Author’s response to reviews

Title: Validation of a new test that assesses functional performance of the upper extremity and neck (FIT-HaNSA) in patients with shoulder pathology

Authors:

Joy C MacDermid (jmacderm@uwo.ca)
Myriam Ghobrial (myriam_ghobrial@hotmail.com)
Karine B Quirion (sea_kwinz@hotmail.com)
Melanie St-Amour (mstamour23@cogeco.ca)
Tanya Tsui (t_tsui@hotmail.com)
Dave Humphreys (dchumphreys@rogers.com)
John McCluskie (mccluski@hhsc.ca)
Eddy Shewayhat (eshewayhat@hotmail.com)
Vickie Galea (galeav@mcmaster.ca)

Version: 2 Date: 5 December 2006

Author’s response to reviews: see over
Reviewer's report
Title: Validation of a new test that assesses functional performance of the upper quarter (FIT-HaNSA) in patients with shoulder pathology
Version: 1 Date: 20 October 2006
Reviewer: Jolanda Luime
Reviewer's report:
General
Very interesting and valuable paper, but need some consideration in introduction, methods and discussion.

Introduction
First paragraph
It is unclear from the introduction what the optimal use of the new tests will be. In the introduction a statement is made about the lack of functional performance tests in shoulder research.
It is well known that there is a lack of valid tests for evaluating shoulder problems, in both diagnostic and evaluation research (and practice).
Would people use the test as diagnostic evaluation, like the methods are suggesting? Or would the test be more appropriately fitted in the evaluation of a new therapy, or should it be used to evaluate whether a patients is able to fulfil his/her duties at work? Or could it serve more than one purpose?

We agree that we have been unclear on whether this test is being used for evaluation or discrimination. The first sub test should be the easiest to perform for most individuals. We suspect that the difficulty of subtest two or three may depend on the nature of the underlying pathology. For example, patients with impingement are expected to have difficulty with subtest 2. However, patients with neck problems, who have difficulty looking overhead or patients with compression of the brachial plexus who have limited endurance may have more difficulty with the third task.
Each task can be completed for up to five minutes providing a broad spectrum of potential responsiveness to detect improvement. For these reasons, we believe that the test will be both discriminative and evaluative. Nevertheless, we have emphasized discrimination in this paper because we are evaluating concurrent validity. We have not considered the test diagnostic in terms of making specific diagnoses, but rather to use discrimination, to identify subgroups for decisions like: who might have difficulty returning to work, what specific functional tasks would be most difficult, etc.

Additional reference: Vd Windt 1996 showed that in GP patients only 50% recovered in 6 months.

Thank you, we have added this additional reference.

Second paragraph
I would suggest to use upper extremity rather than upper quarter, as the latter also includes the upper trunk region.

Agreed, we have changed this terminology.

Third paragraph
Is it necessary to explain the other study in such detail?

We have shortened this description.

Fifth paragraph
The objectives of the study do not logically follow from the first four paragraphs. As the use of the test will become more clear in the introduction it probably helps to understand the rationale of these objectives. Again it will depend on the use of the test. If it is likely that the test will be used to evaluate shoulder problems over time, why should one evaluate the test among healthy controls? Although evaluation of extreme groups (discriminant validity) could be a first step in the evaluation of test performance, it will not be sufficient. Probably all thought over by the authors, but not clearly explained.

Our purpose for evaluating normal healthy controls were to determine:
1. whether our five-minute target was reasonable, as indicated by the fact that the majority of controls should be able to perform it.
2. What is the difference between controls patients with sufficiently explained, to be discriminative

We agree that a variety of additional psychometric tests should be performed in the future crossing different clinical populations, different constructs and testing different psychometric hypotheses. We will proceed to conduct some of these and are hopeful that the publication of this test protocol will allow others to do so as well. We agree with the reviewer that the future emphasis should be placed more distinctly on how this test performs in different clinical populations -now that discrimination from normals has been established.

Methods
2d, 3rd, 4th paragraph
Several steps were taken to come to the three tasks. For readability it would help to start off with saying that. Followed by the explanation of what the steps included.

Agreed, revised as suggested.

5th paragraph
For readability it would help to start with the heading ‘description of the test’, followed by paragraph 6, 7 and 8 and finally describing the stopping rules (paragraph 5)

Agreed, revised and suggested
Participants
This heading leads to confusion.
There is a general problem with the case definition of impingement and moreover the validity of the Hawkins test. This is difficult to solve, but may need some reflection in the discussion. The same applies for the diagnosis mentioned in part 2 – validation.
How many controls were included, and what tests were used to exclude any shoulder pathology?

We agree that sensitivity and specificity of the Hawkins test or, in fact any other clinical tests that can be used to assess impingement is problematic when used in isolation to make a diagnosis about any shoulder condition. All of our patients recruited from the surgical wait list had a detailed physical examination performed by an experienced orthopedic surgeon specializing in shoulder problems as well as any appropriate imaging. These patients were examined once again on the day testing was performed. When presenting for their test session, the Hawkins test was performed to confirm the patients continue to have problems with impingement. However, in fact, we believe that the prior diagnostic process was sufficient to establish the presence of impingement. The subgroup of patients recruited from the University Physical therapy clinic had milder impingement or other shoulder problems and were all currently under treatment for these problems. All patients had a detailed assessment performed by physical therapist specializing in musculoskeletal conditions. This is now clearer in text.

Controls were recruited by convenience sampling. History, physical examination by a physical therapist which included self-reported pain and disability and the Hawkins test were used to confirm the volunteers self-reported absence of shoulder problems.

Analyses
Pearson was used, which assumes normal distribution. It would be more appropriate to use non-parametric tests regarding the small numbers of subjects included. The same applies for anova.

