Systematic investigations into the structure of measurement error of physical activity questionnaires are lacking. We propose a measurement error model for a physical activity questionnaire that uses physical activity level (the ratio of total energy expenditure to basal energy expenditure) to relate questionnaire-based reports of physical activity level to true physical activity levels. The 1999–2006 National Health and Nutrition Examination Survey physical activity questionnaire was administered to 433 participants aged 40–69 years in the Observing Protein and Energy Nutrition (OPEN) Study (Maryland, 1999–2000). Valid estimates of participants’ total energy expenditure were also available from doubly labeled water, and basal energy expenditure was estimated from an equation; the ratio of those measures estimated true physical activity level ("truth"). We present a measurement error model that accommodates the mixture of errors that arise from assuming a classical measurement error model for doubly labeled water and a Berkson error model for the equation used to estimate basal energy expenditure. The method was then applied to the OPEN Study. Correlations between the questionnaire-based physical activity level and truth were modest ($r = 0.32–0.41$); attenuation factors (0.43–0.73) indicate that the use of questionnaire-based physical activity level would lead to attenuated estimates of effect size. Results suggest that sample sizes for estimating relationships between physical activity level and disease should be inflated, and that regression calibration can be used to provide measurement error–adjusted estimates of relationships between physical activity and disease.

Berkson model; bias; energy metabolism; measurement error model; models, statistical; motor activity; self-assessment

Abbreviations: BEE, basal energy expenditure; DLW, doubly labeled water; MET, metabolic equivalent of task; NHANES, National Health and Nutrition Examination Survey; OPEN, Observing Protein and Energy Nutrition; TEE, total energy expenditure; USDA, US Department of Agriculture.

Positive energy balance has been associated with the development and progression of chronic diseases, including diabetes, hypertension, and cancer (1–5). However, accurate assessment of the components of energy balance—energy intake and energy expenditure—is difficult with self-reported data. The measurement error structure of dietary assessment measures of energy intake has been described (6, 7); however, a better understanding of the limitations of current energy expenditure measures is needed. Validation studies indicate that self-reported physical activity assessment instruments are prone to error (8–12), yet systematic investigations into the structure of measurement error are lacking.

Physical activity questionnaires vary in the time period measured, the type and number of activities queried, the way in which intensity and duration of activities are measured, and the methods used to convert item responses into measures of activity. The 1999–2006 National Health and Nutrition Examination Survey (NHANES) physical activity questionnaire queries transportation- and household-related activities, allows the analyst to assign metabolic equivalent of task (MET) values to specific moderate and vigorous
activities, and queries usual daily activity. These characteristics may make this questionnaire more comprehensive than some other questionnaires; however, it has not been validated.

Doubly labeled water (DLW), which is water with enriched deuterium and oxygen-18, is used to estimate carbon dioxide production in a free-living individual. Carbon dioxide production is converted into an unbiased estimate of total energy expenditure (TEE) by using standard equations (13). DLW may be used to estimate a global measure of physical activity, such as physical activity level (the ratio of TEE to basal energy expenditure (BEE)). Both DLW and the NHANES physical activity questionnaire were administered to men and women aged 40–69 years in the Observing Protein and Energy Nutrition (OPEN) Study. BEE was calculated for individuals in this study from age, sex, height, and weight by using ancillary data from the US Department of Agriculture (USDA) Automated Multiple-Pass Method Validation Study (14). This paper presents a measurement error framework for physical activity assessment by using physical activity level and applies this framework to the NHANES physical activity questionnaire from the OPEN Study.

MATERIALS AND METHODS

OPEN Study

Participants in the OPEN Study (15) were recruited by using a random sample from residential telephone listings of 5,000 households with a member aged 40–69 years living in the Maryland metropolitan area of Washington, DC. Exclusion criteria included diabetes, congestive heart failure, kidney failure requiring dialysis, fluid retention, malabsorption, hemophilia, use of supplemental oxygen, poor English literacy, pregnancy, participation in a weight loss/liquid diet, or formal nutrition training. Of the 837 eligible participants, 614 agreed to participate in the study and 484 (261 men and 223 women) attended the first visit. Participants were told that the purpose of the study was to examine daily protein and energy needs. Participants completed 3 visits over 3 months during September 1999, to March 2000, with only 2 participants failing to complete the study.

Objective measure of energy expenditure

DLW was used to measure TEE by using the 2-point method of Schoeller et al. (16, 17). The specific protocol used in the OPEN Study has been described in detail (18). A dose of DLW was given at the first visit, which was followed by a second visit 2 weeks later to complete the DLW assessment. A substudy was conducted on 14 male and 11 female volunteers to quantify within-subject variability of DLW-derived TEE. These participants were given a dose of DLW at the second visit and returned 2 weeks later to complete their DLW assessments.

