Author's response to reviews

Title: Test-Retest Reliability and Sensitivity of the 20-Meter Walk Test among Patients with Knee Osteoarthritis

Authors:

Jillian M Motyl (jmotyl@tuftsmedicalcenter.org)
Jeffrey B Driban (jdriban@tuftsmedicalcenter.org)
Erica McAdams (erica.mcadams@gmail.com)
Lori Lyn Price (lprice1@tuftsmedicalcenter.org)
Timothy E McAlindon (tmcalindon@tuftsmedicalcenter.org)

Version: 2 Date: 8 March 2013

Author's response to reviews: see over
Enclosed is a revised copy of a full-length research article manuscript for *BMC Musculoskeletal Disorders* entitled “Test-Retest Reliability and Sensitivity of the 20-Meter Walk Test among Patients with Knee Osteoarthritis.” This manuscript was prepared by Jeffery Driban, Erica McAdams, Lori Lyn Price, Timothy McAlindon, and myself. All of the authors meet the Uniform requirements for Manuscripts Submitted to Biomedical Journal criteria for authorship. This manuscript has not been submitted and will not be simultaneously submitted to another journal. This study was supported by the NIH/NIAMS (grant RO1AR057802-01).

We have revised the manuscript based on the reviewers’ comments and we are submitting a revised manuscript and a response to the reviewers’ comments that describes how the manuscript was updated (attached on subsequent pages).

I will be serving as the corresponding author on this study. My contact information is below:

**Corresponding Author:** Jillian Motyl  
Division of Rheumatology  
Tufts Medical Center  
800 Washington Street, Box #406  
Boston, MA 02111  
Email: jmotyl@tuftsmedicalcenter.org  
Phone: 617-636-6394  
Fax: 617-636-1542

Thank you in advance for your time and commitment in reviewing this manuscript.

Sincerely,

Jillian Motyl  
Division of Rheumatology  
Tufts Medical Center
Author’s response to reviews

Title: Test-Retest Reliability and Sensitivity of the 20-Meter Walk Test among Patients with Knee Osteoarthritis

Authors:

Jillian Motyl (jmotyl@tuftsmedicalcenter.org)
Jeffrey B Driban (jdriban@tuftsmedicalcenter.org)
Erica McAdams (erica.mcadams@gmail.com)
Lori Lyn Price (lprice1@tuftsmedicalcenter.org)
Timothy E McAlindon (tmcalindon@tuftsmedicalcenter.org)

Version: 2 Date: 8 March, 2013

Author’s response to reviews: see over
Editor's Request:

1.) Please update your ethics statement to include the name of the ethics committee that approved your study.
   Name added to Methods (pg. 6, line 112)

2.) Please add TRN for original study to the methods section.
   TRN added to Methods (pg. 5, line 104)

Reviewer 1 Comments:

Abstract:
1) For reliability the authors used ICC. Please provide this information in the methods of the abstract.
   We have clarified in the methods of the abstract what analyses we performed.
   Based on the reviewer's comments regarding the assumptions of calculating ICCs and limits of agreement we have removed the ICCs and focused primarily on Spearman rank correlations, Wilcoxon sign rank sum test, and empiric 95% confidence intervals of the median (derived from bootstrapping).

Introduction:
2) What was the rationale for choosing the 20 meter walking test?
   From a practical perspective it would be more appropriate to choose a 10-meter walking test since less space and time is required. Are there studies that recommend 20 m walk test over 10 m walk test?
   We have clarified the rationale for the 20-meter walk test in the introduction. (page 4)

Methods:
3) Were other sources of variation/error accounted for? E.g. was the time of the day kept equal between day 1 and day 2?
   Other sources of variation were not accounted for (e.g. time of day). This is a common problem in clinical trials among individuals with knee osteoarthritis and we felt this potential source of variation/error should be permitted. We have noted in the discussion that future research may be needed to determine the contribution of different sources of variation/error.

4) Statistical analyses: it would be more clear if the terms reliability and sensitivity (and how they are defined and determined) are headings within the statistical analyses. Please keep consequent terms with regard to agreement and sensitivity.
   Reliability and sensitivity have been added to the headings of the statistical analysis section.
Results:
5) Bland-Altman plots: how did the authors judge the plots? Is there a cut-off point for lack of agreement?
Based on the reviewer comments we removed the Bland-Altman analyses and focused on the empiric 95% confidence interval of the median (derived from bootstrapping).

