Social Desirability Trait Influences on Self-Reported Dietary Measures among Diverse Participants in a Multicenter Multiple Risk Factor Trial1,2

James R. Hebert, Thomas G. Hurley, Karen E. Peterson, Ken Resnicow, Frances E. Thompson, Amy L. Yaroch, Margaret Ehlers, Doug Midthune, Geoffrey C. Williams, Geoffrey W. Greene, and Linda Nebeling

Abstract

Data collected at 4 Behavioral Change Consortium sites were used to assess social desirability bias in self-reports derived from a dietary fat screener (PFat), a dietary fruit and vegetable screener (FVS), and a 1-item question on fruit and vegetable intake. Comparisons were made with mean intakes derived from up to 3 24-h recall interviews at baseline and follow-up (at 12 mo in 3 sites, 6 mo in the fourth). A social-desirability-related underestimate in fat intake on the PFat relative to the 24HR (percentage energy as fat) was evident in women [baseline b = −0.56 (P = 0.005); follow-up b = −0.62 (P < 0.001)]. There was an overestimate in FVS-derived fruit and vegetable consumption (servings/week) in men enrolled in any intervention at follow-up (b = 0.39, P = 0.05) vs. baseline (b = 0.04, P = 0.75). The 1-item fruit and vegetable question was associated with an overestimate at baseline in men according to SD score (b = 0.14, P = 0.02), especially men with less than college education (b = 0.23, P = 0.01). Women with less than college education expressed a similar bias at follow-up (b = 0.13, P = 0.02). Differences in the magnitude of bias according to gender, type of instrument used, and randomization condition are comparable to what has been seen for other instruments and have important implications for both measuring change in studies of diet and health outcomes and for developing methods to control for such biases. J. Nutr. 138: 226S–234S, 2008.

Introduction

Studies investigating the roles of diet and health in humans almost always rely on subjects’ self-reported dietary intake. Structured instruments, such as the FFQ, have been developed to estimate usual (i.e., habitual) dietary intake of individuals, the exposure of interest in most epidemiologic and intervention studies (1–5). A variety of abbreviated FFQ, e.g., fruit and vegetable and fat screeners, have been designed to focus on specific categories of foods or nutrients (6–10). The FFQ, including these abbreviated versions, is superior to other assessment methods in terms of patient burden, other aspects of “cost,” and feasibility of reflecting the views of the National Institutes of Health. Guest Editors: Shirley A. A. Beresford, University of Washington, Seattle, WA, Lisa M. Klesges, St. Jude Children’s Research Hospital, Memphis, TN, and Helaine R. H. Rockett, Harvard Medical School and Brigham and Women’s Hospital, Boston, MA. Guest Editor disclosure: S. A. A. Beresford, L. M. Klesges, and H. R. H. Rockett will receive compensation from NCI, DCCPS, BRP for editorial services provided for this supplement publication; L. M. Klesges was a member of the BCC.

1 Author disclosures: J. R. Hebert, T. G. Hurley, K. E. Peterson, K. Resnicow, F. E. Thompson, A. L. Yaroch, M. Ehlers, D. Midthune, G. C. Williams, G. W. Greene, and L. Nebeling, no conflicts of interest.

2 To whom correspondence should be addressed. E-mail: jhebert@sc.edu.

0022-3166/08 $8.00 © 2008 American Society for Nutrition.
administration, which are important considerations in large-scale observational studies. Although typically used in high-literacy populations, these instruments can be adapted for use in lower-literacy populations (11–13) and may be more practical for population-based intervention trials (14).

Previous work with longer structured questionnaires indicates that there are biases related to demographic factors, body habitus, and psychosocial factors in relation to either short-term, open-ended (i.e., non-list-constrained) assessment methods such as the 24-h dietary recall interview (24HR)12 (15–19) or estimates of total energy (caloric) intake derived from doubly labeled water (DLW) (20–23). Attempts to understand the psychological basis of self-report errors have focused on specific personality traits, also known as “response sets,” that are known to bias responses on tests or questionnaires perceived by people to be tests (24,25). Prominent among these is social desirability (SD), a response set reflecting the defensive tendency to respond in a manner consistent with perceived social norms (26,27).

