Reviewer's report

Title: The use of individual cut points to assess free-living physical activity by accelerometry: is it useful?

Version: 2 Date: 25 June 2012

Reviewer: Niels Christian Moeller

Reviewer's report:

General:
This research article examines the association between moderate-vigorous accelerometer cut-points, assessed individually through calibration in the laboratory, and minutes of MVPA accumulated during free living conditions. A widely acknowledged limitation in the use of accelerometry is the lack of consensus regarding the definition of count thresholds used to define different PA intensities. Many researchers handle these cut-point problems and limitations less carefully and often a more critical and skeptical approach would be to wish for. Research articles, such as the present one, can help shed light upon the caution needed when processing and interpreting accelerometer data.

Major Compulsory Revisions:
1) Abstract – conclusion: Your conclusions should be seen in the light of the subjects included, the calibration procedure performed, and the variable examined.

2) Background – first paragraph: The authors mention that it is essential to establish accelerometer cut points to separate light, moderate and vigorous PA. What is the reason for not mentioning the relevance of establishing accelerometer cut points to separate sedentary behavior? Reducing time spend on sedentary behavior could be of crucial importance, especially in groups of obese subjects. The authors should include analyses examining the association between sedentary ICPs and times spend on sedentary behavior during free living in the present study. Being able to these out minutes of sedentary in a group of obese individuals might be even more important in a health perspective.

3) Background – fourth paragraph: The description of the aim of the study does not reflect the fact that this study only examines the potential association between individual moderate-vigorous PA cut points and time spend in moderate-vigorous PA intensity during free living. But why is the present study focusing only on moderate-vigorous threshold? The aim of the study should also include cut points defining sedentary behavior and overall mean PA intensity level (i.e. counts/min.), unless it can be explained by the authors why this should not be the case? Including counts/min. in the analyses might provide some interesting information in terms of answering the question whether or not the source of bias responsible for the association observed between ICPs and
MVPA/day have the potential also to course bias when examining the association between ICPs and the overall level of PA - and maybe also indicate whether the bias was due to problems related to the MET classification or biomechanical issues more closely related to accelerometer response. Furthermore, I also miss the information that subjects are obese individuals. In the background it should also be explained more precisely what the hypothesis in this study is – what was your main reason to conduct the study – was the study conducted in the search for a perfect calibration procedure, or was the study conducted in order to, in a more pragmatic manner, illustrate how the use of MET values in a simple and feasible calibration protocol might be problematic when applied to field data obtained in obese individuals?

4) Procedure, fourth paragraph: Minutes of MVPA/day is expected to be related to the total amount of valid registered time. Consequently, MVPA should be expressed as the percentage of valid registered time pr. day in order to account for inter-individual differences in wearing time.

5) Statistical analyses, first paragraph: You could argue that a linear fit might be more robust on the individual level having few observations, however, you could definitely also question the approach of a linear fit. Since MVPA cut points were estimated to be negative in some subjects the use of a linear fit could be questioned, at least in some individuals. I realize that the association between count output and workrate/moving speed often is assumed to be linear (until leveling off occurs) However, Crouter et al. have previously questioned a linear fit inside a range of PA activities including slow and brisk walking. I would like to see the results based on non-linear fit (e.g. quadratic or exponential) included or at least discussed in the manuscript, and if possible more details on the scatterplots, maybe including the most interesting ones, would be appreciated as well.

6) Statistical analyses, third paragraph: I question the inclusion of body weight as a potential confounding variable in the model describing the relationship between ICP and MVPA/day since body weight also is included in the MET definition in your dependent variable. Please explain why controlling for age, sex, body weight and height in the model, and justify the “mutual” adjustment for body weight. Might the inclusion of an estimate of body composition (e.g. body fat) in the model instead be more appropriate?

7) Results: The authors should also provide information on the overall average level of PA (counts/min.) as well as on the absolute level of ICPs in the group of examined subjects (could be included in table 1, or added in another table?) Descriptive information should be given on resting metabolic rate and the walking speed performed at 3 METS.

