Survey Design for Studies of Measurement Error in Physical Activity Assessments

Nicholas K. Beyler[1], Sarah M. Nusser[1], Wayne A. Fuller[1], Alicia L. Carriquiry[1], Gregory J. Welk[2]
Department of Statistics, Iowa State University[1]
Department of Kinesiology, Iowa State University[2]

Abstract

Energy expenditure (EE) data obtained from 24-hour physical activity (PA) recalls and objective activity sensing devices are often subject to considerable measurement error, but limited information is available about the sources and form of the measurement errors. We consider the problem of developing a sample design for a PA survey that includes 24-hour recalls and measurements from an objective sensing device. Our goal is to investigate measurement error models for EE. We propose models for recall-based and objective measures of EE, derive method of moments estimators of model parameters, and use data from a pilot study to calculate preliminary estimates. Finally, we consider the variance estimates for parameters in determining appropriate sample designs.

KEY WORDS: Measurement Error, Physical Activity, Energy Expenditure.

1. Introduction

Physical activity (PA) data are important for studying health-related problems such as cancer, heart disease, and obesity. In PA studies, energy expenditure (EE) estimates measured in calories per day are obtained using a variety of measurement tools, including 24-hour physical activity recalls (24PAR) and objective sensing devices such as the SenseWear Pro2 (SP2) Armband. Recalls estimate a subject's EE on a given day based on information about types of PA completed in the recent past. Sensing devices are worn by subjects and estimate daily EE by monitoring levels of PA over a specified period of time. Unfortunately the data acquired by these measurement tools are subject to considerable measurement error, and the sources and forms of the measurement error are not well understood.

Appropriate measurement error models need to be developed. These models must identify the error structure in PA data, provide useful calibration functions to accurately determine daily EE in study subjects, and aid in sample size determination for large-scale PA studies. In this paper we develop such models to evaluate sample sizes for a proposed PA survey. First we review relevant PA and dietary intake literature and properties of the 24PAR, SP2, and pilot study data from Calabro et al. (2007). Second we present measurement error models for two measurement tools: the 24PAR and the SP2. Third we discuss our design objectives, obtain method of moments estimators for the model parameters, calculate parameter estimates and variances using the pilot study data, and investigate suitable sample sizes for a large-scale PA survey.

2. Background and Significance

Before presenting our measurement error models, we provide some background information on relevant PA and dietary intake literature, the 24PAR and SP2 measurement tools and the pilot study data.

2.1 PA and Dietary Intake Literature

In most PA and dietary intake studies, one of the main goals is to determine the distribution of usual EE or usual dietary intake for a specified group. The usual level of EE or dietary intake for a specific individual in a group is his or her long-run average daily amount of EE or dietary intake (e.g. the average daily EE over a year-long period). Directly observing usual daily EE or dietary intake is difficult. Instead, estimates of usual daily EE or dietary intake from a variety of instruments are considered. In general, estimates come from three types of instruments: inexpensive, reference, and gold standard instruments. Inexpensive instruments estimate usual daily EE or dietary intake on a number of subjects at a relatively low cost, but are subject to large measurement error. Reference instruments are more expensive to use, but offer more accurate estimates than the inexpensive instruments. Gold standard instruments are usually very expensive to use, but provide the most accurate estimates possible.

For PA studies, recalls, objective sensing devices, and doubly-labeled water are often considered to be the inexpensive, reference, and gold standard instruments, respectively. For dietary intake studies, food frequency questionnaires (FFQs), food records, and doubly-labeled water are usually the inexpensive, reference, and gold standard instruments, respectively. The reference and gold standard instruments are used to calibrate the inexpensive instruments that are subject to considerable measurement error. Only a few papers deal with measurement error properties in PA data (Spiegelman et al. 1997, Wong et al. 2004) while many more investigate measurment error models for dietary intake data (Kipnis et al. 2001, Rosner et al. 1989, Freedman et al. 1991, Kipnis et al. 1999, Kipnis et al. 2003).

