Published in final edited form as:

J Appl Stat. 2016 ; 43(9): 1706–1721. doi:10.1080/02664763.2015.1117593.

Methods to Assess Measurement Error in Questionnaires of Sedentary Behavior

Joshua N Sampson^{1,*}, Charles E Matthews⁵, Laurence Freedman³, Raymond J. Carroll⁴, and Victor Kipnis²

¹Biostatistics Branch, DCEG, National Cancer Institute; Rockville, MD

²Biometry Research Group, DCP, National Cancer Institute; Rockville, MD

³Biostatistics Unit, Gertner Institute; Tel Hashomer, Israel

⁴Department of Statistics, Texas A&M University, College Station, TX and School of Mathematical Sciences, University of Technology Sydney, Broadway NSW

⁵Nutritional Epidemiology Branch, DCEG, National Cancer Institute; Rockville, MD

Abstract

Sedentary behavior has already been associated with mortality, cardiovascular disease, and cancer. Questionnaires are an affordable tool for measuring sedentary behavior in large epidemiological studies. Here, we introduce and evaluate two statistical methods for quantifying measurement error in questionnaires. Accurate estimates are needed for assessing questionnaire quality. The two methods would be applied to validation studies that measure a sedentary behavior by both questionnaire and accelerometer on multiple days. The first method fits a reduced model by assuming the accelerometer is without error, while the second method fits a more complete model that allows both measures to have error. Because accelerometers tend to be highly accurate, we show that ignoring the accelerometer's measurement error, can result in more accurate estimates of measurement error in some scenarios. In this manuscript, we derive asymptotic approximations for the Mean-Squared Error of the estimated parameters from both methods, evaluate their dependence on study design and behavior characteristics, and offer an R package so investigators can make an informed choice between the two methods. We demonstrate the difference between the two methods in a recent validation study comparing Previous Day Recalls (PDR) to an accelerometer-based ActivPal.

Keywords

measurement error; PDR; questionnaire; sedentary

1. Introduction

Sedentary behavior has already been associated with mortality, cardiovascular disease, and multiple cancers (Matthews et al., 2012a; Thorp et al., 2011). To refine risk estimates and

*contact: joshua.sampson@nih.gov.

evaluate other potential associations, large epidemiological studies will need to accurately measure the number of hours per day that an individual spends sitting or lying down. As many large studies cannot afford to distribute devices such as the activPal (Dowd et al., 2012) or accelerometer (Esliger et al., 2011) to all participants, there is a need to determine whether previous day recalls (PDRs) (Matthews et al., 2012c), or questionnaires inquiring about an individual's activity over the past 24 hours, can accurately measure sedentary time. For this purpose, validation studies can be designed that measure sedentary behavior by both a PDR and a reliable device on a small group of individuals over time.

Our objective is to quantify the bias and error in a PDR's measure of an individual's sedentary time over *a single day*. Clearly, if the PDR is to be a useful tool, it should accurately reflect, with little error or bias, the true sedentary time of the targeted day. In order to make a statement about its quality, we need to assess its "measurement error". However, unlike the majority of papers discussing measurement error in epidemiology (Ferrari et al., 2007; Kipnis et al., 1999; Kipnis and Freedman, 2008; Sampson et al., 2013; Thiebaut et al., 2007), we are not assessing the level that is directly relevant for estimating effect attenuation. Those papers focus on the error in measuring *the usual, or average, level over time*. To enter that discussion, we would need to address how sedentary time from a single day varies around the individual's average level. Our objective is more modest. We accept that with a sufficient number of accurate daily measurements per individual, we can capture an individual's usual level. Our focus, instead, is to propose a method for determining whether a specific PDR is an appropriate tool for measuring daily sedentary behavior, or equivalently, a method for determining the best possible PDR among a set of options.

In this manuscript, we introduce, discuss and compare two statistical methods that quantify the bias and variability in a PDR's measure of sedentary time for a single day. The first method treats the device's estimate as an error-prone "alloyed gold standard" (Spiegelman et al., 1997), and fits a more complete measurement error model (Plummer and Clayton, 1993a; Plummer and Clayton, 1993b; Rosner et al., 2008) that allows for error in the device. The second method treats the device's estimate of sedentary time as the true value and fits a naïve, or reduced, regression model (Buonaccorsi, 2000; Carroll et al., 2006; Higgins et al., 1997; Tosteson et al., 1998; Wang et al., 1999). The first method produces unbiased estimates, while the second method can produce estimates with lower variance. For a given study, their relative performance, as measured by the Mean Squared Error (MSE, variance + bias²), depends on the study design, characteristics of the measured behavior, and device accuracy. In contrast to nutritional studies (Freedman et al., 2011; Paeratakul et al., 1998), alloyed gold standards for sedentary behavior can be extremely accurate (Grant et al., 2006; Kozey-Keadle et al., 2011; Matthews et al., 2013a) and fitting the full model is no longer the obvious choice.

The full measurement error model, as described in the next section, requires that the true sedentary times for an individual are correlated over the observed days. This correlation is necessary for distinguishing the measurement error from the PDR and that from the device. This correlation also induces a relationship between the PDR and device measures over time, a feature omitted in some more basic models (Plummer and Clayton, 1993a; Plummer

and Clayton, 1993b). Otherwise, our chosen model follows standard form, with the acknowledgement that we assume that the device has neither scale nor person-specific bias (Kaaks et al., 1994; Kipnis et al., 1999), seemingly reasonable assumptions for accelerometer-based measures of sedentary behavior.

In this paper, we offer two sets of estimates, $\hat{\Theta}_F$ (based on the full measurement error model) and $\hat{\Theta}_N$ (based on the naïve model), for the scale-bias, subject-specific bias, and random error of a PDR. We then derive the variance of $\hat{\Theta}_N$, which along with the previously derived bias (Wang et al., 1998) of $\hat{\Theta}_F$ and variance of the unbiased $\hat{\Theta}_F$ allows comparison of the MSE for the two sets of estimates. In the results section, we evaluate the effect of study size, device error, and intra-individual variability, among other factors, on the relative performance of $\hat{\Theta}_F$ and $\hat{\Theta}_N$. Our corresponding R package (<https://github.com/sampsonj74/SEandBias>) allows investigators to compare these two estimates for their own studies. Moreover, through simulations, we consider the potential harm of misspecifying the correlation structure for method 2 and the possibility of letting the data choose the better estimate for a given study. Finally, we conclude with a short discussion section.

