Statistical Considerations in the Analysis of Accelerometry-Based Activity Monitor Data

JOHN STAUDENMAYER[1], WEIMO ZHU[2], and DIANE J. CATELLIER[3]

[1]*Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA;* [2]*Department of Kinesiology and Community Health, University of Illinois at Urbana-Champaign, IL; and* [3]*Department of Biostatistics, University of North Carolina, Chapel Hill, NC*

ABSTRACT

STAUDENMAYER, J., W. ZHU, and D. J. CATELLIER. Statistical Considerations in the Analysis of Accelerometry-Based Activity Monitor Data. *Med. Sci. Sports Exerc.*, Vol. 44, No. 1S, pp. S61–S67, 2012. We review and discuss three statistical aspects of accelerometer-based estimates of physical activity energy expenditure (PAEE): 1) the nature of the relationship between accelerometer output and PAEE; 2) statistical aspects of calibration studies; and 3) two specialized statistical methods that are applicable to the problem of measurement error modeling and missing data methods. We call for a continuing development of statistical methods that use more characteristics of the accelerometer signal to estimate PAEE, advocate the use of bias and SE statistics and inclusion of cross-validation in accelerometer research designs, and encourage more efforts to understand systematic and random errors in accelerometer-based estimates of PAEE. **Key Words:** ACCELEROMETRY, CALIBRATION STUDIES, MEASUREMENT ERROR, PHYSICAL ACTIVITY ENERGY EXPENDITURE

Valid measurement of physical activity (PA) is necessary to understand its relationship to health outcomes, determine the efficacy of interventions to increase individual PA behavior, and reveal the societal determinants of individual PA. Numerous methods have been used to measure PA, including doubly labeled water, room calorimetry, indirect calorimetry, direct observation with trained observers, self-report questionnaires, and portable monitors (e.g., pedometers, accelerometers, HR monitors). The first four methods are often called "criterion measures" (or "gold standards") because of their high accuracy and reliability. They are, however, relatively time-consuming, costly, and technically demanding in assessment practice. In contrast, accelerometers, which will be the focus of this article, have the potential to provide objective and valid measurements of PA without a large subject burden. In addition, accelerometers are empirically important because relatively large-scale studies of free-living populations are currently using them to measure PA. Four examples include the National Health and Nutrition Examination Survey, the Trial of Activity in Adolescent Girls, the Hispanic Community Health Study/Study of Latinos, and Sport, Physical Activity and Eating Behavior: Environmental Determinants in Young People.

This article discusses statistical aspects of estimating PA and related energy expenditure using data from accelerometers. The data example focuses on a particular type of accelerometer measurement (e.g., counts per second from a hip-mounted ActiGraph GT1M (ActiGraph, Pensacola, FL)), but our intention is to make recommendations that generalize to other accelerometer brands, models, and measurement types, such as 30-Hz "raw" measurements of acceleration due to gravity at the hip in three axes. Throughout the article, the phrase "estimating physical activity energy expenditure (PAEE)" refers broadly to three types of measurements: total activity energy expenditure in MET-hours or kilocalories, MET-hours in different activity intensity categories (e.g., sedentary, light, moderate, vigorous), and time spent in those activity categories during the study period. Interest in estimating time spent doing specific activities, such as locomotion or sedentary behaviors, also is growing.

The rest of the article is organized into four sections. First, we present data to illustrate the nature of the relationship between accelerometer measurements and PAEE. These data (and other literature) suggest that the commonly used "cutpoint" algorithm of assigning a PA level to a person based on the total (or mean) acceleration in a unit of time (such as a minute) has inherent and unavoidable limitations when the person does a variety of activities that include both locomotion and upper body movement. We advocate for alternative methods of analysis, such as nonparametric regression methods and more flexible regressions. These alternative algorithms are more statistically sophisticated, but

Address for correspondence: John Staudenmayer, Ph.D., Department of Mathematics and Statistics, University of Massachusetts, Amherst, Lederle Graduate Research Tower, Amherst, MA 01003; E-mail: jstauden@math.umass.edu.

