Imputation of Missing Data When Measuring Physical Activity by Accelerometry

DIANE J. CATELLIER[1], **PETER J. HANNAN**[2], **DAVID M. MURRAY**[3], **CHERYL L. ADDY**[4], **TERRY L. CONWAY**[5], **SONG YANG**[6], and **JANET C. RICE**[7]

1*Department of Biostatistics, University of North Carolina, Chapel Hill, NC*

2*Division of Epidemiology, University of Minnesota, Minneapolis, MN*

3*Department of Psychology, University of Memphis, TN*

4*Department of Epidemiology and Biostatistics, University of South Carolina, Columbia, SC*

5*San Diego State University, CA*

6*National Heart, Lung and Blood Institute, Bethesda, MD*

7*Tulane University, New Orleans, LA*

Abstract

Purpose—We consider the issue of summarizing accelerometer activity count data accumulated over multiple days when the time interval in which the monitor is worn is not uniform for every subject on every day. The fact that counts are not being recorded during periods in which the monitor is not worn means that many common estimators of daily physical activity are biased downward.

Methods—Data from the Trial for Activity in Adolescent Girls (TAAG), a multicenter group-randomized trial to reduce the decline in physical activity among middle-school girls, were used to illustrate the problem of bias in estimation of physical activity due to missing accelerometer data. The effectiveness of two imputation procedures to reduce bias was investigated in a simulation experiment. Count data for an entire day, or a segment of the day were deleted at random or in an informative way with higher probability of missingness at upper levels of body mass index (BMI) and lower levels of physical activity.

Results—When data were deleted at random, estimates of activity computed from the observed data and those based on a data set in which the missing data have been imputed were equally unbiased; however, imputation estimates were more precise. When the data were deleted in a systematic fashion, the bias in estimated activity was lower using imputation procedures. Both imputation techniques, single imputation using the EM algorithm and multiple imputation (MI), performed similarly, with no significant differences in bias or precision.

Conclusions—Researchers are encouraged to take advantage of software to implement missing value imputation, as estimates of activity are more precise and less biased in the presence of intermittent missing accelerometer data than those derived from an observed data analysis approach.

Keywords

EM ALGORITHM; MULTIPLE IMPUTATION; ACCELEROMETER

Within the past several years, accelerometry has emerged as an important means of assessing the duration and intensity of physical activity and has served to define primary outcome measures in several observational (1,4,20) and experimental studies (6,7,11,15,17). It is currently being used in the large group-randomized Trial of Activity in Adolescent Girls (TAAG) to examine the effect of a school- and community-based intervention on physical activity in middle-school girls. The uniaxial accelerometer considered in this trial, the ActiGraph, formerly known as the Computer Science and Applications (CSA) and Manufacturing Technologies Inc. (MTI) (ActiGraph, LLC, Fort Walton Beach, FL) is a small ($5 \times 4 \times 1.5$ cm) and lightweight (45 g) device that captures vertical acceleration. Acceleration is sampled $10\times s^{-1}$, and the data are summed over a user-specified time interval (e.g., 30 s, 1 min) and the summed value or activity "count" is stored in memory.

Although the measurement protocols vary, most studies involve monitoring physical activity over several days to ensure reliable estimates of usual physical activity behavior and to account for potentially important differences in activity patterns on weekdays versus weekend days (19). Participants are typically instructed to wear the accelerometer during waking hours, except when bathing and showering. Data from accelerometers are summarized in numerous ways, including the mean total count per day (11) and the mean minutes per day spent in moderate or vigorous physical activity (using established count cut points to distinguish specific intensity levels). The analysis is complicated in that 1) activity levels vary among days of the week and times of day and 2) over multiple days of monitoring, missing data arising from removal of the monitor are a common occurrence. Although participant noncompliance accounts for a large fraction of the missing data, legitimate reasons for removing the monitor, such as complying with mandated sports league safety regulations or participation in water-related activity (for monitors that are not waterproof), also contribute to data loss. Thus, the timing and amount of data contributed by each individual vary. If summary statistics are computed using the observed data only, these statistics have the potential to be biased. For example, the total count for a given day clearly underestimates the true level of activity on days in which the monitor is worn only part of a day. Some researchers have tried to minimize this bias by computing summary statistics after excluding accelerometer data on days in which the monitor is worn only part of the day. These are called *incomplete* days of observation. This strategy, however, is not without issues. First, even after excluding days with, say, less than 8 h of wearing time, the number of hours the monitor is worn is still likely to vary. Moreover, if included days have intervals when the subject was awake but not wearing the monitor, then total activity will be underestimated. Second, this approach ignores possible differences in activity levels on complete days and incomplete days, making the estimated summary statistics from complete days subject to selection bias.

