Author's response to reviews

Title: Inter-rater reliability of the evaluation of muscular chains associated with posture alterations in scoliosis

Authors:

Carole Fortin (carole.fortin@umontreal.ca)
Debbie E Feldman (debbie.feldman@umontreal.ca)
Clarice Tanaka (cltanaka@uol.com.br)
Michelle Houde (mhoude@santepub-mtl.qc.ca)
Hubert Labelle (hubert.labelle@recherche-ste-justine.qc.ca)

Version: 2 Date: 25 February 2012

Author's response to reviews: see over
February 2012

Doctor Timothy Shipley
In-house Editor
BMC Musculoskeletal Disorders

Dear Doctor Shipley,

We have received the Reviewers’ comments for our manuscript entitled “Inter-rater reliability and validity of the evaluation of muscular chains associated with posture asymmetries”, reference number: MS: 6768700226277520. We have provided answers to the Reviewers’ comments and completed the corrections. The corrections in the manuscript are written in boldface font. We are grateful to the Reviewers for their interest and their thoughtful comments which enhance the quality of our manuscript. Please note that we have modified the title of our manuscript according to the Reviewers’ comments. The new title is “Inter-rater reliability of the evaluation of muscular chains associated with posture alterations in scoliosis”.

We have enclosed the revised version of our manuscript and our answers to Reviewers in a table format.

Best regards,

Carole Fortin, Ph.D., P.T.
Corresponding author

Université de Montréal and
Research Centre of
Sainte-Justine University Hospital Center
3175, Côe-Sainte-Catherine
Montreal (Quebec), Canada, H3T 1C5
Phone: (514) 345-4931, ext : 3277
FAX: (514) 345-4755
E-mail: carole.fortin@umontreal.ca
<table>
<thead>
<tr>
<th>Reviewers Comments</th>
<th>Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reviewer 1</strong></td>
<td><strong>Answers</strong></td>
</tr>
<tr>
<td>Major Compulsory Revisions: You have tested inter-rater reliability not criterion validity as you suggest in the paper. Please re-vise the paper to reflect this.</td>
<td>We have replaced <em>validity</em> by <em>agreement with experts</em> throughout the paper.</td>
</tr>
<tr>
<td><strong>Reviewer 2</strong></td>
<td><strong>Answers</strong></td>
</tr>
<tr>
<td>Of primary concern is the issue of validity.</td>
<td>The term <em>validity</em> has been replaced by <em>agreement with experts</em> throughout the paper.</td>
</tr>
<tr>
<td>The case for the existence of the muscle chain impairment is not clear in the introduction.</td>
<td>We have modified the <em>Introduction</em> accordingly. Please see our answer under point number 2 of Major Compulsory Revisions.</td>
</tr>
<tr>
<td>Of particular interest is the assumption that posture patterns are related to hypertonicity and not hypotonicity.</td>
<td>We agree that hypotonicity can also impaired posture. The causes of posture alterations are different in these cases. According to Souchard, the pathophysiology of tonic muscles is hypertonicity (muscle retraction) whereas pathophysiology of phasic muscles is hypotonicity (weakness).</td>
</tr>
<tr>
<td>Agreement with a consensus of two experts does not ensure the validity of identifying the proper muscle chain impairment, unless the skill of the experts was first validated by a real gold standard.</td>
<td>We modified the third objective as follow: 3) to verify agreement of posture and muscular chain evaluations observed by physical therapists against two experts instructors in global postural re-education.</td>
</tr>
<tr>
<td>Evidence of clinical relevance of the muscle chain impairments would go a long way in supporting the relevance of this study.</td>
<td>There are almost no methodologically sound studies that assess outcomes for GPR. GPR is used a lot for treatment of postural problems, especially in Europe, Brazil and Canada. We sought to first evaluate the inter-rater reliability of the GPR assessment as a first step. A future study will evaluate the GPR method compared to conventional treatments.</td>
</tr>
</tbody>
</table>
In addition, the method for computing the validity statistic is not included, both the intraclass correlation coefficient used and how it was applied to the expert and clinician data.

Reliability and validity are only meaningful when applied to clinically relevant entities. The primary reliability investigation is that for the “muscle chain impairments.” The text says that the reliability of “muscle chain impairments” was computed only for cases of agreement on posture asymmetry. It is not clear how this was implemented, and all ratings should have been included in the analysis at any rate (exception noted below for indeterminate categorization).

The small sample may also make kappa ineffective in some cases; there could be a problem with homogeneity (restriction of range across categories of categories (underrepresented categories) leading to kappa instability. This may explanation of how PTs appear to agree with experts but not with themselves.

**Major Compulsory Revisions:**

1. Patient selection: Identify the study population more

Please see the answer of point 9 under Major Compulsory Revisions.