0195-9131/12/441S-0S61/0
MEDICINE & SCIENCE IN SPORTS & EXERCISE®

DOI: 10.1249/MSS.0b013e3182399e0f

recent evidence suggests that they have the benefit of using more of the information in the accelerometer signal to produce more accurate and precise estimates of PAEE. In the second section, we discuss the statistical aspects of "calibration studies." It is common in the existing literature for investigators to rely on a correlation statistic to evaluate a method, and we discuss drawbacks to that method and the benefits of bias, SE, and their composite "root mean squared error." This section also discusses analysis on the natural log scale to account for the fact that estimates of PAEE are often more variable as PAEE increases. The third section discusses and advocates for the development and use of measurement error models and missing data methods. Finally, the article concludes with a discussion and suggestions for future research.

RELATIONSHIP BETWEEN ACCELEROMETER MEASUREMENTS AND PA ENERGY EXPENDITURE

A common way to estimate PAEE from an accelerometer is to use one of several published equations that calculate METs as a simple linear function of accelerometer counts per minute (e.g., Freedson et al. [10], Swartz et al. [24], Hendelman et al. [12]). Concomitant methods use those linear equations to define accelerometer count cutpoints that categorize a minute as sedentary (<1.5 METs), light (1.5–2.99 METs), moderate (3–5.99 METs), or vigorous (≥6 METs). The data we present in this section suggest that the assumption of a linear relationship may lead to inaccurate estimates. Others also have pointed this out (1), and in fact, the assumption that METs are a function of accelerometer counts per minute is limiting. Instead, we advocate further development of methods that use more of the information in the accelerometer signal than just the total to estimate METs.

Relationship between counts per minute and METs. The data we use come from a recent study conducted in the Physical Activity and Health Laboratory at the University of Massachusetts, Amherst (17). In this study, 277 healthy people (age 20–60 yr, of both sexes, of various races, and with body mass indexes of 17.6–42.0) each performed 11 of a total of 23 activities for 7 min while simultaneously wearing a hip-mounted GT1M accelerometer (ActiGraph) to measure acceleration and a portable metabolic system (Cardinal Health, Yorba Linda, CA) to measure factors from which PAEE can be derived.

Figure 1 shows average METs ("steady state" from last 5 min of the activities) and average accelerometer counts per minute for each activity. This figure demonstrates that METs are not a function of counts per minute in the sense that a unique value on the *x* axis (counts per minute) does not always lead to a unique *y* axis value (MET level). For instance, tennis and cleaning a room both produce about 3500 counts per minute of acceleration, but tennis is a vigorous activity (approximately 9 METs), whereas cleaning a room is moderate (about 5 METs). The figure has other similar examples, too. This demonstrates that the approach of estimating METs as a function of counts per minute will lead to inaccuracies when people do a variety of activities.

Information on the accelerometer signal and alternative methods. Although counts per minute are similar for different activities that have different PAEE requirements, the distribution of accelerometer counts over time differs considerably across different activities. For instance, Figure 2 shows counts per second for tennis, room cleaning, level treadmill walking at 3.0 mph, and descending stairs (with landings) for a representative subject. The PAEE requirements are different, mean counts per minute are similar, and accelerometer signals look different. For instance, the signal from tennis tends not to return to zero as often as room cleaning, treadmill walking produces a signal with a low coefficient of variation, and descending stairs looks quite periodic. From a statistical perspective, this observation suggests using a regression model to estimate METs as a function of more than one statistic that describes the signal in an interval of time. Further, a regression that allows a more flexible relationship than the simple linear model may be useful.

Several authors have pursued this general approach recently. For example, Crouter et al. (7) proposed an intuitive two-step procedure. They observed that locomotion tends to produce counts with a relatively low coefficient of variation as opposed to other less rhythmic activities. As a result, they first classify each minute of counts into locomotion or not, and then, they apply specifically calibrated equations (polynomial and exponential) to estimate METs from accelerometer counts per minute. This approach reduced root mean squared error of estimated METs for an activity by nearly 50% on average compared with the simple linear regression methods.

