We used a biomarker of activity-related energy expenditure (AREE) to assess measurement properties of self-reported physical activity and to determine the usefulness of AREE regression calibration equations in the Women’s Health Initiative. Biomarker AREE, calculated as the total energy expenditure from doubly labeled water minus the resting energy expenditure from indirect calorimetry, was assessed in 450 Women’s Health Initiative participants (2007–2009). Self-reported AREE was obtained from the Arizona Activity Frequency Questionnaire (AAFQ), the 7-Day Physical Activity Recall (PAR), and the Women’s Health Initiative Personal Habits Questionnaire (PHQ). Eighty-eight participants repeated the protocol 6 months later. Reporting error, measured as log(self-report AREE) minus log(biomarker AREE), was regressed on participant characteristics for each instrument. Body mass index was associated with underreporting on the AAFQ and PHQ but overreporting on PAR. Blacks and Hispanics underreported physical activity levels on the AAFQ and PAR, respectively. Underreporting decreased with age for the PAR and PHQ. Regressing logbiomarker AREE on logself-reported AREE revealed that self-report alone explained minimal biomarker variance ($R^2 = 7.6, 4.8,$ and $3.4$ for AAFQ, PAR, and PHQ, respectively). $R^2$ increased to $25.2, 21.5,$ and $21.8,$ respectively, when participant characteristics were included. Six-month repeatability data adjusted for temporal biomarker variation, improving $R^2$ to $79.4, 67.8,$ and $68.7$ for AAFQ, PAR, and PHQ, respectively. Calibration equations “recover” substantial variation in average AREE and valuably enhance AREE self-assessment.

biomarkers; measurement error; physical activity; postmenopausal women

Abbreviations: AAFQ, Arizona Activity Frequency Questionnaire; AREE, activity-related energy expenditure; BMI, body mass index; DLW, doubly labeled water; MET, metabolic equivalent; PAR, 7-Day Physical Activity Recall; PHQ, Physical Habits Questionnaire; REE, resting energy expenditure; TEE, total energy expenditure; WHI, Women’s Health Initiative.
adults in the National Health and Nutrition Examination Survey (NHANES) were adherent to national physical activity recommendations when data were gathered via self-report, but only 5% met the recommended guidelines based on accelerometer data, which suggests that self-assessment is grossly misreported. In a 2008 review of 20 studies that included both self-reported and objective measures of physical activity, Neilson et al. (3) reported that the face validity of self-reported physical activity was very poor compared with that of objective biomarkers. Whether these errors in self-report of physical activity level are random or systematic has been understudied. In one of the few studies in which systematic bias in self-reported physical activity was reported, investigators used standard questionnaires and diaries, as well as objective measures from accelerometry, in 154 adults. The authors reported systematic bias in self-reported data by age, sex, and body mass index (BMI; measured as weight (kg)/height (m)^2) and commented that the results have implications for epidemiologic studies in which researchers test the associations between self-reported physical activity and disease outcomes (5).

We previously reported that measurement error in dietary self-report distorts relative risk estimates for the associations of diet with cancer, cardiovascular disease, frailty, renal function, and diabetes and that the magnitude of measurement error varies by the type of dietary assessment instrument used (6–12). We have developed and applied regression calibration equations utilizing objectively measured nutritional biomarker data to correct the measurement error in dietary self-report and have applied these data to subsequent disease association studies (6, 13). Here we build on this approach by using an objective measure of physical activity level to understand the measurement properties of self-reported physical activity assessment instruments.