Physical activity questionnaire

The 1999–2006 NHANES physical activity questionnaire (http://www.cdc.gov/nchs/data/nhanes/spq-pa.pdf) was administered by interviewers during the first visit of the OPEN Study. The questionnaire queries the type, frequency, intensity, and duration of activities a study participant has engaged in over the past 30 days in 4 categories: transportation-related, yard/housework, moderate recreational activities, and vigorous recreational activities. To calculate questionnaire-based physical activity level, the number of MET minutes spent in specified activities, sleep, and usual activity were computed by multiplying minutes per week spent in the activity by 4.0 for transportation-related activities, 4.5 for household activities, and values from the compendium of physical activities (19) for moderate and vigorous recreational activities. The average amount of time spent sleeping was assumed to be 6.9 hours (20), but was varied in sensitivity analyses. An item that queries usual physical activity (sit, stand, lift) each day was used to estimate activity during nonreported activity or sleep time; for this question, we used the MET scores suggested by NHANES (http://www.cdc.gov/nchs/nhanes/nhanes2005-2006/PAQ_D.htm#Appendix_1_.Suggested_MET_Scores) and varied them in sensitivity analyses. Individual-level data were examined for implausible estimates of activity; participants were excluded if they reported more than 42 hours/week of transportation ($n = 2$) or more than 21 hours/week of household/yard activities ($n = 7$).

Anthropometric and other measures

Height was measured at the first visit; weight was measured under standardized conditions at all visits (15). Demographic and socioeconomic measures were collected at the first visit.

USDA Automated Multiple-Pass Method Validation Study

Because no direct measure of BEE was available from the OPEN Study, indirect calorimetry data from the USDA Automated Multiple-Pass Method Validation Study were used to obtain an estimate of the error associated with the use of an equation to predict BEE in the OPEN Study and the covariance of this error with physical activity level. Details of the USDA Automated Multiple-Pass Method Validation Study have been described (14). Participants were recruited from the greater Washington, DC, metropolitan area. The study sample consisted of 525 participants aged 30–69 years whose weight was stable and who were not actively pursuing weight loss. Of these, 408 were aged 40–69 years and 369 had measures of TEE from DLW and BEE from indirect calorimetry and were included in this analysis. Data collection was completed between July 2002 and June 2004, with each subject participating in 4 visits and 2 telephone interviews over approximately 7 weeks. DLW was used to measure TEE over a 14-day period; 42 subjects completed the DLW protocol twice. Resting energy expenditure was estimated by using indirect calorimetry under a standardized protocol (Beckman Instruments, Inc., Anaheim, California) between 6:30 AM and 9:30 AM after fasting for 12 hours and limiting strenuous activity on the day of measurement. Estimates of BEE were made by using the formula of Weir (21); within-person variation of BEE from
indirect calorimetry was assumed to be 3.5% (22). This assumption was tested in sensitivity analyses. Height was measured at the first visit; weight was measured at the first and last visits.

STATISTICAL METHODS

The structure of measurement error

Measurement error bias may be random or systematic and can have a major impact on the design and interpretation of epidemiologic studies (1, 23, 24). Within-person variation reflects day-to-day variation in activity and other sources of random error, which are not usually distinguishable. Errors may also be systematic and may arise in different ways: they could occur equally for all participants (additive error); they could be related to measured or unmeasured characteristics of individuals (person-specific bias); or they could be related to the true activity level (proportional bias) (25). Proportional bias is a function of true expenditure and is assumed to be identical for individuals with the same energy expenditure. An example of this bias would be a high energy expenditure associated with greater underreporting of physical activities.

Two measures of the quality of questionnaire measurement for relating physical activity to health outcomes

The problem of studying disease or other health outcomes related to dietary intake in the presence of measurement error has been reviewed in the statistical literature (6, 26, 27). It is summarized here and modified for physical activity. Consider, for example, the disease model with a logistic link:

$$\logit(Pr(D = 1|P)) = \alpha_0 + \alpha_1 P,$$

where Pr\((D = 1|P)\) is the probability of disease \(D\), and \(P\) is true long-term physical activity level, a variable that we cannot measure exactly (i.e., it is measured with error). The slope, \(\alpha_1\), is the parameter of interest in epidemiologic studies, representing the relationship between physical activity and the risk of disease. However, because \(P\) is not measured exactly, we are unable to fit this model directly. If we fit model 1 by using a questionnaire-derived measure of physical activity level, denoted \(Q\), to represent the true physical activity level, then the estimate that we will obtain of \(\alpha_1\), denoted \(\hat{\alpha}_1\), will be biased. If we write the relationship between the expected value of the estimate and the true value of the coefficient as

$$E(\hat{\alpha}_1) = \lambda_1 \alpha_1,$$

then the term \(\lambda_1\) represents the multiplicative bias associated with substituting \(Q\) for \(P\). It can be shown that if the regression of \(P\) on \(Q\) is linear, then \(\lambda_1\) is the slope in that regression, or alternatively,

$$\lambda_1 = \frac{\text{cov}(P, Q)}{\text{var}(Q)} = \rho_{QP} \frac{\sigma_P^2}{\sigma_Q^2},$$

where \(\rho_{QP}\) denotes the correlation between \(P\) and \(Q\), \(\sigma_P^2\) denotes the variance of \(P\), and \(\sigma_Q^2\) denotes the variance of \(Q\).