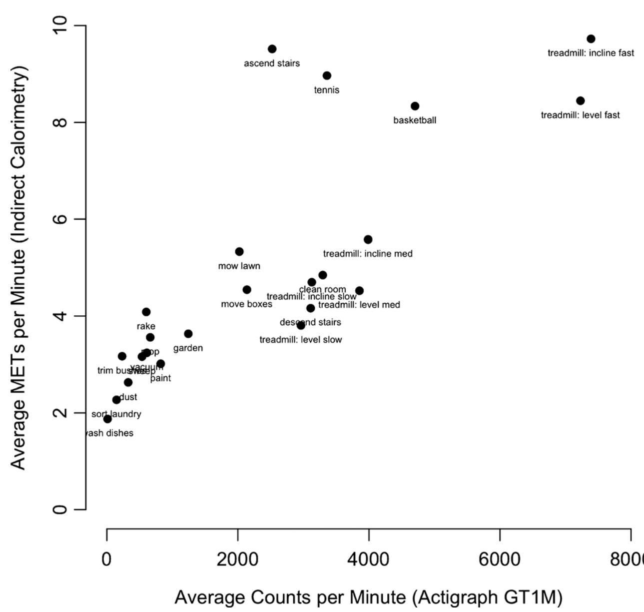

FIGURE 1—METs and average accelerometer counts per minute for a variety of activities.

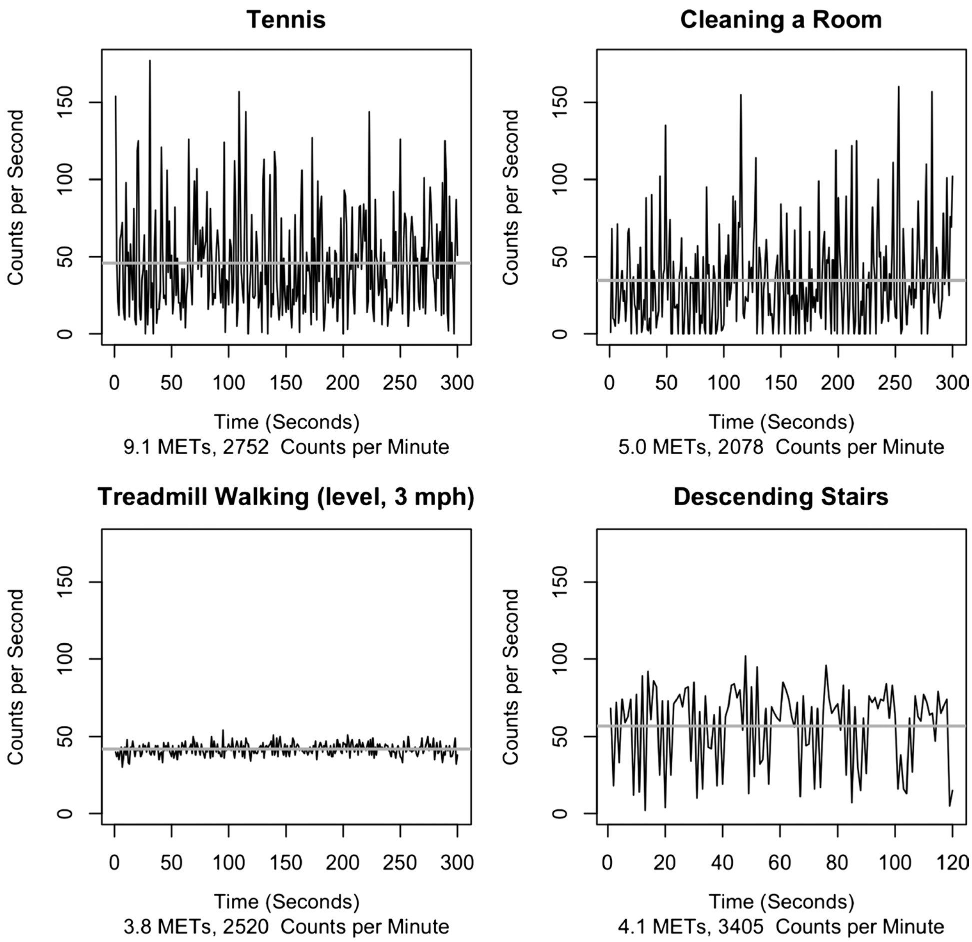

FIGURE 2—Patterns of counts per second for several activities.

Others (15,21,22) have used neural networks to estimate PAEE from statistics derived from the accelerometer signals. Although use of neural networks is prevalent, other non-parametric and flexible regression techniques, such as multivariate adaptive regression splines (28) and trees (2,3,25), are appropriate as well.

Although the detailed methods for those more sophisticated approaches are described in the original articles, the general approach is to use summary statistics from the accelerometer signals (e.g., moments, percentiles, summaries from the frequency spectrum) as covariates from which to predict PAEE (the response). The principle is that the accelerometer signal contains useful information beyond the mean (or total acceleration). The majority of these studies demonstrate substantial improvements (>50% reduction in root mean squared error) over simple linear regression estimates of METs. Some of these studies also use the accelerometer signals to provide some detail about what the person was doing (e.g., locomotion, household activities, sports). A recent general review of activity classification methods can be found in Preece et al. (20). We note that some have found it useful to use signals at higher resolutions than one observation per second.

