The Effect of Social Desirability and Social Approval on Self-Reports of Physical Activity

Swann Arp Adams1,2, Charles E. Matthews3, Cara B. Ebbeling4, Charity G. Moore1, Joan E. Cunningham1,2, Jeanette Fulton5, and James R. Hebert1,2,6

1 Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, SC.
2 South Carolina Cancer Center, Columbia, SC.
3 Division of General Internal Medicine and Vanderbilt-Ingram Cancer Center, Vanderbilt University, Nashville, TN.
4 Division of Endocrinology, Children’s Hospital and Harvard Medical School, Boston, MA.
5 South Carolina Comprehensive Breast Center, Palmetto Health, Columbia, SC.
6 Cancer Prevention and Control Program, Hollings Cancer Center, Charleston, SC.

Received for publication April 27, 2004; accepted for publication October 4, 2004.

The purpose of this investigation was to examine social desirability and social approval as sources of error in three self-reported physical activity assessments using objective measures of physical activity as reference measures. In 1997, women (n=81) living in Worcester, Massachusetts, completed doubly labeled water measurements and wore an activity monitor for 14 days. They also completed seven interviewer-administered 24-hour physical activity recalls (PARs) and two different self-administered 7-day PARs. Measures of the personality traits “social desirability” and “social approval” were regressed on 1) the difference between physical activity energy expenditure estimated from doubly labeled water and each physical activity assessment instrument and 2) the difference between monitor-derived physical activity duration and each instrument. Social desirability was associated with overreporting of activity, resulting in overestimation of physical activity energy expenditure by 0.65 kcal/kg/day on the second 7-day PAR (95% confidence interval: 0.06, 1.25) and overestimation of activity durations by 4.15–11.30 minutes/day (both 7-day PARs). Social approval was weakly associated with underestimation of physical activity on the 24-hour PAR (–0.15 kcal/kg/day, 95% confidence interval: –0.30, 0.005). Body size was not associated with reporting bias in this study. The authors conclude that social desirability and social approval may influence self-reported physical activity on some survey instruments.

Abbreviations: MET(s), metabolic equivalent(s); PAR, physical activity recall.

As with many other human behaviors, self-reporting of physical activity is subject to many sources of error and bias. Existing physical activity assessments capture no more than 50 percent of the variance in free-living physical activity levels, and often much less (1). However, because of the efficiency and utility of this approach, self-reports of physical activity are commonly employed in most large-scale epidemiologic studies.

Certain personality traits may affect self-reporting of physical activity. The traits of “social desirability” and “social approval” have been found to influence participants’ reports of diet (2–5). “Social desirability” is the defensive tendency of individuals to portray themselves in keeping with perceived cultural norms, whereas “social approval” is the need to obtain a positive response in a testing situation (4). It has been found that people, especially women, who score higher on the social desirability scale are more likely to underreport their fat and total energy intake (2–5).

To extend our understanding of systematic errors in self-reports of physical activity, we designed the present investigation to compare three self-reported physical activity assessment approaches commonly used in epidemiologic and clinical studies with objective measures of physical activity and to test for systematic errors that can be ascribed...
to social desirability and social approval. Our objective crite-
ron measures were physical activity energy expenditure
estimated from doubly labeled water and estimated resting
energy expenditure, as well as intensity-specific activity
duration derived by means of the ActiGraph accelerometer
(Manufacturing Technology, Inc., Fort Walton Beach,
Florida).

MATERIALS AND METHODS

A detailed description of this study (The Energy Study),
which was approved by the institutional review board of the
University of Massachusetts Medical School, has been
previously published (3). Briefly, participants \( n = 81 \) were
recruited from June to October 1997, primarily from two
sources: 1) the University of Massachusetts Memorial
Medical Center (Worcester, Massachusetts) and 2) the
general population of the surrounding community. Subjects
agreed to maintain their usual dietary and activity patterns
for the 2-week study period and to be available for seven 24-
hour recall telephone interviews.