Over the past several years, evidence has accumulated showing that SD is associated with biases in dietary self-reports on structured questionnaires in comparison to 24HR (28–30) or DLW (31,32), although 1 DLW study did not see such an effect (23). The effects often differ according to the gender and, sometimes, educational level of participants, indicating an interaction between gender (and sometimes education) and a personal tendency to respond in certain ways on dietary questionnaires (29,30). To date, most of the work in this area has been conducted in European and European-American women. Previous work in multiethnic populations has indicated a bias in fat estimation, concentrated mainly in highly educated (at least college degree) women (33). However, no bias has been observed to be associated with SD in estimates of fruit or vegetable consumption (9,33).

The purpose of the current study is to examine the National Cancer Institute (NCI) Fruit and Vegetable Screener (FVS) (34), the NCI Percentage Energy from Fat Screener (PFat) (35), and the 1- or 2-item fruit and vegetable question (36) for SD bias. Data derived from these 3 methods are compared with 24HR-derived values as an estimate of “true intake.”

Methods

The current study focuses on 4 of the 7 Behavior Change Consortium (BCC) sites: University of Rochester (ROC); University of Rhode Island (URI); Emory University; and the Harvard School of Public Health (HSPH). All 4 had included a dietary intervention component to increase fruit and vegetable intake and, in some instances, to decrease fat intake. Common assessment tools, including the FVS (37), PFat (35), and a simple 1- or 2-item question on fruit and vegetable intake (36) were used in all sites to measure dietary behavior at baseline and follow-up (i.e., either after the intervention or at the corresponding time in the control group). Follow-up measurements were made at 12 mo at all sites except ROC, where those data were collected at 6 mo. The FVS and 1-item produce results in the form of servings per unit time (e.g., day or week). The PFat and its method of scoring, which produces results in the form of percentage energy as fat, are described thoroughly elsewhere (35). All 4 sites (URI, HSPH, Emory, and ROC) also collected up to 3 d of 24HR, by telephone-administered interview, from each participant to evaluate the performance of the shorter assessment instruments used by those sites (36). In 3 of the 4 sites (not Emory), the baseline FVS, PFat, and 1-item were administered in the clinic after enrollment into the intervention study and before randomization to usual care vs. experimental care group. The screener instruments were self-administered in 2 sites (Emory, ROC), and interviewer-administered in 2 sites (URI, HSPH). Analyses are based on subjects with paired screener (FVS, PFat, and the 1-item) and 24HR measures at both the baseline and follow-up time points. A total of 228 subjects had paired 24HR and PFat data; for FVS there were 231 such pairs; and for the 1-item 225.

Instruments

The NCI PFat consists of 16 questions on usual consumption of foods over the past year (36). Estimates of percentage energy as fat are computed from respondent-reported frequency responses, assigned gender-age specific portion sizes, and gender-specific regression coefficients. Details of the tool’s development, scoring, and testing are described thoroughly elsewhere (38), and the application in the context of the BCC is described in another article published as part of this special issue (35). The PFat, described in Thompson et al. in this issue (35) and elsewhere (38), is available electronically (39).

The NCI FVS is a 19-item instrument querying usual consumption of specific fruits and vegetables over the past month. The instrument was developed after cognitive testing and was evaluated in 462 adult men and women living throughout the United States (40). Fruit and vegetable servings are quantified in terms of the 1992 Food Guide Pyramid (41). For fruits, a serving is defined as a whole fruit, 1/2 cup cut-up fruit, or 3/4 cup juice (1 cup ≈ 0.237 L). For vegetables, a serving is defined as 1 cup (0.237 L) raw leafy vegetables such as lettuce, 1/2 cup other vegetables, or 3/4 cup vegetable juice. Frequency and portion size reports were used to estimate self-reported individual Pyramid fruit and vegetable consumption in servings (37).

The FVS was administered at all 4 sites represented in the analysis that forms the basis of this article. As noted, it was administered by an interviewer in 2 sites (URI and HSPH), and completed and returned to Emory staff before collection at churches; at ROC, data were collected at the study site just before randomization on the day of enrollment. In all cases, subjects with either incomplete frequency or quantity data, or both, were excluded from analyses.

A very abbreviated fruit and vegetable screener (1-item) is based either on a single fruit and vegetable question (at ROC, URI, and HSPH) or 1 question each on fruit and vegetables (Emory) that were summed to reflect total fruit and vegetable intake (36).