The methods we have discussed so far have used a universal calibration scheme in which a single estimation algorithm is applied to all subjects. An alternative approach is to calibrate the estimation algorithm individually to each subject. That approach places a larger burden on a subject, but it is likely to be more accurate and precise because it has the potential to account for individual-to-individual variability.

STATISTICAL ASPECTS OF CALIBRATION STUDIES

The methods to estimate PAEE from accelerometer signals referred to in the previous section were developed from calibration studies in which accelerometers were worn while PAEE was measured through a relatively accurate criterion method such as indirect calorimetry. Data from those studies were then used to estimate the equation or algorithm that produces estimates of PAEE from the accelerometer signals. In this section, we discuss three ways in which the quality of those algorithms is estimated and reported. We recommend evaluation statistics, discuss how to estimate "out-of-sample" performance, and consider absolute versus percent (i.e., relative) errors.

TABLE 1. Example of a confusion matrix.

		Criterion Measurements ("Truth") (min)			
		Sedentary	Light	Moderate	Vigorous
Accelerometer estimates (min)	Sedentary	120	20	2	0
	Light	32	30	12	1
	Moderate	3	3	10	2
	Vigorous	0	1	2	7

Evaluation statistics. When discussing evaluation statistics, we distinguish between evaluating methods that produce continuous estimates of PAEE (such as METs in a minute) and those that produce categorical estimates (such as minutes at <3 METs vs minutes at ≥3 METs). We consider continuous estimates first. From a statistical perspective, a commonly used paradigm is to estimate bias (accuracy) and SE (precision).

Bias is the mean of the difference between the accelerometer-based estimate of PAEE and the criterion estimate. SE is the square root of the variance of that difference. Both statistics can be combined into a mean squared error, which is the sum of the squares of the bias and SE. The square root of the mean squared error is root mean squared error. A benefit of these statistics is that bias, SE, and root mean squared error all have the same unit as the original measures of PAEE. This makes the statistics easy to interpret. For instance, suppose a PAEE estimation technique has a bias of -0.05 METs·min^{-1} and an SE of 0.15 METs·min^{-1}. This means that if the method was used for 1 h, MET-hours would be underestimated by $60 \times 0.05 = 3$ MET·h on average. The SE around estimated MET-hours would be $(60 \times 0.15^2)^{0.5} = 1.16$ MET·h, and a 95% confidence interval would be approximately $2 \times 1.96 \times 1.16 = 4.54$ MET·h wide. In this example, bias and SE are estimated as the mean and SD of the difference between the estimate and criterion measure of METs for each minute and each person in the study. In other words, the evaluation statistics are calculated on the same unit of analysis as the method will be evaluated in its use outside the calibration study.

In the literature, it is common to report correlation (or its square, r^2) between continuous estimates of PAEE and the corresponding criterion measures. Correlation as an evaluation statistic has three major problems. First, it is hard to interpret the unit in a clinically relevant way because of its "relative" nature. For instance, although a correlation of 0.80 could tell us that the variation of the criterion measure can be explained by the variation of an accelerometer, it gives us limited information on how well we can estimate PAEE. Second, correlation is only a measure of linearity, and it can be very close to 1 when bias is arbitrarily high. Finally, correlation tends to be higher when the range of criterion PAEE values is greater. As a result, often, it is uninformative to compare correlations across calibration studies. We recommend against using correlation as an evaluation statistic or at least not the major one.

When evaluating categorical estimates of PAEE, a cross-classification of accelerometer-based estimates and the criterion values is useful. Table 1 illustrates a hypothetical example of this cross-classification, which is also known as a "confusion matrix" (apparently without irony). This table cross-classifies minutes according to the accelerometer and criterion measure assignments of whether they were sedentary, light, moderate, or vigorous. Agreement is on the diagonal. Above the diagonal is when the accelerometer underestimated intensity, and below the diagonal is when the accelerometer overestimated intensity.