8) Discussions/strengths and weakness: The author discusses several possible sources of measurement error which might have an impact on the accelerometer output. However, the most important discussion in the present study relates to which systematic bias(es) that might have caused the observed results. This issue is only being poorly discussed. For instance, how was the METS value
scaled according to body size? Different degrees of overweight/obesity could affect the definition of 3 METS if the resting metabolic rate is scaled according to body weight, as traditionally being the case. Body composition (fat mass and fat-free mass) has previously been found to account for 62% of the variance in resting metabolic rate, and it has been suggested that lean body mass or body surface provides the best indexing of resting metabolic rate. Furthermore, Spadano et al. have previously shown that body weight in adolescent girls will explain a large degree of the variation in MET values ranging from sedentary to vigorous activity, and the degree of variation explained by body weight seemed to be increased in particular during walking activities, especially when walking at inclination. Therefore, inter-individual differences in terms of body weight and body composition across the group of obese subjects might cause bias when scaling metabolic rate according to body weight at rest and during different activities. Adding to this potential systematic bias might be differences in weight bearing/none weight bearing activities observed under controlled walking running speeds performed on the treadmill and more complex movement performed in free living conditions, as well as inter-individual differences related to movement efficiency. In conclusion, it is not quite simple in my opinion to figure out if (and how) the classification of ICPs, based on the 3 MET-value, together with the scaling of RMR according to total body weight might introduce a potential source of systematic bias when trying to define MVPA cut-points in a group of obese subjects, and how this potential bias would affect the results observed during free living conditions. However, it would be interesting also to examine if cut-points defined according to relative percentage of VO2 max, or maybe cut-points defined at a fixed walking speed, would lead to the same indications of bias? At least, plausible explanations for the observed association between ICPs and MVPA/day should be provided and discussed.

9) Conclusions: The authors mention themselves in the discussion paragraph that “it is very difficult to perform a valid comparison of precision between the GCPs and the ICPs”, and in the present study this comparison relies on a post hoc analysis with the use of no golden standard measurement technique presented in the discussion. This fact should be mirrored in the overall conclusion of the study, and the conclusion claiming that the GCPs should be used over the ICPs in all studies in all study setting should be considerable less pretentious. The conclusions referring to the more pragmatic issue of how ICPs potential might impact field data should be clearly underlined. The sample enrolled in the present study consists of a highly specific group of obese subjects, and the authors mention themselves in the discussion paragraph that their results not may be valid in a normal weight population. However, as just mentioned, the authors conclude that the GCPs should be used over the ICPs in all studies in all study setting. In conclusion, no conclusions should be drawn unless being supported by analyses examined in the study.

Minor Essential Revisions

1) Title: I find the title a little misleading for three reasons: 1) the study sorely addressed issues related to thresholds defining moderate-vigorous PA intensity through individual calibration, 2) derived cut points were based on a simple
treadmill walking/running protocol and linear regressions analyses (which previously has been suggested to be appropriate by some, but also to be inappropriate by others), 3) the study participants were all obese subjects (i.e. specific group of people likely to be characterized by more extreme physiological responses to body movement/exercise, more pronounced inter-individual variation in terms of instrumental displacement at the hip, as well as by a changed PA level, when compared to normal weights). Consequently, I suggest the title to be changed in order to reflect this essential information (however, see comments raised below related to the inclusion of other PA parameters).

2) Abstract – background: Variation in counts between subjects at a given rate of movement is truly a very important source of error when assessing PA with accelerometers, however, depending on the specific outcome of interest and the methodological design and issues in your study (e.g. how well are seasonal variations and day type variations controlled for - when relevant) other sources of error might be just as important – or even more important.

3) Abstract – methods: According to the abstract cut points were obtained in 42 obese subjects. However, as stated in the text (result paragraph): “44 subjects performed the treadmill protocol, but two were excluded due to accelerometer malfunction, and in the end, only 35 subjects had valid data obtained during calibration in laboratory and in the field”. Description of number of subjects enrolled/analyzed in the study could be improved/clarified. This is also the case regarding the number of participants wearing two accelerometers.

4) Abstract – methods: It is not clear what is meant by the sentence “We analyzed the association between the ICPs and PA field data by linear regression”. Please clarify in more details what you mean by “association between ICPs and field data.”

5) Background – first paragraph: Actually, movements are quantified based on changes in acceleration and reported in the more or less arbitrary unit “counts”

6) Background – first paragraph: The authors mention that health benefits of PA are determined, at least in part, by the work rate or the intensity of the activity. Please describe the difference of work rate and intensity of activity – or are these synonyms?

7) Background – second paragraph: The terms “error” and “variation” are being used less cogently when describing how measurement errors of different origin will be present and how these errors can be attributed (i.e. differences between subjects could also mean true biological variation with less relation to true measurement error).

8) Background – second paragraph: It is correct that the conclusion in the study by Moeller et al. was that individual calibration would have only very limited effect at an individual level when trying to reduce the impact of inter-instrumental variation in field data. However, it was also concluded that all instruments should undergo calibration in order to prevent biased results (e.g. between studies,
between time points, etc.) due to batch effects.

9) Background – third paragraph: Please clarify what you mean by “variation between subjects” when referring to previous studies which have examined the three-way analyses of variance between subjects – “variation between subjects” could mean many things.

10) Subjects – first paragraph: More information on “co-morbidity” would be relevant (e.g. some kind of morbidities could affect metabolic rate and biomechanical characteristics for instance, therefore potentially influencing the results and transmissibility to other populations).

