value of 400m.

<table>
<thead>
<tr>
<th>Results for all 6 minute walk test</th>
<th>Results for 6-min walk test &gt; 400m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>83</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>413.4 (111.0)</td>
</tr>
<tr>
<td>Min</td>
<td>138</td>
</tr>
<tr>
<td>Max</td>
<td>644</td>
</tr>
</tbody>
</table>

The third concern was over the ambiguity of the status of the two further tests recommended in
addition to the minimum core set (timed up and go and six-minute walk test). These tests were
recommended for three reasons: 1) they are commonly used in clinical practice and research in OA
and other populations; 2) they demonstrate reasonable measurement properties; and 3) they
represent relevant activity themes (domains) not covered by the other recommended tests.
Suggested guidelines for incorporating these tests into the core set were provided, including when
the purpose is to: (i) compare outcomes across different population groups (within or beyond OA),
(ii) continue existing research protocols or standard clinical testing that already include these tests,
and (iii) focus on physical function (including the domains of walking long distances or aerobic
capacity) as the main outcome dimension. The six-minute walk test, for example, was suggested as a
useful test to include when the interaction of co-morbidities on walking ability is desired (for
example, in a weight loss study). As highlighted above, the TUG would be useful for detecting short-
term change following surgery and in people at lower levels of functioning.
We hope that the interest in the recommended performance-based measures for people with hip or knee OA continues and that the uptake of these measures assists with the endeavour of the standardization of measurement in research and clinical practice.

Fiona Dobson  
Kim Bennell  
Rana Hinman  
Ewa Roos  
Haxby Abbott  
Paul Stratford  
Aileen Davis  
Rachelle Buchbinder  
Lynn Snyder-Mackler  
Paul Hansen  
Julian Thumboo  
Yves Henrotin

**Corresponding Author**

Dr Fiona Dobson  
Centre of Health, Exercise and Sports Medicine  
Department of Physiotherapy  
The University of Melbourne, Carlton  
Australia  
Email fdobson@unimelb.edu.au

**Author Contribution**

All listed authors were involved in the conception, drafting or revising of the letter critically for important intellectual content and final approval of the version published. Dobson takes full responsibility of the integrity of the work as a whole form inception to completion.

**Conflict of Interest**

The authors declare no competing financial interests

**Declaration of funding**

The corresponding authors time was partly supported by NHMRC Program Grant #631717

**References**

