Author’s response to reviews

Title: Comparison and validation of accelerometer wear time and non-wear time algorithms for assessing physical activity levels in children and adolescents

Authors:

Jérémy Vanhelst (jeremy.vanhelst@chru-lille.fr)
Florian Vidal (florian.vidal@etu.univ-lille2.fr)
Elodie Drumez (edrumezchr@gmail.com)
Laurent Béghin (laurent.beghin@chru-lille.fr)
Jean-Benoît Baudelet (jean-benoit.baudelet@chru-lille.fr)
Stéphanie Coopman (stephanie.coopman@chru-lille.fr)
Frédéric Gottrand (frederic.gottrand@chru-lille.fr)

Version: 2 Date: 28 Feb 2019

Author’s response to reviews:

Dear Editor,

Please find enclosed the revised version of our manuscript, “Comparison and validation of accelerometer wear time and non-wear time algorithms for assessing physical activity patterns in children and adolescents” (BMRM-D-18-00305R2) for resubmission to BMC Medical Research Methodology.

As requested by the reviewers, we have made the appropriate changes in the paper, which have been highlighted in red. Our responses to the reviewers’ comments are included in the following pages.

We thank the reviewers for their constructive remarks, which have helped us to improve the quality of this manuscript. We hope our manuscript will now be suitable for publication in BMC Medical Research Methodology.

Reviewer reports:
Eivind Aadland, PhD (Reviewer 1): I thank the authors for generally providing good answers and making appropriate revisions to my comments. Still, there are some more issues that need to be considered.

1) In line with my previous comment, the authors have highlighted that they extend the findings by Aadland et al (2018) as the diary allows for determining the exact agreement for each period of wear/non-wear page 14 line 7- (not only total wear). However, these analyses are not reported. Compared to the study by Aadland, conducted in a more than 10 times larger sample, this is the only (possibly) novel finding. The exact agreement must be analyzed using the Kappa statistic for every wear/non-wear period as reported in the diary. Please include.

To acknowledge the reviewer’s comment, we modified the presentation of table about the number of non-wear periods as done in Aadland et al. BMC Public Health (2018) 18:323. We now reported the weighted Cohen's kappa.

2) Please use the same scale of all Bland Altman plots to ease comparison between the methods.

As requested, the same scale of all Bland Altman plots has been used.

3) Page 14 line 19-: It is stated that only one previous study, in adults, have compared the effect of the non-wear algorithm on MVPA and the achievement of the PA recommendations (ref 17). As Aadland et al found no difference in MVPA between algorithms, the number of children that achieved the guideline amount of MVPA is obviously similar. Thus, please revise and include Aadland et al's findings in this paragraph.

As requested, this part has been revised.

4) Please refer to the Supplementary material in the text page 9 line 31-.

Done.

5) Agreement and concordance is used more or less interchangeably, and the concordance correlation is used as a measure of agreement, for example in the abstract. Please be aware that these terms, agreement vs reliability measures, (should) have specific meaning. I suggest using these terms in line with the paper by de Vet et al, Journal of Clinical Epidemiology, 59, 2006.

We agree that the term of agreement is more appropriate. However, we could not change the term “concordance correlation coefficient” since it’s the statistical measure name (see King T.S Statist. Med. 2007; 26-3095-3113. “The concordance correlation coefficient is commonly used to assess agreement between two raters or two methods of measuring a response when data are measured as continuous scale”).
6) The number of non-wear periods, now included in the manuscript, have not been discussed. While this study found a maximum of 3 non-wear periods for the 30 min non-wear algorithm, previous studies have found a maximum number of 5-7 periods for this algorithm (Chinapaw, Toftager, and Aadland), which supports the choice of a longer non-wear criteria. Please discuss. Is the difference a result of a smaller sample (compared to the previous studies having around 1000 participants), thus, restricting the information?

As suggested, this finding has been discussed.