2.2 24PAR and SP2

The 24PAR and SP2 are both used to measure EE in Calories per day. The 24PAR is an interviewer-administered assessment that estimates EE from a subject's reported occupational, household, and leisure activities from the previous day (Calabro et al. 2007). The frequency and duration of the reported activities are translated into an EE estimate. This method of measuring EE may provide biased estimates that stem from systematic reporting of bias, recall errors, and the ways in which questions are asked to subjects (Calabro et al. 2007).

The SP2 is a wireless multi-sensor activity monitor worn on a subject's upper arm that keeps track of periods of activity and inactivity over the course of a day (Calabro et al. 2007). Data from a variety of sensors (e.g. heat flux sensor, skin temperature sensor) are recorded and translated into an EE estimate with internal algorithms. Unlike the 24PAR, the SP2 measures EE without subject-induced or interviewer-induced bias because an armband monitor keeps track of and records activity and inactivity instead of the subject and interviewer. Studies also contend that, compared to other devices, the SP2 more accurately assesses EE and PA levels in a variety of human subjects (Zhang et al. 2003, Fruin et al. 2004, King et al. 2004). However, the SP2 and other sensing devices cannot accurately account for all types of physical activities that take place during the course of a day. The SP2 is good at estimating EE from some types of activities (e.g. playing tennis) but not from others in which arm activity is minimal (e.g. biking).

2.3 Pilot Study

In Section 4 we use pilot study data from Calabro et al. (2007) to obtain parameter estimates and determine appropriate sample sizes for a larger PA study. The study subjects were 20 healthy young adults (10 males and 10 females between the ages of 22 and 41) with at least moderate fitness levels (based on self-reported fitness ratings). The subjects wore the SP2 on two different days (at least 1 week apart) and completed the 24PAR on the day following SP2 use. Hence, EE in calories per day was recorded four times (twice from the SP2 and twice from the 24PAR) for each subject. Subjects wore the SP2 during daily activities (excluding swimming and showering) and were asked to include some moderate to vigorous physical activity (MVPA) into their daily schedule.

We examine some preliminary statistics and plots to get a better understanding of the data. Let Y_{ij} be the 24PAR estimate of daily EE for subject i at time j and X_{ij} be the SP2 estimate of daily EE for subject i at time j. Also define $\bar{Y_{i.}}$ and $\bar{X_{i.}}$ as the average daily EE estimates for subject i from the 24PAR and SP2, respectively, $\bar{Y}_{..}$ and $\bar{X}_{..}$ as the overall sample means for the 24PAR and SP2, respectively, and S_Y and S_X as the sample standard deviations for the 24PAR and SP2, respectively. The sample correlation between the 24PAR and SP2 is defined as

$$r_{XY} = \frac{\sum_i(\bar{Y_{i.}} - \bar{Y}_{..})(\bar{X_{i.}} - \bar{X}_{..})}{\sqrt{\sum_i(\bar{Y_{i.}} - \bar{Y}_{..})^2 \sum_i(\bar{X_{i.}} - \bar{X}_{..})^2}}.$$