In validation studies, the correlation between the questionnaire and a reference instrument is usually nonnegative (12); therefore, we expect the correlation with true expenditure (\(\rho_{QP}\)) to be nonnegative, and if \(\sigma_P^2 \leq \sigma_Q^2\), then \(\lambda_1\) is between 0 and 1. When \(Q\) is used in place of \(P\), the estimated slope is therefore attenuated to 0, as can be seen in equation 2. For this reason, \(\lambda_1\) is termed the attenuation factor. Once an estimate of the attenuation factor is obtained, an unbiased estimate of the effect of physical activity level on the risk of disease can be estimated by dividing the estimated coefficient \(\hat{\alpha}_1\) by the estimated attenuation coefficient, \(\hat{\alpha}_1 = \hat{\alpha}_1 / \lambda_1\). In logistic regression, \(\alpha_1\) represents the logarithm of relative risk (RR) and the above relationship transforms to \(\text{RR}_{\text{observed}} = (\text{RR}_{\text{true}})^{\lambda_1}\); The 2 parameters, \(\lambda_1\) and \(\rho_{QP}\), reflect the quality of the measurement \(Q\) as a substitute for \(P\). The closer they are to 1 and the further from 0, the higher the quality of the measurement. In the following section, we develop a method of estimating them for a physical activity questionnaire’s determination of physical activity level.

A measurement error model for physical activity level

We define our measure of physical activity level to be the ratio of true TEE (kcal/day) to true BEE (kcal/day). We find it to be statistically more convenient to work on the natural log scale because variables tend to be closer to normally distributed and regressions closer to linear. We therefore define \(P_i\) to be the natural log of true physical activity level for an individual \(i\) from a sample of \(n\) individuals. From this definition, \(P_i\) is the difference between the true ln(TEE), denoted \(T_i\), and the true ln(BEE), denoted \(B_i\):

$$P_i = T_i - B_i,$$

We assume that the main measurement of physical activity level that we have available to us comes from a physical activity questionnaire, and that its log-transformed value is denoted \(Q_i\). Following Kipnis et al. (6, 25), we assume also that \(Q_i\) is related to \(P_i\) by the following linear measurement error model:

$$Q_i = \beta_{Q0} + \beta_{Q1} P_i + \epsilon_i,$$

which describes a measurement that has additive \((\beta_{Q0})\) and multiplicative \((\beta_{Q1})\) systematic error, as well as random error \((\epsilon_i)\) that represents a combination of person-specific bias and within-person random error (with only 1 administration of the questionnaire, these 2 sources of error cannot be separated), with mean 0 and variance \(\sigma^2_{\epsilon}\).
From model 5, we obtain the following equations for the attenuation factor \( \lambda_1 \) and the correlation coefficient \( \rho_{QP} \):

\[
\lambda_1 = \frac{\text{cov}(P_i, Q_{ij})}{\sigma_Q} \quad \frac{\beta_{Q1}}{\sigma_Q / \sigma_P} = \frac{\beta_{Q1}}{\rho^2 + \sigma^2 / \sigma_P^2} \tag{6}
\]

and

\[
\rho_{QP} = \frac{\text{cov}(P_i, Q_{ij})}{\sqrt{\sigma^2_Q \sigma^2_P}} \quad \frac{\beta_{Q1}}{\sqrt{\rho^2 + \sigma^2 / \sigma_P^2}}.
\]

Thus, we need to estimate the quantities \( \beta_{Q1}, \sigma^2_e, \) and \( \sigma^2_p \) to determine the target parameters \( \lambda_1 \) and \( \rho_{QP} \). To do this, we need to use information from a validation study that includes an unbiased measure of \( P_i \). In our case, this unbiased measure comes from 2 sources: a DLW determination of TEE and an equation-based estimate of BEE. Denoting the measurement error model for the component \( E_i \) as:

\[
M_i = W_i - E_i.
\]

Each of the components of \( M_i \) has its own measurement error structure. Assuming that ln(DLW) measures ln(TEE) with classical measurement error, we use the model:

\[
W_i = T_i + \pi_i,
\]

where \( \pi_i \) is the within-person random error in the DLW measure, including error in the laboratory measurement and day-to-day variation in \( T_i \), and has mean 0 and variance \( \sigma^2_{\pi} \), and is independent of \( T_i \). Its variance \( \sigma^2_{\pi} \) is determined from repeat DLW measurements on the same individual.

The measurement error model for the component \( E_i \) is different from that for \( M_i \). Because \( E_i \) is estimated by an equation, it is assumed to measure true \( E_i \) with Berkson error according to the model:

\[
B_i = E_i + \gamma_i,
\]

where \( \gamma_i \) is the error term with mean 0 and variance \( \sigma^2_{\gamma} \) and is independent of \( E_i \) and of the DLW error \( \pi_{ij} \). Berkson error arises when an average value is assigned to each individual. Because \( E_i \) is a predicted value based on certain personal characteristics of the individual and does not include the deviation from that prediction that may occur for such an individual, it is assumed to exhibit Berkson error. The variance \( \sigma^2_{\gamma} \) may be estimated from a validation study that includes an independent unbiased measure of BEE, for example by using indirect calorimetry, as in the USDA study (indirect calorimetry was not used in the OPEN Study).