MATERIALS AND METHODS

Overview of the Women’s Health Initiative Observational Study

The Women’s Health Initiative (WHI) Observational Study is a prospective cohort study that enrolled 93,676 postmenopausal women from 1994 to 1998 throughout the United States. WHI Observational Study design details have been published previously (14). Briefly, women were eligible for the WHI Observational Study if they were postmenopausal, 50–79 years of age at enrollment, and likely to be residing in the same area for at least 3 years. Women were excluded if they had competing risk factors (medical conditions with a predicted survival of less than 3 years), had adherence/retention issues (alcohol/drug dependency, mental illness, or dementia), or were actively participating in any clinical trial. Participants attended baseline and year 3 in-person clinical visits; all other annual follow-up was completed by mail. Study procedures were approved by the institutional review boards of each clinical center and the Clinical Coordinating Center at the Fred Hutchinson Cancer Research Center, Seattle, Washington. All women signed written informed consent.

The Nutrition and Physical Activity Assessment Study (NPAAS) was an ancillary study to the WHI Observational Study and enrolled 450 participants from 9 of the 40 WHI clinical centers between 2007 and 2009. Recruitment was structured to oversample women who were: 1) black or Hispanic; 2) at the lower and higher ranges of the BMI distribution (<18.5 and ≥30.0 at the WHI Observational Study baseline clinical visit); and 3) 50–59 years of age at WHI enrollment. Letters of invitation were sent to potential participants, and women who responded were screened via a telephone interview. Women were excluded if they had weight instability, travel plans during the study period, or any medical condition that would preclude participation. Of the 2,184 women screened, 20.6% (n = 450) enrolled and completed the study. A subsample of 88 women (19.6%) repeated the entire protocol approximately 6 months later to provide data on the repeatability of study measures. This subsample was a convenience sample of women who were willing and eligible to repeat the protocol. Women who did not successfully complete the first 2 study visits were not eligible (i.e., women who experienced difficulties with timed urine collections or indirect calorimetry-related claustrophobia). All study procedures were approved by the institutional review boards of the 9 participating institutions plus that of the Clinical Coordinating Center at the Fred Hutchinson Cancer Research Center, Seattle, Washington. Participants provided written informed consent and received $100.00 upon study completion.

Study procedures

Activities included 2 visits to local WHI clinics, with additional study activities completed at home (Figure 1). At visit 1, participants supplied informed consent and anthropometry measures and completed the first part of a doubly labeled water (DLW) protocol, a food frequency questionnaire, the WHI Personal Habits Questionnaire (PHQ) (15), and questions on body image and other lifestyle habits. In addition, participants completed either the Arizona Activity Frequency Questionnaire (AAFAQ) (16) or the 7-Day Physical Activity Recall (PAR) (17). At visit 2, participants completed the final part of the DLW protocol, completed either the AAFQ or PAR (whichever was not completed at visit 1), and underwent indirect calorimetry to measure resting energy expenditure (REE). Demographic characteristics had been collected previously in WHI.

Objective measure of physical activity

We obtained an objective measure (biomarker) of physical activity by measuring total energy expenditure (TEE) by DLW and REE by indirect calorimetry. We then computed activity-related energy expenditure (AEE) as AEE = TEE – REE. Analyses were repeated with the biomarker defined as AEE = 0.9(TEE) – REE, with the 0.9 factor intended to provide an adjustment for the thermic effect of food (18).

TEE was estimated from DLW protocol results using our standard procedure (13, 19). DLW is considered the gold standard for assessing short-term energy turnover because in weight-stable individuals, it measures energy intake and
volitional and nonvolitional energy expenditure (i.e., total energy expenditure) over a 2-week period (19–21). A 6.5% quality-control failure rate occurred for the DLW procedure. Of these, half were due to low tracer enrichments or lack of equilibration, whereas the others were due to dilution space or other external reproducibility issues.