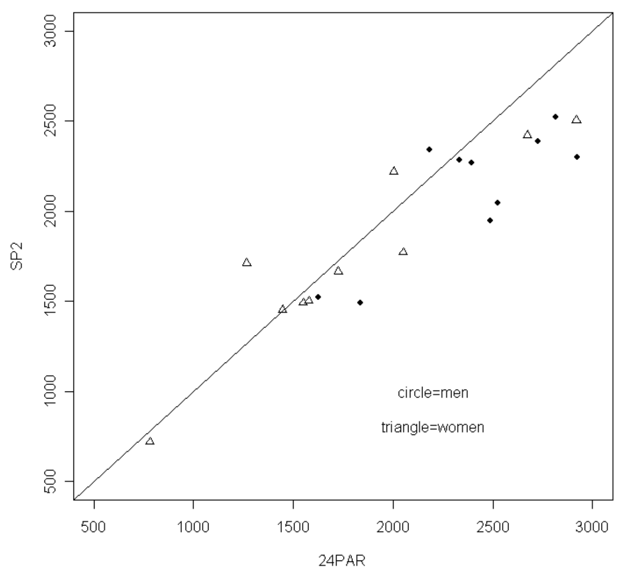
Values for these descriptive statistics are listed in Table 1 for all subjects, men only, and women only. The estimates for mean daily EE from the 24PAR are larger and more variable than the estimates from the SP2. The estimates for men are larger and less variable than the estimates for women. The sample correlation between the 24PAR and SP2 with all subjects is fairly high at 0.91. The correlation for women only is lower (0.81) and the correlation for men only is slightly higher (0.93). The plot of within-person means in Figure 1 confirms that the estimates of daily EE from the SP2 and 24PAR are similar and highly correlated since the data points are plotted closely along the 45 degree line. However, this plot also shows that estimates from the 24PAR are larger than estimates from the SP2 for the most active subjects. The data points in the top right corner are plotted to the right of the 45 degree line. Additionally, in this plot there is a larger range in values for women only, which may explain why the estimates for women only are more variable than the estimates for men only.

We believe the sample correlations in Table 1 may be larger than correlations that would come from the general population. The study subjects are all young adults with good PA awareness. Therefore, the subjects' account of PA used in the 24PAR should be fairly accurate and should closely match the results

Table 1: Descriptive Statistics for Pilot Study Data

	sample size	$\bar{Y}_{..}$	S_Y	$\bar{X}_{..}$	S_X	r_{XY}
All	20	2092.3	649.9	1928.2	538.5	0.91
Men	10	2385.4	414.2	2110.0	357.8	0.93
Women	10	1799.2	639.9	1746.4	529.6	0.81

Figure 1: Within-Person Means



from the SP2. In addition, the subjects were asked to include MVPA into their daily routines. This also may have encouraged good PA awareness and consequently more accurate estimates of EE from the 24PAR.

3. Measurement Error Models

We propose the following measurement error model equations for EE data using the 24PAR and SP2. Let

$$Y_{ij} \quad = \quad \beta_0 + \beta_1(x_i + r_{ij}) + \alpha_i + e_{ij} \qquad (1)$$

and

$$X_{ij} \quad = \quad x_i + r_{ij} + u_{ij}, \qquad (2)$$

where subject is indexed by i, day is indexed by j, Y_{ij} is the EE for subject i on day j using the 24PAR, X_{ij} is the EE for subject i on day j using the SP2, x_i is subject i's usual or habitual daily EE, and $x_i \sim (\mu_x, \sigma_x^2)$, $\alpha_i \sim (0, \sigma_\alpha^2)$, $r_{ij} \sim (0, \sigma_r^2)$, $e_{ij} \sim (0, \sigma_e^2)$, $u_{ij} \sim (0, \sigma_u^2)$ for all i and j. The parameters β_0 and β_1 are intercept and slope coefficients, respectively, included in (1) to account for the possible bias in the measurement of EE from the 24PAR. We do not include such coefficients in (2) because we assume the SP2 provides an unbiased estimate of EE. The error term r_{ij} accounts for the day-to-day deviation in EE for subject i. This term is included in both model equations because the day-to-day deviation in EE for a subject is the same regardless of the instrument used to measure EE. In other words, this error term has nothing to do with instrument error in measuring EE. In (1) this term is multiplied by the slope parameter β_1 because we believe the long run value of EE and the day-to-day deviation in EE both differ from the 24PAR measurement of EE by the same factor. The term α_i is the subject-specific bias from measuring EE with the 24PAR. We do not include a similar term in (2) because the SP2 should provide measurements free of subject-specific bias as long as the SP2 Armband is worn properly. The error terms e_{ij} and u_{ij} account for the random measurement error from using the 24PAR and SP2, respectively.