Substituting equations 8, 9, and 4 into 7 gives:

\[
M_i = T_i - B_i + \pi_i + \gamma_i = P_i + \pi_i + \gamma_i.
\]

We now show how to estimate the quantities \( \beta_{Q1}, \sigma^2_e, \) and \( \sigma^2_p \) by using the validation study data. From models 5 and 10 we can write the following equations for the second moments of \( Q_i \) and \( M_i \):

\[
\sigma^2_Q = \beta_{Q1}^2 + \sigma^2_p \quad \sigma^2_M = \sigma^2_p + \sigma^2_x + \sigma^2_\gamma + 2 \text{cov}(P_i, \gamma_i) \tag{11}
\]

\[
\sigma^2_{QM} = \beta_{Q1}^2 \sigma^2_p + \rho_{Q-P} \sigma_{Q1} \sigma_{P} \gamma_i.
\]

Note that the expressions for \( \sigma^2_M \) and \( \sigma_{QM} \) include the covariance between \( P \) and \( \gamma_i \), which derives from the Berkson error model. Estimates of the second moments of \( Q_i \) and \( M_i \) as well as an estimate of \( \sigma^2_p \) were available from the OPEN Study data. Furthermore, an estimate of \( \sigma^2_{\gamma} \) was available from the USDA study as well as an estimate of \( \text{cov}(P_i, \gamma_i) \) (see below). Therefore, the 3 equations in 11 allow us to estimate the 3 unknown quantities of interest: \( \beta_{Q1}, \sigma^2_e, \) and \( \sigma^2_p \).

We estimated \( \text{cov}(P_i, \gamma_i) \) and \( \sigma^2_{\gamma} \) by using data from the USDA study as follows: 1) denote the measure of indirect calorimetry on the log scale by \( C_i \), and assume that it estimates \( B_i \) with a coefficient of variation of 3.5%; and 2) denote the variance of \( C_i \) by \( \sigma^2_{C_i} \). Thus, the variance of \( B_i \) is estimated by \( \sigma^2_B = \sigma^2_C - 0.035^2 \). We estimated the covariances of the DLW measurement with the indirect calorimetry measurement, \( \text{cov}(W_i, C_i) \), and with the equation-derived measure of BEE, \( \text{cov}(W_i, E_i) \), by using the DLW determinations in the USDA study. We were then able to estimate the following:

\[
\sigma^2_{\gamma} = \sigma^2_B - \sigma^2_{\pi} \text{ and } \text{cov}(P_i, \gamma_i) = \text{cov}(W_i, C_i) - \text{cov}(W_i, E_i) - \sigma^2_{\gamma}, \tag{12}
\]

which comes from

\[
\text{cov}(P_i, \gamma_i) = \text{cov}(T_i - B_i, \gamma_i) = \text{cov}(T_i, \gamma_i) - \text{cov}(B_i, \gamma_i) = \text{cov}(T_i, \gamma_i) - \sigma^2_{\gamma} \text{ and } \text{cov}(T_i, \gamma_i) = \text{cov}(W_i - \pi_{ij}, B_i - E_i) = \text{cov}(W_i, B_i) - \text{cov}(W_i, E_i) = \text{cov}(W_i, C_i) - \text{cov}(W_i, E_i).
\]

From the equations in 11, estimates of the 3 model parameters \( \beta_{Q1}, \sigma^2_e, \) and \( \sigma^2_p \) are given by calculating in sequence:

\[
\sigma^2_p = \sigma^2_M - \sigma^2_{\pi} - \sigma^2_{\gamma} - 2 \text{cov}(P_i, \gamma_i) \quad \beta_{Q1} = \frac{\sigma^2_{QM}}{\sigma^2_p + \text{cov}(P_i, \gamma_i)} \quad \sigma^2_e = \sigma^2_Q - \beta_{Q1}^2 \sigma^2_p. \tag{13}
\]

The attenuation coefficient \( \lambda_1 \) and the correlation coefficient \( \rho_{QP} \) are then obtained from the equations in 6.

**Data analysis**

Descriptive statistics were used to characterize the OPEN Study and the USDA Automated Multiple-Pass Method...
Validation Study samples. A predictive equation for \( E_i \) was obtained from the USDA data by regressing the natural log of resting energy expenditure from indirect calorimetry on age, height, and weight. Interaction terms of sex, age, height, and weight were tested; none of the interaction terms was significant. Model fit was evaluated by examining plots of the jackknife residuals versus the predicted value and the leverage and partial residuals of age, height, and weight. The predictive equation was applied to the OPEN Study data to estimate \( E_i \). The measurement model described above in equation 5 was fit by using the NLMIXED procedure in SAS, version 9.2, software (SAS Institute, Inc., Cary, North Carolina). The code is provided in the Appendix.