11) Procedures – first paragraph: How was participants' compliance when being asked to rehearse on the treadmill prior to the study? Inexperience/insecure in terms of conducting a treadmill exercise protocol could add substantially to the limitations of comparing accelerometer derived data from the laboratory and the field, respectively. In theory, the accelerometer response could be biased across the laboratory and field setting in the people being less experienced with the treadmill and maybe encumbered with poor motor skills, potentially adding to the observed association between ICPs and MVPA/day?

12) Procedures – first paragraph: It should be emphasized that subjects wore two accelerometers during the calibration procedure in the laboratory and that the 22 subjects wearing two accelerometers were a subgroup of the 44?

13) Procedures – second paragraph: Please give more detailed information on the five walking speeds performed - five speeds between 2 and 6 km/h could be numerous different speeds. Was the treadmill speed checked manually in order to validate the electronically derived information, or what do the authors mean when they say that “multiple treadmill speeds were checked by measuring the distance run by the treadmill band”?

14) Procedures – second paragraph: Please provide information in the text that the Metamax 1 analyzer was validated against the Douglas bag technique.

15) Procedures – third paragraph: A short description of the GT1M including instrumental specifications, or reference to other studies which have provide this information should be provided.

16) Procedures – third paragraph: The authors write that the PA level in the field was assessed by accelerometry over a seven day period about one month prior to the start of the lifestyle treatment program. Were any of subjects measured more than one month prior to the program? The calibration procedure was conducted during the fourth week of the program, meaning that there were (at least) a two months gap between the field study and the calibration study performed in the laboratory. A change in the level of aerobic fitness during the two month period among some subjects could potentially introduce bias to the results observed. For instance, if one person moves from one category of aerobic fitness to another (and also experiences a changed efficiency of movement) due to a changes PA behavior during the two month period this person would be
expected to meet the metabolic equivalent of 3 METS during calibration at a
another walking speed as would have been the case if being tested two months
earlier. An increased walking speed at 3 METS performed during calibration due
to increased efficiency of movement, for instance, would in turn be expected to
result in an increased accelerometer output, meaning that a systematic bias
could occur when trying to compare the level of moderate-vigorous ICP with the
minutes spend in moderate-vigorous PA in the field two months earlier. The
opposite situation would (maybe) occur in subjects experiencing a decreased
efficiency of movement during the two month period? A changed level of resting
metabolic rate as a consequence of training could potentially also add to the
source of systematic bias observed in the present study? These issues should at
least be addressed in the discussion paragraph.

17) Procedure, fourth paragraph: It would be highly relevant also to provide some
information on how the conclusions drawn in field data might be affected when
applying/not applying the ICPs. A description of how the number of minutes (or
percentage of total time) spend in different PA intensity intervals would change
when applying ICPs and GCPs, respectively, would be informative.

18) Procedures: Sub-headlines would help improve the overview of the different
procedures paragraphs

19) Results, first paragraph: It was explained in the procedure paragraph that 23
subjects (not 22??) wore accelerometers at both sides of the hip.

20) Results, first paragraph: The number of participants having valid free-living
accelerometer measurements and valid accelerometer calibration data is not quite
clear to me. In one line it says that 42 subjects had valid calibration data and that
40 subjects had valid free-living data, and in another line is says that 35 subjects
of the 40 subjects having valid free-living data also had valid calibration data.
Please clarify the number of subjects (and who) providing valid data.

21) Results, second paragraph: It is not appropriate to report correlation
coefficient assuming that data were independent, since they were not.
Regression analysis assuming dependent data should be performed instead.

22) Results, second paragraph: I would like to see an estimate of the average
difference and two limits of agreement not only one.

23) Figure 3: Is the plot in adjusted for confounding variables?

24) Discussions, third paragraph: the authors state that the findings of a relatively
large SEM/LoA for cut points obtained from the right and left hip indicates that
the attachment of the instrument is a potential source of error. The authors
should hold in mind that any discrepancies observed between accelerometer
outputs from the right and left hip will include measurement error/variation due to
potential differences in terms of attachment, inter-instrumental variations, and
biomechanical variations between the dominant and non-dominant side of the
body.
25) Discussions, third paragraph: Is it plausible to estimate the expected deviation in cut points derived at the right and left hip based on technical variations observed under more standardized conditions? Please clarify what you mean by “because our data were homoscedastic, such a percentage difference is difficult to evaluate”.

26) Discussions, fourth paragraph: The authors dwell on the check of scatterplot and instrumental malfunction – it is not quite clear to me how this relates to the rest of that paragraph?

27) Strengths and weakness, second paragraph: The authors mention that “it is very difficult to perform a valid comparison of precision between the GCPs and the ICPs”. However, Brage et al. have previously shown group calibration to be more effective than individual calibration when trying to reduce measurement error in PAEE measured with combined heart rate and accelerometry (probably because the errors in the calibration procedure are greater than errors resulting from inter-individual variance).

**Level of interest:** An article of importance in its field

**Quality of written English:** Needs some language corrections before being published

**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