This paper proposes an analytical approach whereby the observed data are used to help predict activity levels for segments of the day in which the monitor is not worn. The resultant data set is *pseudo-complete* in the sense that each individual will have either observed or imputed data for all segments of each day in which the monitor was intended to be worn. Summary statistics are then estimated from this pseudo-complete data set. This imputation strategy is analogous to imputing missing item responses on multi-item questionnaires. The literature contains numerous examples in which this treatment of item nonresponse has been found to reduce bias effectively (5,21).

In the following section, accelerometer data collected during the feasibility phase of the TAAG trial are used to demonstrate the potential for bias in estimating physical activity when all observed data are used as well as when a subset of data that excludes incomplete days of monitoring is used. We then describe procedures for filling in missing data using single imputation through expectation maximization (EM) and multiple imputation (MI) (9,14). The remaining sections of the paper describe the design of a simulation study to assess the

effectiveness of the imputation approaches, present its results, and discuss the effectiveness of imputation as a strategy for dealing with missing data in the context of accelerometry.

THE PROBLEM OF MISSING ACCELEROMETER DATA IN TAAG

TAAG is a group-randomized, multicenter trial, sponsored by the National Heart, Lung, and Blood Institute, to assess the impact of a school- and community-based intervention on the physical activity of middle-school girls (16). The six field centers involved in TAAG are the University of Maryland, Baltimore, MD; University of South Carolina, Columbia, SC; University of Minnesota, Minneapolis, MN; Tulane University, New Orleans, LA; University of Arizona Tucson, AZ; and San Diego State University, San Diego, CA. The University of North Carolina is the Coordinating Center.

The data reported here arise from a substudy conducted in the fall of 2002. This substudy was undertaken to inform decisions about measurement protocols and design for the main TAAG trial. In this substudy, each of the six sites recruited two schools and randomly selected 45 eighth grade girls from each school; 80.7% participated ($N = 436$). Girls were asked to wear the monitor for seven complete days, and the epoch of integration was set at 30 s. Because the monitor records the slightest motion as a nonzero count, a sustained (20 min) period of zero counts was judged as a time when the monitor was removed and counts in that period were set to missing. Figure 1 shows the cumulative proportion of girls in our sample classified as wearing the monitor (i.e., nonzero counts were being registered) at specific times of the day.

The primary outcome for assessing the effectiveness of the TAAG intervention was the mean intensity-weighted minutes of MVPA per day. To derive this variable, one must first convert accelerometer counts per 30 s to their MET (multiples of resting metabolic rate (RMR) in kilocalories per kilogram per hour (1) using a calibration equation developed by Treuth et al. (18). Then, the total intensity-weighted minutes (i.e., MET-minutes) of MVPA are computed by summing the MET values above a moderate-intensity threshold (1500 counts per 30 s), and dividing by 2 (to transform from the original 30-s scale to a 1-min scale).

Table 1 shows the average time (h) the monitor was worn by day of the week. As noted earlier, a characteristic of accelerometer data is that activity is not measured over a uniform period each day. The number of hours in which the monitor was worn was much higher on weekdays than on weekend days. Although the genesis of the missing data is not fully understood, many girls reported forgetting to put on the monitor; fewer reported not wearing the monitor because of illness or because of participation in a sporting event where its use was prohibited.

If the primary outcome is computed with an approach that uses observed data, which represents measurement of activity over an average of 12.1 h·d^{-1}, an estimated 136 MET·min of MVPA was accumulated per day. Alternatively, if an approach is used that excludes incomplete days, defined as having less than 8, 10, or 12 h of recorded activity, then the primary outcome was estimated to be 148, 147, or 159 MET·min of MVPA per day, respectively (Table 1). Because the first approach includes days in which only a few hours of activity are measured, the estimate is likely to represent an underestimate of the true mean level of activity in this sample and is not recommended. The approach of excluding incomplete days, on the other hand, involves an interesting trade-off between accurate representation of total activity during a typical day and possible selection bias resulting from differences between days that are dropped and days that are kept in the analysis.

When considering the approaches for dealing with missing data, it is important to understand the possible mechanism that may have given rise to the missing data. Following the terminology of Rubin (12), the missing data are said to be missing completely at random (MCAR) if the "missingness" is independent of all other data, and therefore individuals with incomplete data

have the same distribution for the primary outcome as those with completely observed activity data. When the missingness depends on the observed activity or covariates, but not on the missing (unobserved) pattern of activity, the missing data are said to be missing at random (MAR). When the missingness depends on the missing level of activity, the missing data are not missing at random (NMAR). If the distribution of activity for days that are excluded from analysis is the same as that for days that are included (i.e., the MCAR condition holds), the analysis approach in which incomplete days are excluded would result in a valid estimate of activity (9). On the other hand, if people are less (or more) likely to wear the monitor when they are inactive (i.e., the NMAR condition holds), this approach would lead to a biased estimate of activity.