Study timeline

At the day 0 visit, a fasting urine sample and anthro-
pometric measurements were obtained. Completed question-
naires (mailed 1 week previously) included information on
demographic factors, lifestyle factors, and general health, the
Marlowe-Crowne Social Desirability Scale (6), and the
Martin-Larson Approval Motivation Scale (7). An Acti-
Graph accelerometer was provided to each participant, along
with detailed usage instructions. Each patient was randomly
assigned one of the two types of 7-day physical activity
recalls (PARs), with instructions to complete the instrument
on the evening of day 6. Over the next 14 days (days 1–14),
seven telephone-administered 24-hour PARs and dietary
recalls were obtained, such that one recall was obtained for
each type of day of the week. On day 7, another nonfasting
urine sample and weight measurement were obtained and the
first 7-day PAR questionnaire was collected. All participants
were then given a different 7-day PAR, with instructions to
complete the instrument on the evening of day 13. On day
14, participants provided a final nonfasting urine sample.
The surveys and ActiGraph data were then collected, and
anthropometric measurements were obtained.

Study assessments

Doubly labeled water. Doubly labeled water \( (^2\text{H}_2^{18}\text{O}) \) was
used to assess total energy expenditure, based on an
individual’s clearance of stable (i.e., nonradioactive) hydrogen
deuteronium) and oxygen \( (^{18}\text{O}) \) isotopes administered orally
as water. A more detailed description of this assessment
method can be found elsewhere (3).

Resting metabolic rate. Resting metabolic rate is the
primary determinant of total energy expenditure. To estimate
physical activity energy expenditure, it was necessary to
estimate each person’s resting metabolic rate (8–10). The
equation developed by Arciero et al. (11) in older women
was used for this purpose (resting metabolic rate (kcal/
day) = 21 \times \text{fat-free mass (kg)} + 369). This fat-free mass-
based equation is highly correlated with measured resting
metabolic rate \( R^2 = 0.79; \text{standard error, 46 kcal/day} \) (11).
Fat-free mass was quantified using doubly labeled water-
derived total-body water data, assuming a hydration constant
of 0.73 (12).

Physical activity energy expenditure. As suggested by
Schoeller and Jefford (13), physical activity energy expendi-
ture estimated from doubly labeled water (PAEE<sub>DLW</sub>) was
calculated (kcal/kg/day) as follows: PAEE<sub>DLW</sub> = (total
energy expenditure minus resting metabolic rate)/body mass
(kg).

ActiGraph. A uniaxial ActiGraph accelerometer (for-
merly Computer Science Applications model 7164) was
used to assess motion over the 14-day study period. This
small, lightweight instrument detects acceleration from
0.05g to 2g while rejecting other forms of movement such as
vibration (14). The acceleration signal is filtered by an
analog band-pass filter and digitized by an 8-bit analog/
digital converter at a sampling rate of 10 samples per second,
storing data at 1-minute intervals (15). ActiGraph data are
summarized in counts per minute and have demonstrated
reasonable validity and reliability for the evaluation of phys-
ical activity behaviors against a variety of criterion measures
from direct observation to self-report diaries (16–18).

The following labels and count cutpoints were used to
determine the duration (minutes/day) of time spent in activi-
ties of various levels: inactivity, 0–259 counts/minute; light
activity, 260–759 counts/minute; moderate activity, 760–
2,574 counts/minute; and vigorous activity, ≥2,575 counts/
minute. To ensure the integrity of the data, we used a time
on/off diary and an automated review of monitor wear to
identify periods of noncompliance. Data were excluded if
sustained periods of zero counts or sustained periods of
improbably high counts (>30,000 counts/minute), indicating
accelerometer malfunction, were noted.

7-day PAR 1. The first 7-day PAR was a self-adminis-
tered early version of the Stanford Five-City Project’s 7-day
recall (19). It asked participants to report their amounts of
sleep and moderate and vigorous physical activity for the
previous five weekdays and two weekend days. Moderate,
vigorous, and very vigorous activities were assessed, and
examples of occupational, household, and leisure activities
were provided for these intensity levels. Time spent in light
activities was calculated by subtracting the total time for all
other activities and sleep from 24 hours. Physical activity
energy expenditure (not including sleep) was calculated
using reports of duration for each activity intensity level and
the following metabolic equivalent (MET) weights: light
activity, 1.5 METs; moderate activity, 4.0 METs; vigorous
activity, 6.0 METs; and very vigorous activity, 8.0 METs.
Physical activity energy expenditure was calculated in terms
of kcal/kg/day (1 MET-hour ≈ 1 kcal/kg) using standard
methods (20). The average durations (minutes/day) of light,
moderate, and vigorous (≥6 METs) activities were also
calculated.

