Author's response to reviews

Title: Field assessment of balance in 10 to 14 year old children, reproducibility and validity of the Nintendo Wii board

Authors:

Lisbeth Runge Larsen (lrunge@health.sdu.dk)
Martin G Jørgensen (martin.joergensen@rn.dk)
Tina Junge (tjunge@health.sdu.dk)
Birgit Juul-Kristensen (bjuul-kristensen@health.sdu.dk)
Niels Wedderkopp (nwedderkopp@health.sdu.dk)

Version: 4
Date: 20 May 2014

Author's response to reviews: see over
To the Editor of BMC Pediatrics

Dear Editor

Odense, May 2014

We are pleased to have the opportunity to respond to the additional and useful comments raised by Arnold Huurnink regarding the article “Field assessment of balance in 10 to 14 year old children, reproducibility and validity of the Nintendo Wii board” by Lisbeth Runge Larsen, Martin G Jørgensen, Tina Junge, Birgit Juul-Kristensen and Niels Wedderkopp. The comments, suggestions and questions will clearly condensate and improve the presentation of the key points in the article.

We hereby address the comments giving a point-by-point response to the concerns as well as a revised manuscript. Revised sentences and paragraphs are highlighted in red.

Yours sincerely

Lisbeth Runge Larsen
PhD student, MPH, PT
University of Southern Denmark
Institute of Regional Health Research
Winsløwparken 19, 3rd, DK-5000 Odense C
E-mail: lrunge@health.sdu.dk
Reviewer’s report
Title: Field assessment of balance in 10 to 14 year old children, reproducibility and validity of the Nintendo Wii board
Version: 2 Date: 6 April 2014

Reviewer: Arnold Huurnink

Reviewer’s report:
The authors have done a thorough job to reply on my comments, and, to my opinion, the manuscript has been improved. However, there remain some issues that need to be addressed, but I think that the authors can come up with a revised manuscript that is suitable for publication in the Journal of BMC Pediatrics.

My main concern is that the limitation and implication of the current approach to assess validity of NWB is not always adequately addressed throughout the text. It may help to read the following papers, to better understand my point of criticism:

And I think it is essential to be aware of the criticism of Pagnacco et al. on the paper of Clark et al. (not that I agree with all of his statements...):

Secondly, the results section is unnecessary large, and redundant with Tables en Figures. Readability would be enhanced if only summarized key points of the results are presented in the text, in such a way that it becomes clear how the data should be interpreted.

Answer: Thank you for suggesting papers to improve our manuscript as well as revision of the results section. As you will see from the answers in the following questions, we have taken these suggestions into consideration in our manuscript. Bland and Altman in question 1 (minor essential revisions), Pagnacco in question 6 in the discussion section (major compulsory revisions), revision of the results in question 1 & 2 in the results section.
Minor Essential Revisions

Methods_statistical analyses_paragraph 2:

1) This paragraph is formulated a bit odd, as LOA is actually a quantification method for reproducibility and agreement, and does not visualize it. The Blant-Altman plots will visualize the data points relative to mean difference and LOA. Furthermore, the 95% LOA is not explained. Does the 95% LOA not simply mean that the difference between two sessions (reproducibility), or two methods (validity) will be within the 95% LOA for 95% of all subjects? Please rephrase this part of the paragraph and explain 95% LOA as well.

*Answer: We agree the LOA is not sufficiently explained in the former text.*

*Action: We have revised this paragraph to: “......the mean values from the two measurements are plotted on the mean differences from the two measurements, the plots should be centred around the line of zero difference. The 95% LOA presents the interval containing 95% of the plots and therefore visualizes the spread of the current measurements. Further, the Bland-Altman plots with the 95% LOA indicate systematic differences[1](new reference).*

2) “MDC is also related to limits of agreement, and only if the change in measure is outside the limits of agreement is the change is statistically significant” Please rephrase.