We assume that all error terms are independent across days within subjects and across instruments within subjects and days. These assumptions are reasonable as long as replicate observations are not taken during consecutive recording periods. We further assume that Y_{ij} and X_{ij} are independent because measurements from the 24PAR are subject to bias from the recall process and other subject-specific attributes while measurements from the SP2 are only subject to random measurement error.

Our models offer some improvements to the models presented in Spiegelman et al. (1997) and Kipnis et al. (2001). The intercept and slope coefficients in (1) capture the bias from measuring EE with the 24PAR. Spiegelman et al. (1997) fail to account for the potential bias from measuring PA with the flawed gold standard, which may consequently produce faulty calibration functions. Our models track three main sources of error in measuring long term average EE: subject-specific error (α_i), short term deviation in EE (r_{ij}), and overall random measurement error (e_{ij}, u_{ij}). The models proposed by Spiegelman et al. (1997) only include overall error terms and the models proposed by Kipnis et al. (2001) only include subject-specific error terms and random measurement error terms.

4. Survey Design

Now that we have developed measurement error models, we present our survey objectives and consider appropriate sample sizes for a large-scale PA survey. First we outline our basic survey objectives. Second we develop estimators for the model parameters using method of moments estimation procedures and obtain parameter estimates and variances using the pilot study data and methods from Fuller (1987). Third we present our results and determine suitable sample sizes relative to our survey objectives.

4.1 Survey Objectives

In general, we want to conduct a survey of adults in Iowa to collect PA data. This includes collecting daily EE estimates from the 24PAR and SP3 (new version of the SP2) for our models in Section 3. These data will be used to estimate model parameters and consequently, will help us better understand measurement error properties in PA data. We also want to examine PA properties within subpopulations of individuals based on gender, living environment (urban or rural), and race (Caucasian, Hispanic, or African American). In Subsection 4.3 below we present a suitable sample size for subpopulations based on gender and living environment. Within these subpopulations we will attempt to divide the sample across ethnic groups.

4.2 Parameter Estimation

Using method of moments procedures, we estimate the parameter vector,

$$\theta = \left(\mu_x, \beta_0, \beta_1, \sigma_x^2, \sigma_r^2, \sigma_\alpha^2, \sigma_e^2, \sigma_u^2\right)^T,$$

from functions of sample moments that are calculated using the pilot study data. The components of θ are defined in the previous section. First we create a new set of variables, $\{\mathbf{Z_i}\}$, from the EE data, $\{Y_{ij}\}$ and $\{X_{ij}\}$, to produce mean and covariance matrices that are functions of the model parameters. Let

$$\mathbf{Z_i} = \begin{pmatrix} \frac{(Y_{i1}+Y_{i2})}{2} \\ \frac{(X_{i1}+X_{i2})}{2} \\ Y_{i1}-Y_{i2} \\ X_{i1}-X_{i2} \end{pmatrix} = \begin{pmatrix} \bar{Y}_{i.} \\ \bar{X}_{i.} \\ Y_{i1}-Y_{i2} \\ X_{i1}-X_{i2} \end{pmatrix}$$

for all i, where $\bar{Y}_{i.}$ is the average EE estimate from the 24PAR for subject i and $\bar{X}_{i.}$ is the average EE estimate from the SP2 for subject i. We define $\mathbf{Z_i}$ in this manner because it provides algebraically simpler mean and covariance matrices than the observed data vector $(Y_{i1}, Y_{i2}, X_{i1}, X_{i2})^T$. Based on the model assumptions mentioned in the previous section we have a mean vector

$$E(\mathbf{Z_i}) = E\begin{pmatrix} \bar{Y}_{i.} \\ \bar{X}_{i.} \\ Y_{i1}-Y_{i2} \\ X_{i1}-X_{i2} \end{pmatrix} = \begin{pmatrix} \beta_0+\beta_1\mu_x \\ \mu_x \\ 0 \\ 0 \end{pmatrix}$$

and symmetric covariance matrix

$$Cov(\mathbf{Z_i}) = Cov\begin{pmatrix} \bar{Y}_{i.} \\ \bar{X}_{i.} \\ Y_{i1}-Y_{i2} \\ X_{i1}-X_{i2} \end{pmatrix}$$

that both simplify to functions of parameters from θ.