7-day PAR 2. The second 7-day PAR was developed for
use in this investigation and was modeled after the approach
used in the Minnesota Leisure Time Physical Activity
Survey to assess the information on frequency (per week)
and duration (per day) of activity. The 7-day PAR 2 was expanded to capture six domains of activity and focused on the previous 7 days. The activity domains evaluated were household (indoor), household (outdoor), child-care, occupational and volunteer, leisure and sport, and miscellaneous. Each activity domain contained a list of 5–41 common activities; respondents were asked to report the number of days and average amount of time per day spent in each activity. The miscellaneous category included six sedentary activities. MET estimates for each line item were made on the basis of example activities provided in the text of the instrument (20, 21). As described above for 7-day PAR 1, physical activity energy expenditure (kcal/kg/day) and overall and intensity-specific activity durations (minutes/day) were calculated.

24-hour PAR. The 24-hour PAR was administered by trained interviewers, either immediately prior to or after the 24-hour dietary recall (determined by the participant). The method employed has been previously described by Matthews et al. (22) and has been shown to have reasonable relative validity for assessment of short-term physical activity energy expenditure using the Baecke questionnaire and activity monitoring as criterion measures. Briefly, participants were asked, for the previous day, to recall the amount of time they had spent in bed and the amount of time they had spent in light, moderate, vigorous, and very vigorous activities in each of three activity domains (household, occupational, and leisure). Domain-specific example activities were provided for each activity intensity. As described above for 7-day PAR 1, physical activity energy expenditure (kcal/kg/day) and overall intensity-specific activity durations (minutes/day) were calculated. The average of the seven recalls was used for analyses.

Social desirability and social approval. The 33-item Marlowe-Crowne Social Desirability Scale was used to ascertain a participant’s tendency “to avoid criticism” and display herself in a favorable social image (6). The 20-item Martin-Larson Approval Motivation Scale was used to assess the social approval trait (7). Both scales have been shown to have good validity and reliability over time and were administered only at baseline (6, 7).

Statistical methods

Complete data from doubly labeled water measurements were available for 80 of the 81 women recruited into the study. SAS, version 8.1 (SAS Institute, Inc., Cary, North Carolina), was used for all analytic procedures (23). Descriptive statistics were computed for all variables. Student’s t tests were used to assess differences in mean physical activity energy expenditure as estimated from doubly labeled water and each survey instrument. Continuous variables were assessed for evidence of linear model assumptions, including nonnormality. Spearman correlation coefficients were used to assess the rank correlation among the energy expenditure estimates, ActiGraph counts, social approval scores, social desirability scores, and various other potentially confounding or effect-modifying variables. Social desirability or social approval scores were plotted by the difference in physical activity energy expenditure between each instrument and doubly labeled water measurements. We calculated Bland-Altman plots to compare each survey instrument with doubly labeled water assessments. To assess the degree of bias from social desirability or social approval, we fitted regression models using the PROC GLM procedure in SAS, using the difference in physical activity energy expenditure between the self-reported measure of interest and doubly labeled water as the dependent variable (24). We included the social approval and social desirability scores simultaneously as independent variables. The regression coefficient for the social desirability or social approval score reflects the degree of bias, with a positive beta coefficient indicating overestimation of energy expenditure on the self-report instrument and a negative beta coefficient indicating underestimation of energy expenditure.

Previous findings in the literature on diet and physical activity have shown evidence for effect modification by education and body mass index (weight (kg)/height (m)²) (1, 3). Thus, the social approval/social desirability models were stratified by educational status (less than a college education, college education or higher) and body mass index (<27, ≥27). The cutpoint used for body mass index stratification was the median value for the study population. The cutpoint for education was based on prior work from this study (3). Confounding by body mass index, educational status, menopausal status, and age was assessed.