6) ICC should be given with Confidence Intervals
Based on the reviewer comments we removed ICC. We opted to use Spearman Correlations and confidence intervals have been provided.

7) Figure 3: please provide units
Figure 3A and 3B were removed from manuscript.

Discussion:
8) Please also discuss the sensitivity in relation to the Minimal Important Difference (MID).
While this is an interesting research question our study did not involve the Minimal Important Difference (MID) and this has yet to be determined in prior studies.

9) In the conclusion the authors state that one rater is recommended however in the present study they did not investigate inter-rater reliability. Therefore this statement is hypothetical and not empirical.
The authors agree and have removed this from the conclusion.

Reviewer 2 Major Compulsory Revisions (MCR)

MCR 1: Background, second paragraph. This paragraph needs to be re-written with a common thread leading to a clear-cut purpose. In its present form it seems a bit incoherent, particularly the first sentences.
This paragraph has been revised and the author’s feel these changes improve the clarity and purpose of the study. (pg 4, line 75)

MCR 2: Background, third paragraph. You state “test administration in OA studies show poor standardization”. This statement should be supported by references or changed.
This statement has been removed.

MCR 3: Background, third paragraph. You state “the reliability and sensitivity of the test remains unclear”. However reliability of 20-meter walk test in OA patients has been investigated in a previous study by Villadsen et al. 2012, http://www.ncbi.nlm.nih.gov/pubmed/22311054. It is important that you are clear on the additional contribution of your study compared to the previous study.
Villadsen et. al. 2012 explored the test re-test reliability of the 20-meter walk in OA patients, but the authors do not feel their sample is representative of the OA
population because they included only those with end-stage OA awaiting knee replacements. End-stage OA patients are a unique sample that is not typically included in studies evaluating symptoms and disease modification. Our aim is to instruct researchers and clinicians on the reliability of the 20-meter walk test in a OA patient population with mild to moderate disease. We have clarified this aim in the purpose statement and reflected on Villadsen et al in our discussion. (pg 14, line 283)

MCR 4: Report your study according to the “Guidelines for Reporting Reliability and Agreement Studies (GRRAS)” and state this in the manuscript. This will ensure homogeneity and completeness in studies on reliability and agreement and potentially heighten the quality of your manuscript. **We have reviewed the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) and have reported the results according to these guidelines and stated so in the manuscript.**

MCR 5: Methods, 20-Meter Walk (Self-Selected Pace). Please comment on, why the time between visit 1 and visit 2 was not standardized and also comment on possible implications for the study results. **The study visits had already been scheduled for the main study, thus the variability. The time between Day 1 and Day 2 was based on a 2 week follow-up with a one week window on either side (7 to 21 days). We have noted in the discussion that it would be interesting for future research to determine potential sources of variation (e.g., time between visits) in walking times and how much they contribute to the variation.**

MCR 6: Statistical Analyses, Comparisons and Agreement between Trials. You write that “Walk times and differences between trials were not normally distributed”. As far as I know normal distribution is an (at least indirect) assumption of both ICC and LOA. Please check this with an experienced statistician. **We agree that ICC and LOA assume a normal distribution (or at least a normal distribution of the residuals). Based on our data we have opted to focus on nonparametric methods (e.g., Spearman Rank Correlation Coefficients, Wilcoxon Signed Rank Tests, and empiric 95% CI of the median with bootstrapping).**

MCR 7: Results section and Figure 2. The mix between mean and median makes the interpretation a bit confusing, please revise. Referring also to MC6, I would advise you to consult a statistician (if you have not already done so). **In consultation with our experienced statistician we opted to remove the mean from the results to eliminate confusion.**

MCR 8: Discussion. It is important that you discuss and compare your findings with the previous findings by Villadsen et al. **Discussion of our findings in comparison to Villadsen et. al was added to Discussion.** (pg 14, line 283)
MCR 9: Discussion. In paragraph one and two several repetitions exists (e.g. practice trials, significant impact on the results). I would advise you to shorten the summary of your results in the first paragraph of the discussion and then discuss in the following paragraphs instead.