Then for $\{\mathbf{Z_i}\}$ we define two sets of sample moments $\mathbf{m_1}$ and $\mathbf{m_2}$. Let

$$\mathbf{m_1} = \begin{pmatrix} \bar{Y}_{..} \\ \bar{X}_{..} \end{pmatrix},$$

where $\bar{Y}_{..}$ is the average of the sample means for the 24PAR data and $\bar{X}_{..}$ is the average of the sample means for the SP2 data. Also let

$$\mathbf{m_2} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ & m_{22} & m_{23} & m_{24} \\ & & m_{33} & m_{34} \\ & & & m_{44} \end{pmatrix}$$

be the symmetric sample covariance matrix of $\{\mathbf{Z_i}\}$. We focus only on the upper 2x2 and lower 2x2 diagonals of this matrix because the off diagonal components $(m_{13}, m_{14}, m_{23}, m_{24})$ are assumed to be zero. In deriving values for $\mathbf{m_2}$ we wish to remove gender effects. Thus, the sample moments are calculated from residuals that estimate the error terms, ϵ_{ij} and η_{ij}, in the simple linear regressions

$$Y_{ij} = \mu_0 + \mu_1 G_i + \epsilon_{ij}$$

and

$$X_{ij} = \gamma_0 + \gamma_1 G_i + \eta_{ij},$$

where Y_{ij} and X_{ij} are still the EE values from the 24PAR and SP2, respectively, for individual i on day j, and G_i is equal to one if subject i is female and is zero if subject i is male. From the four sets of residuals $\{\hat{\epsilon_{i1}}\}$, $\{\hat{\epsilon_{i2}}\}$, $\{\hat{\eta_{i1}}\}$, and $\{\hat{\eta_{i2}}\}$, we create a residual vector for $\{\mathbf{Z_i}\}$, $(\hat{\bar{\epsilon_{i.}}}, \hat{\bar{\eta_{i.}}}, \hat{\epsilon_{i1}} - \hat{\epsilon_{i2}}, \hat{\eta_{i1}} - \hat{\eta_{i2}})^T$, where

$$\hat{\bar{\epsilon_{i.}}} = \frac{\hat{\epsilon_{i1}} + \hat{\epsilon_{i2}}}{2}$$

and

$$\hat{\bar{\eta_{i.}}} = \frac{\hat{\eta_{i1}} + \hat{\eta_{i2}}}{2}$$

for all i. Using these sets of values we define $\mathbf{m_2}$ more concretely as the sample moments of the residual vector for $\{\mathbf{Z_i}\}$ adjusted for gender effects.

The method of moments estimators are then determined by solving the equations

$$\mathbf{m_1} = E(\mathbf{Z_i})$$

and

$$\mathbf{m_2} = Cov(\mathbf{Z_i})$$

for the parameters in θ in terms of the sample moments (considering only the 2x2 diagonals of $\mathbf{m_2}$). The method of moments estimators are

$$\hat{\mu_x} = \bar{X_{..}}, \tag{3}$$

$$\hat{\beta_0} = \bar{Y_{..}} - \frac{(m_{12} - 0.25m_{34})}{(m_{22} - 0.25m_{44})}\bar{X_{..}}, \tag{4}$$

$$\hat{\beta_1} = \frac{m_{12} - 0.25m_{34}}{m_{22} - 0.25m_{44}}, \tag{5}$$

$$\hat{\sigma_x^2} = m_{22} - 0.25m_{44}, \tag{6}$$

$$\hat{\sigma_r^2} = \frac{m_{34}(m_{22} - 0.25m_{44})}{2(m_{12} - 0.25m_{34})}, \tag{7}$$