REE was estimated by indirect calorimetry using a standard protocol (22) with either a DeltaTrac II Respiratory Gas Analyzer (Datex-Ohmeda, Inc., Madison, Wisconsin) or a SensorMedics VMAX metabolic cart (SensorMedics, Inc., Yorba Linda, California). All metabolic carts were calibrated each day according to the manufacturer instructions, and gasses were carefully monitored during each test. Participants arrived on day 15 after a 12-hour fast and rested in a semi-reclined position in a thermally neutral room for 30 minutes before a 30-minute test under a canopy (8 clinics) or using a nonrebreathing fitted mouthpiece (1 clinic). Previous research has shown no apparatus-related differences in REE measurements (23, 24). Data points were obtained every minute, but the first 10 minutes of collected data were excluded from data analysis because 10 minutes are needed to achieve steady-state metabolism. Steady state was defined as 10 minutes during which the oxygen consumption, minute ventilation, and respiratory exchange ratio did not vary by more than 10% (22). Participants who did not reach a steady state or did not have at least 10 minutes of useable data (n = 16) were not included in AREE biomarker analyses.

Self-reported assessment of physical activity

Participants completed 3 self-report physical activity instruments: the interviewer-administered PAR (17, 25), the self-administered AAFQ (16), and the self-administered WHI PHQ (15). The PAR included questions about time spent in sleep and performing moderate, hard, and very hard intensity activities for each segment of each day over the 7 days before the interview. Total daily energy expenditure was estimated by assigning to each activity category standard values of intensity expressed as multiples of resting metabolic rate or metabolic equivalents (METs). For example, 1 MET is approximately equal to 1 kilocalorie per kilogram per hour in sleep, and 1.5, 4.0, 6.0, and 10.0 METs were assigned to light, moderate, hard, and very hard activities, respectively. We next multiplied the METs by reported time spent in each activity, summed over all intensities, and multiplied by body weight to yield a summary score for energy expenditure in kilocalories per day (17, 26). Hours spent in light activities were derived by summing the time spent in the other reported activity categories and subtracting the total number of hours from 24. Final values represent 24-hour energy expenditure, including sleep. However, for the analyses presented in the present article, we omitted sleep (computed as reported weekly sleep hours × 1 MET × kg of body weight) by subtracting the sleep kilocalories from total activity.

The AAFQ is a self-administered mark-sense booklet adapted from the Minnesota Leisure Time Physical Activity Questionnaire (27) and was previously validated using DLW in a small sample of adults (16). The AAFQ includes specific activities grouped by domain (e.g., occupation, leisure, recreation, home maintenance), with categorical responses for frequency and duration of leisure, recreational, personal care, and household chore activities over the previous 4 weeks. Response options for each activity include the number of days and the length of time in hours.
The average daily energy expenditure for each recorded activity was calculated as the average number of hours per day reported for each activity multiplied by an estimated REE (computed using the Mifflin equation) and then multiplied by the METs intensity value assigned to the activity. Similar to what was seen with the PAR, AAFQ data output also yielded a summary score in kilocalories per day (including sleep). As done for the PAR, we omitted sleep by subtracting sleep kilocalories, which were computed as [sleep hours × (REE/24)]/1.84.

The WHI PHQ is a short, self-administered questionnaire that inquires about the usual frequency and duration of walking activity outside the home, other mild recreational activity (e.g., slow dancing, bowling), moderate recreational activity (e.g., outdoor biking, easy swimming), and strenuous recreational activity (e.g., aerobics, jogging). To characterize duration of physical activity, women chose 1 out of 5 frequency categories that ranged from never to 5 or more days per week. In addition, participants were asked how often they walked outside the home for more than 10 minutes (ranging from never to ≥7 times/week), the usual duration of their walking episodes, and their usual walking speed (casual strolling, normal, fairly fast, or very fast). Similar to how we scored the PAR and AAFQ, standard intensity values were expressed as METs, were assigned to each activity item, multiplied by reported duration, and then summed to compute AREE in MET-hours per week. Because the PHQ only obtains data on recreational activity, we used data previously collected in WHI (www.whiscience.org) (years 3 and baseline for the primary and reliability studies, respectively) to obtain estimates for housework, yard work, sitting, sleeping, and all other activities to be able to create total daily AREE. We assigned METs to each of the activity categories using standard algorithms and 1.5 METs to all other nonspecified activities.