2. Methods

2.1 Model Assumptions

We consider a validation study that measures the number of hours spent sitting using both a PDR and device on multiple days. For individual i , $i \in \{1, \dots, n\}$, in the validation study, let X_{ij} be the true number of hours spent sitting during day j , $j \in \{1, \dots, N\}$. We consider both the device's, W_{ij} , and the PDR's, V_{ij} , measure of this quantity. We let an arrow over the variable name indicate a column vector, with $\vec{X}_i = [X_{i1}, X_{i2}, \dots, X_{iN}]^t$, $\vec{W}_i = [W_{i1}, W_{i2}, \dots, W_{iN}]^t$, and $\vec{V}_i = [V_{i1}, V_{i2}, \dots, V_{iN}]^t$. Let $T_i = E[X_{ij}]$ be an individual's average, or usual, number of hours per day sitting during the observational period. We assume that $\{T_i, \vec{X}_i, \vec{W}_i, \vec{V}_i\}$ follows a multivariate normal distribution that can be described by

$$V_{ij} = \beta_0 + \beta_1 X_{ij} + r_i + \varepsilon_{ij} \quad (1)$$

$$W_{ij} = X_{ij} + U_{ij} \quad (2)$$

$$X_{ij} = T_i + \delta_{ij} \quad (3)$$

where r_i , ε_{ij} , U_{ij} , and T_i are independent normally distributed variables with mean = 0, and respective variances of σ_r^2 , σ_ε^2 , σ_U^2 , and σ_T^2 . Finally, $\vec{\delta}_i \sim N(\vec{0}, \sigma_\delta^2 \Sigma)$ and is independent of the other variables. Equation 1 includes the assumption that the correlation in the PDR's measurement error is not a function of time. Equation 2 includes a similar assumption of the device's measurement error and the additional assumption that there is neither scale- nor reporting- bias in the device's measure. These assumptions will be discussed later.

2.2 Methods to estimate $\Theta \equiv \{\beta_0, \beta_1, \sigma^2, \alpha^2\}$

Because our interest is in the performance of the PDR, our objective is to obtain estimates, $\hat{\Theta}$, of $\Theta \equiv \{\beta_0, \beta_1, \sigma_r^2, \alpha_\varepsilon^2\}$. Our first option is to consider the full model (equations 1 – 3). Here, we would obtain the maximum likelihood or full model estimates,

$\hat{\Theta}_F \equiv \{\hat{\beta}_{F0}, \hat{\beta}_{F1}, \hat{\sigma}_{Fr}^2, \hat{\sigma}_{F\varepsilon}^2\}$ for the likelihood defined in the appendix, although estimates could be calculated using regression calibration or one of the other available options (Bartlett et al., 2009; Tosteson et al., 1998; Wang et al., 1998; Zhong et al., 2002; Wang et al., 2000; Wang et al., 1999). For example, we could also have estimated the parameters by a methods of moments approach. In this approach, we first calculate the empirical covariance matrix of V and W [Ξ in appendix A.1] and then solve the equations (A.1.1 – A.1.6) relating the terms in this covariance matrix to the parameters of interest. This approach shows that all quantities are identifiable, with the caveat that using an autocorrelation or Toeplitz structure for Σ requires at least three observations per person. We know that $\hat{\Theta}_F$ is asymptotically unbiased and its variance can be estimated by the inverse of the information matrix. Our second option is to treat the device's measure, W , as the truth and fit the naïve model.

$$V_{ij} = \beta_{N0} + \beta_{N1} W_{ij} + r_i^N + \varepsilon_{ij}^N \quad (4)$$

where r_i^N and ε_{ij}^N are presumed to be independent and normally distributed with means 0 and variances σ_{Nr}^2 and $\sigma_{N\varepsilon}^2$. We let $\hat{\Theta}_N \equiv \{\hat{\beta}_{N0}, \hat{\beta}_{N1}, \hat{\sigma}_{Nr}^2, \hat{\sigma}_{N\varepsilon}^2\}$ be the solutions to the score equations corresponding to equation 4 (details in the appendix). The estimates $\hat{\Theta}_N$ are known to be biased (Buonaccorsi, 2000; Tosteson et al., 1998). The asymptotic bias for $\hat{\Theta}_N$ has been recently evaluated (Wang et al., 1998), and the delta method can be used to calculate the asymptotic variance of $\hat{\Theta}_N$ (details in the appendix).

For correcting the observed effect, $\hat{\gamma}_{obs}$, of a questionnaire's measure of sedentary behavior on an outcome, the observed effect can be divided by the attenuation factor, λ (Matthews et al., 2012b).

$$\lambda(\beta_1, \sigma_T^2, \sigma_\delta^2, \sigma_r^2, \sigma_\varepsilon^2) = \frac{\beta_1(\sigma_T^2 + \sigma_\delta^2)}{\beta_1^2(\sigma_T^2 + \sigma_\delta^2) + \sigma_r^2 + \sigma_\varepsilon^2}$$

This attenuation factor only adjusts for the questionnaire measurement error at a given time point, and does not adjust for a single time point's possibly poor reflection of an individual's usual level. Given the potential importance of λ , we also compare the MSE for

$\hat{\lambda}_N = \lambda(\hat{\beta}_{N1}, \hat{\sigma}_{NT}^2, \hat{\sigma}_{N\delta}^2, \hat{\sigma}_{Nr}^2, \hat{\sigma}_{N\varepsilon}^2)$ and $\hat{\lambda}_F = \lambda(\hat{\beta}_{F1}, \hat{\sigma}_{FT}^2, \hat{\sigma}_{F\delta}^2, \hat{\sigma}_{Fr}^2, \hat{\sigma}_{F\varepsilon}^2)$. When the naïve model is assumed true, W_{ij} is presumed to equal $T_i + \delta_{ij}$, $\hat{\sigma}_{NT}^2$ and $\hat{\sigma}_{N\delta}^2$ can be estimated from the distribution of W_{ij} , and $E[\hat{\sigma}_{N\delta}^2] = \sigma_\delta^2 + \sigma_U^2$. Because $\hat{\lambda}_N$ and $\hat{\lambda}_F$ are not normally distributed, we calculate their MSE by (i) generating 100,000 sets of $\hat{\Theta}_N$ and $\hat{\Theta}_F$ from their respective asymptotic normal distributions (ii) applying the function $\lambda(\cdot)$ to each set of parameters (iii)

calculating the MSE. Another metric of PDR quality is the correlation,

$\rho_F^{PDR}(\beta_1, \sigma_T^2, \sigma_\delta^2, \sigma_r^2, \sigma_\varepsilon^2)$, between the PDR measure and the truth:

$$\rho_F^{PDR}(\beta_1, \sigma_T^2, \sigma_\delta^2, \sigma_r^2, \sigma_\varepsilon^2) = \frac{\beta_1(\sigma_T^2 + \sigma_\delta^2)}{\sqrt{\beta_1^2(\sigma_T^2 + \sigma_\delta^2) + \sigma_r^2 + \sigma_\varepsilon^2} \sqrt{(\sigma_T^2 + \sigma_\delta^2)}}$$

To estimate the MSE for $\hat{\rho}_{N\wedge}^{PDR}$ and $\hat{\rho}_F^{PDR}$, we use a simulation approach similar to the approach used for $\hat{\lambda}_{N\wedge}$ and $\hat{\lambda}_F$.