IMPUTATION OF MISSING DATA

The fundamental idea of imputation is to use observed data values to assist in predicting missing values. How close the estimate of the missing value is to its true value depends on how many predictors are used and their correlation with the missing variable. In general, the greater the number of predictors and the higher the correlations among variables, the more precise the estimates of missing values will be. The quality of estimates is also affected by how much data are observed versus missing for an individual and by the pattern of missing values.

Table 2 shows the correlation matrix for measures of physical activity for each day of the week estimated from girls with seven completely observed days of data in the TAAG substudy (N = 181). Here, "complete" was defined as having nonmissing counts over at least 80% of a standard measurement day; with a standard measurement day defined as the length of time in which at least 70% of sample participants were wearing the monitor. For example, the minimum number of hours of nonmissing data for a weekday to be deemed complete would be 11.2 h (= 0.80*(70th percentile of *off-time − on-time*) = 0.80*(21:25h − 7:25h)) (Fig. 1). For weekend days, at least 7.2 h of nonmissing data would be required.

Before we describe the imputation procedures used in this paper, a little notation is needed. Let $y_i = (y_{i1}, …, y_{ip})'$ denote the set responses from subject i (i = 1,…, N) on P days of monitoring. The vector y_i can be partitioned into values that are observed and missing, $y_i = (y_{i,obs}, y_{i,mis})$, where the dimensions of each component depends on the number of observed and missing data values. Let $z_i = (z_{i1},…, z_{ip})'$ denote the ideal vector of responses for the ith subject in which all data have been observed, and Z be the matrix of observed responses, z_i, from all subjects. Assume z_i follows a multivariate normal distribution with parameter $\theta = (\mu, \Sigma)$. One can generate plausible versions of Z in many ways. The most straightforward approach replaces an individual's missing values with the mean of observed values. Although mean substitution is easy to use, it tends to decrease variance estimates as more means are added to the data. In turn, covariance estimates also are attenuated (8,10). Nevertheless, mean substitution may be a reasonable approach when correlations between variables (y_{ij} and y_{ik}) are low, and missing data are less than 10% (3). A more refined approach that uses probability density estimates of the missing values instead of point estimates, is known as the EM algorithm.

Starting with an initial parameter value $\theta^{(0)}$, the EM algorithm (2) repeats the following two steps. The jth iteration of the expectation or E-step consists of imputing missing values of Y by their conditional expected values given the observed data and the current parameter value $\theta^{(j-1)}$:

$$Y_{mis}^{(j)} = E[Y_{mis}|Y_{obs}, \theta^{(j-1)}].$$

[1]

For a detailed description of the computations involved, see Schafer (14). Any convenient starting value $\theta^{(0)}$ will do; the maximum likelihood estimate for θ computed using only the

complete rows of Y was used here. The maximization or M-step uses both Y_{obs} and current imputed data $Y_{mis}^{(j)}$ to reestimate θ, which is used in the next E-step to generate new imputations of Y_{mis}. This process is repeated until convergence, specified to occur when the successive log-likelihood values differ by less than 0.00001. The ideal data matrix Z is taken to be the observed and imputed data matrix produced from the final E-step.

It is important to note that single imputation methods treat imputed values as though they were the actual values. Therefore, the uncertainty about the correct value to impute is not taken into account and the variance of the summary statistic will be underestimated. To properly reflect this uncertainty, the MI (13) procedure replaces each missing value with m > 1 plausible values. For the present work m = 5 imputations were performed. The final result is m versions of Z, where the missing data are replaced by independent random draws from $P(Z \mid Y_{obs})$, the predictive distribution of Z given the observed data. The resulting data sets $Z^{(1)}$, $Z^{(2)}$, …, $Z^{(m)}$ are analyzed separately using standard complete-data methods, and the results combined in a manner that takes the imputation variability into account. In the current setting of multivariate normal data with arbitrary patterns of missing values, the form of $P(Z \mid Y_{obs})$ is complicated, making the distribution hard to sample directly. Thus, we employed a more convenient strategy whereby missing data are replaced by random draws from a simulated distribution of $P(Z \mid Y_{obs})$ obtained by the Markov chain Monte Carlo (MCMC) method. The MCMC method constructs a Markov chain that is long enough for the distribution to stabilize to a stationary distribution, which is the distribution of interest.