Statistical analysis

Our analytic goals included 1) understanding measurement properties of self-reported physical activity; 2) studying whether a biomarker of AREE could be used to calibrate (correct) the corresponding self-report assessments; and 3) characterizing the reliability of the self-reported data and biomarkers in a subsample of participants. We used descriptive statistics to characterize the study population and to assess the distributions of the self-reported measures of physical activity from the 3 instruments (PAR, AAFQ, and PHQ) and the objective biomarkers. To compare energy expenditure from the self-reported questionnaires with that from the biomarker using consistent units, we converted MET values to kilocalories for the PAR and PHQ (AAFQ output includes kilocalories). We computed daily PHQ kilocalories as total METs/week × body weight/7. For the PAR, we computed daily activity kilocalories as PAR TEE – PAR REE, where PAR REE = 1 kcal/kg/hour × body weight in kg × 24 hours. The biomarkers and self-reported values were transformed by the natural logarithm to improve the normal distributional approximations, and geometric means and 95% confidence intervals are presented. We examined the reliability of the biomarker and self-reported data by using unadjusted correlations to assess the agreement between the primary and reliability samples. Our principal analyses do not include an adjustment for the thermic effect of food because participants were in the fasting state.

Our measurement error analysis relied on the assumption that the log-transformed AREE biomarkers adhered to a classical measurement error model with “errors” that are independent of the corresponding log-transformed self-report assessment. We defined reporting error as log(self-report AREE) – log(biomarker AREE). To understand the extent to which participant characteristics influence self-reported physical activity reporting error, we regressed the reporting error on pertinent characteristics (age, race/ethnicity, BMI) for each of the physical activity instruments.

We next conducted a series of linear regression models to understand the association of the biomarker with the self-reported data and to estimate the fraction of the total variance in the log-transformed biomarker that could be explained by the self-report assessment and other pertinent variables. In other words, the R^2 provides information about the strength of the AREE “signal” from the self-report and from other study subject characteristics. These regression analyses included age, race, and BMI, and results included an R^2 for each predictor variable, which estimates the contribution of each predictor to the total variability. Statistical analyses were conducted with R software, version 2.13.0 (R Development Core Team, Vienna, Austria) and SAS, version 9.2 (SAS Institute, Inc., Cary, North Carolina).

RESULTS

The distributions of the demographic and lifestyle characteristics of the study participants are presented in Table 1. We met the recruitment goal to oversample women who were younger (<60 years of age at WHI enrollment), women who were at the extremes of the BMI distribution, and women who were black or Hispanic.

Self-reported and objective biomarker estimates of AREE for the primary and reliability study are provided in Table 2. REE and AREE comprised, on average, 66.7% and 33.3% of TEE, respectively. Compared with the biomarker, the AAFQ and PAR both slightly underestimated AREE, whereas the PHQ substantially underestimated AREE by approximately 40%. Unadjusted Pearson correlation coefficients estimating the reliability between primary and reliability measures were 0.59, 0.75, 0.42, and 0.42 for TEE (from DLW), REE (from indirect calorimetry), AREE (TEE – REE), and calculated AREE (substituting the Mifflin equation for measured REE), respectively (Figure 2). Intraclass correlation coefficients comparing primary measures with reliability measures of TEE, REE, AREE, and calculated AREE were 0.58, 0.77, 0.42, and 0.42 respectively.

Table 3 provides results from linear regression of reporting error on participant characteristics in which separate models are given for the 3 self-report instruments. Each physical activity assessment instrument demonstrates some
systematic bias by personal characteristics, but the extent of the bias for related participant characteristics varies by instrument. For the AAFQ, AREE underreporting was greater for women with a higher BMI than for women with a normal BMI and for women who were black than for white participants. We found that in the PAR, activity underreporting decreased with age. Underreporting on the PAR was also predominant among those who were Hispanic and suggestive for black women compared with white women. The direction of the AREE misreporting in relation to BMI differed for the PAR, wherein women with a higher BMI tended to overreport physical activity on the PAR compared with those with lower BMI. Only the PHQ had systematic bias in reporting by age in which underreporting decreased with age. In each of these models, the $R^2$ was small (ranging from 4.6 to 6.8); only a minor fraction of the reporting error variance for the 3 self-report instruments was explained by these selected participant characteristics.