2.3 Comparison of Methods

We can calculate the mean-square-error (MSE), $E[(\hat{\phi} - \phi)^2]$, where $\phi \in \{\beta_0, \beta_1, \sigma_r^2, \sigma_\varepsilon^2\}$ for the naïve and full model estimates. We then evaluate how changing the study design and true parameter values affects the relative performance of $\hat{\Theta}_F$ and $\hat{\Theta}_{N\wedge}$. Without closed form solutions for the bias and variance of these estimates, we consider specific scenarios. As our foundation, we start with two primary scenarios, a large error scenario where, as observed in our own validation study, the error of the PDR is 40x the error of the device, $\sigma_\varepsilon^2 = 40\sigma_U^2$, and a small error scenario where we are more optimistic and assume the error of the PDR is only 5x the error of the device, $\sigma_\varepsilon^2 = 5\sigma_U^2$. In our base models, all other parameters reflect those from our own validation studies: $\beta_0 = 0$, $\beta_1 = 1$, $\sigma_T^2 = 4$, $\sigma_\delta^2 = 1.5$, $\sigma_U^2 = 0.1$, $\sigma_r^2 = 0.8$, and Σ is known to be the identity matrix. Furthermore, we assume that the study includes 50 individuals and each individual has four measurements ($N=4$). Using these models as points of reference, we evaluate the effects of changing the values of key parameters.

We will allow the correlation matrix, Σ , to have a structure indicative of independence ($\Sigma_{kl} = 0$ for $k \neq l$), autocorrelation ($\Sigma_{kl} = \rho^{|k-l|}$), or toeplitz ($\Sigma_{ij} = \rho$ if $|k-l| = 1$, $\Sigma_{kl} = 0$ if $|k-l| > 1$). We omit exchangeable ($\Sigma_{kl} = \rho$ for $k \neq l$) because, by letting $\sigma_{T*}^2 = \sigma_T^2 + \rho\sigma_\delta^2$ and $\sigma_{\delta*}^2 = \sigma_\delta^2(1 - \rho)$, exchangeable can be reformulated as independence. With only two measurements per each individual ($N=2$), ρ cannot be identified. All other parameters from the model specified in equations 1 – 3 are identifiable.

2.4 Simulation Studies

Although most comparisons between $\hat{\Theta}_F$ and $\hat{\Theta}_{N\wedge}$ can be performed based on theoretical calculations of means and variances, we rely on simulations for to evaluate our approximations, assess the effect of model misspecification, and to test our two step procedure. First, we generate 100,000 datasets based on equations 1–3 and the primary models, to obtain empirical estimates of biases and variances for $\hat{\Theta}_F$ and $\hat{\Theta}_{N\wedge}$ to ensure that our theoretical calculations are appropriate for small samples. Datasets were also generated assuming autocorrelation and toeplitz ($\rho = 0.2, 0.4$, and 0.6) correlation matrices. Results for $\rho = 0.6$ are presented in table 1, while results for $\rho = 0.2$ and 0.4 are presented in supplementary tables 1–4. Second, we consider the effect of model misspecification on the full model estimates. For this analysis, we generate a similar set of datasets. We then calculate the MSE resulting from fitting the full models that assume each of the three

possible correlation structures. Finally, simulations allowed us to evaluate a data-directed two-step procedure designed to choose the best estimates for a given dataset. In the first step, we obtained the estimates $\{\hat{\beta}_{F0}, \hat{\beta}_{F1}, \hat{\sigma}_{Fr}^2, \hat{\sigma}_{F\epsilon}^2, \hat{\sigma}_{Fu}^2, \hat{\sigma}_{Ft}^2, \hat{\sigma}_{F\delta}^2\}$ from fitting the full model and, treating those estimates as the truth, calculated the MSEs expected for the naïve and full model estimates. In the second step, we chose those estimates that were expected to produce the lower MSE.

2.5 Validation Study

The measurement properties of a Previous Day Recall (PDR) were evaluated as part of a validation study which has been previously described (Matthews et al., 2013b). Briefly, as part of a larger study, 40 men (18 – 71 years old) from Amherst, MA and Nashville, TN wore an activPal and completed a telephone-administered PDR on three different days during a one-week period. The PDR was an updated version of a 24-Hour Physical Activity Recall that has been previously used as a reference instrument (Matthews et al., 2005) and the activPal (PAL Technologies, Glasgow, Scotland) is an accelerometer-based posture and activities logger that is worn on the thigh. Sedentary time was defined as the time sitting or laying during waking hours.

3. Results

We calculated the biases and variances for the three parameters $\hat{\beta}_1$, $\hat{\sigma}_{\epsilon}^2$ and $\hat{\sigma}_r^2$ estimated using both the full and naïve models. Tables 1 and 2 shows that these asymptotic values were accurate, in that they were nearly identical to those estimated from simulation. In addition, Tables 1 and 2 illustrate our general observation that the structure of the within-individual variability had an insignificant effect on the biases and standard errors of the estimates of $\hat{\beta}_1$, $\hat{\sigma}_{\epsilon}^2$ and $\hat{\sigma}_r^2$. Moreover, we also examined the effect of incorrectly specifying the correlation structure when fitting the full model. In practice, this structure will be unknown and the specified structure may be incorrect. However, simulations suggest that the MSE of estimates based on a model that assumes the wrong structure will not be significantly higher than estimates based on assuming the correct model. Table 3 suggests that even when the true structure of Σ is Toeplitz(1) or Autocorrelation, fitting a model where Σ is assumed to be the identity matrix (and the overall correlation matrix for an individual's measurements is exchangeable) will not significantly increase the MSE.