*Answer: We agree this paragraph could be presented more clearly*

*Action: We have rephrased the paragraph to: “MDC is also related to limits of agreement as a true change in measure is only statistically significant and not due to measurement error, if the change in measure is outside the 95% LOA”*

3) “CCC point estimates were interpreted as poor (0-0.39), modest (0.4 – 0.74), or excellent (0.75-1)” Why do the authors stick to these point estimates? The readability throughout the text will be considerably increased to simply state that CCC (or ICC) > 0,7 is considered sufficient or acceptable. In the discussion CCC values can still be compared with other studies (Clark et al), and the limitation of the point estimate is already discussed.
Answer: Thank you for this suggestion. The alterations in the manuscript have led to less importance of the varying cut-points of interpretation of the CCC.

Action: The paragraph has been rephrased to: “Interpretations of CCC or ICC point estimates are not yet agreed upon [2-5]. In the current study CCC point estimates ≥ 0.70 were interpreted as satisfactory.”

Major Compulsory Revisions:

Results Test-retest reproducibility of NWB and AMTI:

1) There exists redundancy between the text and figures/tables. It is sufficient to highlight the key points of the results as it would greatly enhance readability. For example, presenting both SEM and MDC in the text is unnecessary, as both represent absolute reproducibility (and are 100% correlated) and are both presented in the tables. In my opinion the important results of the reproducibility analyses can be summarized in approximately 5 sentences, for example:

- Out of the eight Bland-Altman plots only one reveals a systematic bias (5%) for the unilateral non-dominant eyes open test on the AMTI.
- For NWB, the MDC varied between 16.9 and 36.9 cm (26.5 – 28.6 % of mean outcome) in the four tests, which is comparable to the results of the AMTI with MDC values between 14.7 and 36.1 cm (26.3 - 28.1%).
- Additionally, all CCC values were above 0.7, ranging from 0.76 to 0.83 for NWB and from 0.79 to 0.86 for AMTI, with all 95% lower bounds being # 0.65. Although confidence intervals of the CCC values largely overlap between NWB and AMTI, the values for NWB are consistently slightly lower.

Answer: Thank you for the suggestions on revision of the results section. We agree it could be condensed. However, we do like the systematics in presenting one platform at a time for further comparison. However, we believe that we have met your suggestions by reducing the section, and summarizing the results in the last paragraph.

Action: The result section has been rephrased and shortened as:

Test-retest reproducibility of NWB and AMTI

Regarding the NWB, Bland-Altman plots of the average COPL (Figure 1) demonstrated no systematic bias. The line of observed agreement was approximately similar to the line of perfect agreement. The range of LOA was largest in the test for the dominant leg (Table 1).
For the NWB, the CCC was ≥ 0.70, ranging from 0.76 to 0.83 (Table 1). The MDC varied between 16.9 and 36.9 cm (26.5-28.6% of the mean COPL). The mean COPL difference was highest for the unilateral test on the non-dominant leg.

For the AMTI, Bland-Altman plots (Figure 2) demonstrated no systematic bias in three of the four tests. In the unilateral test on the non-dominant leg, however, the differences increased with larger values, and the observed agreement indicated longer COPL on retests. The range of LOA was largest in the test for the dominant leg.

For the AMTI, CCC values for COPL were ≥ 0.70 in all four tests, ranging from 0.79 to 0.86 (Table 1). MDC varied between 14.7 and 36.1 cm (26.3 - 28.1% of the mean COPL). The highest mean differences were seen in the unilateral tests.

In summary, among the eight Bland-Altman plots only one revealed a systematic bias (AMTI, unilateral test on the non-dominant leg), the CCC coefficients were slightly higher in AMTI, whereas MDC and LOA were comparable for the NWB and AMTI.

Results_concurrent validity:

2) “LOA showed larger variation in the unilateral tests than in the bilateral tests and the line of observed agreement indicated that the NWB gave longer measurements in bilateral tests, but shorter measurements in unilateral tests” This information is not really helpful. What the reader want to know is:

- Is there a systematic bias? (you address that correctly in sentence 1).
- Does the 95% LOA between NWB and AMTI differ from the reproducibility results (95% LOA or MDC)? If so, what is the extent of the difference, as this information will give the extent of measurement error due to NWB
- Does the concurrent validity CCC’s differ from the concurrent reproducibility CCC’s, as this information will give the extent of measurement error due to NWB. In the present measurement setup the 95% LOA between NWB and AMTI and the validity CCC’s per se are of little meaning, unless it is compared with the reproducibility analyses. Therefore it is also impossible to formulate point estimates for validity CCC’s on forehand, when the current validity design is used.