Multiple nonconsecutive 24HR were administered by telephone to respondents by trained interviewers. Data collection for Emory and HSPH was performed by the Diet Assessment Unit of the Cancer Prevention and Control Program at the University of South Carolina; for ROC by the Diet Assessment Center at Pennsylvania State University; and for URI by their staff. Interviews were conducted by trained interviewers using the automated Minnesota Dietary Coding System. All sites followed the same protocol, including a preinterview mailing of a 2-dimensional food portion guide (42) and 3 nonconsecutive, unannounced 24HR over a 3-wk period, including 1 weekend day. Each subject was intended to complete 3 24HR at each time point, and the vast majority did so. Overall, 80.2, 15.7, and 4.1% of the participants contributed 3, 2, and 1 24HR, respectively, at baseline; and 75.7, 18.4, and 6.0% of the participants contributed 3, 2, and 1 24HR, respectively, at follow-up.

The Minnesota Nutrient Data System for Research (versions 4.05_33 and 4.06_34) was used to conduct, code, and process the 24HR. Interviews were reviewed by supervisors, and all missing items were added with consultation from the Minnesota Nutrition Coordinating Center. Coding quality was checked with built-in systems that flag extreme values. Data processing was conducted at the University of South Carolina (for Emory and HSPH), Penn State University (ROC), or URI. All individual recalls defined as unreliable by the interviewer were reviewed and exclusion confirmed by a dietitian with experience in conducting/supervising NDS recalls as well as the Principal Investigator.
from the URI coordinating center. In most cases, reasons for determining that a recall was unreliable were listed in the interviewers’ notes and included examples such as “subject appeared confused,” and “sick/vomiting all day.” Only subjects with at least 1 valid recall were included in analyses.

Pyramid servings of fruits and vegetables were estimated for the 24HR through the United States Department of Agriculture (USDA)’s Continuing Survey of Food Intakes by Individuals (CSFII) 1994–96 survey database [see the article by Yaro et al. (36) in this supplement]. This database provides the number of servings of fruits and vegetables per 100 g for each of >5000 food codes. Foods reported in the BCC and coded using NDS were linked to identical or similar food codes in the USDA database. The estimated percentage energy as fat was used as the major criterion variable for the fat screener analyses.

SD trait was measured using a 10-item version of the Marlowe-Crowne Social Desirability scale [M-C 2(10)], a true-false questionnaire that assesses the tendency to respond in a manner consistent with social norms or beliefs in a testing situation. In piloting in an elderly population and in adults recruited in African-American churches, the M-C 2(10) was found to be acceptable (43–45). The M-C 2(10) was used at all BCC sites that measured SD (and are represented here).

Sociodemographic variables were assessed at baseline using demo-graphic variables common to all BCC sites (gender, age, education, race/ethnicity, income, and marital and employment status). Race/ethnicity options included White-European American, not of Hispanic origin; Black/African American, not of Hispanic origin; Hispanic; Asian or Pacific Islander; American Indian/Alaskan Native; or other. For analyses, the 2 latter categories were collapsed into “other.”

**Intervention condition.** At baseline, subjects were randomized into treatment and control arms. Treatment arms included fat reduction (less fat, less red meat), FV increase, physical activity, smoking cessation, or some combination [see Yaro et al. (36) for details]. In this article we categorize subjects who are in an explicit fat-reduction intervention when percentage energy as fat is the outcome and those who are in an explicit fruit and vegetable intervention when the number of fruits and vegetable servings is the outcome.

**Statistical methods**

Descriptive statistics were computed for each site and over all 4 sites, by gender. The susceptibility of the FVS and PFat screeners to SD bias was evaluated by comparison to the 24HR in a measurement error (ME) model. Exploratory analyses included examining correlations between SD and potential confounders such as relative weight [as estimated by BMI = weight (kg)/height (m)^2], age, and ethnicity. Race/ethnicity options included White-European American, not of Hispanic origin; Black/African American, not of Hispanic origin; Hispanic; Asian or Pacific Islander; American Indian/Alaskan Native; or other. For analyses, the 2 latter categories were collapsed into “other.”

For all tests, exact P-values are given unless they are below a probability of 0.0001 (i.e., P < 0.0001). The nominal level for judging “significance” is α = 0.05.