Similar analyses were performed by regressing social desirability or social approval score on the difference in duration of activity between the self-reported measure (24-hour PAR, 7-day PAR 1, or 7-day PAR 2) and the ActiGraph measure. Duration models were stratified by intensity of activity (light, moderate, or vigorous). For analyses involving the 24-hour PAR (n = 72), only those ActiGraph data corresponding to the same day as the 24-hour PAR were included. For the 7-day PAR analyses, subjects were included if they had at least 3 days of ActiGraph data from the observation period of the 7-day PAR (for 7-day PAR 1, n = 68; for 7-day PAR 2, n = 71). The outcomes modeled for these analyses represent the difference in average daily physical activity between each instrument and the ActiGraph (difference = average minutes/day from the instrument minus average minutes/day from the ActiGraph).

RESULTS

Descriptive statistics for the study population have been previously published (3). Average physical activity energy expenditure as measured by doubly labeled water was 12.07 kcal/kg/day (table 1). The average differences between physical activity energy expenditure as estimated by doubly labeled water and each survey instrument were 1.51 kcal/kg/day (standard deviation, 8.76; p = 0.14) for 7-day PAR 1, 13.62 kcal/kg/day (standard deviation, 14.95; p < 0.0001) for 7-day PAR 2, and –1.11 kcal/kg/day (standard deviation, 5.40; p = 0.07) for the 24-hour PAR. The 7-day PAR 2 instrument largely overestimated physical activity energy expenditure, as evidenced by the proportion of positive difference scores (84 percent). In contrast, the difference scores for the 24-hour PAR and 7-day PAR 1 were more equally distributed above and below zero (36 percent and 49...
percent positive difference scores, respectively). Because data for most of the activity duration measures were highly skewed, the median value was used as the measure of central tendency for these variables.

Physical activity energy expenditure estimated from doubly labeled water was significantly correlated with energy expenditure estimated from the 24-hour PAR but not with any other self-report measure (table 2). The ActiGraph measure showed significant moderate correlation with energy expenditure estimated from doubly labeled water, energy expenditure estimated from the 24-hour PAR, and social desirability score but not with energy expenditure estimated from either of the 7-day PARs.

Regression analyses showed that social desirability was associated with a significant overestimation of physical activity energy expenditure on 7-day PAR 2 (table 3). After adjustment for body mass index, social approval was associated with a marginally significant ($p = 0.06$) underestimation of energy expenditure on the 24-hour PAR, with an average difference of $-0.15$ kcal/kg/day for every 1-unit increase in this index. We also examined educational level, body mass index, and age as independent predictors of reporting bias (data not shown). None of these variables was associated with systematic reporting differences in energy expenditure on any instrument.

In an effort to further understand the expression of these biases, we performed stratified analyses by levels of education and body mass index. The power of these analyses was limited by the small numbers, and no interaction term was statistically significant. However, a significant effect of social desirability was found on 7-day PAR 2 in women with less than a college education ($\beta = 1.18$ kcal/kg/day, 95 percent confidence interval: 0.36, 2.01) but not in women with a college education or more. Stratification by body mass index resulted in a significant effect of social approval on the 24-hour PAR among women with a body mass index of 27 or higher ($\beta = -0.25$, 95 percent confidence interval: $-0.45$, $-0.04$) but not among women with a body mass index less than 27.

The Bland-Altman plots for the comparison of each study instrument with doubly labeled water measures are depicted in figures 1, 2, and 3. As evidenced by these graphs, both 7-day PAR 1 and 7-day PAR 2 demonstrated proportional error. No instrument demonstrated an absolute systematic bias.