Table 4 gives regression coefficients from linear regression of log(AREE biomarker) on log(self-report AREE) and other participant characteristics for each self-report instrument. Even though age, race/ethnicity, and BMI were not consistently associated with systematic bias across assessment instruments (Table 3), we included all covariates in these models to help explain the variation in the biomarker. In general, for these models, the self-report explains a small amount of biomarker variation. The AAFQ, PAR, and PHQ assessments explained 7.6%, 4.8%, and 3.4% of the biomarker variation in these equations, respectively, but adding the covariates increased these values substantially to 25.2%, 21.5%, and 21.8%. We next used biomarker AREE data from the reliability study to create an adjusted $R^2$ that targeted the variation in average AREE over a 6-month period, assuming the measurement error correlation between primary and reliability biomarker assessments to be zero. These adjusted $R^2$ values were calculated by dividing the unadjusted $R^2$ by the correlation between log(AREE) for the primary and reliability measurements (0.317). Here we see that correcting for temporal variation in the biomarker added valuably to explaining biomarker variation, improving the $R^2$ estimates to 79.4, 67.8, and 68.7 for the AAFQ, PAR, and PHQ, respectively. To confirm the robustness of this approach, we conducted a sensitivity analysis in which the primary and reliability measurement error correlations took respective values of $\rho = 0$, $-0.1$, and $-0.2$, with the negative values allowing for the fact that the primary and reliability samples were drawn at the extremes of the targeted 6-month period and the zero value corresponding to our primary adjusted $R^2$ analyses. The negative values for $\rho$ yielded slightly larger adjustment factors (0.361 and 0.433, respectively) and adjusted $R^2$ values that were 16.7% and 26.7% smaller than when $\rho = 0$ (data not shown).

**DISCUSSION**

In the present study of 450 postmenopausal women participating in the WHI Observational Study, we used a biomarker of AREE to characterize the measurement properties of 3 commonly used self-reported measures of physical activity. The PAR and AAFQ slightly underestimated activity, whereas the PHQ substantially underestimated activity compared with the biomarker. All 3 self-reported physical activity assessment tools captured only a small fraction (4.2%–8.5%) of AREE biomarker variation. However, when participant
Table 2. Biomarker and Self-Reported Estimates of Energy in the Primary and Reliability Studies \((n=450)\), Women’s Health Initiative Nutrition and Physical Activity Assessment Study, 2007–2009

<table>
<thead>
<tr>
<th>Variable</th>
<th>Primary Study</th>
<th>Reliability Study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of Participants</td>
<td>Geometric Mean, kcal/day</td>
</tr>
<tr>
<td>Biomarker</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEE</td>
<td>415</td>
<td>2,023</td>
</tr>
<tr>
<td>REE(^a)</td>
<td>413</td>
<td>1,222</td>
</tr>
<tr>
<td>AREE(^b)</td>
<td>396</td>
<td>735</td>
</tr>
<tr>
<td>Self-report</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAFQ</td>
<td>450</td>
<td>683</td>
</tr>
<tr>
<td>PAR</td>
<td>444</td>
<td>721</td>
</tr>
<tr>
<td>PHQ</td>
<td>426</td>
<td>462</td>
</tr>
</tbody>
</table>

Abbreviations: AREE, activity-related energy expenditure; AAFQ, Arizona Activity Frequency Questionnaire; REE, resting energy expenditure; PAR, Physical Activity Recall; PHQ, Personal Habits Questionnaire; TEE, total energy expenditure.

\(^a\) REE values from one clinic were slightly larger than those from other clinics. A sensitivity analysis excluding the higher values showed no impact on overall patterns or subsequent analyses, so all data are included.

\(^b\) Not adjusted for the thermic effect of food; indirect calorimetry was conducted in fasting state.