We have revised the discussion to reduce repetition.

MCR 10: Discussion, third paragraph. Please elaborate this to make it more understandable, why there may be a difference between the mean- and the median-based approach.

After reading the reviewers comments and consulting with an experienced statistician we removed the mean-based approached and reported solely the median.

MCR 11: Discussion, fifth paragraph. You state that "While we had only 15 participants they are representative of patients with knee OA." This statement must be supported in some way.

We revised to specify that they are “representative of patients with knee OA who participate in clinical trials.” Sources include clinical trials with similar participant characteristics.

MCR 12: Discussion/Conclusions. In the abstract you give specific numbers of SDD. The discussion and/or conclusions would benefit from actual numbers translated and discussed. This would help both clinicians and researchers. In its current state the discussion is not strong enough.

We have added the actual numbers to the discussion and conclusion to clarify the implications of these findings to researchers and clinicians. (pg 15, line 258)

MCR 13: Conclusions. You conclude that "We advocate that one assessor, or as few assessors as possible, should administer this standardized protocol to all 299 study participants." Since this is a study of intratester reliability you cannot conclude on the number of assessor.

We agree and have removed this from the conclusion.

Minor Essential Revisions (MER)
MER 1: Title and the whole manuscript. Your study is on intratester (and not intertester) reliability. Please state this clearly throughout the manuscript.

The authors agree and have removed this from the conclusion. We have also changed the title to reflect this.

MER 2: Methods, participants. Assuming you refer to the criteria found here: http://www.rheumatology.org/practice/clinical/classification/oaknee.asp

You need to be more clear on which of the three (clinical and laboratory, clinical and radiographic or clinical) the participants met.

Clarification added. Participants met clinical and radiographic criteria for osteoarthritis. (pg 5, line 105)
MER 3: Methods, participants. Do you have a reference for your definition of knee synovitis? If so refer to it.

In those participants meeting the ACR criteria for knee osteoarthritis, we included only those with synovitis (as defined by synovial pouch depth of >2mm at the suprapatellar recess). References added. (pg 5, line 107)

MER 4: Methods, participants. Why did you choose a WOMAC pain subscore of over or equal to 2?

This inclusion criterion was used based on the author's previous research experience to eliminate a floor effect [McAlindon et al. Effect of vitamin D supplementation on progression of knee pain and cartilage volume loss in patients with symptomatic osteoarthritis: a randomized controlled trial. JAMA. 2013 Jan 9;309(2):155-62.]

MER 5: Methods, participants. Please state the number of participants in the study in this section.

Revised to include number of participants. (pg 5, line 101)

MER 6: Methods, 20-Meter Walk (Self-Selected Pace). I would advise you to change “8-inch” to “approx. 20cm” since you are using the metric system for the 20-meter walk test.

Revised from “8-inch” to “20 cm” (pg 6, line 127)

MER 7: Results, first paragraph. You refer to “figure 1”. You must mean “figure 2”?

Correct, it should read Table 2. Revised. (pg 10, line 199)

MER 8: Figure 2. In the figure legend, you write “Lines under each set of trials indicate…” I cannot see the line. Please make them more obvious.

Figure 2 has been edited to make the lines more obvious.

MER 9: Results, Comparisons and agreement between Trials. You refer to “table 1”. You must mean “table 2”?

Correct, it should read Table 2. Revised. (pg 10, line 205)

Reviewer 2 Minor issues not for publication (MIP)

MIP 1: Abstract, Methods. "elected" should be "were elected"

Revised. (pg 2, line 47)

MIP 2: Background, second paragraph. “may be” could be changed to “can be” in the first line.

Revised. (pg 4, line 76)

MIP 3: Background, third paragraph. The phrase “clinical studies and rehabilitation clinics” have already been used in the first paragraph of the
background. Consider changing to another phrase.

Revised.

MIP 4: Discussion, first paragraph. In line six you write “Increased session 1 walking times”. Perhaps you should another word than increased since it was the first measurement and therefore can not increase.