**RESULTS**

One participant did not complete the interview, and 8 participants had items that were missing from the questionnaire so that physical activity level could not be determined. Nine additional participants were excluded for implausible reports on the questionnaire. Thirty-three DLW measures were excluded because of unacceptable internal agreement, failure to isotopically equilibrate on dosing day, isotopic dilution space ratios outside the range of 1.00–1.08, or missing specimens. The final number of participants included in the analysis was 433.

The regression of log of resting energy expenditure from indirect calorimetry on age, height, and weight in the USDA data set yielded the following predictive equation:

\[
\ln(\text{BEE}) = 6.64 + 0.11 \times \text{male} + 0.00589 \times \text{weight (kg)} + 0.00206 \times \text{height (cm)} - 0.00310 \times \text{age (years)}.
\]

The \( r^2 \) value was 0.75, and the root mean squared error was 0.085.

Characteristics of the study samples for the OPEN Study and the USDA Automated Multiple-Pass Method Validation Study are given in Table 1. The participants from both studies were from the same geographic area and were similar in terms of sex, age, and race distributions. The OPEN Study sample had a greater proportion of obese participants.

### Table 1. Sample Characteristics, the OPEN Study, 1999–2000, and the USDA AMPM Validation Study, 2002–2004, Both in the Washington, DC, Metropolitan Area

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>OPEN Study</th>
<th>USDA AMPM Validation Study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Women ((n=201))</td>
<td>Men ((n=232))</td>
</tr>
<tr>
<td>Age, years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40–49</td>
<td>80 39.8</td>
<td>85 36.6</td>
</tr>
<tr>
<td>50–59</td>
<td>74 36.8</td>
<td>76 32.8</td>
</tr>
<tr>
<td>60–69</td>
<td>47 23.4</td>
<td>71 30.6</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic white</td>
<td>154 76.6</td>
<td>200 86.2</td>
</tr>
<tr>
<td>Other/unknown</td>
<td>47 23.4</td>
<td>32 13.8</td>
</tr>
<tr>
<td>Smoking status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current</td>
<td>27 13.5</td>
<td>20 8.6</td>
</tr>
<tr>
<td>Former</td>
<td>54 27.0</td>
<td>78 33.6</td>
</tr>
<tr>
<td>Never</td>
<td>119 59.5</td>
<td>134 57.8</td>
</tr>
<tr>
<td>Educational level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school or less</td>
<td>39 19.9</td>
<td>20 8.6</td>
</tr>
<tr>
<td>Some college</td>
<td>52 26.5</td>
<td>45 19.4</td>
</tr>
<tr>
<td>College graduate</td>
<td>54 27.6</td>
<td>82 35.3</td>
</tr>
<tr>
<td>Post graduate</td>
<td>51 26.0</td>
<td>85 36.6</td>
</tr>
<tr>
<td>BMI(^a)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;25.0</td>
<td>78 38.8</td>
<td>57 24.6</td>
</tr>
<tr>
<td>25.0–29.9</td>
<td>63 31.3</td>
<td>108 46.6</td>
</tr>
<tr>
<td>&gt;29.9</td>
<td>60 29.9</td>
<td>67 28.9</td>
</tr>
</tbody>
</table>

Abbreviations: AMPM, Automated Multiple Pass Method; BMI, body mass index; OPEN, Observing Protein and Energy Nutrition; USDA, US Department of Agriculture.

\(^a\) Body mass index: weight (kg)/height (m)\(^2\).
but physical activity level values were similar between the 2 samples (Table 2).

A plot of physical activity level from the questionnaire versus physical activity level from DLW and BEE from equation 14 (Figure 1) illustrates the bias in questionnaire-based determinations of physical activity level. Figure 1 illustrates the combination of additive systematic error and the proportional bias. In particular, physical activity level demonstrates the “flattened slope” effect often seen in energy intake data; persons with higher physical activity level as estimated by the biomarker (DLW and BEE from an equation) have lower levels of questionnaire-based physical activity, and persons with lower levels of biomarker-based physical activity level have higher estimates from the questionnaire. The proportional bias is quantified in Table 3; those who had the highest true physical activity level had lower levels of estimated physical activity level from the NHANES questionnaire. The correlation coefficients between truth and estimated physical activity level were 0.41 for men and 0.32 for women. The attenuation factor was higher for men than for women, with values of 0.73 and 0.43, respectively. Sensitivity analysis indicated that the results were not sensitive to the amount of time spent sleeping, the MET values assumed for “usual” activity, or the coefficient of variation assumed for indirect calorimetry.