Our primary validation sample and the overall group was normally distributed, the isolated surgical sample group was too small to assure this is the case, but we preferred to use parametric statistics across all analyses as we wish to use means and averages and felt that the parametric tests were sufficiently robust. To be assured of this we cross ran correlation as Spearman rho and found similar results.

------------------------------------------------------------------------------------------------------------------------
Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)
------------------------------------------------------------------------------------------------------------------------
Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

Discretionary Revisions (which the author can choose to ignore)

What next?:
Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

Reviewer's report
Title: Validation of a new test that assesses functional performance of the upper quarter (FIT-HaNSA) in patients with shoulder pathology
Version: 1 Date: 6 November 2006
Reviewer: Sally Green
Reviewer's report:
General
This is an interesting body of work and the authors should be encouraged to continue and to publish more substantive results as they come available

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

The work is premature, perhaps too premature to be published, and I believe at the least the following revisions need to be considered.

It is our opinion that it is the correct time to publish this work and that an open access journal is the correct format. We believe the timing is perfect because we have undergone sufficient revisions of the test protocol to refine it to its final composition, and there is also sufficient data at this time to suggest strong psychometric properties. We believe open access is an important, vehicle for publishing this kind of study because it allows developers to retain copyright and the electronic format allows us to share the specific details of the test protocols so that others might readily adopt our standardized protocol.

We agree that our findings must be considered preliminary, but they also represent a substantial body of work. Rarely are new outcome measures or test methodologies ever published with comprehensive psychometric data because this entails testing the instrument in a variety of circumstances, populations and assessing different psychometric constructs. It was our goal to provide sufficient psychometric data at this time, to suggest that this test was ready to be adopted and evaluated in the spectrum of circumstances by others. As these additional studies are performed, they will more clearly define where this test is useful and where it is not. This is the evolution of many outcome instruments.
1. Background: The introduction states that these has been relatively little work developing standardized functional tests for the shoulder that may be used to assess patients over time, yet this paper does not address this need and is instead reporting a measure which, at this stage, can be used only to discriminate between those with shoulder pathology and those without. This needs to be more clearly laid out in the introduction and objectives.

We have clarified this section and agree that we have not been specific at this time whether this test should be used for discrimination or evaluation over time. Because of the nature of our test, which includes graded difficulty in the subtests and a five minute time frame for each subtests, we believe that the test will be useful for both purposes. However at this time, we have assessed concurrent validity, and so are unable to comment on responsiveness. This longitudinal work is being conducted as a master's thesis and may require additional two-three years to reach publication. We have clarified these distinct roles in the manuscript.

2. The background should also include more discussion on why this observed measure is needed in addition to the well researched self reported function scales for the shoulder (of which there are many, two of these have been used as construct validity references in this paper). The developed test is relatively complex compared to a questionnaire, and the rationale for why it is needed should be covered in the introduction, particularly as the paper demonstrates good correlation between the developed tool and the SPADI and the DASH. What is this tool adding over use of the SPADI and the DASH?

There is a substantial body of literature showing that self-report and performance-based measures provide different information. Some of these papers are now cited to make this point more clearly and to clarify a justification for the test to propose. When it comes to self-report measures to measure shoulder function, there is an overabundance. However, there is a substantial gap in similar measures that can be used to assess functional performance. Our correlation suggests that both self-report and are performance-based tests have sufficient concordance to substantiate validity. However, they are not interchangeable.

3. The method of tool development is interesting and I would like to see more detail about this in the paper.

We have added some additional detail about tool development.

4. It is not surprising that those with shoulder pathology have less endurance than those with shoulder pathology. The greatest omission from this paper is assessment of reliability. Until the test is demonstrated to be consistent on repetition it will be of limited use as a discriminatory measure and cannot be used as an evaluative measure. This needs to be much more explicitly discussed.
We agree that reliability is an important aspect of establishing the psychometric properties of any outcome measure or assessment procedure. We also agree that this was an omission from our paper, and prior to implementing a test users would want some level of confidence that measures are stable. We would actually expect a performance-based measure that assesses endurance to have some inherent variability as endurance-based tests have typically less reliability than physical parameters like maximal strength. To respond to this concern we have tested a small sample of subjects and found high reliability. We include subjects similar to those evaluated for validity testing. The sample while small does verify that the test can provide reliable results, and that the validity estimates made in this particular paper have internal validity. We do not think our sample size is sufficient to provide strong statements at this time on reliability and have noted the small sample size in our discussion. Reliability testing in different populations is needed and under investigation in a larger sample.

5. The first sentence of the discussion states that this study provides preliminary support for the use of a new functional performance test... I am not convinced that this statement is backed up by the data as the sample size is small, not all results are significant, there are many untested assumptions (as stated by the authors) and there is no report of reliability.

We have addressed the reliability issue (ICCs>0.95). While the sample size is small in some cases many significant results have been obtained and confidence intervals indicate sufficient precision to substantiate our stated conclusions. We agree that additional testing is warranted. In particular, we think it is important that independent authors evaluate this test under different circumstances and in different clinical populations. We ourselves will be performing some additional reliability and validity testing in larger samples and more homogeneous populations to provide the detailed information needed for clinical decision-making. However, we think it is incredibly important that independent validation also be conducted, once a protocol has been sufficiently developed and psychometric testing indicates a viable test protocol. Application in an open access journal with details of the test protocol is an important step towards that aim. Furthermore, because of the current gap and availability of a similar test, we have demands to use this test in pending clinical studies (our own and others). It is important that the information be communicated in an open format before widespread implementation proceeds.

-------------------------------------------------------------------------------

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

Discretionary Revisions (which the author can choose to ignore)

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions
Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable
Statistical review: No
Declaration of competing interests: I declare that I have no competing interests