In general, we found that $\text{var}(\hat{\sigma}_{N\epsilon}^2) < \text{var}(\hat{\sigma}_{F\epsilon}^2)$, $\text{var}(\hat{\beta}_{N1}) < \text{var}(\hat{\beta}_{F1})$ and, by simulation, $\text{var}(\hat{\rho}_N^{PDR}) < \text{var}(\hat{\rho}_F^{PDR})$, supporting our belief that the variance for the naïve estimates would be smaller. However, that trend did not hold for $\hat{\sigma}_r^2$, where the naïve estimate often had the same or slightly larger variances. Because of this reversal, the Mean Squared Error (MSE) for $\hat{\sigma}_{Nr}^2$ was often larger than the MSE for $\hat{\sigma}_{Fr}^2$, especially in small-error scenario (figures 1 and 2). In terms of trends or generalizations, we found that the relative performance of the estimates from the full model improved as (i) the number of subjects increased (ii) the variance of the device (σ_v^2) increased (iii) the intra-individual variability (σ_δ^2) decreased (figures 1 and 2). Note, that $\sigma_\delta^2=0$ is equivalent to repeated measures of a

single quantity and the full model becomes similar to standard models designed for repeated measurements. Although determining when the full model estimates would be superior requires consideration of all parameters, our simulations offer some suggestions. When $n < 50$, the variance of the device is less than 5% of the between-individual variability, and the intra-individual variability is greater than 25% of the total variability, the naïve estimates tend to be preferable. The importance of n was limited to the small-error scenario where $\sigma_\varepsilon^2 = 0.5$. This scenario also seemed to favor the full model, in the sense that the parameters would require more extreme values for the reduced model to become superior.

The errors, defined as the square-root of the MSE divided by the parameter's value ($\sqrt{MSE(\phi)/\phi}$), tend to be low for the attenuation factor. For both primary models, the error of $\hat{\lambda}_N$ and $\hat{\lambda}_F$ were less than 10%. Across most scenarios, with the exception of the small-error scenarios with high $\sigma_{U^*}^2$, we found $\hat{\lambda}_N$ and $\hat{\lambda}_F$ to behave similarly. Even in scenarios where $MSE(\hat{\sigma}_{N\varepsilon}^2) < MSE(\hat{\sigma}_{F\varepsilon}^2)$ and $MSE(\hat{\beta}_{N1}) < MSE(\hat{\beta}_{F1})$, the strong negative correlation between $\hat{\beta}_{F1}$ and $\hat{\sigma}_{F\varepsilon}^2$ seemed to offset their effects on $\hat{\lambda}_F$ so that it performed no worse than $\hat{\lambda}_N$ (figures 3, 4). In attempt to choose the best estimates for a given dataset, we created a data-driven two step procedure. In the first step, we obtained the maximum likelihood estimates $\{\hat{\beta}_{F0}, \hat{\beta}_{F1}, \hat{\sigma}_{Fr}^2, \hat{\sigma}_{F\varepsilon}^2, \hat{\sigma}_{Fu}^2, \hat{\sigma}_{Ft}^2, \hat{\sigma}_{F\delta}^2\}$ and, treating those estimates as the truth, calculated the MSEs expected for the naïve and maximum likelihood estimates. In the second step, we chose those estimates that were expected to produce the lower MSE.

Unfortunately, the choice of estimates is usually driven by $\hat{\sigma}_{Fu}^2$ and in those scenarios where we question whether or not we can use the maximum likelihood estimates, we often have a poor estimate of $\hat{\sigma}_{Fu}^2$. For example, in both primary models, the standard deviation of $\hat{\sigma}_{Fu}^2$ is 0.2, resulting in the probability of selecting the inferior parameter being as high as 0.3.

Moreover, $\hat{\sigma}_{Fu}^2$ is highly correlated with estimates of the desired parameters. From simulation, the $\text{cor}(\hat{\sigma}_{Fu}^2, \hat{\beta}_{F1})$, $\text{cor}(\hat{\sigma}_{Fu}^2, \hat{\sigma}_{F\varepsilon}^2)$, and $\text{cor}(\hat{\sigma}_{Fu}^2, \hat{\sigma}_{Fr}^2)$ in the small-error primary model were 0.8, -0.8, and -0.1. Although the reduced model was expected to perform better for the small-error primary model, this correlation seemed to result in the full-model estimates being selected when they were farthest from the truth. Hence, even though the naïve estimates are selected more often, this two step procedure offered little benefit. Ultimately, across all parameter values tested, we found that our two-step estimates performed similarly to the full-model estimates.

As a practical demonstration of two models, we considered a study measuring hours of sedentary behavior by both PDR and activPal. Studies comparing the activPal to the truth, as determined by an observer, have suggested that the correlation between the two measures in a population is as high with $r > 0.95$, in both laboratory studies (Grant et al., 2006) and during free living activities (Kozey-Keadle et al., 2011; Lyden et al., 2012), suggesting that the reduced estimates will likely have lower MSE. In our study, we found the estimates of $\hat{\beta}_{N1}$, $\hat{\sigma}_{N\varepsilon}^2$ and $\hat{\sigma}_{Nr}^2$ to be 0.98, 3.02, and 0.81. We omit confidence intervals as standard estimates do not account for the bias. In contrast, had we chosen the full-model estimates, our estimates (and confidence intervals) would have been $\hat{\beta}_{F1}$, $\hat{\sigma}_{F\varepsilon}^2$ and $\hat{\sigma}_{Fr}^2$ to be 1.18 (0.81

– 1.55), 2.05 (0.32 – 3.77), and 1.26 (0.14 – 2.39). Although these results are similar, this example still demonstrates that both methods can be applied to this type of validation study.

4. Discussion

Epidemiological studies will aim to measure the “usual” level, or long-term patterns, of sedentary behavior. This aim can be accomplished by a questionnaire inquiring about usual behavior or by using accurate measures of sedentary behavior on one or more specific days as an approximation of usual levels. Because questionnaires of long-term behavior, especially sedentary behavior, can be highly inaccurate due to the implicit difficulty of retrieving and organizing a great deal of information, there is mounting preference for the latter option (Matthews et al., 2012c; Patterson, 2000; Sallis and Saelens, 2000). PDRs offer an affordable option for measuring sedentary time during a given day. However, both the accuracy and precision of a PDR need to be tested in a validation study before being incorporated into a larger study. As it is nonsensical to record multiple PDR’s on the same day in a given individual, validation studies often record multiple PDRs over time. If the validation study also measures sedentary behavior by an unbiased device on those same days, as is the case in many studies, then we can assess the bias and variability of the PDR. In the manuscript, we consider two methods for such an evaluation. The first method treats the device’s estimate of sedentary time as the true value, whereas the second acknowledges that the device’s estimate also includes error.

We can obtain estimates of the desired parameters, specifically those describing the accuracy and bias of the PDR, by either fitting the full or naïve model. The better option, as defined by the lower MSE, depends on the exact characteristics in the truth. Nevertheless, we found some basic trends worth reporting. The relative performance of the estimates from the full model improved as (i) the number of subjects increased (ii) the variance of the device (σ_v^2) increased (iii) the intra-individual variability (σ_δ^2) decreased. Although far from a definitive rule, we found that the naïve estimates tend to be preferable when $n < 50$, the variance of the device is less than 5% of the between-individual variability, and the intra-individual variability is less than 25% of the total variability.