$$\hat{\sigma_\alpha^2} = m_{11} - 0.25m_{33} - \frac{(m_{12} - 0.25m_{34})^2}{m_{22} - 0.25m_{44}}, \tag{8}$$

$$\hat{\sigma_e^2} = 0.5m_{33} - \frac{m_{34}(m_{12} - 0.25m_{34})}{2(m_{22} - 0.25m_{44})}, \tag{9}$$

and

$$\hat{\sigma_u^2} = 0.5m_{44} - \frac{m_{34}(m_{22} - 0.25m_{44})}{2(m_{12} - 0.25m_{34})}. \tag{10}$$

Numerical estimates for these estimators will be examined more closely in the next subsection.

Using methods from Fuller (1987) we derive a covariance matrix for parameter estimates, $\hat{V}\{\hat{\theta}\} = D\{\hat{\theta}\}\hat{\mathbf{\Sigma}}D\{\hat{\theta}\}^T$, for sample size n. $D\{\hat{\theta}\}$ is an 8x8 matrix of derivatives constructed using the method of moments equations (3)-(10). $\hat{\mathbf{\Sigma}}$ =block diag($\hat{\mathbf{\Sigma_{11}}}, \hat{\mathbf{\Sigma_{22}}}, \hat{\mathbf{\Sigma_{33}}}$) is the 8x8 covariance matrix for the sample moments constructed using formulas from Fuller (1987), where for sample size n,

$$\hat{\mathbf{\Sigma_{11}}} = \frac{1}{n}\begin{pmatrix} m_{11} & m_{12} \\ m_{12} & m_{22} \end{pmatrix},$$

$$\hat{\mathbf{\Sigma_{22}}} = \frac{1}{n-1}\begin{pmatrix} 2m_{11}^2 & 2m_{11}m_{12} & 2m_{12}^2 \\ 2m_{11}m_{12} & m_{11}m_{22} + m_{12}^2 & 2m_{12}m_{22} \\ 2m_{12}^2 & 2m_{12}m_{22} & 2m_{22}^2 \end{pmatrix},$$

and

$$\hat{\mathbf{\Sigma_{33}}} = \frac{1}{n-1}\begin{pmatrix} 2m_{33}^2 & 2m_{33}m_{34} & 2m_{34}^2 \\ 2m_{33}m_{34} & m_{33}m_{44} + m_{34}^2 & 2m_{34}m_{44} \\ 2m_{34}^2 & 2m_{34}m_{44} & 2m_{44}^2 \end{pmatrix}.$$

Variances of parameter estimates for various sample sizes will be examined in the next subsection.

4.3 Results

In this subsection we present the parameter estimates and variances and determine a suitable sample size for our subpopulations of interest. The numeric values for the sample moments we consider are

$$\mathbf{m_1} = \begin{pmatrix} 2,092 \\ 1,928 \end{pmatrix}$$

and

$$\mathbf{m_2} = \begin{pmatrix} 275,243 & 205,914 & & \\ & 193,506 & & \\ & & 262,959 & 232,685 \\ & & & 246,461 \end{pmatrix}.$$

We obtain parameter estimates and variances for various proposed sample sizes using formulas from the previous subsection. Parameter estimates (including $\bar{Y_{..}}$ for $\mu_Y = \beta_0 + \beta_1\mu_x$), standard errors for sample sizes 100, 200, 300, and 600, and corresponding coefficients of variation (CVs) are displayed in Table 2.

The average EE estimates from the 24PAR ($\bar{Y_{..}} = 2092$) and SP2 ($\bar{X_{..}} = 1928$) are similar. The estimate for β_1 is close to 1 and the estimate for β_0 is slightly larger than 0 (relative to the size of the average EE estimates), suggesting that for these study subjects there is close to a one-to-one relationship between usual daily EE and EE estimated from the 24PAR. This might be expected with the high PA awareness of the study subjects. In larger studies with a less homogeneous group of subjects, the estimates for β_1 and β_0 may be further away from 1 and 0, respectively. It is difficult to directly interpret the other parameter estimates without any means of comparison to estimates from other data. We do however hypothesize that the variance estimates will be larger using data from more general PA studies with less homogeneous groups of subjects.