DISCUSSION

We have presented a measurement error model for self-reported physical activity level that accommodates the mixture of classic error (in measurements of TEE) and Berkson error (in measurements of BEE) from the nonquestionnaire measure of physical activity level. We applied the model to the physical activity level as calculated from the 1999–2006 NHANES physical activity questionnaire and found moderate levels of correlation and attenuation. Men had higher correlation and attenuation factors than did women, indicating that their reports were less biased. The

| Table 2. Energy Expenditure in the Study Sample, the OPEN Study, 1999–2000, and the USDA AMPM Validation Study, 2002–2004, Both in the Washington, DC, Metropolitan Area |
|---------------------------------------------------------------|---------------------------------------------------------------|
| **OPEN Study** | **USDA AMPM Validation Study** |
| **Women**, Mean (SD) | **Men**, Mean (SD) | **Women**, Mean (SD) | **Men**, Mean (SD) |
| Weight, kg | 73.4 (16.9) | 87.5 (15.6) | 69.6 (14.2) | 84.3 (14.1) |
| Height, m | 1.63 (0.06) | 1.76 (0.07) | 1.62 (0.07) | 1.76 (0.06) |
| TEE, kcal/day | 2,311 (394) | 2,899 (530) | 2,190 (406) | 2,877 (498) |
| BEE as measured by IC, kcal/day | 1,377 (179) | 1,719 (214) | 1,370 (131) | 1,716 (164) |
| BEE from equation, kcal/day | 1,408 (160) | 1,749 (183) | 1,370 (164) | 1,716 (164) |
| TEE – BEE from equation, kcal/day | 903 (310) | 1,151 (423) | 159 (0.24) | 1.68 (0.25) |
| Physical activity level from IC | 1.64 (0.21) | 1.65 (0.21) | 1.60 (0.26) | 1.68 (0.27) |
| Physical activity level from equation | 1.62 (0.14) | 1.61 (0.13) |

Abbreviations: AMPM, Automated Multiple Pass Method; BEE, basal energy expenditure; IC, indirect calorimetry; OPEN, Observing Protein and Energy Nutrition; TEE, total energy expenditure; USDA, US Department of Agriculture.

a There were 201 women and 232 men in the OPEN Study and 180 women and 189 men in the USDA AMPM Validation Study.

b From doubly labeled water.

c Estimated by using the following USDA equation: ln (BEE) = 6.64 + 0.11 × male + 0.00589 × weight (kg) + 0.00206 × height (cm) − 0.00310 × age (years).

d Physical activity is expressed as a ratio of kcal/day to kcal/day.

Figure 1. Physical activity level for men as measured on the log scale based on questionnaire versus doubly labeled water (DLW) and calculated basal energy expenditure (BEE), Observing Protein and Energy Nutrition (OPEN) Study, Washington, DC, metropolitan area, 1999–2000. (Plot was similar for women.)

<table>
<thead>
<tr>
<th>Sex by Sleep, hours</th>
<th>METs for Usual Activity</th>
<th>Coefficient of Variation of BEE From Indirect Calorimetry, %</th>
<th>Attenuation Factor ( \rho_{QP} )</th>
<th>Correlation of Physical Activity Level From Questionnaire and True Physical Activity Level ( p_{QP}^{\text{est}} )</th>
<th>Mean of Estimate of Log True Physical Activity Level ( \bar{Q}_{Q1}^{\text{est}} )</th>
<th>Variance of Log True Physical Activity Level ( \sigma_Q^{2,\text{P}} )</th>
<th>Slope in Regression of Log Physical Activity Level From Questionnaire on Log True Physical Activity Level ( \beta_{Q1}^{\text{est}} )</th>
<th>Variance of Within-Person Error in Physical Activity Level From Questionnaire ( \sigma_e^{2,\text{Q1}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>( 6.9 ) NHANES ( ^b )</td>
<td>3.5</td>
<td>0.73 (0.12)</td>
<td>0.41 (0.07)</td>
<td>0.50 (0.008)</td>
<td>0.019 (0.002)</td>
<td>0.23 (0.05)</td>
<td>0.005 (0.0005)</td>
</tr>
<tr>
<td></td>
<td>( 8.0 ) NHANES</td>
<td>3.5</td>
<td>0.73 (0.12)</td>
<td>0.41 (0.07)</td>
<td>0.50 (0.008)</td>
<td>0.019 (0.002)</td>
<td>0.23 (0.05)</td>
<td>0.005 (0.0005)</td>
</tr>
<tr>
<td></td>
<td>( 6.9 ) Higher ( ^c )</td>
<td>3.5</td>
<td>0.73 (0.12)</td>
<td>0.42 (0.07)</td>
<td>0.50 (0.008)</td>
<td>0.019 (0.002)</td>
<td>0.24 (0.05)</td>
<td>0.005 (0.0005)</td>
</tr>
<tr>
<td></td>
<td>( 6.9 ) NHANES</td>
<td>5.0</td>
<td>0.68 (0.11)</td>
<td>0.40 (0.07)</td>
<td>0.50 (0.008)</td>
<td>0.018 (0.002)</td>
<td>0.23 (0.05)</td>
<td>0.005 (0.0005)</td>
</tr>
<tr>
<td>Female</td>
<td>( 6.9 ) NHANES</td>
<td>3.5</td>
<td>0.43 (0.12)</td>
<td>0.32 (0.09)</td>
<td>0.49 (0.009)</td>
<td>0.012 (0.002)</td>
<td>0.23 (0.07)</td>
<td>0.006 (0.0007)</td>
</tr>
<tr>
<td></td>
<td>( 8.0 ) NHANES</td>
<td>3.5</td>
<td>0.43 (0.12)</td>
<td>0.32 (0.09)</td>
<td>0.49 (0.009)</td>
<td>0.012 (0.002)</td>
<td>0.23 (0.07)</td>
<td>0.006 (0.0007)</td>
</tr>
<tr>
<td></td>
<td>( 6.9 ) Higher ( ^c )</td>
<td>3.5</td>
<td>0.40 (0.12)</td>
<td>0.31 (0.09)</td>
<td>0.49 (0.009)</td>
<td>0.012 (0.002)</td>
<td>0.24 (0.08)</td>
<td>0.007 (0.0007)</td>
</tr>
<tr>
<td></td>
<td>( 6.9 ) NHANES</td>
<td>5.0</td>
<td>0.38 (0.11)</td>
<td>0.30 (0.08)</td>
<td>0.49 (0.009)</td>
<td>0.011 (0.002)</td>
<td>0.23 (0.07)</td>
<td>0.006 (0.0006)</td>
</tr>
</tbody>
</table>