Our research adds to the existing, rich, literature describing statistical methodology for assessing the accuracy of questionnaires. This literature includes multiple methods that assess accuracy when only an error-prone surrogate of the truth is available. Fitting some version of equations 1–3 is a common approach (Nusser et al., 2012; Rosner et al., 2008). However, the methodology has primarily focused on fields, such as diet (Freedman et al., 1991; Kaaks et al., 1994) and physical activity (Atkinson and Nevill, 1998; Bowles, 2012; Chasan-Taber et al., 1996; Matthews et al., 2012c; Nusser et al., 2012; Ferrari et al., 2007), where there is comparatively poor correlation between the surrogate and truth. Here, we address the issue of questionnaire accuracy for sedentary behavior, where accurate surrogates are likely available. To address this issue, we extended the statistical evaluation of bias and variance in parameters from both the naïve and full random-effects models (Bartlett et al., 2009; Tosteson et al., 1998; Wang et al., 2000; Wang et al., 1998; Wang et al., 1999; Zhong et al., 2002). In this context, our discussion is one of the first to focus on a

comparison of the Mean-Square Errors between the two types of parameters and is the first to calculate the variance of the naïve parameter estimates.

Our method was designed for studies where there is a single, unbiased, measure from a device accompanying each PDR measure. If individuals had used multiple devices on each day, then these replicate measures would have been sufficient for estimating σ_U^2 . In turn, we would have had alternative options, requiring fewer assumptions, for estimating the desired parameters. Our method also required the device to be unbiased. Given the data collected, we would not have the ability to distinguish bias from the device and PDR. Although we believe this to be a realistic assumption, it is also a necessary assumption. The objective in epidemiological studies is to capture an individual's usual sedentary behavior. Therefore, the PDR will likely be administered at multiple times during the year. The next step should be to evaluate the correlation structure of daily sedentary behavior and the correlation structure of its measurement error over longer periods of time, such as a year. Such information will help determine the number of measurements needed to get an accurate estimate of usual levels. A more general limitation is that our method was designed to assess how well a PDR captures one specific quantity, such as total number of hours in a sitting or lying position. However, a PDR can measure an extensive range of activities, and therefore its true value should not be assessed by a single metric.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This research was supported by the intramural research program at the NIH (JNS, CEM) and by a grant from the National Cancer Institute, U01-CA057030 (RJC).

References

- Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports medicine*. 1998; 26:217–238. [PubMed: 9820922]
- Bartlett JW, De Stavola BL, Frost C. Linear mixed models for replication data to efficiently allow for covariate measurement error. *Statistics in Medicine*. 2009; 28:3158–3178. [PubMed: 19777493]
- Bowles HR. Measurement of active and sedentary behaviors: closing the gaps in self-report methods. *Journal of physical activity & health*. 2012; 9(Suppl 1):S1–S4. [PubMed: 22287442]
- Buonaccorsi JD. Eugene; Tosteson, Tor. ESTIMATION IN LONGITUDINAL RANDOM EFFECTS MODELS WITH MEASUREMENT ERROR. *Statistica Sinica*. 2000; 10:885–903.
- Carroll, RJ.; Ruppert, D.; Stefanski, LA., et al. *Measurement Error in Nonlinear Models*. Boca Raton: Chapman and Hall, CRC; 2006.
- Chasan-Taber S, Rimm EB, Stampfer MJ, et al. Reproducibility and validity of a self-administered physical activity questionnaire for male health professionals. *Epidemiology*. 1996:81–86. [PubMed: 8664406]
- Dowd KP, Harrington DM, Donnelly AE. Criterion and Concurrent Validity of the activPAL™ Professional Physical Activity Monitor in Adolescent Females. *PLoS ONE*. 2012; 7:e47633. [PubMed: 23094069]
- Eslinger DW, Rowlands AV, Hurst TL, et al. Validation of the GENE Accelerometer. *Medicine & Science in Sports & Exercise*. 2011; 43:1085–1093. 10.1249/MSS.1080b1013e31820513be. [PubMed: 21088628]

- Ferrari P, Friedenreich C, Matthews CE. The Role of Measurement Error in Estimating Levels of Physical Activity. *American Journal of Epidemiology*. 2007; 166:832–840. [PubMed: 17670910]
- Freedman LS, Carroll RJ, Wax Y. Estimating the Relation between Dietary Intake Obtained from a Food Frequency Questionnaire and True Average Intake. *American Journal of Epidemiology*. 1991; 134:310–320. [PubMed: 1877589]
- Freedman LS, Schatzkin A, Midthune D, et al. Dealing with dietary measurement error in nutritional cohort studies. *J Natl Cancer Inst*. 2011; 103:1086–1092. [PubMed: 21653922]
- Grant PM, Ryan CG, Tigbe WW, et al. The validation of a novel activity monitor in the measurement of posture and motion during everyday activities. *Br J Sports Med*. 2006; 40:992–997. [PubMed: 16980531]
- Higgins K, Davidian M, Giltinan D. A Two-Step Approach to Measurement Error in Time-Dependent Covariates in Nonlinear Mixed-Effects Models, With Application to IGF-I Pharmacokinetics. *Journal of the American Statistical Association*. 1997; 92:436–448.
- Kaaks R, Riboli E, Estève J, et al. Estimating the accuracy of dietary questionnaire assessments: Validation in terms of structural equation models. *Statistics in Medicine*. 1994; 13:127–142. [PubMed: 8122049]
- Kipnis V, Carroll RJ, Freedman LS, et al. Implications of a New Dietary Measurement Error Model for Estimation of Relative Risk: Application to Four Calibration Studies. *American Journal of Epidemiology*. 1999; 150:642–651. [PubMed: 10490004]
- Kipnis V, Freedman LS. Impact of exposure measurement error in nutritional epidemiology. *J Natl Cancer Inst*. 2008; 100:1658–1659. [PubMed: 19033567]
- Kozey-Keadle S, Libertine A, Lyden K, et al. Validation of wearable monitors for assessing sedentary behavior. *Med Sci Sports Exerc*. 2011; 43:1561–1567. [PubMed: 21233777]
- Lyden K, Kozey Keadle SL, Staudenmayer JW, et al. Validity of two wearable monitors to estimate breaks from sedentary time. *Med Sci Sports Exerc*. 2012; 44:2243–2252. [PubMed: 22648343]
- Matthews CE, Ainsworth BE, Hanby C, et al. Development and Testing of a Short Physical Activity Recall Questionnaire. *Medicine & Science in Sports & Exercise*. 2005; 37
- Matthews CE, George SM, Moore SC, et al. Amount of time spent in sedentary behaviors and cause-specific mortality in US adults. *The American Journal of Clinical Nutrition*. 2012a; 95:437–445. [PubMed: 22218159]
- Matthews CE, Keadle SK, Sampson J, et al. Validation of a previous-day recall measure of active and sedentary behaviors. *Med Sci Sports Exerc*. 2013a; 45:1629–1638. [PubMed: 23863547]
- Matthews CE, Keadle SK, Sampson J, et al. Validation of a Previous-Day Recall Measure of Active and Sedentary Behaviors. *Medicine & Science in Sports & Exercise*. 2013b; 45:1629–1638. 1610.1249/MSS.1620b1013e3182897690. [PubMed: 23863547]
- Matthews CE, Moore SC, George SM, et al. Improving Self-Reports of Active and Sedentary Behaviors in Large Epidemiologic Studies. *Exercise and Sport Sciences Reviews*. 2012b; 40
- Matthews CE, Moore SC, George SM, et al. Improving Self-Reports of Active and Sedentary Behaviors in Large Epidemiologic Studies. *Exercise and Sport Sciences Reviews*. 2012c; 40:118–126. 110.1097/JES.1090b1013e31825b31834a31820. [PubMed: 22653275]
- Nusser SM, Beyler NK, Welk GJ, et al. Modeling errors in physical activity recall data. *Journal of Physical Activity and Health*. 2012; 9:S56. [PubMed: 22287449]
- Paeratakul S, Popkin BM, Kohlmeier L, et al. Measurement error in dietary data: implications for the epidemiologic study of the diet-disease relationship. *Eur J Clin Nutr*. 1998; 52:722–727. [PubMed: 9805218]
- Patterson P. Reliability, validity, and methodological response to the assessment of physical activity via self-report. *Res Q Exerc Sport*. 2000; 71:S15–S20. [PubMed: 10925820]
- Plummer M, Clayton D. Measurement error in dietary assessment: An investigation using covariance structure models. Part I. *Statistics in Medicine*. 1993a; 12:925–935. [PubMed: 8337549]
- Plummer M, Clayton D. Measurement error in dietary assessment: An investigation using covariance structure models. Part II. *Statistics in Medicine*. 1993b; 12:937–948. [PubMed: 8337550]
- Rosner B, Michels KB, Chen YH, et al. Measurement error correction for nutritional exposures with correlated measurement error: Use of the method of triads in a longitudinal setting. *Statistics in Medicine*. 2008; 27:3466–3489. [PubMed: 18416440]