As noted previously (35), true usual intake is not observable in free-living populations (46). However, we can estimate the distribution of true intake in the population by use of appropriate reference data and by use of statistical methods. Our reference instrument is multiple nonconsecutive 24HR. We applied the ME model, described by Friedman et al. (47), to estimate relationships between true intake and the test instruments, assuming that the reference instrument (24HR) is unbiased at the individual level and contains only within-person error. Using this model, we estimated the overall level of agreement, as described by the slope of the regression line (β) and the correlation coefficient (ρ) between the screeners and 24HR. This model was extended by the addition of a fixed covariate to estimate the bias in the screener relative to true intake that was associated with SD. The mixed model is specified as \( W = \beta_0 + \beta_1X + \beta_2SD + \epsilon \); where W is intake from the screener, X is true intake from the 24HR (a latent construct), and SD is social desirability (assumed to be measured without error). Because there are multiple repeat observations of unknown \( X \) \((24HR_1, 24HR_2, \text{ and } 24HR_3)\) measured with classical error, the ME can be used to fit the normal distribution of a vector \((W, X, 24HR_1, 24HR_2, \text{ and } 24HR_3)\). From its variance-covariance matrix, one can then calculate an estimate of \( \beta_2 \) and, using the Delta method, estimate its standard error (48).

Models were stratified by gender and gender-by-education, as previous research has shown SD bias to vary by these factors (29,30,33). We also included other potential sources of bias and modifiers such as BMI, age, and race/ethnicity. In the baseline analyses of PFat, FVS, and the 1-item in this supplement (35,37), ME models were stratified by site and gender because marked differences in performance were observed among these strata. However, when site was fit as a covariate in stratified models to control for site differences, estimates of SD bias were distorted. Sites in this study have unique, predominantly nonoverlapping gender, age, and education distributions (36). Because of the high correlation between site and these effect modifiers, adding this factor into the model resulted in model instability because of collinearity. Thus, results are presented without stratification for, or explicit control of, site. Also, small numbers, particularly in men, precluded any stratification beyond gender and education. Indeed, given the inherent variability in the measures, there were too few men in educational-level strata for analyses of percentage energy as fat to produce stable estimates.

Analyses were first conducted on the baseline values and then repeated, employing the same analytic procedures, using follow-up values (i.e., postintervention or time concordant in control subjects). Because this study was conducted in the context of a dietary intervention, data also were analyzed to determine whether there were differences between the control and intervention groups, particularly in comparing the baseline to the follow-up assessment. This entailed a simple stratification of models by intervention condition.

**Results**

Consistent with the results for the larger BCC dataset (36), there was considerable variability on most characteristics of the analytic sample (Table 1). Results on the general performance of the 3 test instruments, the PFat, FVS, and 1-item, in relation to 24HR are provided in other articles contained in this *Journal of Nutrition* supplement (35,37,49,50). We also compared means of the 4 measurement instruments by gender, time, and high or low SD category (Table 2). In general, the PFat underestimated relative to the 24HR, but the differences were much larger for women with high SD scores and in follow-up vs. baseline measures in both genders. The FVS produced large overestimates, and these were not altered results (did not modify or confound the effect of the other variables). Therefore, these variables were not fit in the models when we compared 24HR-derived values with those from the PFat (Table 3), the FVS (Table 4), and the 1-item (Table 5). Based on results from previous work, we stratified analyses presented by gender and education (<college or ≥college degree completed). The small number of men who completed the PFat instrument coupled with the variability associated with that measure precluded our being able to stratify analyses of those data by education. For percentage energy as fat (Table 3), there was no suggestion of an effect in men. By contrast, in women there was a consistent, and virtually identical, downward estimate of the PFat-derived estimate of fat intake (as a percentage of energy) at both baseline and follow-up. Unlike previous work, the bias did not appear to attenuate at follow-up.
Women, in an intervention targeting dietary fat intake (n = 34), had an apparent attenuation of the SD effect at both baseline (units = percentage of energy as fat) \[b = -0.03 (P = 0.97)\] and follow-up \[b = 0.25 (P = 0.65)\]. Of interest, those women not receiving an intervention targeting dietary fat intake (i.e., other intervention and control group; n = 146) evinced an SD effect that was roughly equal to the overall group average [baseline \(b = -0.62 \) (P = 0.01) and follow-up \(b = -0.70 \) (P < 0.01)].