Abbreviations: BEE, basal energy expenditure; MET, metabolic equivalent of task; NHANES, National Health and Nutrition Examination Survey.

\(^a\) Value is estimate (standard error).

\(^b\) Estimated physical activity level from the NHANES questionnaire: sit during the day and do not walk much = 1.4; stand or walk a lot but do not carry or lift things = 1.5; lift or carry light loads or have to climb stairs often = 1.6; do heavy work or carry heavy loads = 1.8.

\(^c\) Higher: sit during the day and do not walk much = 1.5; stand or walk a lot but do not carry or lift things = 1.6; lift or carry light loads or have to climb stairs often = 1.8; do heavy work or carry heavy loads = 2.

Observed attenuation factors indicate that a true protective relative risk of 0.50 would be observed as a relative risk of 0.60 for men and 0.76 for women when the questionnaire was used to estimate physical activity level. It is likely that the beneficial effects of physical activity from epidemiologic studies that rely upon self-report estimates are underestimates of the true effects. In addition, the correlation coefficients estimated in this study indicate that there is a loss of statistical power for estimating relationships between disease and physical activity. In particular, the sample size required to detect an effect if using \( Q \) is \( 1/\rho_{QP}^2 \) times the sample size needed if using \( P \). Therefore, for the correlations estimated in this study of 0.32 for women and 0.41 for men, sample sizes would need to be inflated by 6–10 times what would be required if we could measure physical activity level exactly.

We observed both additive systematic error \( \beta_{Q0} \) and proportional bias \( \beta_{Q1} \) (equation 5). Additive systematic error shifts the estimated mean activity level but does not affect the relationship between true physical activity level and disease. However, proportional bias does affect this relationship. In particular, in this study, those with the highest true physical activity levels tended to have lower estimated physical activity levels on the questionnaire, with an estimated slope of 0.23 on the log scale for both men and women. For example, a 50% higher value of true physical activity level would be associated with a 10% higher value of questionnaire-based physical activity level. The overall impact of measurement error on the relationship between diet and activity as measured by the attenuation factor is the product of this proportional error and the ratio of the variance of \( P \) to the variance of \( Q \), as seen in equation 6. It also follows from equation 6 that if \( \sigma_Q^2 \) is less than \( \sigma_P^2 \) (specifically, if the ratio is less than \( \beta_{Q1} \)), then the attenuation factor will be greater than 1, leading to an exaggerated effect of physical activity level on disease. In this study, the ratio of \( \sigma_Q^2/\sigma_P^2 \) was less than 1 (0.32 for women and 0.55 for men) but greater than \( \beta_{Q1} \) (0.23), which led to an underestimated impact of physical activity level on disease, as evidenced by the estimated attenuation factors of less than 1.

This finding is in contrast to studies of dietary intake. In the OPEN Study, the ratios of these variances were always greater than 1 for energy, protein, and protein density, and the attenuation factors were always smaller than the proportional bias. That is, the measure of diet had so much “noise” that its impact was much greater than the effect of systematic intake-related bias error on estimating the relationship between diet and disease. In fact, with a ratio of the variances of \( F \) for energy (as was observed for the food frequency questionnaire), the estimated relationship between the questionnaire and disease would always be biased, even in the absence of systematic intake-related error (6). However, in this analysis of the questionnaire-based physical activity level, \( \sigma_Q^2 \) was estimated to be smaller than \( \sigma_P^2 \), which led to an attenuation factor that was greater than the estimated proportional bias and correlation. The bias demonstrated by the questionnaire.
based physical activity level, therefore, comes from the proportional error and $\sigma^2_Q/\sigma^2_P$, but is not dominated by $\sigma^2_Q$ in the same manner as in the assessment of dietary intake.

The correlations estimated in this study of true physical activity level with the questionnaire-based physical activity level are moderate and are comparable with what is observed for food frequency questionnaire-based protein intake after energy adjustment. Although they indicate that there is a loss of power when using the questionnaire-based measure, the correlation and attenuation factors found in this study support the use of calibration studies for questionnaire-based physical activity level.