- Sallis JF, Saelens BE. Assessment of physical activity by self-report: status, limitations, and future directions. *Res Q Exerc Sport*. 2000; 71:S1–S14. [PubMed: 10925819]
- Sampson JN, Boca SM, Shu XO, et al. Metabolomics in epidemiology: sources of variability in metabolite measurements and implications. *Cancer Epidemiol Biomarkers Prev*. 2013; 22:631–640. [PubMed: 23396963]
- Spiegelman D, Schneeweiss S, McDermott A. Measurement Error Correction for Logistic Regression Models with an “Alloyed Gold Standard”. *American Journal of Epidemiology*. 1997; 145:184–196. [PubMed: 9006315]
- Thiebaut AC, Freedman LS, Carroll RJ, et al. Is it necessary to correct for measurement error in nutritional epidemiology? *Ann Intern Med*. 2007; 146:65–67. [PubMed: 17200225]
- Thorpe AA, Owen N, Neuhaus M, et al. Sedentary Behaviors and Subsequent Health Outcomes in Adults: A Systematic Review of Longitudinal Studies, 1996–2011. *American Journal of Preventive Medicine*. 2011; 41:207–215. [PubMed: 21767729]
- Tosteson TD, Buonaccorsi JP, Demidenko E. Covariate measurement error and the estimation of random effect parameters in a mixed model for longitudinal data. *Statistics in Medicine*. 1998; 17:1959–1971. [PubMed: 9777689]
- Wang CY, Wang N, Wang S. Regression Analysis When Covariates Are Regression Parameters of a Random Effects Model for Observed Longitudinal Measurements. *Biometrics*. 2000; 56:487–495. [PubMed: 10877308]
- Wang N, Lin X, Gutierrez R, et al. Bias analysis and SIMEX approach in generalized linear mixed measurement error models. *Journal of the American Statistical Association*. 1998; 93:249–261.
- Wang N, Lin X, Guttierrez RG. A bias correction regression calibration approach in generalized linear mixed measurement error models. *Communications in Statistics - Theory and Methods*. 1999; 28:217–232.
- Zhong X-P, Fung W-K, Wei B-C. Estimation in Linear Models with Random Effects and Errors-in-Variables. *Annals of the Institute of Statistical Mathematics*. 2002; 54:595–606.

Appendix

A.1 Likelihood

If V and W are defined by equations 1 – 3, then they follow a normal distribution:

$$\begin{bmatrix} V_{i1} \\ \vdots \\ V_{iN} \\ W_{i1} \\ \vdots \\ W_{iN} \end{bmatrix} \sim N \left(\begin{bmatrix} \beta_0 \\ \vdots \\ \beta_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \Xi \right)$$

Where the covariance, Ξ , is defined by the following equations

$$\text{var}(V) = \beta_1^2 \sigma_T^2 + \beta_1^2 \sigma_\delta^2 + \sigma_r^2 + \sigma_\varepsilon^2$$

$$\text{var}(W) = \sigma_T^2 + \sigma_U^2 + \sigma_\delta^2$$

$$\text{cov}(V_{ij}, V_{ik}) = \sigma_r^2 + \beta_1^2(\sigma_T^2 + \sigma_\delta^2 \rho_{jk})$$

$$\text{cov}(W_{ij}, W_{ik}) = \sigma_T^2 + \sigma_\delta^2 \rho_{jk} \quad \text{A.1.1} - \text{A.1.6}$$

$$\text{cov}(V_{ij}, W_{ij}) = \beta_1(\sigma_T^2 + \sigma_\delta^2)$$

$$\text{cov}(V_{ij}, W_{ik}) = \beta_1(\sigma_T^2 + \sigma_\delta^2 \rho_{jk})$$

Calculating the Variance of Naïve Estimates

Recall, that the naïve model presumes

$$V_{ij} = \beta_0^N + \beta_1^N W_{ij} + r_i^N + \varepsilon_{ij}^N$$

However, we are going to greatly simplify notation if remove the superscript “N”. Therefore, for the appendix, unless otherwise stated, all parameters will refer to those from the reduced model. The presumed likelihood for V conditioned on W is then

$$L(V|\theta^N) = (2\pi)^{\frac{n}{2}} |\Sigma_{\text{red}}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(V - \beta_0 - \beta_1 W) \Sigma_{\text{red}}^{-1} (V - \beta_0 - \beta_1 W)\right]$$

where

$$\Sigma_{\text{red}} = \begin{bmatrix} \sigma_r^2 + \sigma_\varepsilon^2 & \sigma_r^2 & \dots & \sigma_r^2 \\ \sigma_r^2 & \sigma_r^2 + \sigma_\varepsilon^2 & & \sigma_r^2 \\ \vdots & & \ddots & \\ \sigma_r^2 & \sigma_r^{N2} & & \sigma_r^2 + \sigma_\varepsilon^2 \end{bmatrix}$$