Paul McCrorie, Ph.D (Reviewer 2): Overall, the authors have improved the manuscript in certain places, e.g. in response to reviewer 1’s comments re: agreement visualisation and Bland Altman plots. There are also some improvements in the clarity throughout (particularly methods), however, I am still left unconvinced that the authors have demonstrated (in their narrative or analyses) their contribution over and above the newly cited papers - and therefore how this paper advances our knowledge.

The included section within the background section does make reference to the other papers that have been published in the field, however what this paragraph lacks is a genuine critique of why the current study provides new insight: reviewer 1 explicitly identifies the ‘one superior point' that the current study has in its favour yet this is still poorly conveyed in the background section and - from interpreting the comment from reviewer 1 - hasn't been included as a separate analysis that specifically explores the intricacies and exact agreement between each period and not just total wear time. It is this point that should be the focus of the 'adequate advance in knowledge' mentioned by the editor. I think the authors try to make reference to it in the first of the two sentences on page 6, lines 36-44 (as I can see it) but it isn't completely clear what is meant. Either way, I believe this is still under developed and needs to be strengthened.

page 6, lines 36-44 (as I can see it):

"No specific time points for wear and non-wear were provided in the logbook that could lead to an error in the ability to identify the better algorithm. These conflicting results lead confusion for researchers and practitioners who have to make better decisions in their data collection using accelerometer, in order to obtain more valid and comparable data."

As requested, sentences have been modified and more developed.

Additionally, the final sentence in the above states….. "conflicting results lead to confusion…". The authors don’t really frame their discussion in the terms of why their study findings alleviate those concerns (or at least strongly defend the use of 30min classification as superior to all other algorithms) - particularly considering they promote of a non-wear time algorithm that is in disagreement with those already published. How does this improve confusion? I fully appreciate that if your results suggest 30 minutes is the better option then you should present these findings
but this needs to be presented alongside a cogent and nuanced discussion with reference to previous work in the field (as mentioned by reviewer 1). In general, the discussion needs to be elevated from the descriptive to the more critical.

As requested, the discussion has been more developed. In addition, our conclusion has been modified also in order to be more nuanced on our findings.

Some finer points:

* There are a quite a few occasions where reference is made to adults and older adults literature; however, there is a vast amount of children and young people literature (of which you now cite) that is more appropriate and could frame a substantial portion of the discussion.

As requested, discussion has been more developed according to children and young people literature.

* The added section regarding 'mean number of wear time periods' should be used in the discussion to compare and contrast referencing previous work. You have added a small paragraph in the discussion but it doesn't really add anything to the wider argument that the paper relates.

As requested by reviewer 1, a section has been added referencing previous work.

* Figure 2 suggests that the 10min algorithm significantly underestimates % of those meeting 10 hours on 3 out of 4 days (with the other day prob close to significance). The 60 minute algorithm significantly over-estimated on Day 2 only. This seems really interesting. Why on one day only when the other days are much closer to the log book reference? Specific analyses looking at classification accuracy of individual non-wear periods (as suggested by reviewer 1) could prove useful to explore this.

As requested by reviewer 1, the table 5 has been modified and new analysis has been performed using the Kappa statistic.

* What do your findings mean for the field? Does it add to the confusion as you mention in the background or should researchers working with children and adolescents use a 30min non-wear algorithm to process their data? Why would your results provide stronger justification of processing decisions than the other papers cited (i.e. a 60min classification)?

As requested, the discussion has been more developed.
Other comments

The paper needs to be checked thoroughly for grammar and typo's, including the new sections that have been added. There are quite a few issues so it might be a good idea to have someone proof-read before re-submission.

Done.

Tables and Figures

All tables and figures could benefit from further description/labelling either at the bottom of tables or embedded into figures.

Done.

Bland Altman plots should be labelled for mean diff and limits of agreement. Y axis should have unit of measurement (mins) in label. They seem to be different sizes and scales. I think it would be beneficial for the reader to see them all on the same scale and size.

Done.

Figure legends for Figure 1 and 2 are reversed (page 25).

Change has been made.

Figure 2 has no x or y axis labelling. When it stands alone, it is unclear what the y axis refers to.

Done.