DESIGN OF A SIMULATION STUDY TO EVALUATE THE EFFECTIVENESS OF IMPUTATION

We selected girls from the TAAG substudy who had seven complete days of monitoring (as defined in the previous section) to assess whether imputation could effectively estimating missing accelerometer data. This resulted in a sample of 181 girls. Although the primary outcome of interest was the average daily MET-minutes of MVPA, important secondary outcomes were physical activity during weekdays and weekends and during specific periods of the day. In particular, each weekday was partitioned into periods roughly corresponding to before school (6–9 a.m.), during school (9 a.m.–2 p.m.), after school (2–5 p.m.), early evening (5–8 p.m.), and evening (8 p.m.–12 a.m.). On weekends, the first two time periods were segmented at 11 a.m. rather than 9 a.m. to reflect differences in sleep patterns. The relative effectiveness of imputation methods was examined under four conditions resulting from two different missing data mechanisms (MCAR, NMAR), and level of data aggregation (e.g., entire day, or before/during/after school). Thus, data were missing either for the entire day or missing for specific segments of the day.

Missing data were created as follows. Let $P_y(0.75)$ denote the 75th percentile for the mean level of activity across all days and $P_y(0.75)$ the 75th percentile for the mean level of activity for the jth day. Missing data were generated from a logistic regression model with three covariates: body mass index (BMI) in tertiles ($x_{i1} = 1, 2,$ or 3); mean level of activity across all days y_i; and activity level on the jth day (y_{ij}). The model had the form:

$$p(r_{ij}=1|y_i,x_i,\phi)=$$
$$1/[\,1+\exp\{-[\,\alpha_0+\alpha_1 x_{il}+\alpha_2 I(\bar{y}_i<P_{\bar{y}}(0.75))+\alpha_3 I(y_{ij}<P_{\bar{y}_j(0.75)})]\}],$$
[2]

where r_{ij} is an indicator variable for missingness and I(.) denotes the indicator function. In all missing data models, the missing data rate was chosen to match that observed in the TAAG substudy (Table 1), and the following imposed restrictions:

$$\text{MCAR:}\alpha_0=-1.386,\alpha_1=\alpha_2=\alpha_3=0$$
[3]

(probability that y_{if} is missing is unrelated to observed or unobserved data)

$$\text{NMAR}: \alpha_0 = -4.116, \alpha_1 = 0.235, \alpha_2 = 1.386, \alpha_3 = 1.386. \qquad [4]$$

At one extreme, the MCAR condition imposes a random pattern of missingness; at the other, the NMAR case assumes that the odds that the ith individual's response on the jth day (or segment of the day) is missing is 1.6 times greater if the individual is in the highest versus the lowest tertile of the BMI tertile ($e^{2*0.235} = 1.6$), and 4 times greater if their observed mean activity level is below the 75th percentile ($e^{1.386} = 4$) or if their response on the jth day is below the 75th percentile for the mean responses on that day.

The bias and precision of the estimates of primary and secondary outcomes were both considered important criteria for assessing the performance of the imputation techniques. Bias represents systematic error (e.g., over- or underestimation of parameter estimates), and precision random error (e.g., the spread or variability in estimates about the true values). The mean and SD of the outcome variables were computed before creating the missing values and after imputing the values set to missing, with smaller differences between the two set of parameters suggestive of less bias. In addition, bias in estimation ("prediction bias") was summarized with the mean signed difference (MSD):

$$\text{MSD} = \sum_{i=1}^{\mathbb{N}_{\text{mis}}} (\text{true value} - \text{imputed value})/n_{\text{mis}} \qquad [5]$$

where N_{mis} is the number missing data values. The precision in estimation of the missing data values was summarized with root mean square difference (RMSD):

$$\text{RMSD} = \sqrt{\sum_{i=1}^{\mathbb{N}_{\text{mis}}} (\text{true value} - \text{imputed value})^2/n_{\text{mis}}} \qquad [6]$$

With the MI procedure, MSD and RMSD were computed for each of the five imputations and the average was reported. Smaller values of these criterion measures indicate greater accuracy (MSD) and precision (RMSD).

RESULTS OF IMPUTATION IN TAAG

The estimated means and SD for MET-minutes of MVPA based on the completely observed data are provided in column 3 of Table 3. The proportion of missing data created under a MCAR mechanism and the parameter estimates computed from the observed data, and both imputation approaches are given in the next four columns. When estimating the mean, we see that there was minimal bias regardless of level of aggregation and analytic approach (observed data only, EM imputation, MI). When estimating the variance, the observed data analysis showed a positive bias while both imputation procedures were essentially unbiased for the weekday, weekend, and daily activity averages. For example, the SD of the weekend average MET-minutes of MVPA based on the completely observed data was 105, 133 based on the observed data, and 103 and 100 based on the pseudo-complete data sets created from EM imputation and MI, respectively. The same conclusions can be drawn from the MSD in Table 4, which show that the average prediction bias when imputing for activity across the entire day was minimal (−5.9 vs −3.7 MET·min of MVPA on weekdays; −2.3 vs −1.1 MET·min of MVPA on weekend days using EM imputation and MI, respectively). The prediction bias on Mondays and Fridays was considerably higher, with the missing activity values underestimated by more than 25 MET·min on Mondays and overestimated by more than 20 MET·min on Fridays. The average prediction bias when imputing for activity in segments of the day was low, with an average discrepancy between imputed and true activity of less than 2.1 MET·min of MVPA.