In contrast to what was observed in women for self-reports of energy as fat, there was no suggestion of a bias of a FVS-derived self-report of fruit and vegetable intake in women relative to their 24HR-derived values (Table 4). Overall, results were null in men, as they had been for intake of fat (percentage of energy as fat). Roughly 10% more men completed the FVS (n = 53) than the PFat (n = 48). Despite somewhat small numbers, stable results were obtained for fruit and vegetable intake when data were stratified by education in men. Although no result achieved statistical significance in any stratum of gender-by-education, there was a suggestion of an increase in the regression coefficient, \(b\), for men with <college education at follow-up. In analyzing the data to determine whether there were differences by intervention status, we found that FVS-derived intake (units = servings/wk) was associated with a SD-related overestimate in men receiving any intervention \(n = 39\) at follow-up \(b = 0.39, P = 0.05\) vs. baseline \(b = 0.04, P = 0.75\). Intervention condition-stratified results in women were absent any effect even remotely close to significant (all \(P\)-values between 0.31 and 0.98).

The most prominent biases in the 1-item were observed in subjects with <college education (Table 5). Less-educated men overestimated most at baseline, whereas less-educated women overestimated most at follow-up. Women randomized to intervention expressed a bias at follow-up \(b = 0.13, P = 0.05\) relative to those in the control group \(b = 0.06, P = 0.38\).

### Discussion

The observed SD-related underestimate (downward bias) in PFat-derived fat intake in women, but not men (Table 3), is consistent with observations from other studies (28,29,51). It appears that the biased report is more specifically focused on fat energy (the numerator) than total energy (the denominator). The bias was evident across the regional, racial/ethnic, age, and gender diversity of the populations contributing data to these analyses; and it was roughly of the same magnitude, but slightly larger, at follow-up. In contrast to previous studies, however, the difference at either time was not concentrated in highly educated women. The point estimates observed represent large downward biases (i.e., negative \(b\) values) in women. After the relative size of this SD scale in relation to that used previously (about one-third, i.e., 10 vs. 33 questions) has been taken into account, the observed results are of approximately the same magnitude (i.e., −0.6% of energy as fat/point vs. 0.2% of energy as fat/point) (28,29,33). Biases of this magnitude may distort findings from epidemiologic studies of diet and health (32,33,52).

The results for FVS-derived fruit and vegetable intake (Table 4) at baseline also are broadly consistent with earlier observations; i.e., no SD bias in the structured questionnaire (the FVS) vs. the open-ended comparison method (the 24HR) in women or men. Results at follow-up indicating the relative overestimate attributable to SD trait in men in the intervention (as opposed to the control) group have implications for future studies.

The results obtained for the 1-item-derived fruit and vegetable intake (Table 5) were interesting in that the largest SD biases were observed in both male and female subjects with low educational attainment, and these differed by gender and time, with men expressing the bias at baseline and women expressing it at follow-up. It appears that women at follow-up who were randomized to intervention reported intake with increased SD bias, whereas men seemed to be less prone to bias at follow-up.

In the absence of a true criterion for dietary fat or fruits or vegetables (53–58), validation studies of dietary assessment instruments are usually relegated to using a relative criterion measure of exposure. As in most of the studies that have examined the relation between specific dietary assessment methods (59,60) or have tested for bias (28,29,33), this study used the 24HR as its criterion measure. Until the last decade or so, it was assumed that short-term, open-ended instruments have different error structures than longer-term closed-ended structured instruments such as the FFQ (61,62). It has been shown by various groups (42,63,64) that the 24HR is associated with smaller total error