One needs to be careful when creating an overall evaluation of the quality of a categorical estimate. Simply computing the percentage of items that are misclassified may be misleading depending on the distribution of the data. The Cohen κ (e.g., Fleiss (9)) is a measure of the degree of agreement between the predicted and criterion category that adjusts for the agreement probability that is due to chance. For ordinal classification problems, a weighted κ statistic can be used. The weight applied to the errors increases as the severity of the disagreement increases. It is suggested that a linear or quadratic weighting scheme be used unless information is available to suggest a different weighting scheme. The κ statistic ranges from -1 (total disagreement) through 0 (random classification) to 1 (total agreement).

Cross-validation and out-of-sample performance. The goal of the evaluation statistics discussed above is to predict how a method will estimate PAEE from accelerometer data that were collected in a different study or to estimate out-of-sample performance. When the development and evaluation of the method are done on the same subjects, the evaluation statistics tend to be biased and are often overly optimistic. Therefore, there is a need to cross-validate the prediction equations to another sample or sometimes another population (e.g., to a different age, sex, or ethnicity group). This process is known as "cross-validation" or "out-of-sample" performance.

Ideally, another sample from the same or targeted population should be used, which, however, could be costly. One alternative is to use the "split-sample" approach, in which the study sample is split into half, one for calibration and one for cross-validation. One should realize that the generalizability of the split-sample method could be low, especially when the sample sizes are small. One way to address this problem is to develop the estimation algorithm on all but one of the subjects and evaluate it on the "held-out" subject. That procedure is then repeated where each subject is used as the held-out subject, and the mean of the evaluations is reported. This is called leave-one-out cross-validation (e.g., Hastie et al. [11]). A computationally less taxing version is k-fold cross-validation, in which groups of subjects of size k are held out in each round. In either case, it should be noted that the estimation method that is published and disseminated to others should be developed on all subjects, not the subsets that were used for cross-validation.

Finally, we note that cross-validation estimates how a method will perform on a population that is similar to the one in the calibration study. The statistics are estimates, and confidence intervals should be provided. Even more importantly though, the performance estimates (and confidence intervals) often will not be very informative about how the method will perform on people who are demographically different or who do a mix of activities different from that done in the calibration study.

Percent error versus absolute error. Figure 3 illustrates a final point about calibration studies. Using data introduced earlier (17) and the method of Crouter et al. (7), panel A of this figure shows that the SD of the estimation error increases with the level of the PAEE. A consequence of this is that a statistic such as SE or root mean squared error that is reported by a calibration study is influenced by the mix of activities in the calibration study. A study that includes a higher proportion of high-MET activities will tend to have a higher overall SE and root mean squared error. That is a problem when one wants to compare methods across studies. An alternative, which addresses this problem, is to report the SE as a percentage of the mean. One way to approximately accomplish that is to compute and report the SD of the difference between the natural logarithms of the accelerometer estimate of PAEE and the criterion measure (13). The percentage errors for the example are shown in panel B of Figure 3. A measurement error that increases with the mean suggests a multiplicative error model, and we discuss measurement error models in the next section.