The last four columns of Table 3 show that when the missing data are NMAR, an analysis of observed data only resulted in mean daily activity values (e.g., Monday–Sunday) that were significantly larger than those from the completely observed data. This was to be expected, as the missing data were generated assuming a greater probability of missingness associated with lower activity levels. The imputation approaches also yielded positively biased estimates, but the magnitude of error was lower. Interestingly, the bias in weekday, weekend, and daily average activity estimates were similar for the observed data analysis and imputation procedures. This follows from the fact that the least active individuals are selectively excluded from the estimation of activity on a given day of the week, while they are not excluded from estimation of weekday, weekend, and daily averages as long as they have at least 1 d observed.

In general, EM imputation and MI performed similarly. Paired *t*-tests did not reveal significant relative biases between estimates of means and SD derived from the two imputation procedures. However, EM imputation showed a slight advantage over MI with regard to the precision with which the imputed values approximated the true values and higher correlations between true and imputed values. For example, under the NMAR condition, the correlations between imputed and true activity during the prespecified hour-blocks were 0.37, 0.38, 0.34, 0.00, and 0.29 using EM imputation versus 0.16, 0.28, 0.13, 0.19, and −0.04 using MI (Table 4).

DISCUSSION

The analyses and simulations described in this paper focus on the problem of obtaining unbiased estimates of physical activity with accelerometer data collected over multiple days. Although we used data collected during a TAAG substudy, the potential issues of bias and the performance of imputation techniques are expected to generalize to other studies using accelerometry.

Analysis of accelerometer data with intermittent missing data can be handled in a number of ways. One common approach is to restrict the analysis to days in which a sufficient amount of data were recorded. The problem arises in choosing an appropriate cut point for what is to be considered sufficient. In the TAAG data, we showed that the estimate of physical activity was remarkably different depending on the choice of cut points. One possible solution to this problem might to be to propose a standard cut point for researchers working with accelerometer data. This approach has two problems. First, the same cut point may not be appropriate for all populations (e.g., young and old), or for all days of the week (e.g., weekday vs weekend). The risk in leaving the choice of cut point to data analysts is that it is very difficult to strike the right balance between setting the cut point high enough to eliminate days when the monitor was clearly not worn long enough to accurately represent that day's activity versus making it so high that a significant number of individuals are excluded from analysis (thereby introducing the potential for selection bias). This suggests a need for guidelines to determine cut points that are appropriate for each individual analysis. The approach described in this paper, and that being used in the TAAG trial, involves defining a standard measurement day as the period over which at least 70% of the study population have recorded accelerometer data and making the cut point 80% of that observation period.

An alternative approach to dropping days with insufficient accelerometer data is to use imputation techniques to predict activity on those days. The advantages and limitations of imputation have been widely considered (e.g,, see Schafer (14)). The effectiveness of an imputation technique depends not only on its ability to obtain unbiased estimates of missing values, but also on the ability to reduce bias in summary statistics.

The results of the simulation study clearly show that performance of either imputation technique was affected by the proportion of missing data, the correlation of activity across days

of the week, and the missing data mechanism. Imputation performed better on weekdays than weekends, due to the lower percent of missing records and the higher correlation between activity levels on those days.

Both imputation methods yielded unbiased estimates of the mean daily physical activity under conditions in which the data were MCAR. The relative differences between the imputation methods' ability to reduce bias in summary statistics were small and nonsignificant. The variance of the summary statistic from MI is derived from two parts: the variance of the summary statistic computed from each multiple imputed data sets (within-imputation variance) and the variance in the summary statistic across the multiple imputed data sets (between-imputation variance). The EM approach will overstate the precision of the summary statistic because it is unable to estimate the second component of variation. In this particular application, the precision was only slightly overstated. The *ad hoc* analysis approach of dropping incomplete days also resulted in an unbiased estimate of the mean daily activity, but the SD was larger than for the original (complete) data and either imputation approach. Thus, a key limitation of the *ad hoc* approach is a loss of efficiency. In turn, the resulting reduction in precision leads to a loss of power for testing hypotheses.

When the missingness mechanism is NMAR (not missing at random), the imputation techniques failed to eliminate the bias in estimating the overall summary measure of activity, but both performed better than the *ad hoc* procedure when estimating activity on individual days. The mean MET-minutes of MVPA estimated from the complete data was about five units (less than a tenth of a SD) lower than that for the estimates obtained using either the *ad hoc* or imputation approach. However, when examining the estimated activity on each day of the week, the magnitude of this bias was markedly higher using the *ad hoc* method but relatively unchanged with either imputation approach.