Please rephrase this results section in such a way that the abovementioned points are addressed.

Answer: We agree we should address this issue more thorough. We have decided that it should be briefly presented in the result section and in the discussion section we have given more
attention to this issue, to make a more conclusive paragraph (please see answer to question 6)

**Action:** We changed few wordings and added a paragraph to the result section: “Point estimates of concurrent validity were satisfactory (CCC=0.74-0.87) (Table 2). The mean difference was highest for the unilateral test on the non-dominant leg.

**Overall, both the 95% LOA and the CCC coefficients in the validity study were comparable to the results from the test-retest study.**

**Discussion:**

3) Please make it more clear in the first paragraph what the main findings of the paper are, related to the aims formulated in the introduction. Aren’t the main findings that (1) it is possible to measure sway of children in a field setting with sufficient reliability (>0.7), (2) that NWB and AMTI possess equal reproducibility of COPL in children (based on 95% LOA, MDC and CCC), and (3) that a possible measurement error of NWB compared to AMTI is small compared to intra-subject variability, as 95% LOA and CCC of NWB vs AMTI is comparable to 95% LOA and CCC of test-retest comparison.

**Answer:** Thank you for giving us the opportunity to resume the results more clearly.

**Action:** The following paragraph has been added: “The main findings of this study were that NWB is a reproducible and valid tool for measuring sway of children in a field setting, and that NWB and AMTI possess almost equal reproducibility of COPL in children (based on 95% LOA, MDC and CCC>0.70), the AMTI presenting a slight tendency of systematic bias in the reproducibility study. Furthermore, a possible measurement error in the validity of the NWB towards AMTI is small compared to the intra-subject variability, since 95% LOA and CCC of NWB when compared to the AMTI, is comparable to 95% LOA and CCC of the test-retest study.”

4) For comparison of your CCC and MDC values of children in bipedal stance with previous studies in adults, the following references may be used:


**Answer:** Thank you for your nice reference suggestions. We have incorporated a reference to
Ruhe et al. in the discussion section. However, due to differences in equipment, measurements and measurement units in Salaveti, it was very difficult to compare our study to this study, and therefore this reference was not incorporated. The fact, that only very few studies from the review by Ruhe have calculated MDC, a direct comparison to the current study is difficult. However, the reference is mentioned in the discussion section as an important study within this field, along with the other suggested study (please see answer to question 5) by Chang et al.

Action: We have added a paragraph in the discussion along with new references and slight changes in the existing wording: “As this was the first study to examine reproducibility and concurrent validity of the NWB in a population of children, comparisons of CCC estimates, MDC values and conclusions are made to studies of sway measures in adult populations. Comparisons of MDC values to previous studies are limited, as only few studies on COPL as a balance measure, reported MDC [6].

In line with previous studies [7-10], reproducibility and concurrent validity of the NWB were found to be satisfying. Bland-Altman plots illustrating the reproducibility of the NWB and AMTI showed almost similar COPL, confirmed by $\text{CCC} > 0.70$ ($\text{CCC} 0.76-0.86$). The MDC of NWB in percentage was relatively high (26-28%) in the current study, but in line with a previous study [8], and was similar to that of the AMTI. The relatively large LOA and MDC indicated large variation between trials, however, which questions the validity of the CCC. The importance of this variation in determining the appropriateness of using NWB and AMTI to measure sway is unknown, but needs examination in future studies as it might influence the usefulness of NWB and AMTI measures as predictor of injuries or risk of falls. The NWB was found sensitive enough to detect postural changes associated with subtle variations in visual tasks in elderly people [11], and despite the indication of systematic bias in one test of the AMTI, the AMTI has shown to be sensitive enough to predict injuries from sway measures [12].

5) Besides Clark et al., there is another study that has focused on reproducibility and validity of the NWB and AMTI (this should also be added in the introduction):


Answer: Thank you for giving us the opportunity to refer to this new study.

Action: We have added the reference in the introduction: “…comparable with sway measures obtained from laboratory force platforms (ICC of 0.77-0.89) [7-11] for…” and corrected the text
in the paragraph in the discussion: "... to make this field study in a child population comparable with the two other studies that evaluated both concurrent validity and reproducibility of the NWB [7, 8]."