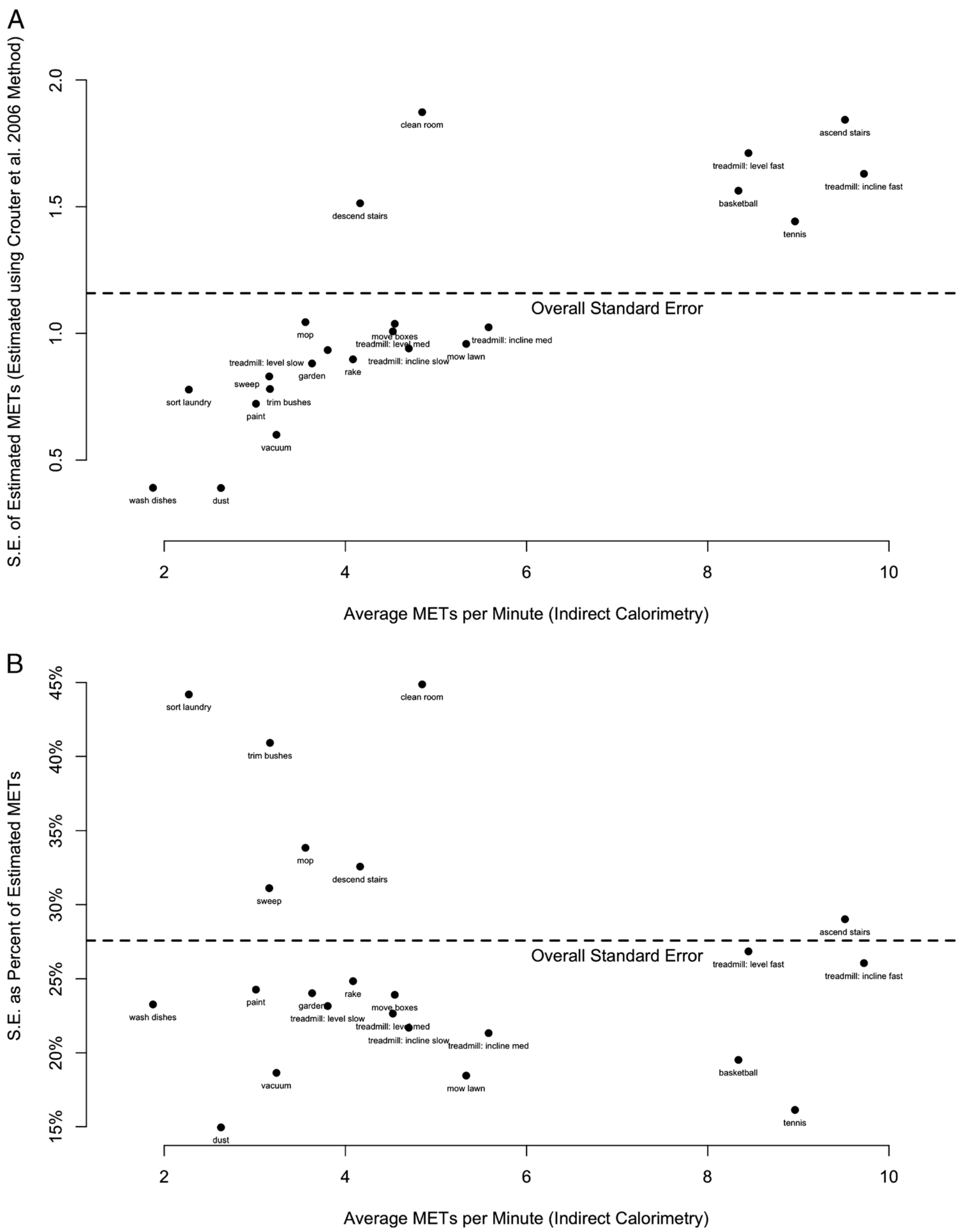

FIGURE 3—A, Error SD increases with level of PAEE. B, SD as a percentage of PAEE.

SPECIALIZED STATISTICAL METHODS

Two specialized statistical methods can contribute to the analysis and understanding of accelerometer-based methods of estimating PAEE: 1) measurement error and misclassification modeling and 2) missing data methods. Measurement error modeling and misclassification modeling are concerned with modeling the relationship between imperfect measures of PAEE (such as accelerometer-based methods) and criterion measures. These models can be used to improve estimates of PAEE, and they can be used to correct bias and accurately assess variability when estimates of PAEE are used as covariates in regression models with health outcome responses. Missing data methods address the situation in which a subject's accelerometer data are missing for part of the time when the subject is being observed. Flexible models are used to identify patterns in the rest of the data and impute the missing values.

Measurement error and misclassification models. Although accelerometers and calibrated PAEE estimation algorithms can produce valid estimates, an error relative to criterion measures always exists. The purpose of measurement error modeling (or misclassification modeling for categorical measures of PAEE) is to describe those errors for the purpose of improving estimates of PAEE and to correct for the problems that ensue from using an imperfect measure of PAEE as a covariate in a regression. Recently published book-length treatments of measurement error modeling are in the works of Buonaccorsi (4) and Carroll et al. (5), and this technique has been used extensively in nutritional epidemiology. For example, Kipnis et al. (16) and Sugar et al. (23) describe the development and application of sophisticated measurement error models to the Observing Protein and Energy Nutrition study. Parallels between that work and how measurement error modeling could be applied to the field of PAEE estimation are evident, and Ferrari et al. (8) provide a start at measurement error modeling in PAEE in their consideration of measurement error in self-report estimates of PAEE versus linear regression and accelerometer-based estimates of PAEE.

As an illustration of how measurement error modeling has the potential to be useful, consider Troiano et al. (27), who used accelerometers and linear regression methods to estimate PAEE from 2003–2004 National Health and Nutrition Examination Survey data. They concluded that less than 5% of adults were meeting a recommendation of 30 min·d^{-1} of at least moderate PA. That estimate contrasts with another national self-report–based estimate that 25% to 33% of adults were meeting the recommendation (26). It is reasonable to wonder whether both the 5% and the 25%–33% estimates suffer from the effects of measurement error and if neither is an unbiased estimate.