Although imputation has emerged as a popular strategy to deal with missing data, an important limitation must be noted. Both the EM algorithm and MI assume the missing data are either MCAR or MAR. Thus, inherent in the problem of missing value imputation is the fact we have no objective way of knowing whether the missing data are MCAR, MAR, or NMAR. However, some analyses can be performed that explore whether known factors are related to the probability of missingness. For example, one could assess whether respondents with the lowest observed activity levels were most likely to have incomplete data. In the present TAAG substudy, the probability of having one or more incomplete days of monitoring was not found to be associated with race, age, or average activity level on completely observed days.

With the growing use of accelerometers to measure physical activity, some appreciation for the potential bias that might result from using various methods for computing summary statistics is needed. Because missing value imputation is never worse and often superior to the *ad hoc* procedure of deleting incomplete days, those analyzing accelerometer data should consider using the available software to undertake imputations. The EM algorithm was found to have slightly greater precision in predicting missing values than MI for the measure of activity used in this analysis (MET-minutes of MVPA) and was the procedure chosen for imputation in TAAG. This may have been due to the slight skewness in the distribution of the outcome variables. Although both the EM algorithm and MI assume that the data come from a multivariate normal distribution, the EM algorithm has been shown to provide consistent estimates under weaker assumptions about the underlying distribution (9). We do not believe, however, that this conclusion is necessarily true for other physical activity outcome variables, such as the average total count per day or even for a transformation of the MET-minutes of MVPA, which may better satisfy the multivariate normality assumption.

References

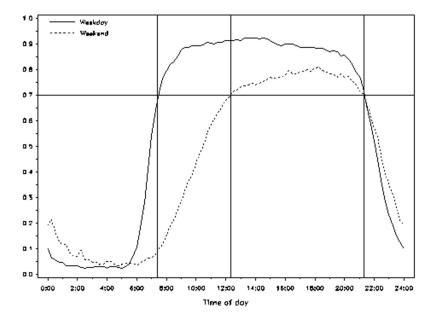1. Cooper AR, Page AS, Foster LJ, Qahwaji D. Commuting to school: are children who walkmore physically active? Am J Prev Med 2003;25:273–276. [PubMed: 14580626]

2. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Series B 1977;39:1–38.

3. Donner A. The relative effectiveness of procedures commonly used in multiple regression analysis for dealing with missing values. Am Stat 1982;36:378–381.

4. Epstein LH, Paluch RA, Coleman KJ, Vito D, Anderson K. Determinants of physical activity in obese children assessed by accelerometer and self-report. Med Sci Sports Exerc 1996;28:1157–1164. [PubMed: 8883004]

5. Gmel G. Imputation of missing values in the case of a multiple item instrument measuring alcohol consumption. Stat Med 2001;20:2369–2381. [PubMed: 11468769]

6. Going S, Thompson J, Cano S, et al. The effects of the Pathways Obesity Prevention Program on physical activity in American Indian children. Prev Med 2003;37:S62–S69. [PubMed: 14636810]

7. Kien CL, Chiodo AR. Physical activity in middle school-aged children participating in a school-based recreation program. Arch Pediatr Adolesc Med 2003;157:811–815. [PubMed: 12912788]

8. Little RJA. Missing data adjustments in large surveys. J Bus Econ Stat 1988;6:1–15.

9. Little, RJA.; Rubin, DB. Statistical Analysis with Missing Data. New York: Wiley; 1987.

10. Malhotra NK. Analyzing marketing research data with incomplete information on the dependent variable. J Market Res 1987;24:74–84.

11. Meijer EP, Westerterp KR, Verstappen FT. Effect of exercise training on total daily physical activity in elderly humans. Eur J Appl Physiol Occup Physiol 1999;80:16–21. [PubMed: 10367718]

12. Rubin DB. Inference and missing data. Biometrika 1976;63:581–592.

13. Rubin, DB. Multiple Imputation for Survey Nonresponse. New York: Wiley and Sons; 1987.

14. Schafer, JL. Analysis of Incomplete Multivariate Data. New York: Chapman and Hall; 1997. p. 163-181.

15. Selvadurai HC, Blimkie CJ, Meyers N, Mellis CM, Cooper PJ, Van Asperen PP. Randomized controlled study of in-hospital exercise training programs in children with cystic fibrosis. Pediatr Pulmonol 2002;33:194–200. [PubMed: 11836799]

16. Stevens J, Murray DM, Catellier DJ, et al. Design of the Trial of Activity in Adolescent Girls (TAAG). Contemp Clin Trials 2005;26:223–233. [PubMed: 15837442]

17. Story M, Sherwood NE, Himes JH, et al. An after-school obesity prevention program for African-American girls: the Minnesota GEMS pilot study. Ethn Dis 2003;13(1 suppl 1):S54–S64. [PubMed: 12713211]

18. Treuth MS, Schmitz K, Catellier DJ, et al. Defining accelerometer thresholds for activity intensities in adolescent girls. Med Sci Sports Exerc 2004;36:1259–1266. [PubMed: 15235335]

19. Trost SG, Pate RR, Freedson PS, Sallis JF, Taylor WC. Using objective physical activity measures with youth: how many days of monitoring are needed? Med Sci Sports Exerc 2000;32:426–431. [PubMed: 10694127]

20. Trost SG, Pate RR, Ward DS, Saunders R, Riner W. Determinants of physical activity in active and low-active, sixth-grade African-American youth. J School Health 1999;69:29–34. [PubMed: 10098116]

21. Winglee M, Kalton G, Rust K, Kasprzyk D. Handling item nonresponse in the U.S. component of the IEA Reading Literacy Study. J Educ Behav Stat 2001;26:343–359.

FIGURE 1.
Proportion of girls wearing the accelerometer (i.e., nonzero activity registered), by time of day and day of the week.

TABLE 1

Mean MET-minutes of MVPA computed from observed data or subsets of the data after excluding days in which the monitor was not worn for a minimum wearing time.

Day of the Week	Accelerometer Wearing Time (h)		MET-minutes of MVPA							
	Mean	SD	Observed data		8 h of data		10 h of data		12 h of data	
			No.	Mean	No.	Mean	No.	Mean	No.	Mean
Mon	12.0	4.4	430	129	363	145	332	149	300	152
Tues	12.7	3.6	432	151	397	162	374	164	329	167
Weds	13.7	2.7	435	159	416	164	402	166	367	172
Thurs	13.8	2.8	435	154	420	158	401	161	369	164
Fri	13.7	3.6	436	180	403	189	383	195	348	201
Sat	9.9	4.5	432	102	313	125	257	127	181	136
Sun	9.2	4.3	426	73	303	90	234	95	121	100
Daily average[a]	12.1	4.1	435	136	366	148	319	147	231	159

[a] Weighted average with weekdays being given weight 5/7 and weekend days weight 2/7.

TABLE 2

Spearman correlations of MET-minutes of MVPA, for completely observed data used in the simulation experiment.

Day of the Week	Spearman Correlations						
	Mon	Tues	Weds	Thurs	Fri	Sat	Sun
Mon	1.00						
Tues	0.54	1.00					
Weds	0.60	0.63	1.00				
Thurs	0.57	0.52	0.56	1.00			
Fri	0.53	0.48	0.60	0.51	1.00		
Sat	0.41	0.36	0.35	0.36	0.30	1.00	
Sun	0.36	0.35	0.36	0.32	0.26	0.27	1.00

TABLE 3

Mean MET-minutes of MVPA computed before data were deleted and after imputation of missing values.

| Day of the Week | Complete Sample | | Artificially Generated Missing Data (MCAR) | | | | Artificially Generated Missing Data (NMAR) | | | |
| | | | Observed Sample | | EM-Imputed | Mean across Multiple (5) Imputation(s) | Observed Sample | | EM-Imputed | Mean across Multiple (5) Imputation(s) |
	No.	Mean (SD)	% Missing	Mean (SD)	Mean (SD)	Mean (SD)	% Missing	Mean (SD)	Mean (SD)	Mean (SD)
Mon	181	147 (113)	23	143 (106)	140 (98)	139 (97)	25	162 (120)	150 (107)	150 (108)
Tues	181	162 (109)	11	160 (108)	162 (108)	163 (111)	11	170 (112)	166 (107)	164 (109)
Weds	181	162 (120)	6	162 (121)	163 (119)	163 (119)	8	168 (123)	166 (119)	166 (119)
Thurs	181	154 (141)	9	155 (144)	153 (140)	154 (139)	8	160 (146)	157 (140)	155 (141)
Fri	181	191 (150)	13	194 (155)	194 (149)	195 (148)	13	205 (155)	198 (147)	197 (148)
Sat	181	127 (154)	22	126 (157)	125 (142)	125 (142)	27	147 (173)	134 (152)	134 (155)
Sun	181	84 (103)	25	83 (106)	85 (98)	85 (95)	26	95 (114)	90 (101)	90 (103)
Weekday average	181	163 (96)		163 (100)	162 (96)	163 (95)		165 (97)	167 (94)	167 (94)
Weekend average	181	106 (105)		108 (133)	105 (103)	105 (100)		113 (117)	112 (104)	112 (106)
Daily average*	181	147 (90)		146 (96)	146 (89)	146 (88)		152 (93)	152 (89)	151 (89)
Weekdays										
6–9 a.m.	114	27 (28)	12	26 (26)	26 (25)	26 (26)	11	28 (28)	29 (27)	29 (28)
9 a.m.–2 p.m.	167	41 (28)	13	42 (29)	42 (28)	42 (28)	13	42 (29)	42 (28)	42 (28)
2–5 p.m.	132	44 (37)	12	44 (39)	44 (37)	44 (37)	13	45 (38)	46 (37)	46 (37)
5–8 p.m.	140	35 (39)	12	35 (38)	34 (34)	34 (33)	13	36 (40)	37 (39)	36 (40)
8 p.m.–12 a.m.	119	23 (27)	12	24 (29)	24 (27)	25 (27)	13	23 (27)	25 (27)	24 (26)
Weekend days										
6–11 a.m.	114	13 (21)	25	12 (19)	12 (15)	13 (15)	30	16 (28)	15 (21)	16 (23)
11 a.m.–2 p.m.	167	22 (30)	25	22 (32)	22 (26)	22 (26)	28	24 (33)	24 (28)	21 (30)
2–5 p.m.	132	29 (41)	24	29 (46)	30 (39)	30 (39)	25	31 (47)	32 (39)	30 (41)
5–8 p.m.	140	29 (35)	26	30 (47)	27 (30)	28 (29)	28	35 (47)	34 (34)	31 (38)
8 p.m.–12 a.m.	119	23 (33)	26	22 (36)	23 (29)	24 (29)	24	27 (38)	26 (31)	27 (31)

Weighted average with weekdays being given weight 5/7 and weekend days weight 2/7.

TABLE 4

Root mean square difference (RMSD), mean signed difference (MSD), and correlation between true and imputed MET-minutes of MVPAAU.

| | Artificially Generated Missing Data (MCAR) | | | | Artificially Generated Missing Data (NMAR) | | | |
| | EM-Imputed | | Mean across Multiple (5) Imputations | | EM-Imputed | | Mean across Multiple (5) Imputations | |
Day of the Week	RMSD (MSD)	Spearman Correlation	RMSD (MSD)	Spearman Correlation	RMSD (MSD)	Spearman Correlation	RMSD (MSD)	Spearman Correlation
Mon	101 (−26)	0.63	108 (−32)	0.64	64 (14)	0.54	67 (15)	0.41
Tues	92 (−4)	0.72	108 (7)	0.80	65 (30)	0.33	63 (13)	0.55
Weds	76 (11)	0.75	80 (8)	0.87	65 (39)	0.61	87 (44)	0.37
Thurs	129 (−6)	0.21	133 (5)	0.34	60 (35)	0.56	62 (12)	0.44
Fri	119 (19)	0.53	119 (26)	0.42	89 (52)	0.36	87 (45)	0.32
Sat	139 (−8)	0.29	156 (−9)	0.13	76 (26)	0.21	98 (25)	0.09
Sun	92 (3)	0.45	92 (5)	0.40	60 (25)	0.29	69 (23)	0.32
Weekday average	105 (−5.9)	0.58	111 (−3.7)	0.60	69 (30)	0.48	72 (24)	0.41
Weekend average	114 (−2.3)	0.38	122 (−1.1)	0.27	68 (25)	0.25	83 (24)	0.20
Daily average[a]	109 (−4.4)	0.49	116 (−2.6)	0.46	69 (28)	0.38	77 (24)	0.32
Weekdays								
6–9 a.m.	41 (−8)	0.59	42 (−8)	0.49	20 (14)	0.37	21 (10)	0.16
9 a.m.–2 p.m.	31 (3)	0.42	33 (6)	0.38	28 (10)	0.38	32 (10)	0.28
2–5 p.m.	44 (5)	0.38	46 (6)	0.26	30 (17)	0.34	41 (16)	0.13
5 p.m.–8 p.m.	81 (−4)	0.09	80 (−3)	0.21	25 (16)	0.00	34 s(10)	0.19
8 p.m.–12 a.m.	26 (6)	0.44	28 (9)	0.11	20 (8)	0.27	29 (2)	−0.04
Weekday average	44 (0.3)	0.37	46 (2.1)	0.29	25 (13)	0.27	31 (9.4)	0.14
Weekend days								
6–11 a.m.	31 (−2)	0.12	32 (0)	0.11	12 (10)	0.12	23 (12)	0.06
11 a.m.–2 p.m.	36 (1)	0.05	36 (0)	0.05	20 (10)	0.24	26 (−1)	−0.08
2–5 p.m.	38 (3)	0.22	41 (4)	0.17	26 (12)	0.33	34 (4)	0.15
5–8 p.m.	39 (−4)	0.23	45 (−1)	−0.11	24 (18)	0.09	34 (10)	0.10
8 p.m.–12 a.m.	36 (0)	0.23	37 (3)	−0.05	22 (11)	0.28	41 (15)	0.15
Weekend average	36 (−0.5)	0.17	38 (1.1)	0.12	21 (12)	0.21	32 (8.1)	0.08
Daily average[a]	42 (0.1)	0.32	44 (1.8)	0.24	23 (13)	0.25	31 (9.0)	0.12