**Revised. “Increased” changed to “greater”.**

MIP 5: Discussion, fifth paragraph. You write “Another limitation of or study”. However, the first limitation, though present, is not clear. Consider making it obvious what the first limitation is.

**Discussion revised to clarify the first limitation of the study.**

Reviewer 2 Discretionary Revisions (DR)

DR 1: Abstract, Background. You state “and therefore limit our ability to evaluate changes in gait speed”. To further emphasize the purpose of your study, you could change this to “and therefore limit our ability to evaluate real changes in gait speed not attributable to normal variability”.

**Revised.**

DR 2: Methods, 20-Meter Walk (Self-Selected Pace). You could report the number of steps during the 20-meter walk, since this is also a part of the test.

**Number of steps was not recorded in this study and not included in the proposed protocol.**

Reviewer 3 Major comments:

1) The sample size of the study is too small for precise estimation of 95% limits of agreement. The 95% confidence intervals for the limits of agreement can be calculated and these will be wide with a sample size of 15. Altman recommended a sample size of at least 50, but preferably rather larger, for a method comparison study (practical statistics for medical research, Page 402).

**We agree that the sample size may lead to a large estimate of the 95% limits of agreement. To calculate a more precise estimate of the SDD we have conducted bootstrapping to better estimate the distribution. Therefore, we report the SDD based on the median and empiric 95% confidence interval of the median from a bootstrapping sample.**

2) Page 3, line 54: "Changes in walk time between -2.59 seconds (walking slower) and 1.65 seconds (walking faster) should be considered within the range of normal variability of 20-meter walking speed." This is an overstatement: the estimates of the limits of agreement (i.e., -2.59,1.65) are imprecise due to small sample size of the study.

**We agree that the small sample size has an impact on precision. To obtain a better estimate of the distribution, we performed 200 bootstrapped samples and**
calculated the median difference for each sample. To be conservative, we used the percentile method to calculate a 95% empiric confidence interval by taking the 2.5% and 97.5% percentiles of the 200 bootstrapped median differences. This provides a wider confidence limit than that generated by using the regular 95% CI of the mean.

3) The SDD (usually defined as mean difference ± 1.96 times the SD of the differences) is computed based on 95% confidence interval (CI) for the median difference. This is a non-parametric alternative for 95% CI for the mean difference, but not for 95% limits of agreement. In fact, the 95% CI for the median difference converges to the median difference in the population as the sample size approaches infinity.

We have removed all references to limits of agreement, including the Bland-Altman plot and have focused on the bootstrapped sample results as described above.

4) The Bland-Altman plots for differences between sessions and between trials within sessions 3 and 4 are not presented in the manuscript.

Bland-Altman plots were removed from manuscript.

5) The ranges of agreement instead of 95% limits of agreement have been reported in Table 3. The range is not a satisfactory measure of variation, because it is determined by only two of the data points (which are the extreme values and perhaps outlier) and depends on sample size.

To calculate a more precise estimate of the SDD we have conducted bootstrapping. Therefore, we report the empiric 95% confidence interval of the median from a bootstrapping sample as described above.

6) Page 10, line 208: "There was a lack of agreement between session 1 and sessions 3 or 4 with no discernible pattern. In contrast, session 2 had agreement with sessions 3 or 4." Table 3 of the paper does not support the statement e.g., the range of agreement between sessions 1 and 3 and between sessions 2 and 3 are not markedly different.

We agree. We have removed the Bland-Altman plots from the results. Therefore, we are more focused on Table 3 in the results. This table shows a difference between sessions 1 and 2. Furthermore, qualitatively session 1 had lower correlations with sessions on the second day (r = 0.78) compared to session 2 (r = 0.94 to 0.95).

Reviewer 3 Minor comments:
1) Page 8, lines 145 and 160: P value should be replaced with alpha level.
Revised.

2) Page 9, line 184: The degree of freedom (df) for the t statistic is required.
Revised to include df = 14.
3) Page 10, lines 193 and 194: "Table 2" is correct. **Revised.**

4) Page 11, line 212: Why did the authors report p value for the Spearman correlation? **P-value removed.**

5) Page 21, Table 2: Please report 25th and 75th percentiles of walk time instead of minimum and maximum values. **Table 2 has been revised to include and 75th percentiles of walk time.**