The resulting score equations are

$$\ell'(\theta) = \begin{bmatrix} \ell'_{\beta_0}(\theta) \\ \ell'_{\beta_1}(\theta) \\ \ell'_{\sigma_r^2}(\theta) \\ \ell'_{\sigma_\varepsilon^2}(\theta) \end{bmatrix} = \begin{bmatrix} 1' \Sigma_{\text{red}}^{-1} (V - \beta_0 - \beta_1 W) \\ W' \Sigma_{\text{red}}^{-1} (V - \beta_0 - \beta_1 W) \\ (V - \beta_0 - \beta_1 W) \Sigma_{\text{red}}^{-1} \frac{d\Sigma_{\text{red}}}{d\sigma_r^2} \Sigma_{\text{red}}^{-1} (V - \beta_0 - \beta_1 W) - \text{tr}(\Sigma_{\text{red}}^{-1} \frac{d\Sigma_{\text{red}}}{d\sigma_r^2}) \\ (V - \beta_0 - \beta_1 W) \Sigma_{\text{red}}^{-1} \frac{d\Sigma_{\text{red}}}{d\sigma_\varepsilon^2} \Sigma_{\text{red}}^{-1} (V - \beta_0 - \beta_1 W) - \text{tr}(\Sigma_{\text{red}}^{-1} \frac{d\Sigma_{\text{red}}}{d\sigma_\varepsilon^2}) \end{bmatrix}$$

We use $\frac{d\Sigma_{\text{red}}}{d\phi}$ to denote an entry $\times$ entry derivative of Σ_{red} relative to ϕ and we use 1 to denote a column of 1's. Our objective is to find the variance of the solutions to those four

equations. Let's denote the true solutions by $\tilde{\theta}$, where $E[\ell(\tilde{\theta})]=0$ and let $\hat{\theta}$ be the solutions from our data.

By a Taylor series expansion, we can approximate

$$\text{var}(\hat{\theta} - \tilde{\theta}) = E[\ell''(\tilde{\theta})]^{-1} \text{var}(\ell'(\tilde{\theta})) (E[\ell''(\tilde{\theta})]^{-1})^t$$

Where $\ell''(\theta)$ is the 4×4 matrix containing the second derivatives. Now, we only need to evaluate the two terms, $\text{var}(\ell(\tilde{\theta}))$ and $E[\ell''(\tilde{\theta})]^{-1}$, on the right. Both are easy. We first note that $\ell(\tilde{\theta})$ includes only the products of central moments of normal random variables with a known covariance matrix, $\Sigma_{\ell} = \text{var}([V^t - \beta_0^N - \beta_1^N W^t \quad W^t])$ and

$$\Sigma_{\ell} = \begin{bmatrix} (\sigma_R^2 + \sigma_{\varepsilon}^2) + (\sigma_T^2 + \sigma_{\delta}^2)(\beta_1^* - \beta_1) + \beta_1^2 \sigma_{\delta}^2 & (\sigma_T^2 + \sigma_{\delta}^2)(\beta_1^* - \beta_1) - \sigma_{\delta}^2 \beta_1 \\ (\sigma_T^2 + \sigma_{\delta}^2)(\beta_1^* - \beta_1) - \sigma_{\delta}^2 \beta_1 & \sigma_T^2 + \sigma_{\delta}^2 + \sigma_U^2 \end{bmatrix}$$

Because β_1 now represents the naïve parameter, we let the parameter defined in equation (1) be denoted by β_1^* . Now, it only remains to calculate $E[\ell''(\tilde{\theta})]^{-1}$. We can do this in two steps. First, we provide formulas for $\ell''(\tilde{\theta})$ for each of the 16 terms. Then, we note that all of their expectations can be calculated from the trick $E[X'BX] = \text{tr}(B\Sigma_X) + \mu_X' B \mu_X$ when X is a normally distributed variable with mean μ_X and variance Σ_X , and B is a matrix.

For simplifying the notation of $\ell''(\tilde{\theta})$, we shall let $W_0=1$ and $W_1=W$. In an abuse of notation, we further let $W=[W_0 \ W_1]$. Let $\beta = [\beta_0 \ \beta_1]$ and $a, b \in \{0, 1, r, \varepsilon\}$ as appropriate. Then

$$\frac{d(W_a^t \Sigma_{\text{red}}^{-1} (V - W\beta))}{d\beta_b} = W_a^t \Sigma_{\text{red}}^{-1} W_b$$

$$\frac{d(W_a^t \Sigma_{\text{red}}^{-1} (V - W\beta))}{d\sigma_b^2} = W_a^t \frac{d\Sigma_{\text{red}}^{-1}}{d\sigma_b^2} (V - W\beta)$$

Moreover, when $\frac{d^2 \Sigma_{\text{red}}}{d\phi} = 0$ for all $\phi \in \theta$, we also have

$$\Sigma_{\text{red}} \frac{d((V - W\beta) \Sigma_{\text{red}}^{-1} \frac{d\Sigma_{\text{red}}^{-1}}{d\sigma_a^2} \Sigma_{\text{red}}^{-1} (V - W\beta))}{d\beta_b} = 2W_b^t \frac{d\Sigma_{\text{red}}^{-1}}{d\sigma_a^2} (V - W\beta)$$

$$\frac{d(\text{tr}(\Sigma_{\text{red}}^{-1} \frac{d\Sigma_{\text{red}}^{-1}}{d\sigma_a^2}))}{d\beta_b} = \text{tr}(\Sigma_{\text{red}}^{-1} \frac{d\Sigma_{\text{red}}}{d\beta_b} \Sigma_{\text{red}}^{-1} \frac{d\Sigma_{\text{red}}}{d\sigma_b^2})$$

$$\frac{d((V - W\beta)\Sigma_{\text{red}}^{-1} \frac{d\Sigma_{\text{red}}^{-1}}{d\sigma_a^2} \Sigma_{\text{red}}^{-1} (V - W\beta))}{d\sigma_b^2} = (V - W\beta)^t (\Sigma_{\text{red}}^{-1} \frac{d\Sigma_{\text{red}}}{d\sigma_a^2} \Sigma_{\text{red}}^{-1} \frac{d\Sigma_{\text{red}}}{d\sigma_b^2} \Sigma_{\text{red}}^{-1} + \Sigma_{\text{red}}^{-1} \frac{d\Sigma_{\text{red}}}{d\sigma_b^2} \Sigma_{\text{red}}^{-1} \frac{d\Sigma_{\text{red}}}{d\sigma_a^2} \Sigma_{\text{red}}^{-1}) (V - W\beta)$$

$$\frac{d(\text{tr}(\Sigma_{\text{red}}^{-1} \frac{d\Sigma_{\text{red}}^{-1}}{d\sigma_a^2}))}{d\sigma_b^2} = \text{tr}(\Sigma_{\text{red}}^{-1} \frac{d\Sigma_{\text{red}}}{d\sigma_a^2} \Sigma_{\text{red}}^{-1} \frac{d\Sigma_{\text{red}}}{d\sigma_b^2})$$

We now have all the information needed to calculate the variance of θ^N for any given set of parameters.