In analyses assessing the duration of physical activity, social desirability also appeared to result in overestimation

### TABLE 1. Descriptive data on physical activity variables, The Energy Study ($n = 80$), Worcester, Massachusetts, June–October 1997

<table>
<thead>
<tr>
<th>Measure and variable</th>
<th>Mean</th>
<th>Median</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Criterion measures</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAEE* from doubly labeled water (kcal/kg/day)</td>
<td>12.07 (4.49)†</td>
<td>1.28</td>
<td>23.13</td>
<td></td>
</tr>
<tr>
<td>Activity duration (minutes/day) from ActiGraph ($n = 72$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Light activity</td>
<td>232</td>
<td>148</td>
<td>402</td>
<td></td>
</tr>
<tr>
<td>Moderate activity</td>
<td>132</td>
<td>48</td>
<td>262</td>
<td></td>
</tr>
<tr>
<td>Vigorous activity</td>
<td>0</td>
<td>0</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td><strong>Self-report measures</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7-day PAR 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAEE (kcal/kg/day)</td>
<td>13.55 (8.14)</td>
<td>2.45</td>
<td>43.32</td>
<td></td>
</tr>
<tr>
<td>Activity duration (minutes/day) ($n = 68$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Light activity</td>
<td>200</td>
<td>51</td>
<td>665</td>
<td></td>
</tr>
<tr>
<td>Moderate activity</td>
<td>76</td>
<td>0</td>
<td>546</td>
<td></td>
</tr>
<tr>
<td>Vigorous activity</td>
<td>4.5</td>
<td>0</td>
<td>147</td>
<td></td>
</tr>
<tr>
<td>7-day PAR 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAEE (kcal/kg/day)</td>
<td>25.65 (14.43)</td>
<td>5.38</td>
<td>78.70</td>
<td></td>
</tr>
<tr>
<td>Activity duration (minutes/day) ($n = 72$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Light activity</td>
<td>320</td>
<td>82</td>
<td>2,829</td>
<td></td>
</tr>
<tr>
<td>Moderate activity</td>
<td>90</td>
<td>14</td>
<td>841</td>
<td></td>
</tr>
<tr>
<td>Vigorous activity</td>
<td>0</td>
<td>0</td>
<td>180</td>
<td></td>
</tr>
<tr>
<td>24-hour PAR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAEE (kcal/kg/day)</td>
<td>10.95 (3.75)</td>
<td>4.78</td>
<td>20.46</td>
<td></td>
</tr>
<tr>
<td>Activity duration (minutes/day) ($n = 73$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Light activity</td>
<td>229</td>
<td>115</td>
<td>468</td>
<td></td>
</tr>
<tr>
<td>Moderate activity</td>
<td>45</td>
<td>0</td>
<td>199</td>
<td></td>
</tr>
<tr>
<td>Vigorous activity</td>
<td>0</td>
<td>0</td>
<td>80</td>
<td></td>
</tr>
</tbody>
</table>

* PAEE, physical activity energy expenditure; PAR, physical activity recall.
† Numbers in parentheses, standard deviation.
DISCUSSION

In this study, social desirability appeared to influence self-reports of physical activity on both 7-day PARs. The impact of social approval was less evident and only reached marginal significance ($p = 0.06$) on the 24-hour PAR after adjustment for body mass index. However, two of the three survey instruments performed relatively well. It is notable that physical activity energy expenditure estimated from two of the three instruments examined in this study was not significantly different from energy expenditure estimated from doubly labeled water. Thus, on average, reporting errors did not systematically bias physical activity reports at the group level for some of the instruments examined. Although individual errors were large for each instrument examined in comparison with energy expenditure derived from doubly labeled water, an interviewer-administered instrument derived from multiple 24-hour PARs had the lowest overall error. It is interesting that the percentage of women overreporting physical activity was greater for those instruments with the longer recall period (7 days vs. the previous day). To our knowledge, this is the first study that has examined the potential effect of personality traits on reporting errors in physical activity self-reports using both doubly labeled water estimates of physical activity energy expenditure and ActiGraph estimates of intensity-specific activity duration.