Figure 2. Comparisons between the primary and reliability study measures \((n=82)\) for total energy expenditure (TEE) from doubly labeled water, measured resting energy expenditure (REE) from indirect calorimetry, biomarker-assessed activity-related energy expenditure (AREE), and calculated activity-related energy expenditure from the Mifflin equation (AREE2), 2007–2009. All units are kilocalories per day.

A) Comparison of the primary study TEE with the reliability study TEE \((p=0.59, R^2 = 0.35)\). B) Comparison of the primary study measures of REE with reliability study measures of REE \((p=0.75, R^2 = 0.56)\). C) Comparison of the primary study measure of AREE with the reliability study measure of AREE \((p=0.42, R^2 = 0.18)\). D) Comparison of the primary study measures of AREE2 with the reliability study measures of AREE2 \((p=0.42, R^2 = 0.17)\).

characteristics, such as age, race/ethnicity, and BMI were considered in linear regression models, larger but still modest fractions of the biomarker variation could be explained: 25.2% for the AAFQ, 21.5% for the PAR, and 21.8% for the PHQ. Large and meaningful improvements in variance were observed when using the reliability study data to correct for the measurement error in the biomarker, even when using a more conservative approach to computing the adjusted $R^2$. It should be noted that the adjusted $R^2$ values assume uncorrelated measurement errors between primary and reliability samples and can be expected to involve some overadjustment if such errors are positively correlated. Overall, there appears to be only a modest signal from the self-reported physical activity, suggesting that participant characteristics should be included when using calibrated AREE estimates in disease association studies.

Few other studies have examined the measurement properties of self-reported physical activity and specifically whether systematic reporting bias exists. Several studies have compared self-reported physical activity from standardized questionnaires and diaries to various objective measures assessed by pedometer, accelerometer, heart rate monitoring with or without concurrent accelerometry, or DLW. Many of these studies concluded that self-reported assessment is fair to poor compared with objective measures when studied under free-living conditions (2, 4, 5, 27, 29–33). Of these, only one study examined the role of systematic bias in reporting (5). Ferrari et al. used a measurement error model to examine 2 self-reported assessments (questionnaire and diaries) compared with an objective criterion measure obtained by accelerometer. They reported that measurement error varied by study participant characteristics, wherein greater misreporting occurred among persons who were older than 50 years of age and those above the median for sex-specific BMI (5).

These findings are in general agreement with those we reported here. However, our results suggested some differential bias by instrument; for example, in the subset of WHI Observational Study women who completed the present study, BMI was associated with overreporting physical activity on the PAR but underreporting physical activity on the AAFQ and the PHQ.

Our regression calibration models explained only modest fractions of the variation in the AREE biomarker. Previous studies have shown that body mass is a strong determinant of energy expenditure that may account for up to half of the variance in total daily energy expenditure, and most of this is driven by the fat-free mass component (31, 34). Measures of body composition that are able to partition fat and lean muscle may be more informative in regression calibration models than BMI alone. Alternative explanations may be that our biomarker assessed TEE and AREE over a relatively short period of time (2 weeks), whereas the recall time period for the questionnaires in this study were the past 7 days and past 28 days for the PAR and AAFQ, respectively, and no reference time period was provided for the PHQ.

Many plausible reasons exist to explain the differences in physical activity patterns and related energy expenditure between self-reported and objectively measured physical activity. First, many commonly used physical activity assessment instruments measure only certain domains, such as recreation or sport, and do not assess occupational, transportation, or nonvolitional related activity. The AAFQ measures all of these domains (excluding nonvolitional), which may partly explain the slightly higher $R^2$ in our regression models. Most questionnaires are limited by their ability to assess light activity, which is common among the postmenopausal women enrolled in the WHI. Second, most questionnaires ask respondents to self-assess a measure of