Clearly standard errors decrease as sample size increases because a larger sample provides more information about the population, and hence more precise parameter estimates. As expected, we see smaller and smaller differences in standard errors as n increases.

We are primarily concerned with having small standard errors for the parameters β_1 and β_0 since

Table 2: Parameter Estimates, Standard Errors and Coefficients of Variation

Parameter	μ_Y	μ_x	β_0	β_1	σ_x^2	σ_r^2	σ_u^2	σ_e^2	σ_α^2
Estimate	2,092	1928	-67.31	1.12	132,000	104,000	19,400	1,200	44,000
Std. Error	117	98	429	0.191	66,000	67,000	39,700	28,400	19,000
Standard errors for various sample size n									
$n = 100$	52	44	189	0.084	29,000	29,000	17,400	12,400	8,000
$n = 200$	37	31	133	0.059	20,000	21,000	12,300	8,800	6,000
$n = 300$	30	25	109	0.048	17,000	17,000	10,000	7,100	5,000
$n = 600$	21	18	77	0.034	12,000	12,000	7,100	5,000	3,000
Coefficients of variation for parameter estimates for various sample size n									
$n = 100$	0.02	0.02		0.07	0.22	0.28	0.90	10.56	0.18
$n = 200$	0.02	0.02		0.05	0.15	0.20	0.63	7.45	0.14
$n = 300$	0.01	0.01		0.04	0.13	0.16	0.52	6.08	0.11
$n = 600$	0.01	0.01		0.03	0.09	0.11	0.37	4.29	0.07

they are used to calibrate the 24PAR from the SP2. A sample size of $n = 300$ seems to provide small enough standard errors for these parameters, since doubling the sample size to $n = 600$ does not reduce the errors much further. Also, CVs for β_1 and μ_x are low on the order of 5% for samples of $n = 300$. CVs are somewhat larger for parameters σ_α^2, σ_x^2, and σ_r^2 with $n = 300$, ranging from 11% to 16%. CVs are quite a bit larger for the variance parameters of the overall error terms. This suggests that it may be difficult to precisely estimate some of the variance components (particularly σ_e^2 and σ_u^2) in large-scale PA studies even with large sample sizes.

We must be careful about over-interpreting results from these data because they are calculated using estimates from a small, homogeneous group of study subjects. We expect the standard errors for parameter estimates to be larger when calculated with data from large-scale PA surveys. Based on the results we have and other cost factors, we contend that a sample size of $n = 300$ for the subpopulations of interest (based on gender and living environment) will be large enough to provide suitable parameter estimates and variances.

5. Discussion

In the sections above we have provided background information on PA studies and measurement error models, proposed our own measurement error models for 24PAR and SP2 data, derived parameter estimates using pilot study data, and determined an appropriate sample size for subpopulations in a large-scale PA survey. Our models are constructed to calibrate the 24PAR from the SP2 and incorporate random measurement error, subject-specific error from the 24PAR, and short term deviation in EE for each subject. The model parameter estimates for β_1 and β_0 suggest a linear relationship between EE estimates from the 24PAR and SP2. The standard errors and CVs in Table 2 help us better understand variability properties of the estimates across different sample sizes. Using these estimates and taking other cost restrictions into account, we propose a sample size of 300 subjects for subpopulations based on gender and living environment for a large-scale PA survey.