It is important to note that there are limitations to these findings. The OPEN Study population lived in the Washington, DC, area and was well-educated, English-speaking, predominately non-Hispanic white. We used an external sample to obtain estimates of the correlation of true physical activity level with the error in estimating BEE; ideally we would have estimated BEE through indirect calorimetry in the OPEN Study. The structure of measurement error that was estimated in this population may not extend to other populations. However, the general measurement error modeling approach that is outlined in this article could be used in the validation of other studies.

We are aware of only 1 study, by Ferrari et al (28), that has examined bias of physical activity data in a measurement error framework. This study used a 7-day physical activity log as an unbiased measure of MET hours per week and compared it with estimates from the Past Year Total Physical Activity Questionnaire (29) in 154 men and women aged 35–65 years. The estimated correlation of the measure with truth was 0.26 for the questionnaire, and the attenuation factor was 0.13, which were both lower than what we observed. One explanation for this difference may be the time reference of the questionnaires; the NHANES questionnaire queried the past 30 days compared with the past year in the study by Ferrari et al. (28). Another reason may be due to our use of physical activity level as a measure of physical activity compared with MET hours per week in the earlier study and the different reference measures (physical activity log vs. DLW/estimated BEE) used in the 2 studies. The differences in the way in which physical activity was measured in the 2 studies highlight the importance of considering the impact of physical activity measures when examining measurement error properties of these assessment tools.

In the dietary assessment literature, some researchers have suggested using regression calibration to calculate adjusted estimates of relative risk and reporting these with the usual relative risks when possible (30–32). This study supports that regression calibration methods may be useful for self-reporting of physical activity as well. If one were to do this for physical activity level, it would require unbiased biomarkers of TEE and BEE; ideally one would use indirect calorimetry to measure the latter as opposed to relying on an equation as in this study. If one were to use indirect calorimetry, which exhibits classical measurement error, the model presented in this paper could be easily modified to accommodate these errors. Whether accelerometers or other objective monitors could be used to estimate TEE for this purpose remains an unanswered question; future research should address the measurement error properties of accelerometers.

In conclusion, we found moderate levels of correlation and attenuation by using a measurement error model we developed for physical activity level. Physical activity researchers should consider the impact of measurement error when designing and analyzing studies of physical activity and disease.

ACKNOWLEDGMENTS

Author affiliations: Department of Biostatistical Sciences, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina (Janet A. Tooze); Risk Factor Monitoring and Methods Branch, Applied Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, Maryland (Richard P. Troiano); Department of Statistics, Intercollegiate Faculty of Nutrition, and Interdisciplinary Faculty of Toxicology, Texas A&M University, College Station, Texas (Raymond J. Carroll); Beltsville Human Nutrition Research Center, US Department of Agriculture, Beltsville, Maryland (Alanna J. Moshfegh); and Gertner Institute for Epidemiology and Health Policy Research, Sheba Medical Center, Tel Hashomer, Israel (Laurence S. Freedman).

This work was supported by the National Cancer Institute at the National Institutes of Health (grant R21CA139291 for J. A.T., R.J.C., and L.S.F.; and grant R27CA057030 for R.J.C.). Conflict of interest: none declared.

REFERENCES


(Appendix follows)
APPENDIX

The following code was used in the NLMIXED procedure in SAS, version 9.2, software (SAS Institute, Inc., Cary, North Carolina) to fit the statistical model.

```sas
proc nlmixed data=open4; by gender;
parms b=7 p=0.5 varD=0.029 covDBEE=0.012 varBEE=0.01 beta0star=0.2256 beta1star=0.25 varqw=0.005;
  varpi=0.002579;
  vareta=0.001225;
  if gender=0 then do;
    varc=0.0153124816;
    cov_d_c=0.0111094129;
  end;
  if gender=1 then do;
    varc=0.0164051743;
    cov_d_c=0.0140271075;
  end;
  cov_t_gam=cov_d_c-covDBEE;
  vargam=varc-vareta-varBEE;
  mean=dlw*(P+B+uD) + eqn*(B+uBEE) + paq*(beta0star+(beta1star*(P+uD-uBEE)));
  model energy ~ normal(mean,((dlw*0.0000000001) + (eqn*0.0000000001) + (paq*varqw)));
  random uD uBEE ~ normal([0], [varD, covDBEE, varBEE]) subject=nid;
  varP=varM-varpi+vargam-(2*cov_t_gam);
  varT=varD-varpi;
  varM=varD+varBEE-(2*covDBEE);
  varB=varBEE+vargam;
  covTB=covDBEE+cov_t_gam;
  covqm=betalstar*varm;
  betal=(betalstar*varm)/(varm-varpi-cov_t_gam);
  varQ=betalstar**2*(varD+varBEE-(2*covDBEE))+varqw;
  vare=varq-(betal**2*varp);
  covpgam=cov_t_gam-vargam;
  denom=betal**2+(vare/varp);
  corr=betal/sqrt(denom);
  atten=betal/denom;
run;
```