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>AAFQ$^a$</th>
<th>PAR$^b$</th>
<th>PHQ$^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$-0.033$</td>
<td>$0.038$</td>
<td>$-0.459^*$</td>
</tr>
<tr>
<td>Age</td>
<td>$0.001$</td>
<td>$0.016^*$</td>
<td>$0.014^*$</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>$-0.138^*$</td>
<td>$-0.111$</td>
<td>$-0.078$</td>
</tr>
<tr>
<td>Hispanic</td>
<td>$-0.071$</td>
<td>$-0.219^*$</td>
<td>$0.092$</td>
</tr>
<tr>
<td>Body mass index$^d$</td>
<td>$-0.013^*$</td>
<td>$0.008^*$</td>
<td>$-0.008$</td>
</tr>
<tr>
<td>Model $R^2$ %</td>
<td>5.1</td>
<td>6.8</td>
<td>4.6</td>
</tr>
</tbody>
</table>

Abbreviations: AAFQ, Arizona Activity Frequency Questionnaire; AREE, activity-related energy expenditure; METs, metabolic equivalents; PAR, 7-Day Physical Activity Recall; PHQ, Physical Habits Questionnaire; REE, resting energy expenditure; SE, standard error; TEE, total energy expenditure.

$^a$ Reporting error = log(self-report) – log(biomarker), where the biomarker AREE = TEE – REE.

$^b$ AAFQ AREE = AAFQ TEE – AAFQ REE.

$^c$ PAR AREE = PAR TEE – PAR REE, where PAR TEE = PAR total METs x body weight. PAR REE = 1 MET/hour x 24 hours x body weight (assuming 1 MET/hour).

$^d$ PHQ AREE = (PHQ total METs – 24) x body weight.

$^e$ Weight (kg)/height (m)$^2$. 

582 Neuhouser et al.
Percent of Biomarker Variation Explained From Regression Calibration of Biomarker on Self-Reported Physical Activity and Participant Characteristics (n = 450), Women’s Health Initiative Nutrition and Physical Activity Assessment Study, 2007–2009

<table>
<thead>
<tr>
<th>Variable</th>
<th>AAFQ β Coefficient</th>
<th>SE</th>
<th>Adjusted R²</th>
<th>PAR β Coefficient</th>
<th>SE</th>
<th>Adjusted R²</th>
<th>PHQ β Coefficient</th>
<th>SE</th>
<th>Adjusted R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>4.789*</td>
<td>0.395</td>
<td></td>
<td>5.353*</td>
<td>0.508</td>
<td></td>
<td>5.632*</td>
<td>0.271</td>
<td></td>
</tr>
<tr>
<td>Log(self-report AREE)</td>
<td>0.273*</td>
<td>0.060</td>
<td>7.6</td>
<td>24.0</td>
<td>0.184*</td>
<td>0.076</td>
<td>4.8</td>
<td>15.1</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>−0.018*</td>
<td>0.004</td>
<td>9.1</td>
<td>28.7</td>
<td>−0.024*</td>
<td>0.004</td>
<td>8.6</td>
<td>27.1</td>
<td>−0.022*</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>0.086</td>
<td>0.055</td>
<td></td>
<td>0.075</td>
<td>0.058</td>
<td></td>
<td>0.095</td>
<td>0.059</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.108*</td>
<td>0.054</td>
<td></td>
<td>0.141*</td>
<td>0.055</td>
<td></td>
<td>0.055</td>
<td>0.053</td>
<td></td>
</tr>
<tr>
<td>Body mass index</td>
<td>0.015*</td>
<td>0.004</td>
<td>6.1</td>
<td>19.2</td>
<td>0.012*</td>
<td>0.004</td>
<td>5.7</td>
<td>18.0</td>
<td>0.015*</td>
</tr>
<tr>
<td>Total</td>
<td>25.2</td>
<td>79.4</td>
<td></td>
<td>21.5</td>
<td>67.8</td>
<td></td>
<td>21.8</td>
<td>68.7</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: AAFQ, Arizona Activity Frequency Questionnaire; AREE, activity-related energy expenditure; PHQ, Personal Habits Questionnaire; SE, standard error.