To more fully quantify the magnitude of the effect of systematic reporting errors on physical activity energy expenditure, we examined the association between reporting errors and social desirability and social approval using linear regression models. The results are presented in Table 3. The models suggest that social desirability and social approval are both negatively associated with reporting errors, with estimates indicating that for every one unit increase in social desirability and social approval, there is a decrease of 0.12 and 0.16 kcal/kg/day in reporting errors, respectively. These effects were statistically significant ($p < 0.05$) in Model 1 but not in Model 2 after controlling for body mass index.

**TABLE 2.** Correlations among study variables pertaining to measures of physical activity energy expenditure, social desirability, and social approval ($n = 80$), The Energy Study, Worcester, Massachusetts, June–October 1997

<table>
<thead>
<tr>
<th>Measure of PAEE†</th>
<th>Doubly labeled water</th>
<th>7-day PAR† 1</th>
<th>7-day PAR 2</th>
<th>24-hour PAR</th>
<th>Body mass index‡</th>
<th>Social desirability score</th>
<th>Social approval score</th>
<th>Age</th>
<th>Resting metabolic rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAEE from doubly labeled water</td>
<td>0.19</td>
<td>0.05</td>
<td>0.24*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ActiGraph average counts§ (counts/minute/day)</td>
<td>0.30*</td>
<td>0.18</td>
<td>−0.02</td>
<td>0.23*</td>
<td>−0.03</td>
<td>−0.29*</td>
<td>0.03</td>
<td>−0.12</td>
<td>0.02</td>
</tr>
<tr>
<td>Body mass index</td>
<td>0.13</td>
<td>0.04</td>
<td>−0.05</td>
<td>−0.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social desirability score</td>
<td>−0.02</td>
<td>0.12</td>
<td>0.21</td>
<td>0.06</td>
<td>0.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social approval score</td>
<td>−0.02</td>
<td>−0.18</td>
<td>−0.08</td>
<td>−0.21</td>
<td>−0.12</td>
<td>−0.28*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.01</td>
<td>0.02</td>
<td>0.20</td>
<td>0.12</td>
<td>0.13</td>
<td>0.16</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resting metabolic rate</td>
<td>0.04</td>
<td>0.11</td>
<td>0.04</td>
<td>−0.04</td>
<td>0.43</td>
<td>0.02</td>
<td>−0.22*</td>
<td>0.28*</td>
<td></td>
</tr>
</tbody>
</table>

* $p < 0.05$.
† PAEE, physical activity energy expenditure; PAR, physical activity recall.
‡ Weight (kg)/height (m)$^2$.
§ $n = 72$ because of missing values generated from the instrument or participant error.

**TABLE 3.** Models of the association between reporting difference and the covariates social desirability and social approval ($n = 80$), The Energy Study, Worcester, Massachusetts, June–October 1997

<table>
<thead>
<tr>
<th>Model and covariate</th>
<th>7-day PAR† 1</th>
<th>7-day PAR 2</th>
<th>24-hour PAR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta$†</td>
<td>95% CI†</td>
<td>$\beta$†</td>
</tr>
<tr>
<td>Model 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social desirability</td>
<td>0.12</td>
<td>−0.24, 0.48</td>
<td>0.65</td>
</tr>
<tr>
<td>Social approval</td>
<td>−0.16</td>
<td>−0.41, 0.09</td>
<td>−0.15</td>
</tr>
<tr>
<td>Model 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social desirability</td>
<td>0.12</td>
<td>−0.25, 0.48</td>
<td>0.70</td>
</tr>
<tr>
<td>Social approval</td>
<td>−0.16</td>
<td>−0.41, 0.10</td>
<td>−0.17</td>
</tr>
<tr>
<td>Body mass index‡</td>
<td>0.06</td>
<td>−0.45, 0.57</td>
<td>−0.43</td>
</tr>
</tbody>
</table>

* PAR, physical activity recall; CI, confidence interval.
† Regression coefficient obtained in a general linear model of the difference between physical activity energy expenditure derived from self-reported physical activity and physical activity energy expenditure derived from doubly labeled water.
‡ Weight (kg)/height (m)$^2$.  


expenditure values, we calculated the possible range of systematic error using our regression results, the interquartile range of social desirability among women in this study (interquartile range = 7), and the average body mass of women in this study (69.7 kg). Calculations based on the average body mass and the interquartile range would result in a 317-kcal/day (4.55-kcal/kg/day) overestimation of physical activity on 7-day PAR 2 for women in the 75th percen-


