A Method to Minimize Loss of Accelerometer Data to Improve Assessment of Device-Based Measurement of Physical Activity

Patty Freedson
University of Massachusetts, Amherst

In this issue we highlight an article by Dr. Hotaka Maeda et al. entitled, “Comparing Methods for Using Invalid Days in Accelerometer Data to Improve Physical Activity Measurement.” Maeda and colleagues examine methods to maximize the use of as much accelerometer data as possible to assess device-based physical activity. This is a new approach for determining which days of activity monitoring with an accelerometer are used to quantify physical activity. We typically use rules (e.g., 10 hrs/day of ‘good data’) to determine which days of multi-day monitoring are included in device-based data processing of physical activity, and then usually discard several days of data that do not satisfy the data inclusion criteria. In this study the authors evaluate methods “that used data” typically discarded to maximize data usage without compromising the meaning and interpretation of the data.

Dr. Maeda conducted this study at the University of Wisconsin–Milwaukee where he is a postdoctoral fellow in the Center for Aging and Translational Research. Dr. Maeda agreed to be interviewed and his responses to several questions appear below.

A reaction to this study is then provided by Dr. John Sirard (jsirard@kin.umass.edu), an Associate Editor for the Journal for the Measurement of Physical Behaviour and an Assistant Professor at the University of Massachusetts, Amherst.

Interview With Dr. Hotaka Maeda, Postdoctoral Fellow, Center for Aging and Translational Research, University of Wisconsin–Milwaukee

What was your motivation for performing this study?

As researchers, we spend a great deal of time, money, and effort collecting data. We are often attached to our data and treat it like it is a treasure. Also, we want to squeeze out every bit of the information contained within the available data to ensure our conclusions are as accurate as possible. From both perspectives, I am discomfited by the fact that the gold standard of free-living daily physical activity (PA) accelerometer analysis includes throwing away a large fraction of the data. Eventually, I hope that methods that can use and interpret almost all accelerometer data become accepted and accessible to all kinesiologists. I wanted to contribute to this change, because I am certain that it will help propel the field to move forward.

This article offers researchers who are monitoring physical activity levels using a hip worn uniaxial Actigraph, a method to maximize sample size. You state that this method can only be used when counts per minute is the variable of interest. Do you think this method of reducing invalid monitoring days will work for variables that require further data processing (e.g., minutes of moderate-to-vigorous activity, sedentary time)? Why or why not?

There are dozens of published cut points used to categorize counts into sedentary, light, moderate, and vigorous PA, which typically vary by age. For this article, rather than restricting to one set of cut points, we decided to only include counts per minute to maximize the simplicity and generalizability of the analyses. With that being said, I still think it would be useful to know if the methods we compared works with variables that require further processing, such as moderate-to-vigorous PA or energy expenditure. That will be an area where we can expand upon in the future.

What challenges did you confront while conducting this study? How were those challenges overcome?

Well, it was a lot of data to process. Writing the simulation program in R was a challenge, but running it took time as well. Although I tried to optimize the speed of the program, it was still very slow. The program ran for multiple days straight, so I had to keep it running on my computer. I also had to re-run the simulation every time we decided to modify the study design. Finally, there were computer memory issues that I had to work around, for example, by breaking large data sets into smaller pieces.

Besides that, I had some trouble effectively communicating our findings to the target audience, which we assumed were not used to reading about semi-simulation studies. This was one area where my co-authors were very helpful. For example, we suspected correlations were more commonly known and easier to interpret than root mean square or bias, which was one of the reasons why we included it in the article. We also considered adding some sort of significance testing in the article because people in the field are used to seeing p-values. However, we were unable to find a way to do so without over complicating the analysis or violating some critical statistical assumptions. Ultimately, we decided that focusing on a graphical approach might be the most effective mode of communication.

What are the primary take home messages from this study?

Our results showed that the conventional method of calculating PA levels discarded 26.0% to 28.6% of the data. We suggest practitioners to use the within-minute average method for adults over 18 years of age, and the day-level imputation method for children and teenagers. To note, the day-level imputation method may be unstable for sample sizes of less than 50. These methods are particularly useful when the number of valid days is three or less.

---

Freedson (psf@kin.umass.edu) is Editor-in-Chief of the Journal for the Measurement of Physical Behaviour, and Professor Emerita, Dept. of Kinesiology, University of Massachusetts, Amherst.
Potentially, these methods can allow some participants with an insufficient number of valid days to be included in the final analysis.

Any additional comments?

The methods we presented in the article have the potential for further improvement. We need methods that are effective, reliable, flexible, and convenient. Therefore, I hope we can motivate readers to think about what they can do to utilize all available data, and possibly develop new approaches they can test to save data from becoming wasted.