* P < 0.05.

a Adjusted R² value corrected for biomarker measurement error, that is, R² divided by log(biomarker) correlation in the reliability sample.

b Reference group for race is white.

c Weight (kg)/height (m)².

d Total percent of variation explained by all variables. R² for a specific variable was calculated from the simple linear regression model with the variable only, with subsequent rescaling so that these R² values added to the total regression R².

Notable strengths of this study include the large, diverse study population. Our ability to recruit the large, diverse population may allow us to examine systematic bias in self-reported physical activity, smoking, and consuming alcohol in moderation. Results may differ among younger and mixed gender populations.
Finally, there might be other unmeasured covariates with systematic measurement error that were not assessed in the WHI that could contribute to the reporting error.

Physical activity is an important component of health promotion and disease prevention. In 2008, the US Department of Health and Human Services issued the first ever Physical Activity Guidelines for Americans (36), wherein all adults were advised to engage in at least 150 minutes/week of moderate intensity aerobic activity or 75 minutes/week of vigorous intensity aerobic activity. To fulfill this goal, credible advice must be provided by health care providers, scientists, and others with vested interests in the physical activity of the nation. Accurate assessment of physical activity to support such advice is important. The use of biomarker-calibrated estimates of self-report should be considered for use in population-based studies to reduce measurement error and to provide unbiased estimates of physical activity with disease outcomes. Future work will examine associations of calibrated AREE using the PHQ with various clinical outcomes in the WHI.

ACKNOWLEDGMENTS

Author affiliations: Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington (Marian L. Neuhaus); Chongzhi Di, Lesley F. Tinker, Rebecca Seguin, Ross L. Prentice); Canyon Ranch Center for Prevention & Health Promotion, University of Arizona, Tucson, Arizona (Cynthia Thomson); Division of Research, Kaiser Permanente, Oakland, California (Barbara Sternfeld); Department of Epidemiology and Population Health, Albert Einstein College of Medicine, New York, New York (Yasmin Mossavar-Rahmani); Department of Preventive Medicine, Stanford University, Stanford, California (Marcia L. Stefanick; Stacy Sims); School of Medicine, University of Hawaii, Honolulu, Hawaii (J. David Curb); Department of Preventive Medicine, University at Buffalo, State University of New York, Buffalo, New York (Michael Lamonte); and Department of Preventive Medicine, University of Tennessee Health Science Center, Memphis, Tennessee (Karen C. Johnson).

Funding was provided by the National Heart, Lung, and Blood Institute (grant N01WH22110) and the National Cancer Institute (CA119171).

Program Office: National Heart, Lung, and Blood Institute, Bethesda, Maryland: Jacques Rossouw, Shari Ludlam, Joan McGowan, Leslie Ford, and Nancy Geller. Clinical Coordinating Center: Fred Hutchinson Cancer Research Center, Seattle, Washington: Garnet Anderson, Ross Prentice, Andrea LaCroix, and Charles Kooperberg. Investigators and Academic Centers: Brigham and Women’s Hospital, Harvard Medical School, Boston, MA: JoAnn E. Manson; MedStar Health Research Institute/Howard University, Washington, DC: Barbara V. Howard; Stanford Prevention Research Center, Stanford, California: Marcia L. Stefanick; The Ohio State University, Columbus, Ohio: Rebecca Jackson; University of Arizona, Tucson/Phoenix, Arizona: Cynthia A. Thomson; University at Buffalo, Buffalo, New York: Jean Wactawski-Wende; University of Florida, Gainesville/Jacksonville, Florida: Marian Limacher; University of Iowa, Iowa City/Davenport, Iowa: Robert Wallace; University of Pittsburgh, Pittsburgh, Pennsylvania: Lewis Kuller; and Wake Forest University School of Medicine, Winston-Salem, North Carolina: Sally Shumaker.

Conflict of interest: none declared.

REFERENCES