<table>
<thead>
<tr>
<th>Instrument and covariate</th>
<th>Light activity</th>
<th>Moderate activity</th>
<th>Vigorous activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>7-day PAR† 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social desirability</td>
<td>2.25</td>
<td>–3.99, 8.48</td>
<td>4.81</td>
</tr>
<tr>
<td>Social approval</td>
<td>–2.06</td>
<td>–6.40, 2.29</td>
<td>–1.03</td>
</tr>
<tr>
<td>7-day PAR 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social desirability</td>
<td>11.30</td>
<td>1.87, 20.73</td>
<td>4.15</td>
</tr>
<tr>
<td>Social approval</td>
<td>–0.005</td>
<td>–6.51, 6.52</td>
<td>1.05</td>
</tr>
<tr>
<td>24-hour PAR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social desirability</td>
<td>2.22</td>
<td>–1.82, 6.27</td>
<td>0.74</td>
</tr>
<tr>
<td>Social approval</td>
<td>2.68</td>
<td>–0.27, 5.64</td>
<td>0.47</td>
</tr>
</tbody>
</table>

* Regression coefficient obtained in a general linear model of the difference between self-reported activity duration and ActiGraph activity duration. The total sample size may vary for each instrument because of the elimination of outlier values in highly skewed and kurtotic distributions.
† CI, confidence interval; PAR, physical activity recall.
tile of social desirability score as compared with the 25th percentile. Increased social desirability score also was associated with a systematic overestimation of duration of activity for light activity (7-day PAR 2 only) and moderate activity (both 7-day PAR 1 and 7-day PAR 2). Using the same interquartile range for social desirability, the durations of light and moderate activity would be overreported by 79 minutes/day and 29 minutes/day, respectively, on 7-day PAR 2.

The greater social desirability effect observed on 7-day PAR 2 compared with 7-day PAR 1 could be a result of the survey format. The first 7-day PAR was much less structured than the second and simply asked respondents to report the total amount of time spent in moderate, vigorous, and very vigorous activities. The participant was not queried directly about light activity. Time spent in light activity was calculated from reports of sleep and moderate-to-vigorous activity. These findings are consistent with our experience with dietary data in that biases may be more concentrated in response to more structured questionnaires (3, 25).

In contrast to 7-day PAR 1, 7-day PAR 2 was structured such that activities were grouped by activity domain and intensity in an effort to systematically assess the full range of activities encountered in daily living. On the basis of our examination of reported activity durations, overreporting appeared to occur for both light and moderate activities. For these women, household activities comprised the majority of reported time spent in light activity (~70 percent). There are at least two possible explanations for this social desirability effect. First, the effect could be mediated through the societal norm for women to be “good caretakers” of the home. Second, the bias may have been expressed more strongly in reports of highly prevalent routine light- and moderate-intensity activities, because it may be that persons who are prone to overreporting may inflate reports of activities they engage in regularly rather than overreport activities they engage in less frequently or not at all. For example, when asked about their past week of both household and exercise-related activities in the structured survey, women with high social desirability scores may have reported spending more time in household activity, particularly when faced with leaving the leisure and sports sections empty. In contrast, on 7-day PAR 1, women were asked about all domains of activity; thus, less emphasis was placed on the types of activities performed. Further investigation of these initial findings appears to be warranted.

Our finding of a marginally significant negative bias associated with social approval on the 24-hour PAR was unexpected. We originally hypothesized that women with higher social approval scores would want to “please” study staff by reporting relatively high levels of activity. The stratified analyses suggested that this effect may have been concentrated among women with a body mass index of 27 or higher. Future research is needed to replicate this finding and to attempt to differentiate the possible influence of the interviewer’s presence in eliciting reporting bias.

