Author's response to reviews

Title: Impact of accelerometer data processing decisions on the sample size, wear time and physical activity level of a large cohort study

Authors:

Sarah Kozey Keadle (sarah.keadle@nih.gov)
Eric J Shiroma (ericshiroma@gmail.com)
Patty S Freedson (psf@kin.umass.edu)
I-Min Lee (ilee@rics.bwh.harvard.edu)

Version: 3
Date: 22 October 2014

Author's response to reviews: see over
We would like to thank the reviewers for their thoughtful comments and suggestions, which have improved the manuscript. A point-by-point response is provided below.

Title: Impact of accelerometer data processing decisions on sedentary time and physical activity measures from a large cohort study

Reviewer: Melvyn Hillsdon

Reviewer's report:
This study examines the effect of different accelerometer data processing decisions on estimates of the amount of time participants wore their accelerometer and how much time per day was spent being sedentary and at different intensities of physical activity in a large sample of older women.

Major
In general I think the results section could be written with greater clarity. It would help readers if the text followed the layout of table 2. Sentences from 144 onwards just seem to run into each other.

Thank you for the suggestion. We now re-organize the results section to parallel the tabular presentation.

Title
The title does not represent the fact the much of the focus of the paper is on wear time estimates when I think it should.

We have changed the title.

A 'consort' type diagram showing the flow of participants who were invited to wear an accelerometer through to the denominators for Table 2 (1 day and 4 day) would be really helpful and will make it clear where numbers are lost up to the point of analysis. See Figure 1, Sabia et al Am J Epidemiol. Mar 15, 2014; 179(6): 781–790 for an example.

We have added a figure that is similar to Sabia et al., 2014.

Line 193 highlights that the sample size for analysis varies according to the various rules applied. I feel that this could feature more in the analysis as this is an important factor if the lost participants introduce bias into the sample (supplementary file suggest it does). Therefore, I would request that the authors report differences in sample characteristics as well as the N when applying the different methods for data processing.

We appreciate the reviewer’s concerns - differences in sample sizes can potentially result in bias in the “included” sample. However, in this particular cohort we have not found evidence of differences in sample characteristics among the following groups: the entire eligible cohort, those invited as of August 2013 (and thus eligible for the present analyses), those who returned a monitor among those invited as of August 2013, and those with at least 1 valid day (based on limited log + Choi). All groups were similar with regard to descriptive characteristics, shown in the table below (in the interest of space, we have not shown the group with at least 4 valid days, since the number of women with at least 1 and at least 4 valid days are very similar).
Because of the similarity among groups, and because the existing tables have a lot of data, we respectfully have chosen not to include the data below. However, if the editor feels these data should be included, we will be glad to do so.

<table>
<thead>
<tr>
<th>Eligible for sub-study</th>
<th>Invited as of Aug 2013</th>
<th>Returned monitor</th>
<th>1 log+choi day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Characteristics</td>
<td>28,987</td>
<td>15,250</td>
<td>8,373</td>
</tr>
<tr>
<td>Age at entry, mean, y</td>
<td>53.8</td>
<td>53.9</td>
<td>53.3</td>
</tr>
<tr>
<td>BMI, mean, kg/m²</td>
<td>26.8</td>
<td>25.8</td>
<td>26.1</td>
</tr>
<tr>
<td>Current smoking, (%)</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Energy intake, mean, kcal/d</td>
<td>1735</td>
<td>1730</td>
<td>1740</td>
</tr>
<tr>
<td>Fruits and vegetables, mean, servings/d</td>
<td>6.1</td>
<td>6.1</td>
<td>6.2</td>
</tr>
</tbody>
</table>

Minor
Line 64 Reference required for Lee and Shiroma
We have added the reference.

Line 66 I’m not entirely sure it is true to say that data processing decisions and their effect on sample size and summary estimates has not been quantified. See Toftager et al IJBNPA 2013,10:140
We have clarified that this has not been quantified, to our knowledge, among adults.

Line 93 It would be helpful for external validity to have a brief summary of the 8,373 women who returned accelerometers and how they compared to the 39,876 who took part in the main summary. Also, how representative were the women in Table 2 compared to those that were invited to wear an accelerometer.