Commentary by John R. Sirard, PhD, Assistant Professor, Department of Kinesiology and Commonwealth Honors College, University of Massachusetts, Amherst

In the article, “Comparing Methods for Using Invalid Days in Accelerometer Data to Improve Physical Activity Measurement,” Dr. Maeda and colleagues address the topic of missing accelerometer data due to excessive non-wear time; an extremely important vexing problem for all investigators that process free-living accelerometer data. While this article focuses on post-data collection processing, we should not lose sight of the need to continue to reduce the amount of missing accelerometer data at the outset. Using log books, making phone calls or sending text messages, and, when possible, providing monetary incentives for wearing the device have all been demonstrated to increase wear time (Matthews, et al., 2012). Obviously, study resources are variable and, thus, we do what we can to obtain adequate accelerometer wear time. Collecting accelerometer data from wrist-worn devices is an additional promising strategy to improve compliance; how to make sense of that data is a topic for another commentary.

Once you have done all you can do to ensure adequate accelerometer wear time, you will almost inevitably be left with some missing data and need to decide how to proceed. These decisions essentially center around how much missing data is acceptable while maintaining reliable estimates of a person’s usual physical behaviors. The conventional practice is to require at least four days of data, with at least 10 hours of wear time per day (Miguel, et al., 2017). These minimum requirements are themselves an issue for debate and continued research. If a participant does not supply this minimal amount of data, they are typically removed from the analyses, weakening statistical power and potentially biasing results. Therefore, valid data processing methods to preserve sample size are still needed.

In this issue of the Journal of Measurement of Physical Behaviour, Dr. Maeda and colleagues present timely and innovative research that tests several imputation models to address the issue of missing accelerometer data. The imputation models were compared to each other and to a conventional approach (i.e., removing days with less than 10 hours of wear time), using average counts per day as the outcome of interest. One of the imputation methods, originally published by Alhassan et al. (Alhassan, et al., 2008), is actually not a true statistically based imputation technique. Rather, it is a within-subject “use all available data” approach to produce an aggregate day of data, based on within-minute averages across days, including incomplete days of data that would be excluded using the conventional method. This relatively simple approach is compared with two more sophisticated “group-information centered” imputation methods published by Lee and colleagues (Lee, 2013; Lee & Gill, 2016). These imputation methods rely on inputs from age- and sex-matched participants. Data for the study were obtained from the NHANES accelerometer subsample, and divided into “recipients,” who provided seven days of accelerometer data, and “donors,” who provided less than seven days of data. The recipients received realistic patterns of missing data from an age- and sex-matched donor. Using a simulation process, performance of the four methods (i.e., conventional approach plus three imputation methods) was compared using root mean square error (RMSE), bias, and correlation coefficients, stratified by age group and sample size.

Interestingly, it seems the mathematically simplest approach (Alhassan, et al., 2008) produced the lowest RMSE estimates within each strata, compared to the other imputation methods (Lee, 2013; Lee & Gill, 2016). This was especially apparent when there were only one to three days of valid data; methods tended to converge when there were four to six days of valid data. When comparing age groups, the RMSE values were greater for children and adolescents, compared with adults and older adults. This difference emphasizes the need to consider longer data collection protocols beyond the one-week paradigm for children and adolescents. Interestingly, the within-minute averaging approach (Alhassan, et al., 2008) was the only method to produce a positive bias (over-estimation) for the youth populations when there were greater than three days of valid data. The implications of this finding are not readily intuitive. Dr. Maeda and colleagues indicate that this may be a true over-estimation when using the Alhassan approach, assuming the seven days of accelerometer data actually represents the “truth.” In reality, seven days of accelerometer data may not be enough to capture the “truth” of young people’s usual physical behavior. Therefore, this apparent over-estimation may be an artifact that would decrease in magnitude if the imputed data were compared to an “alternative truth” based on eight or more days of accelerometer data (Trost, et al., 2000). Because children’s and adolescents’ lives, schedules, and sleep patterns can be extremely variable, especially compared with older individuals, a two-week data collection protocol may be more appropriate to capture usual movement in these populations. With the switch to collecting accelerometer data from wrist-worn devices, this may be more feasible than ever. Still, how well seven or 14 days of wrist-worn accelerometer captures usual physical behavior remains to be tested.

The methods presented in this article are not without limitations. One such limitation is that these imputation methods produce one composite day to represent usual physical behaviors, which limits our ability to look at patterns of activity throughout a given observation period, such as a week. Identifying such behavioral patterns may be important depending on the research question.

Dr. Maeda and colleagues have presented research on a timely and important topic. And, like most good research, leads us to consider additional research questions in our quest to better understand physical behavior and health across the lifespan. Could each of these methods be used with just weekday or weekend data? Will they work for other accelerometer-derived metrics, such as minutes spent in moderate-to-vigorous PA or bouts of sedentary behavior? Would they be improved by incorporating data from each of the orthogonal axes and vector magnitude? Would similar imputation performance be achieved when applied to data collected at the wrist? How strongly are the resultant metrics from each
method associated with other health behaviors, risk factors, or disease states? Keep calm and process on.

References