Figure 1.

(large-error scenario): The % error (y axis), defined to be the square-root of the Mean Square Error divided by the parameter value, for estimates of β_1 (1st column), σ_e^2 (2nd column), and σ_r^2 (3rd column), as a function of changing n (number of subjects, 1st row), σ_u^2 (variance of gold standard, 2nd row), and σ_δ^2 (intra-individual variability, 3rd row). The % error for estimates based on the full model are unbroken black lines, while the % error for estimates based on the reduced model are broken red lines. The evaluated scenario assumes

N=4 measures per individual, an independence correlation matrix, and with the exception of the variable being changed, we let $n=50$, $\beta_1=1$, $\sigma_T^2=4$, $\sigma_\delta^2=1.5$, $\sigma_U^2=0.1$, 4.0, and $\sigma_r^2=0.8$

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 2.

(small error scenario): The % error (y axis), defined to be the square-root of the Mean Square Error divided by the parameter value, for estimates of β_1 (1st column), σ_ϵ^2 (2nd column), and σ_r^2 (3rd column), as a function of changing n (number of subjects, 1st row), σ_u^2 (variance of gold standard, 2nd row), and σ_δ^2 (intra-individual variability, 3rd row). The % error for estimates based on the full model are unbroken black lines, while the % error for estimates based on the reduced model are broken red lines. The evaluated scenario assumes

N=4 measures per individual, an independence correlation matrix, and with the exception of the variable being changed, we let $n=50$, $\beta_1=1$, $\sigma_T^2=4$, $\sigma_\delta^2=1.5$, $\sigma_U^2=0.1$, $\sigma_\varepsilon^2=0.5$, and $\sigma_r^2=0.8$

Figure 3.

(large error scenario): The % error (y axis), defined to be the square-root of the Mean Square Error divided by the parameter value, for estimates of λ (1st column) and ρ^{pdr} (2nd column), as a function of changing n (number of subjects, 1st row), σ_u^2 (variance of gold standard, 2nd row), and σ_δ^2 (intra-individual variability, 3rd row). The % error for estimates based on the full model are unbroken black lines, while the % error for estimates based on the reduced model are broken red lines. The evaluated scenario assumes $N=4$ measures per individual, an

independence correlation matrix, and with the exception of the variable being changed, we let $n=50$, $\beta_1=1$, $\sigma_T^2=4$, $\sigma_\delta^2=1.5$, $\sigma_U^2=0.1$, $\sigma_\varepsilon^2=4$, and $\sigma_r^2=0.8$

Figure 4.

(small error scenario): The % error (y axis), defined to be the square-root of the Mean Square Error divided by the parameter value, for estimates of λ (1st column) and ρ^{PDR} (2nd column), as a function of changing n (number of subjects, 1st row), σ_u^2 (variance of gold standard, 2nd row), and σ_δ^2 (intra-individual variability, 3rd row). The % error for estimates based on the full model are unbroken black lines, while the % error for estimates based on the reduced model are broken red lines. The evaluated scenario assumes $N=4$ measures per

individual, an independence correlation matrix, and with the exception of the variable being changed, we let $n=50$, $\beta_1=1$, $\sigma_T^2=4$, $\sigma_\delta^2=1.5$, $\sigma_U^2=0.1$, $\sigma_\varepsilon^2=0.5$, and $\sigma_r^2=0.8$

Table 1

Accuracy of Estimates of Variance from Full Model

	Independence		Autocorrelation		Toeplitz	
	Theoretical	Simulated	Theoretical	Simulated	Theoretical	Simulated
β	0.094	0.084	0.081	0.079	0.074	0.074
σ_e^2	0.422	0.385	0.373	0.363	0.357	0.355
σ_r^2	0.327	0.325	0.327	0.326	0.327	0.328

The theoretical values for the standard deviation of $\hat{\beta}$, $\hat{\sigma}_e^2$ and $\hat{\sigma}_r^2$ when estimated from the full model are compared to values obtained from simulation. Theoretical values are based on the inverse of the information matrix. The comparison was performed for each of the three types of correlation structures using the parameters of the large-error primary scenario.

Table 2

Accuracy of Estimates of Bias and Variance from Reduced Model

	Bias						Variance					
	Independence		Autocorrelation		Toeplitz		Independence		Autocorrelation		Toeplitz	
	Theory	Sim	Theory	Sim	Theory	Sim	Theory	Sim	Theory	Sim	Theory	Sim
β	-0.026	-0.027	-0.028	-0.029	-0.027	-0.028	0.068	0.070	0.071	0.073	0.070	0.071
σ_e^2	0.096	0.096	0.095	0.095	0.095	0.095	0.358	0.358	0.358	0.357	0.358	0.357
σ_r^2	0.003	0.004	0.004	0.006	0.003	0.004	0.329	0.333	0.329	0.336	0.329	0.335

The theoretical values for the bias and standard deviation of $\hat{\beta}$, $\hat{\sigma}_e^2$ and $\hat{\sigma}_r^2$ when estimated from the reduced model are compared to values obtained from simulation. The comparison was performed for each of the three types of correlation structures using the parameters of the large-error primary scenario.

Table 3

Error from Model Misspecification

	β			σ_e^2			σ_r^2		
	I	A	T	I	A	T	I	A	T
I	1	1.008	1.009	1	1.006	1.011	1	1.000	1.001
A	0.987	1	1.018	0.996	1	1.004	1.001	1	1
T	0.963	1.035	1	0.997	1.058	1	0.999	1.002	1

Comparison of Mean Squared Errors for $\beta, \hat{\sigma}_e^2$ and $\hat{\sigma}_r^2$ when assuming different correlation structures. The rows indicates the correlation structure used to generate the data (I=Independence, A=Autocorrelation, T=Toeplitz(1)) while the column indicates the assumed correlation structure in the data analysis. The entries show the MSE based of the maximum likelihood estimates from the model that used the assumed correlation structure, as compared to MSE based on the true correlation structure (by definition, the diagonals in each 3×3 box are equal to 1).