One way to address that question would be to include a validation sample in the study design. PAEE could be measured by both accelerometers (and perhaps self-report) and a criterion measure such as doubly labeled water or direct observation by trained observers on a subsample of the subjects. The data from that validation sample then could be used to fit a model that predicts the criterion measurement as a function of the accelerometer-based estimate, and that model could be used in turn to correct the accelerometer-based estimates for the subjects who were not in the validation sample. Several challenges would need to be overcome to execute that type of study, and we discuss these in the "Discussion and Conclusions" section below.

Missing data methods. One disadvantage of activity monitors is that subjects may remove them periodically (during sleep, bathing, and noncompliance), and this has been shown to affect the prediction of both total accumulated and average PA (6). Imputation is an established statistical procedure to reduce the biases caused by missing data (18), and these methods are appropriate for accelerometer-based estimates of PAEE (6). Recent work has shown that losing as little as 1 h of data during waking hours results in significant biases and variability (coefficients of variation between 7% and 21%) in the estimates of PA (19). Inserting a constant value for sleep and imputing values for missing data during waking hours significantly improved the estimates of PA (19). We note that replacing missing values with their expectations tends to underestimate the true variance of the complete data set, had there been no missing values. Full multiple imputation, as suggested by Catellier et al. (6), avoids that problem. A pedometer study also showed that individual-centered imputation is more accurate than a group-centered approach (14).

DISCUSSION AND CONCLUSIONS

This article reviewed and discussed several statistical aspects of estimating PAEE from accelerometer data. We provided data to show that estimating PAEE as a function of counts per minute often will result in inaccurate estimates, and we support continuing development of statistical methods that use more characteristics of the accelerometer signal (such as percentiles of the counts and the frequency spectrum) than just the total to estimate PAEE. The "cost" of these methods is that they will require close collaboration with statistical experts to develop, but the benefits are accuracy and validity of estimates. We encourage method developers to make open-source computer code available, but we also believe the community would see enormous benefit from easy-to-use "point-and-click" implementations of newly developed methods. Development of the methods often will require people with graduate degrees in statistics, but validation and use of the methods should not.

We also considered some statistical aspects of calibration studies that develop and validate methods to estimate PAEE from accelerometer signals. We advocated for the use of bias and SE as opposed to correlation, recommended cross-sample cross-validation as opposed to split-sample methods, and suggested that analysis on a log scale might be appropriate. In addition, we encourage researchers to share their calibration data in enough detail that the community can develop and validate methods using data from many calibration studies performed in different laboratories and locations. It seems likely that a multisite validation approach has the potential to produce methods that perform more robustly when applied to accelerometer data that are not part of a validation study.

We concluded by discussing two specialized statistical methods that can contribute to the effort to estimate PAEE from accelerometers: measurement error modeling and missing data methods. Measurement error modeling has the potential to improve the accuracy of accelerometer-based estimates of PAEE, but it presents many challenges that must be overcome to realize this potential, especially with respect to the criterion measure. For instance, if interest is in the pattern of PAEE over time (such as the number of "bouts" of 30 min of >3 METs per week), then DLW methods may not be applicable, and one may need to pursue direct observation methods or alternatives, such as room calorimetry. Further, times when the accelerometer typically is not worn but doubly labeled water (DLW) is still measuring energy expenditure (such as sleeping or early or late in the day) must be accounted for. Nonetheless, because accelerometer-based methods are used to measure PAEE in large national studies, understanding the systematic and random errors in those methods is critically important.

The authors report no conflicts of interest.