6) In line with the abovementioned comments for the results section, the discussion should explicitly state that validity of NWB cannot be assessed directly, but that it can be estimated to compare LOA of NWB vs AMTI with the test-retest LOA. If these values are comparable, than the LOA of NWB vs AMTI is mostly due to intra-subject variability. Than it can be concluded that measurement error due to NWB is probably small compared to intra-subject variability.

Answer: Thank you for resuming this important issue and for suggesting an additional paragraph.
Action: We have added a paragraph and changed wording in some sentences in the discussion: “The validity of NWB is difficult to assess directly, but by comparing LOA from the validity study with LOA in the reproducibility study it is possible to have an indication of the size of measurement error of the NWB. As the LOA and CCC coefficients in the reproducibility study are comparable to the LOA and the coefficients in the validity study, the measurement error due the NWB is probably small compared to the intra-subject variability. CCC for COPL was satisfactory (CCC 0.74-0.86). Thus, if the variable of interest is COPL, the results for the NWB are comparable to those for the AMTI, confirming previous studies [7, 8].” ......... Bland-Altman plots of concurrent validity revealed that NWB seemed to produce longer COPL measures in bilateral tests and shorter COPL measures in the unilateral tests than the AMTI, indicating systematic bias. However, as the differences between NWB and AMTI were small and the CCC coefficients from the validity analysis were satisfactory, we consider this issue to be of minor importance. The use of NWB as a tool to measure sway, and the comparison of NWB with an AMTI platform has been debated by e.g. Pagnacco et al. [13] because of too much noise in the NWB measures, when it was compared to a platform manufactured by Pagnacco, and because the AMTI measures not only COPL but also three-dimensional measures as rambling and trembling. However, in the current study, since comparisons of the NWB with AMTI was only made to the COPL measures, and the noise of measurement primarily was found to be due to intra-subject variability, the indicated differences seem to be of minor importance. Overall the concurrent......”

7) It should be stated in the limitation section that the validity of the NWB was assessed without putting the NWB on top of the AMTI, therefore intra-subject variability did
obscure the validity methods. Although this limitation was to some extent compensated by taking the intra-subject variability into account, the current method is probably not able to detect small systematic errors or the assess the exact extent of the possible bias of NWB measurements.

Answer: Thank you for reminding us to take this issue into account.

Action: We have added a paragraph: “The validity study was performed with two single measures of each sway platform, and not by putting NWB on top of the AMTI as seen in a previous study by Huurnink et al.[10]. Although we tried to take intra-subject variability into account in the discussion of the results, the current method is probably not able to detect small systematic errors, and the possibility of an unknown sized bias of NWB measurements remains.”

Abstract.

- Please find a way to present actual (ranges of) values of 95% LOA or MDC, so it is clear for the reader that NWB vs AMTI is comparable to test-retest comparisons, instead of only presenting CCC values.

Answer: We agree the abstract could have a more precise wording.

Action: We have rephrased the result and conclusion paragraphs to: “Results: Bland-Altman plots supplemented by Minimum Detectable Change (MDC) and concordance correlation coefficient (CCC) demonstrated satisfactory reproducibility for both the NWB and the AMTI (MDC: 26.3-28.2%, CCC: 0.76-0.86) using Centre Of Pressure path Length as the measurement parameter. Bland-Altman plots demonstrated satisfactory concurrent validity between the NWB and the AMTI, supplemented by satisfactory CCC in all four tests (CCC: 0.74-0.87). The ranges of the limits of agreement in the validity study were comparable to the limits of agreement of the reproducibility study.

Conclusion: Both NWB and AMTI have satisfactory reproducibility for testing static balance in a population of children. Concurrent validity of NWB compared with AMTI was satisfactory. Furthermore, the results from the concurrent validity study were comparable to the reproducibility results of the NWB and the AMTI. Thus, NWB has the potential to replace the AMTI in field settings in studies including children. Future studies are needed to examine intra-subject variability and to test the predictive validity of NWB.

Level of interest: An article whose findings are important to those with closely related
research interests

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests

References: