Proceedings of Statistics Canada Symposium 2014
Beyond traditional survey taking: adapting to a changing world

Properties of measures of usual daily energy expenditure

Wayne A. Fuller[1] and Dave Osthus

Abstract

Monitors and self-reporting are two methods of measuring energy expended in physical activity, where monitor devices typically have much smaller error variances than do self-reports. The Physical Activity Measurement Survey was designed to compare the two procedures, using replicate observations on the same individual. The replicates permit calibrating the personal report measurement to the monitor measurement and make it possible to estimate components of the measurement error variances. Estimates of the variance components of measurement error in monitor-and self-report energy expenditure are given for females in the Physical Activity Measurement Survey.

Key words: Measurement error, calibration, usual energy expenditure, self-reports.

1. Introduction

Physical activity is an important component of the long-run average daily energy expenditure, called usual energy expenditure (usual EE). Two instruments used to measure daily energy expenditure in large samples are self-report instruments and monitoring devices. A personal activity report (PAR) is practical for large scale surveys, but most monitor procedures are too expensive or (and) too burdensome to be used for all respondents in a large survey. A practical procedure is to use the PAR for all respondents and to select a subsample (or a separate sample) in which both PAR and monitor are used to collect data from the same respondent for the same period. The replicate data can then be used to construct a calibration function relating PAR to the monitor.

Both monitors and PAR are known to have measurement error. Measurement errors for monitoring devices are due to the inability of monitors to accurately capture the full range of activities and due to the imperfect conversion of monitor data into energy expenditure estimates. See Welk (2002) and Welk et al (2004). Measurement errors from self-report instruments are due to factors such as social desirability effects, difficulty in understanding concepts of survey questions, and cognitive limitation for recalling activity from the past. See Adams et al. (2005), Troiano et al. (2008) and Mathews (2002).

We give estimation procedures appropriate for studies in which more than one observation per measurement method per individual is available. We calibrate the personal report of physical activity to a monitor measure and estimate properties of the measurement errors.

2. The Iowa Physical Activity Measurement Survey (PAMS)

The PAMS was conducted continuously over two years in four Iowa counties (Black Hawk, Dallas, Marshall, and Polk), starting in the Fall of 2009. The sample was a multi-stage stratified probability sample with two strata per county. In each county, one stratum is a "high minority" stratum defined by Census tracts that have relatively high percentages of minorities and the other stratum is a "low minority" stratum defined by Census tracts that have relatively low percentages of minorities. The "high minority" strata were oversampled to achieve a higher percentage of minorities in the sample. Households in each stratum were systematically selected from a white pages

[1] Wayne A. Fuller, Department of Statistics, Iowa State University, Ames, IA, 50011, waf@iastate.edu.
Dave Osthus, Department of Statistics, Iowa State University, Ames, IA, 50011.

listing of telephone numbers. A screening interview was used to randomly select one eligible adult in each household to participate in the survey. To be eligible, the adult had to be between the ages of 21 and 70, capable of physical activity, and competent to be interviewed. After agreeing to participate in the study, the height and weight of each respondent was determined. Each respondent provided data for two days. The measurement days were randomly assigned approximately two weeks apart. On the assigned measurement days, the respondent wore the SenseWear armband monitor for the full 24 hours of the day, except for water activities such as swimming and showering. The respondent recorded activities for the period when not wearing the armband. The day following the measurement day, the respondent was contacted and a 24 hour (PAR) completed by telephone. The interview instrument was developed for a telephone sample survey and was based on Mathews et al (2002).

Our analyses is for the 785 females in the PAMS sample who provided both energy expenditure measurements for both measurement days. Over 90% of the responding females (735 out of 785) in the sample wore the monitor for more than 90% of the day. Activity during time unaccounted for by the monitor was estimated on the basis of a log kept by the respondent. One observation with extreme deviations was excluded from the analyses.

Measurements of daily energy expenditure (EE) in kilocalories per day (kcal/d) were provided by the monitor. The activities reported using the PAR were placed in one of 270 activities. Each of these activities is assigned a metabolic equivalent (CMET) intensity level using a modified version of the Compendium of Physical Activities. See Ainsworth et al (1993) and Ainsworth et al (2000). A version of the Compendium is available on the Web under "The Compendium of Physical Activities Tracking Guide". A CMET is the energy required for an activity by a "standard individual" relative to the baseline activity of sitting quietly. A MET is a measure of energy expenditure relative to weight, where 1 MET = 0.0175 kcal/kg/min. The total of CMETS for a day is the sum of the products $CMET_a \times time_a$, where $time_a$ is the time spent in activity a, and the sum is over activities.

3. Model

The basic data are the monitor EE for the 24-hour day and the CMET total for the same day. Each individual provides two pairs. Calibration of PAR to monitor requires the two measures to be in common units. The EE is the unit of interest in physical activity research and is directly obtained from the monitor. Therefore, the observed CMETS calculated from the personal responses must be transformed into EE. One procedure is to multiply CMET by a function of weight, height and age, where the most common function is the Harris-Benedict equation. The original Harris-Benedict equation for women is

$$\text{REE} = 655 + (9.6 \times \text{ weight in kg}) + (1.8 \times \text{ height in cm}) - (4.7 \times \text{ age in years}),$$

where REE is resting energy expenditure, and REE is the calories required by a non-active person for a 24-hour period. Because there are many other formulas, and because the Harris-Benedict formula may not be directly applicable to EE as measured by the monitor, we estimate an equation analogous to the Harris-Benedict equation as part of the calibration estimation.

Personal characteristics affect activity, energy expenditure and personal reporting. In some studies overweight individuals tended to over report activity relative to normal weight individuals. Also, there is evidence to suggest that some of the same factors affect the monitor. In developing a calibration equation, the personal characteristics are of interest in the way that they affect personal reporting, but we can only estimate the reporting effect relative to another instrument. Therefore a coefficient for a personal characteristic estimated in a function relating PAR CMETs to monitor EE can be composed of the three effects; reporting effects, monitor effects, and conversion-of-activity-to-EE effects. Non-the-less, we call our estimated function an estimator of REE.

Let C_{ij} denote the total CMET value for individual *i* on day *j* and let

$$EE_{C,ij} = (\sum_{h=1}^{4} \alpha_h z_{hi}) C_{ij}, \tag{1}$$

where $z_i = (z_{1i}, z_{2i}, z_{3i}, z_{4i}) = (1$, weight in kg, height in cm, age in years) for individual i and the vector $\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4)$ is to be estimated.

The variances of the measurement errors for the two procedures are correlated with the value and the distributions are skewed. The log data are much less skewed than the original data and variances appear less correlated with the value. For examples of the use of logarithms in physical activity research, see Ferrai et al. (2007) and Tooze et al. (2013). Our model employs logarithms but in a somewhat different manner than that of Tooze et al. (2013).

Assume the observations satisfy the model

$$\begin{aligned} EE_{M,ij} + \zeta_M &= k_{ij}^* u_{ij}^*, \\ EE_{C,ij} + \zeta_C &= \beta_0^* k_{ij}^* r_i^* e_{ij}^*, \end{aligned} \tag{2}$$

where $k_{ij}^* - \zeta_M$ is the unobservable "true" energy expenditure of individual *i* on day *j*, $(\zeta_M, \zeta_C, \beta_0^*)$ is a vector of parameters and $(u_{ij}^*, r_i^*, e_{ij}^*)$ is a vector of random variables. We let * denote exponentiation so that, for example, $u_{ij} = \log u_{ij}^*$. The proposed measurement error model is

$$\begin{aligned} x_{ij}(\zeta_M) &= \log(EE_{M,ij} + \zeta_M) = \mu_x + t_i + d_{ij} + u_{ij}, \\ y_{ij}(\zeta_C) &= \log(EE_{C,ij} + \zeta_C) = \mu_y + t_i + d_{ij} + r_i + e_{ij}, \end{aligned} \tag{3}$$

where

μ_x is the population mean of *x*,
μ_y is the population mean of *y*,
$t_i \sim (0, \sigma_t^2)$ is the deviation of the mean of individual *i* from the population mean of *x*,
$d_{ij} \sim (0, \sigma_d^2)$ is the deviation of day *j* from the mean of individual *i*,
$u_{ij} \sim (0, \sigma_u^2)$ is the measurement error in the monitor,
$r_i \sim (0, \sigma_r^2)$ is the mean reporting error for individual *i*, and
$e_{ij} \sim (0, \sigma_e^2)$ is the individual measurement error in the self-report instrument.

The $k_{ij} = \mu_x + t_i + d_{ij}$ is the unobservable "true" value of *x* in the transformed scale. The random variables $(t_i, d_{ij}, u_{ij}, e_{ij}, r_i)$ are assumed to be mutually independent. Note that the back transform of a predicted value from (3) is a linear function of the original value.

4. Estimation

The vector $\boldsymbol{\alpha}$ of (1) was estimated by the estimated generalized least squares(GLS) regression of $C_{ij}^{-1} EE_{M,ij}$ on $\mathbf{z}_i$. The estimated coefficients given in Table 1 are similar to those of the Harris-Benedict equation, but differ significantly. The differences between the estimates and the Harris-Benedict coefficients are given in the third column of Table 1. The standard errors are the estimated GLS standard errors. The F-test of the hypothesis that $(\alpha_2, \alpha_3, \alpha_4)$ equals the Harris-Benedict vector is 8.52 with 3 and 781 degrees of freedom.

Table 1.
Estimated coefficients of REE equation

Term	Estimate	Estimate - $\alpha(HB)$	Standard error
α_1 (Intercept)	-78.31	-733.32	177.78
α_2 (weight)	8.70	-0.91	0.35
α_3 (height)	6.29	4.49	1.07
α_4 (age)	-3.18	1.51	0.62

In estimation, the estimated $\boldsymbol{\alpha}$ was used to construct $EE_{C,ij}$ of (1). We fixed ζ_C at -350, a value such that the variance of $\log(EE_{C,ij})$ is nearly a constant function of $\log(EE_{C,ij})$. The estimate of ζ_M is the value of ζ_M such that the estimated β for an extended model is one, where the extended model is

$$\begin{aligned} x_{ij}(\zeta_M) &= \mu_x + t_i + d_{ij} + u_{ij}, \\ y_{ij}(\zeta_C) &= \mu_y + \beta(t_i + d_{ij}) + r_{ij} + e_{ij}. \end{aligned} \tag{4}$$

Given ζ_C and ζ_M, the method of moments was used to estimate the parameter vector $(\beta, \sigma_t^2, \sigma_d^2, \sigma_u^2, \sigma_e^2, \sigma_r^2)'$. Moments were constructed for the vector of individual observations,

$$\mathrm{a}_i = (\bar{x}_i, \bar{y}_i, x_{i1} - x_{i2}, y_{i1} - y_{i2})', \tag{5}$$

where $(\bar{x}_i, \bar{y}_i) = [0.5(x_{i1} + x_{i2}), 0.5(y_{i1} + y_{i2})]$, $x_{ij} = \log(EE_{M,ij} + \hat{\zeta}_M)$, $y_{ij} = \log(\hat{E}E_{C,ij} + \zeta_C)$, and $\hat{E}E_{C,ij}$ is defined by (1) with $\hat{\boldsymbol{\alpha}}$ replacing $\boldsymbol{\alpha}$. Given the model assumptions, and known parameters, $E\{\mathrm{a}_i\} = 0$ and the variance of $\mathbf{a}_i$ is

$$V\{\mathrm{a}_i\} = \begin{pmatrix} \sigma_t^2 + 2^{-1}(\sigma_d^2 + \sigma_u^2) & \beta\sigma_t^2 + 2^{-1}\beta\sigma_d^2 & 0 & 0 \\ & \beta^2(\sigma_t^2 + 2^{-1}\sigma_d^2) + \sigma_r^2 + 2^{-1}\sigma_e^2 & 0 & 0 \\ & & 2(\sigma_d^2 + \sigma_u^2) & 2\beta\sigma_d^2 \\ \textit{symmetric} & & & 2(\beta^2\sigma_d^2 + \sigma_e^2) \end{pmatrix}. \tag{6}$$

The simple sample variance of $\mathbf{a}_i$ is

$$\hat{V}\{\mathrm{a}_i\} = (n-1)^{-1}\sum_{i=1}^{n} (\mathrm{a}_i - \bar{\mathrm{a}})(\mathrm{a}_i - \bar{\mathrm{a}})', \tag{7}$$

where $\bar{\mathrm{a}} = n^{-1}\sum_{i=1}^{n} \mathrm{a}_i$ is the sample mean of $\mathbf{a}_i$ and n is the sample size. The method of moments estimators are given by setting $\hat{V}\{\mathrm{a}\} = V\{\mathrm{a}\}$ and solving for the parameters of (6). There are six model parameters and six unique moments in these equations, which permits expression of each estimated model parameter as a function of the sample moments.

Estimates of the model parameters are given in Table 2. The variances of estimates of Table 2 were computed using a delete-a-group jackknife procedure, as described in Kott (2001) and Section 4.2.2 of Fuller (2009). Twenty-five partitions were formed by first classifying individuals according to their race and sampling stratum. Within each race/stratum combination the individuals were ordered by age, and partition numbers 1-25 systematically assigned to the ordered individuals. A replicate was created by deleting one of the partitions. The jackknife estimated variance for an estimator $\hat{\theta}$ is

$$\hat{V}\{\hat{\theta}\} = \frac{1}{B(B-1)} \sum_{b=1}^{B} (\hat{\theta}_b - \hat{\theta}_{JK})(\hat{\theta}_b - \hat{\theta}_{JK})'$$

where $\hat{\theta}_{JK} = B^{-1}\sum_{b=1}^{B} \hat{\theta}_b$, $\hat{\theta}_b = B\hat{\theta} - (B-1)\hat{\theta}^{(b)}$, $\hat{\theta}^{(b)}$ is the estimate calculated for the *b*th replicate, and *B* = 25. All calculations, including the estimation of $\boldsymbol{\alpha}$, were carried out for each replicate. As in the original estimation, ζ_C was treated as fixed.

On the basis of the estimates of Table 2 about 15% of the variation in observed EE_M is due to measurement error and about 17% is due to day-to-day variation in individual energy expenditure. The person-to-person variation in usual *EE* represents about 68% of the total variance. The σ_u of 0.664 is approximately equivalent to a standard deviation of 179 in the original scale.

The measurement error in the PAR is the sum of the variances of *r* and *e* of model (3). The sum of the two variances is about 95% of the variance of daily *EE* (about 1.18% of usual *EE*). Thus an individual monitor determination has a relative standard deviation of about 32% as an estimator of usual *EE* and the calibrated PAR *EE* has a relative standard deviation of about 1.18% as an estimator of usual *EE*.

The interpretation of the variances of Table 2 must recognize the fact that both $EE_{C,ij}$ and $EE_{M,ij}$ are functions of weight, height, and age. If the errors in $EE_{M,ij}$ are correlated with those factors, then the estimated error variance for EE_M and the estimated covariance between EE_C and EE_M are affected.

The estimate of $\exp(\mu_y - \mu_x)$ is an estimate of the β_0^* of model (2). That parameter is the slope in the original scale of the regression of EE_C on the true *EE*, where true *EE* is the expected value of the monitor measure.