We added the following sentences in **Data analysis** of the **Methods** section to better explain this point: “Among those who identified the same posture alteration, we determined whether there was agreement on the muscular evaluation. For example, among those who recorded knee valgus as an alteration, we assessed the inter-rater reliability of the muscular chain assignment associated with knee valgus.”

We agree that the sample of youths was small but the goal of this study was to assess inter-rater reliability of a large group of PTs. As mentioned in the discussion (**Study Limitation** section), the choice to include a large number of categories affects negatively the Kappa coefficient. The agreement with the experts had only two categories (agree vs disagree) which contribute to higher coefficient. We added the following sentences under **Study Limitation** in **Discussion** section: “The small sample of youths included in this study is also a limitation since Kappa coefficients are more favourably influenced by sample size magnitude than by large number of raters [28]. Moreover, some posture alterations were not present among these youths and thus muscular chain impairment could not be determined.”

The following information have been added under **Participants** in the **Methods** section: “Youths with idiopathic scoliosis were...
completely. How were the five patients selected besides having scoliosis?

chosen because they typically display posture asymmetries [18,23]. These youths (4 females, one male) were recruited from a previous study on posture assessment performed at the Sainte-Justine University Hospital Center in Montreal. We selected youths with different morphological characteristics (anterior, posterior or mixed) with clear photographs.”

2. Introduction: Make a case for the validity of the muscle chain impairment construct. Include evidence for the existence of muscle chain impairment and association of posture patterns with muscle dysfunction patterns. Otherwise, there is no clear point for the current study.

We modified the second paragraph of the Introduction section: “Souchard [6] describes muscles as being organized into two main static postural chains: the anterior and posterior muscular chains. Muscular chains are an ensemble of muscles defined according to their localization as well as their functional role which can explain posture alterations and movement dysfunctions [6,10-12]. Specific posture patterns caused by muscle chain retractions have been associated with lower back or neck pain among elite athletes in muscular power competitions [12] and functional disabilities in an adult with hemiparesis [13]. Despite the lack of studies linking muscular chain impairments to abnormal posture patterns and dysfunction, it seems that global muscular chain stretching is more effective than analytic muscle stretching to improve function and quality of life for several pathologic conditions including respiratory, musculoskeletal and neurological problems [7-9,13,14,15].”

3. Expert agreement needs to be elaborated under methods, since there are two methods of measurement that are confounded with rater. This is also a limitation of the study.

We added the following sentence under Procedure of Methods section to clarify this point: “After consensus, agreement between PTs and experts was calculated.”