REFERENCES

1. Bassett DR Jr, Ainsworth BE, Swartz AM, Strath SJ, O'Brien WL, King GA. Validity of four motion sensors in measuring moderate intensity physical activity. *Med Sci Sports Exerc*. 2000;32(9 suppl): S471–80.
2. Bonomi AG, Plasqui G, Goris AH, Westerterp KR. Improving assessment of daily energy expenditure by identifying types of physical activity with a single accelerometer. *J Appl Physiol*. 2009;107(3): 655–61.
3. Brage S, Brage N, Franks PW, et al. Branched equation modeling of simultaneous accelerometry and heart rate monitoring improves estimate of directly measured physical activity energy expenditure. *J Appl Physiol*. 2004;96(1):343–51.
4. Buonaccorsi JP. *Measurement Error: Models, Methods, and Applications*. New York (NY): CRC; 2010.
5. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. *Measurement Error in Nonlinear Models: A Modern Perspective*. 2nd ed. New York (NY): CRC; 2006.
6. Catellier DJ, Hannan PJ, Murray DM, et al. Imputation of missing data when measuring physical activity by accelerometry. *Med Sci Sports Exerc*. 2005;37(11 suppl):S555–62.
7. Crouter SE, Clowers KG, Bassett DR Jr. A novel method for using accelerometer data to predict energy expenditure. *J Appl Physiol*. 2006;100(4):1324–31.
8. Ferrari P, Friedenreich C, Matthews CE. The role of measurement error in estimating levels of physical activity. *Am J Epidemiol*. 2007;166(7):832–40.
9. Fleiss JL. *Statistical Methods for Rates and Proportions*. 2nd ed. New York (NY): John Wiley; 1981.
10. Freedson P, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. *Med Sci Sports Exerc*. 1998;30(5):777–81.
11. Hastie T, Tibshirani R, Friedman JH. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed., New York (NY): Springer; 2009.
12. Hendelman D, Miller K, Baggett C, Debold E, Freedson P. Validity of accelerometry for the assessment of moderate intensity physical activity in the field. *Med Sci Sports Exerc*. 2000;32(9 suppl): S442–9.
13. Hopkins WG. Percent effects from log-transformed variables. In: *A New View of Statistics* [Internet]. 2003. Available from: http://sportsci.org/resource/stats/logtrans.html.
14. Kang M, Zhu W, Tudor-Locke C, Ainsworth B. Experimental determination of effectiveness of an individual information-centered approach in recovering step-count missing data. *Meas Phys Educ Exerc Sci*. 2005;9:233–50.
15. Kiani K, Snijders CJ, Gelsema ES. Recognition of daily life motor activity classes using an artificial neural network. *Arch Phys Med Rehabil*. 1998;79(2):147–54.
16. Kipnis V, Subar AF, Midthune D, et al. Structure of dietary measurement error: results of the OPEN biomarker study. *Am J Epidemiol*. 2003;158(1):14–21.
17. Kozey SL, Lyden K, Howe CA, Staudenmayer JW, Freedson PS. Accelerometer output and MET values of common physical activities. *Med Sci Sports Exerc*. 2010;42(9):1776–84.
18. Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. New York (NY): J. Wiley & Sons; 1987.
19. Paul DR, Kramer M, Stote KS, et al. Estimates of adherence and error analysis of physical activity data collected via accelerometry in a large study of free-living adults. *BMC Med Res Methodol*. 2008;8;38.
20. Preece SJ, Goulermas JY, Kenney LP, Howard D, Meijer K, Crompton R. Activity identification using body-mounted sensors—a review of classification techniques. *Physiol Meas*. 2009;30(4): R1–33.
21. Rothney MP, Neumann M, Beziat A, Chen KY. An artificial neural network model of energy expenditure using nonintegrated acceleration signals. *J Appl Physiol*. 2007;103(4):1419–27.
22. Staudenmayer J, Pober D, Crouter S, Bassett DR Jr, Freedson PS. An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. *J Appl Physiol*. 2009;107(4):1300–7.
23. Sugar EA, Wang CY, Prentice RL. Logistic regression with exposure biomarkers and flexible measurement error. *Biometrics*. 2007;63(1):143–51.
24. Swartz AM, Strath SJ, Bassett DR Jr, O'Brien WL, King GA, Ainsworth BE. Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. *Med Sci Sports Exerc*. 2000; 32(9 suppl):S450–6.
25. Tapia EM, Intille SS, Haskell W, et al. Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor. In: *Proceedings of the International Symposium on Wearable Computers*. New York: IEEE Press; 2007. p. 37–40.
26. Troiano RP. A timely meeting: objective measurement of physical activity. *Med Sci Sports Exerc*. 2005;37(11 suppl):S487–9.
27. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc*. 2008;40(1):181–8.
28. Zakeri I, Adolph AL, Puyau MR, Vohra FA, Butte NF. Multivariate adaptive regression splines models for the prediction of energy expenditure in children and adolescents. *J Appl Physiol*. 2010;108: 128–36.