Other investigators have attempted to characterize reporting error by participant demographic characteristics such as age, body fat, and physical activity level (26). Irwin et al. (26) reported a significant correlation between body mass index, percentage of body fat, and reporting error for physical activity records but not for a 7-day PAR. Similarly, in our investigation, we did not observe a significant independent association of age, body mass index, or menopausal status with our 7-day PAR. Our finding of no independent association between body mass index and the 24-hour PAR, which is akin to physical activity records, further emphasizes the need to investigate reporting biases in the use of short-term recall methods.

This investigation had a number of limitations that should be considered when interpreting its findings. The study population was heavily scrutinized, with some type of contact being made at least 3–4 times per week and multiple activity assessments being completed by each subject. With this amount of observation, reporting accuracy in this population may have been greater than usual. In this case, the relation between social desirability or social approval and reporting accuracy may have been attenuated. In addition, the study population was comprised of predominately European-American women, thereby limiting the applicability of these findings to men and to minority populations.

In comparing the 24-hour PAR with the 7-day PAR, it was not possible to differentiate between interviewer effects and the effect of recall interval (i.e., the past 24 hours vs. the past 7 days). Future studies should be designed to evaluate the effects of differences between recall-interval effect and different modes of administration on reporting errors. In addition, our reliance on estimated resting energy expenditure values in our calculation of energy expenditure from doubly labeled water certainly resulted in the introduction of some error in our criterion measure. To minimize this loss of precision in our doubly labeled water energy expenditure values, we employed a prediction equation that used our measured lean body mass values derived from the doubly labeled water procedure. The most likely effect of this loss of precision would be a loss of statistical power in our analyses and attenuation of the effects observed.

Similarly, use of the ActiGraph as the criterion measure for the duration analyses may have introduced some bias into the results. While it demonstrates good relative validity against a variety of criterion measures from direct observation to activity diaries (15, 18, 27), there are some activities that ActiGraph activity counts cannot adequately capture (e.g., bicycling or weight-lifting). Consequently, the reporting differences in durations may actually be smaller than calculated; that is, the ActiGraph may not record activity that the participant actually engages in and might ultimately report. Nevertheless, use of the ActiGraph remains one of the few feasible ways to objectively estimate the intensity and duration of physical activity in free-living populations.

This investigation also had several strengths that should be considered. This is one of the first studies to quantify the direct effect of social approval and social desirability on physical activity energy expenditure and duration in a relatively large group of women. The combined use of doubly labeled water and the ActiGraph as criterion measures enabled us to evaluate both bias in absolute physical activity energy expenditure and bias in different activity intensities.
Although this work is an important first step in examining specific sources of bias in self-reported physical activity, clearly much additional research is needed in this area. The stratified analyses suggested that the effects of this bias may be modified by demographic characteristics and body habitus. Investigators will need larger stratum-specific sample sizes to fully understand the relation of these variables. Given the prohibitive cost of doubly labeled water studies, less expensive measures of activity that overcome some of the limitations of waist-mounted accelerometers (i.e., multiple sensors and/or heart rate measurements) could be employed to replicate and extend our findings (28, 29). Because there is some evidence for differential expression of social desirability bias by ethnicity (30), future investigations should focus recruitment on minority populations.

In conclusion, we have described a possible source of systematic biases in certain self-reports of physical activity that are attributable to the personality traits social desirability and social approval. The presence of these biases may depend largely on the type of survey instrument employed. These results suggest that reporting biases may be minimized through survey and questionnaire design. Further study is required to confirm these findings and to better characterize differences in the expression of bias by mode of administration, length of recall, questionnaire structure, and type and intensity of activity reported. As with dietary intake (31, 32), fitting social desirability or social approval scores in regression equations may improve overall model explanatory ability. Additionally, this avenue of inquiry may aid researchers in the creation of new physical activity assessment methods that are less prone to biased reporting.

ACKNOWLEDGMENTS

The authors thank Dr. Patty S. Freedson for her insight in helping to develop the physical activity assessments evaluated in this research and for providing the ActiGraphs used for data collection. Susan Druker, as study manager, also provided invaluable assistance with subject recruitment and data collection.

REFERENCES

24. Irwin ML, Ainsworth BE, Conway JM. Estimation of energy expenditure from physical activity measures: determinants of