Table 2.
Estimated model parameters

Parameter	Estimate	Standard error
ζ_M	346	150
$\beta_0 = \mu_y - \mu_x$	-0.299	0.017
$\exp(\mu_y - \mu_x)$	0.742	0.043
$100\sigma_t^2$	1.975	0.255
$100\sigma_d^2$	0.478	0.063
$100\sigma_u^2$	0.441	0.135
$100\sigma_e^2$	1.130	0.123
$100\sigma_r^2$	1.200	0.093

Moving from the log scale to the original scale requires care because the log of the expectation of a random variable is not equal to the expectation of the logs. Our representation makes the transformation easier than most because the elements of (2) and (3) are relatively simple functions. Assume one desires calibration in the original scale and desires the conditional expectation, conditional on the true value, of the calibrated PAR to be equal to the conditional expected value of the monitor. Using (2), the calibrated PAR EE is

$$EE_{PAR} = (\beta_0^*)^{-1}\zeta_C - \zeta_M + (\beta_0^*)^{-1} EE_{C,ij}\,. \tag{8}$$

The estimates from Table 2 give

$$EE_{PAR} = -817.7 + 1.348\, EE_{C,ij}. \quad (9)$$

An alternative estimator of $(\beta_0^*)^{-1}$ is

$$(\tilde{\beta}_0^*)^{-1} = \left[\sum_{i=1}^{n} \sum_{j=1}^{2} (EE_{C,ij} + \hat{\zeta}_C) \right]^{-1} \sum_{i=1}^{n} \sum_{j=1}^{2} (EE_{M,ij} + \zeta_M) = 1.338. \quad (10)$$

The second estimator is computed in the original scale and should be less biased than the transformed estimator from the log scale. In this case the two estimates are very similar.

References

Adams, S. A., C. E. Matthews, C. B. Eberling, C. G. Moore, J. E.Cunningham, J. Fulton, and J. R. Hebert (2005). The effect of social desirability and social approval on self-reports of physical activity. *American Journal of Epidemiology*; **161**: 389-398.

Ainsworth, B. E., W. L. Haskel, M. C. Whitt, M. L. Irwin, A. M. Swartz, S. J. Strath, W. L. O'Brien, D. R. Bassett, K. H. Schmitz, P. O. Emplaincourt, D. R. Jacobs, A. S. Leon (2000). Compendium of physical activities: An update of activity codes and MET intensities. *Medicine and Science in Sports and Exercise*; **32**: S498-S516.

Ainsworth B. E., W. L. Haskell, A. S. Leon, D. R. Jacobs, H. J. Montoye, J. F. Sallis, R. S. Paffenbarger (1993). Compendium of physical activities: Classification of energy costs of human physical activities. *Medicine and Science in Sports and Exercise*; **25**: 71-80.

Ferrari, P., C. Friedenreich, C. E. Matthews (2007). The role of measurement error in estimating levels of physical activity. *American Journal of Epidemiology*; **166**: 832-840.

Fuller, W. A. (1987). *Measurement Error Models.* Wiley: New York, 20-25.

Fuller, W. A. (2009). *Sampling Statistics*. Wiley: New York; 251-260.

Kott, P. S. (2001). The delete-one-group jackknife. *Journal of Official Statistics*: **17**: 521-526.

Matthews, C. E. (2002). Use of self-report instruments to assess physical activity. In Welk, G. J. *Physical Activity Assessments for Health Related Research.* Human Kinetics: Champaign, 107-123.

Tooze, J., R. Troiano, R. Carroll, A. Moshfegh, L. Freedman (2013). A measurement error model for physical activity level as measured by a questionnaire with application to the 1999-2006 NHANES questionnaire. *American Journal of Epidemiology;* **177**: 1199-1208.

Troiano, R. P., D. Berrigan, K. W. Dodd, L. C. Masse, T. Tilert, M. McDowell (2008). Physical activity in the United States measured by accelerometer. *Medicine and Science in Sports and Exercise;* **40**: 181-188.

Van Remoortel, H. et al. (2012). Validity of activity monitors in health and chronic disease: a systematic review. *International Journal of Behavioral Nutrition and Physical Activity*; **9**: 84.

Welk, G. J. (2002). Introduction to physical activity research. In Welk G. J. *Physical Activity Assessments for Health Related Research.* Human Kinetics: Champaign, 3-18.

Welk, G. J., J. A. Schaben, J. R. Morrow (2004). Reliability of accelerometry-based activity monitors: A generalizability study. *Medicine and Science in Sports and Exercise;* **36**: 1637-1645.