4. Clinical relevance: The primary reliability analysis should be identified as that for the muscle chain impairment.

We did the modification in the Results section and we modified the Tables as well.
<table>
<thead>
<tr>
<th></th>
<th>Posture assessments are an intermediate step and reliability of any intermediate step is secondary.</th>
<th>We have replaced the term validity by agreement throughout the text. We also added the following sentences under Participants in the Methods section to make clear that no true gold standard was used: “Two physical therapists instructors in GPR served as experts for determining muscular chain impairments associated with posture alterations, in the absence of any objective “gold standard” criterion for this assessment. Muscular chain impairments were determined by the two experts according to standards taught in GPR which can be found in GPR literature [6,7,13,14].”</th>
</tr>
</thead>
<tbody>
<tr>
<td>5. Validity: The term implies agreement with a clinical entity identified by a gold standard. This study is looking more at agreement with expert performance. Do not use the term validity in the title or text, and make it clear that no true gold standard was used.</td>
<td>We agree with the reviewer that addressing this point separately will be clearer for the reader. We added a specific section entitled “Muscular chain impairment associated with posture alterations”. We also modify this paragraph p.11-12: “The muscular chain could be determined for the majority of posture alterations (Table 1, fourth column). However, both the physical therapists and the two experts did not attribute specific muscular chain impairment for head lateral bending, head rotation and knee flexion. No muscular chain impairment has been identified for elevated shoulder and adducted scapulae since no such asymmetries were reported in the five cases that were evaluated and the term “no answer” is thus written in the table under these posture indices.” We also added under Study Limitations in Discussion section, the following sentence: “Moreover, some posture alterations were not present among these youths and thus muscular chain impairment could not be determined.”</td>
<td></td>
</tr>
<tr>
<td>6. The “no answer” category: The problem with this category is that if it predominates, then you can have a high reliability/validity index that merely indicates muscle chain impairment is indeterminate. This must clearly be explained, since you have highly positive findings that may support that lack of clinical usefulness of the index. This needs to be addressed separately from the reliability/validity of muscle chain impairment categorization.</td>
<td>The global mean score have been removed.</td>
<td></td>
</tr>
<tr>
<td>7. No known meaning of the global mean score is given. This</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ad hoc index should be excluded.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>8. Kappa:</strong> Identify the kappa used more clearly. It is a generalized kappa for multiple raters and multiple categories.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>We better identified the Kappa used in our study under <em>Data analysis</em> of the <em>Methods</em> section. Please see the following sentences: “<em>We used Fleiss’ Kappa coefficients (for categorical data) and percentage of agreement to assess inter-rater reliability of muscular chain evaluation and associated posture alterations (objective 1).”</em> and “<em>Fleiss’s Kappa is used to measure the overall agreement between several raters and is adapted for nominal scales with multiple categories [28-30].”</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The reliability derivation in the appendix is unnecessary. Those who are interested will have the reference or can go to the cited online site. Make it clear that the Wikipedia is presenting the Fleiss kappa.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>This appendix has been removed. The reference is written as follow under <em>References</em> section: “30. Fleiss’ Kappa Equations. Available from: <a href="http://en.wikipedia.org/wiki/Fleiss'_kappa.%E2%80%9D">http://en.wikipedia.org/wiki/Fleiss'_kappa.”</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>9. Intraclass Correlation Coefficient:</strong> The ICC used is not identified. There are many of them with different applications. It is not clear if categorical data were used or some index vaguely alluded to in the text. It is not unknown how the average expert rating and PT ratings were included in the ICC used. An ICC is often a wrong choice for a validity statistic; commonly used ICCs are measures of agreement among equals (equal error variances assumed for all raters). This does not apply to experts assumed to have superior skill to the study population of PTs. Look for appropriate indices that allow for error-free measures such as likelihood ratios at Pearson’s r (through regression analysis which assumes the independent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>We used ICC because it is considered similar to coefficient of concordance for assessing agreement with experts and also because ICC is well applicable in a reliability study using a two categorical data model (agree or disagree) for a large number of raters. We added the following explanation under <em>Data analysis</em> in <em>Methods</em> section (paragraph 2): “For our third objective, we examined agreement between PTs and experts regarding muscular chain evaluation and their associated posture alterations, using intra-class correlation coefficients (ICC$_{3,k}$) for categorical data (agree or disagree). PTs’ answers were recoded as agree versus disagree with experts’ answers (after consensus) and were averaged for each possible choice of posture indices of the grid (example: right knee flexum – R, left knee flexum – L, R &gt; L or R &lt; L; see Appendix).”</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
variable is error-free in its computation).

10. Table 2. You need to explain how the PTs can agree with experts well but not with each other. If raters do not agree with each other well, then some cannot agree with any standard such as experts. This could be a problem with the choice of statistics or how it is applied.

11. The Discussion seems to stray a bit from reliability/validity to treatment strategy of the GPR technique. The discussion should focus on the level of reliability/validity and the relevance to practice.

12. Limitation: The study looked at the reliability of reading a photograph. It did not take into consideration positioning error. This would have required a second photograph.

13. Limitation: There is no small sample standard error for a generalized kappa to my knowledge that can be used to perform a t-test. There is no standard error for constructing a confidence interval.

14. Conclusion: It is not clear that reliability is greater with experience. No significance testing was conducted comparing independent kappa values between the three experience subgroups. It is apparent that experts are more reliable than non-experts for posture evaluation.

We modified the Discussion accordingly.

We agree that positioning can affect the perspective and yield to errors. However, the photographs of all youths were taken in the same position and distance from the camera.

We agree with the reviewer. We performed a chi-squared test on posture indices with Kappa coefficient ≥ 0.40 to determine if there was a significant difference between the 3 groups. Please see Data analysis in Methods section.

As mentioned in point 13, we added a chi-squared test to verify this second objective and found that Group 3 has a significantly higher level of reliability than the other groups ($\chi^2, p=0.005$). This sentence has been added in Results section.

Minor Essential Revisions:
1. None.
2. Discretionary Revisions:

1. Title: Add in scoliosis. The generalizability is limited to this population.

| Authors | We have replaced the term *posture asymmetry (ies)*

We have removed the following sentences under *Description of the posture analysis grid* in Methods section because the information can be found in the Appendix.

|  | By *posture alteration (s)*

We removed:

*Posture indices evaluated for the head and neck region are: head protraction, head lateral bending, head rotation and cervical lordosis (normal, hypo or hyper). For the shoulder and scapula regions, the posture indices are shoulder elevation, shoulder protraction, rounded shoulder, and adducted or abducted scapula. The thoracic kyphosis and lumbar lordosis posture indices (normal, hypo or hyper) were examined respectively for the thoracic and lumbar regions. The pelvic frontal tilt and left and right sagittal pelvic tilts were assessed at the pelvis level. For the lower limbs, the posture indices at the knee are valgus, varus, recurvatum or flexum and at the feet are decreased or increase plantar arch and valgus or varus of the rearfoot.*

And we added:

*(see Appendix for more details).*