The sample who were eligible for this analysis (N= 28,987) as of Aug 12, 2013, are not expected to be representative of the 39,876 who completed the trial in 2004, as previous work shows those who are in studies tend to be healthier than those not in studies (For example, Lee IM et al., JAMA 2010;303:1173-9).
However, as shown in Table and response above, the women with at least one valid day are very similar to those eligible and those who were invited to wear the monitor in the present analysis.

Line 120 A short note explaining why the algorithm for VM data was altered would help the readers with less expertise in this area.
We added a sentence to indicate this was for comparability with algorithms for the vertical axis which use a sedentary cut-point of 100 cpm. For VM, we use the sedentary cut-point of 200 cpm from Aguiler-Faris et al.

Line 125 The full link to the physical activity package in R would be helpful as the link in the reference is too general and does not work.

We added in the link to the specific package.

Line 149 It would be beneficial if % as well as N values were presented in Table 2 with the denominator clearly shown (presumably the 7,650 eligible women minus those who returned devices that failed).

We added in the percentages.

Line 163 It needs to be made clear how the wear-times in Table 2 and reported in the text were derived. E.g., Are they for all valid days and if so how were any differences in the number of days included taken account of? Wear times appear to differ by day of week and between weekday and weekends making direct comparison with differences in days tricky.

We now clarify that the wear times are for valid days only in the methods section. We did not do any additional analyses to adjust or account for the number of days because our main goal was to show how making these data processing decisions impacts the summary estimates and the fact that number of days included varies is a major consideration. We agree that the questions raised are of interest, and do plan to do more detailed examination of e.g., weekday versus weekend day in another paper.

We now additionally note the sample overall was very compliant. Of the participants who wore the monitor at least one day, 98% had 4 or more valid days and 93.4% had 6 or more valid days. Therefore, we wouldn’t expect major differences if we adjusted for the days worn.

Line 174 Same point as above for Table 3 results. How were differences in number of valid days dealt with?

We simply reported the variables based on the number of valid days available for a particular method. The log data for V and VM are based on the exact same wear-time and number of days, so a reader can see the impact of sedentary time/PA metrics without differences in wear-time or number of days by comparing those values directly in the table.

Line 201 You say the amount of missing accelerometer data did not vary from 1-7 days yet the N’s for women with # 1day and # 4 days is different suggesting fewer women complying with the greater number of days required.

We agree this is confusing and have removed this sentence to avoid generalization.

Line 214 You recommend a combined approach for mail out methods on this line and the conclusion yet I think the data really only indicates that you need to know when the participant first wears the accelerometer and when they finally take it off for returning by post. These are the only occasions you cannot infer wear time from an algorithm. This minimizes the burden on the participant and the researcher.
At this time, we respectfully feel most comfortable recommending the inclusion of the log dates because we do not know of any studies that have tested a protocol using only the first day (and we also did not examine our data using only the first and last days). We have added that future research should examine whether a single day is sufficient.

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

Title: Impact of accelerometer data processing decisions on sedentary time and physical activity measures from a large cohort study

Reviewer: Lisa Phillips
This study looks at the impact of processing algorithms upon number of valid days and estimates of PA intensities. The manuscript presents important information for the physical activity literature and helps move the area towards a consensus regarding accelerometry data reduction. On the whole it is well written, though certain areas need to be addressed. Please see below for specific comments.

Minor essential revisions
Abstract:
1. Line 37, remove ‘we’, and write in the third person.
   We changed this to the third person.

2. Line 45: This line is confusing as it states using log dates only plus choi algorithms … this would imply that it is not only the logs. Please consider how this is worded.
   Agreed; we removed the word “only”.

Background:
3. Line 67: remove ‘we’ and write it in the third person
   We changed this to the third person.

Method:
4. Line 104: when was the accelerometer initialised to begin recording?
   Monitors were mailed by first class post (versus next day delivery) due to the cost and logistics of mailing approximately so many monitors. The monitor was initialized to start recording on the day BEFORE the estimated arrival time (based on US Postal Service estimated delivery time) to ensure the accelerometer was ‘on’ upon arrival at the participant’s home.