There are several issues that we still need to address. First and foremost we must collect data from a larger PA survey with a more general group of study subjects and fit our models to the data. This will provide us with more realistic parameter estimates from a more general sample of subjects with varying levels of PA. With these data and our models, we hope to develop reliable calibration methods for EE measurements from the 24PAR and similar recall instruments. We may also want to conduct some model comparisons using PA data to see if our models perform better than some of the ones mentioned in Section 2.1. Of course the models used for model comparisons in Kipnis et al. (2001) may have to be modified for PA data before directly comparing them to our models in Section 3. Our models could also be expanded to include EE data measured with other types of instruments. The models in Section 3 are specifically designed for the 24PAR and SP2 measuring devices. We could include other measurement error model equations into our model structure for EE measurements from doubly-labeled water, other types of recalls and other sensing devices.

The most important, overall goal is to develop appropriate measurement error models for PA data in order to more accurately estimate characteristics of the distribution of usual EE from inexpensive instruments such as the 24PAR. With better parameters describing the distribution of usual daily EE, we will be more able to investigate the relationships between EE and prominence of certain fatal diseases such as cancer, heart disease, and obesity.

REFERENCES

Calabro, M.A., Welk, G.J., Beyler, N.K., Carriquiry, A.L., Nusser, S.M., Matthews, C.E., (2007) Validation of a Self-Report Instrument Using Two Pattern Recognition Activity Monitors. Iowa State University. Reference Type: Thesis/Dissertation.

Freedman, L.S., Carroll, R.J., Wax, Y., (1991) Estimating the Relationship Between Dietary Intake Obtained From a Food Frequency Questionnaire and True Average Intake. *American Journal of Epidemiology*, **134**, 310-320.

Fruin, M.L., Rankin, J.W., (2004) Validity of a Multi-Sensor Armband in Estimating Rest and Exercise Energy Expenditure. *Medicine and Science in Sports and Exercise*, 1063-1069.

Fuller, W.A., (1987) *Measurement Error Models*. John Wiley & Sons Inc., Hoboken, New Jersey.

King, G.A., Torres, N., Potter, C., Brooks, T.J., Coleman, K.J., (2004) Comparison of Activity Monitors to Estimate Energy Cost of Treadmill Exercise. *Medicine and Science in Sports and Exercise*, 1244-1251.

Kipnis, V., Carroll, R.J., Freedman, L.S., Li, L., (1999) Implications of a New Dietary Measurement Error Model for Estimation of Relative Risk: Application to Four Calibration Studies. *American Journal of Epidemiology*, **150**, 642-651.

Kipnis, V., Midthune, D., Freedman, L.S., Bingham, S., Schatzkin, A., Subar, A.F., Carroll, R.J., (2001) Empirical Evidence of Correlated Biases in Dietary Intake Assessment Instruments and Its Implications. *American Journal of Epidemiology*, **153**, 394-403.

Kipnis, V., Subar, A.F., Midthune, D., Freedman, L.S., Ballard-Barbash, R., Troiano, R.P., Bingham, S., Schoeller, D.A., Schatzkin, A., Carroll, R.J., (2003) Structure of Dietary Measurement Error: Results of the OPEN Biomarker Study. *American Journal of Epidemiology*, **158**, 14-21.

Rosner, B, Willett, W.C., Spiegelman, D., (1989) Correction of Logistic Regression Relative Risk Estimates and Confidence Intervals for Systematic Within-person Measurement Error. *Statistics in Medicine*, **8**, 1051-1069.

Spiegelman, D., Schneeweiss, S., McDermott, A., (1997) Measuremnt Error Correction for Logistic Regression Models With an Alloyed Gold Standard. *American Journal of Epidemiology*, **145**, 184-196.

Wong, M.Y., Day, N.E., Luan, J.A., Wareham, N.J., (2004) Estimation of Magnitude in Gene-environment Interactions in the Presence of Measurement Error. *Statistics in Medicine*, **23**, 987-998.

Zhang, K, Werner, P., Sun, M., Pi-Sunyer, F.X., Boozer, C.N., (2003) Measurement of Human Daily Physical Activity. *Obsesity Research*, 33-40.