5. Line 136 – 141: can you specific whether your estimates of PA / sed are based on one criterion of valid days please.
   We now clarify that we based the estimate of physical activity and sedentary time based on the sample with at least one day of valid data.
Results:

6. Line 145 – 147: this needs re-writing, it could be made clearer.

Thank you. The other reviewer also expressed a similar concern. We now have re-organized the results to parallel data presented in the tables for clarity.

7. Line 150: change ‘about’ to ‘approximately’

This change has been made.

8. Table 2: the median and IQR do not provide much detail, it would be more beneficial to report the proportion of people achieving each number of days. Please amend this.

We added the percentage of the eligible sample that had 1 and 4 valid days to this table. We respectfully elected to keep the median, IQR for number of days and wear time, to provide as much data as possible.

9. Line 157: This is more an oversight in the protocol of the study, the accelerometer could have been posted and the recording beginning at a later date, when the women were instructed to wear it, you can also set a ‘stop date’. Although the result demonstrates a misclassification of non-wear as wear during transit, this is easily overcome through different protocols.

We agree, and when designing the study, we considered this approach. However, we ended up using the current protocol because as described above, it is not possible to predict precisely when the monitor will arrive as participants are distributed across the United States, and time of year may affect delivery also (e.g., longer transit times during the holiday season). We further cannot predict whether the participant will start wearing the device on the day it arrives; for example, a participant may not collect mail daily or may be away. Thus, if we initialized the device to start recording on the expected arrival day and to stop 7 days later, we might not capture the true 7 days of wear. Therefore, we chose a protocol where we initialized the device to start recording on the day before anticipated arrival, and to continued recording until data were downloaded or the battery ran out (~21 days after initialization).

10. Line 157: remove ‘we’ try, thus for any subsequent analysis... or something similar.

We changed this to the third person.

11. Line 165: there appears to be an ‘e’ that has no place.

Thank you; this typo has been fixed.

12. Line 178 - 186: Although different cut-points are required for vertical / vector magnitude, the wide intervals between the VM lower bound light and lower bound MVPA is substantially larger than the vertical, this may account for a portion of the substantial differences seen at this intensity.

We agree; it is possible the wider interval for VM accounts for a portion of the differences in intensity. We elected to use the Freedson et al. (vertical axis) and Sasaki et al. (VM) cut-points because they were derived using an identical study protocol; thus, we would expect them to be more similar than cut-points derived using different protocols. The reviewer raises an excellent
point highlighting the need for more methodological work to maximize the potential of data collected using the VM axis.

We changed this to the third person.

Discussion:
14. Line 196: remove ‘we’ and ‘our’ throughout the discussion
We changed all instances of “we” and “our”

15. Line 214: You could start and stop accelerometers without including mailing days, this may be a more appropriate than using the algorithms.

Please see the response to #4 and #9 above.

16. Line 217: Can you indicate this in your table please?
We have re-worded this sentence for clarity. Because of the extremely large sample size, differences with little practical significance (e.g., 19 vs 20 min of MVPA), are statistically significant. This sentence now reads; “Across different wear-time methods, the estimates of sedentary time differed by over an hour, while differences in MVPA were less than 1 min.”

Conclusions:

17. Line 272: Surely the loss of power from a loss of 10% is more substantial for a smaller study. A study of n =20000 losing 10% should still have substantial power to detect small changes whereas this may not be the case in smaller studies?
We apologize for the lack of clarity. We were primarily referring to the loss of resources/cost involved with collecting the accelerometer data that is then unusable because of the log, rather than the power for detecting associations with health. We removed this sentence from the conclusion to clarify.

Discretionary revisions:
1. Line 43, this line could be altered; it may be preferable to remove ‘commonly used’ from this sentence, with logs and algorithms removed from brackets.
We removed “commonly used” from this sentence.

2. Line 128 – 133: Can you create a sentence out of this please rather than a list.
We have now made this into a sentence.

Level of interest: An article of importance in its field
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