---

**TABLE 1** Demographic and SD data (at baseline) by gender \((n = 267)^1\)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Male</th>
<th>Female</th>
<th>(P)-value^2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n)</td>
<td>%</td>
<td>(n)</td>
</tr>
<tr>
<td>Age, y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18–39</td>
<td>5</td>
<td>9</td>
<td>75</td>
</tr>
<tr>
<td>40–59</td>
<td>19</td>
<td>32</td>
<td>57</td>
</tr>
<tr>
<td>60 or older</td>
<td>35</td>
<td>59</td>
<td>76</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Up to HS/TS grad or GED</td>
<td>27</td>
<td>47</td>
<td>116</td>
</tr>
<tr>
<td>Some college</td>
<td>16</td>
<td>28</td>
<td>43</td>
</tr>
<tr>
<td>College grad or more</td>
<td>14</td>
<td>25</td>
<td>44</td>
</tr>
<tr>
<td>Race/ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White/European American</td>
<td>42</td>
<td>73</td>
<td>98</td>
</tr>
<tr>
<td>Black/African American</td>
<td>9</td>
<td>16</td>
<td>63</td>
</tr>
<tr>
<td>Hispanic</td>
<td>1</td>
<td>2</td>
<td>36</td>
</tr>
<tr>
<td>Other^3</td>
<td>5</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>Employment status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full time</td>
<td>16</td>
<td>28</td>
<td>60</td>
</tr>
<tr>
<td>Part time</td>
<td>2</td>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>Unemployed</td>
<td>7</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>Homemaker</td>
<td>0</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>Retired</td>
<td>32</td>
<td>55</td>
<td>65</td>
</tr>
<tr>
<td>Other</td>
<td>3</td>
<td>5</td>
<td>33</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married or cohabit</td>
<td>43</td>
<td>74</td>
<td>110</td>
</tr>
<tr>
<td>Single</td>
<td>9</td>
<td>16</td>
<td>33</td>
</tr>
<tr>
<td>Widowed</td>
<td>3</td>
<td>5</td>
<td>33</td>
</tr>
<tr>
<td>Separated/divorced</td>
<td>3</td>
<td>5</td>
<td>28</td>
</tr>
<tr>
<td>Smoking status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>22</td>
<td>38</td>
<td>37</td>
</tr>
<tr>
<td>No</td>
<td>36</td>
<td>62</td>
<td>167</td>
</tr>
<tr>
<td>BMI (kg/m²) category^4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal (18.5 &lt; \text{BMI} &lt; 25.0)</td>
<td>13</td>
<td>22</td>
<td>50</td>
</tr>
<tr>
<td>Overweight (25 \leq \text{BMI} &lt; 30.0)</td>
<td>26</td>
<td>45</td>
<td>70</td>
</tr>
<tr>
<td>Obese (\text{BMI} &gt; 30)</td>
<td>19</td>
<td>33</td>
<td>82</td>
</tr>
<tr>
<td>Mean Std^5 Mean Std^5</td>
<td>7.1</td>
<td>2.3</td>
<td>7.3</td>
</tr>
</tbody>
</table>

^1 This table presents frequencies or mean and standard deviation, as appropriate.

^2 \(P\)-value based on \(\chi^2\), Fisher's exact, or t test, as appropriate.

^3 Other includes Asian, Pacific Islander, American Indian, Alaskan Native, and other groups.

^4 BMI = body mass index = weight (kg)/height (m²).

^5 Std = standard deviation.
TABLE 2  Comparison of percentage energy from fat and fruit and vegetable intake by assessment method, gender, time, and SD category in the BCC sample1

<table>
<thead>
<tr>
<th>Subject category/time / SD score1</th>
<th>Percentage calories as fat2,3</th>
<th>Fruit and vegetable intake,2,4 servings/wk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>24HR</td>
<td>PFat</td>
</tr>
<tr>
<td>Males Baseline</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low SD score</td>
<td>31.9 (n = 26)</td>
<td>31.5 (n = 26)</td>
</tr>
<tr>
<td>High SD score</td>
<td>32.2 (n = 22)</td>
<td>30.9 (n = 22)</td>
</tr>
<tr>
<td>Follow-up</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low SD score</td>
<td>32.5 (n = 26)</td>
<td>30.6 (n = 26)</td>
</tr>
<tr>
<td>High SD score</td>
<td>31.4 (n = 22)</td>
<td>29.6 (n = 22)</td>
</tr>
<tr>
<td>Females Baseline</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low SD score</td>
<td>31.5 (n = 88)</td>
<td>31.4 (n = 88)</td>
</tr>
<tr>
<td>High SD score</td>
<td>33.5 (n = 92)</td>
<td>30.5 (n = 92)</td>
</tr>
<tr>
<td>Follow-up</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low SD score</td>
<td>32.2 (n = 88)</td>
<td>31.4 (n = 88)</td>
</tr>
<tr>
<td>High SD score</td>
<td>33.1 (n = 92)</td>
<td>30.0 (n = 92)</td>
</tr>
</tbody>
</table>

1 Data shown are for both genders at both measurement times and with SD score cut at the median (I = 7.5), where low scores are <median and high scores are ≥median; value in parentheses are the n contributing data to a particular analysis or the P-value, as appropriate.

2 Values shown are the mean for each category, as defined by the column label.

3 Dietary fat as a percentage of energy derived from the 24HR or the NCI PFat Screener.

4 Fruit and vegetable intake in servings/wk derived from the 24HR, the FVS, or the 1-item Screener.

5 Diff = the difference between the measures with 24HR subtracted from the comparison value (e.g., PFat – 24HR); shown in parentheses is the P-value for the test H0: difference = 0 (which takes into account the standard deviation of each measure).

TABLE 3 Effect of SD bias in PFat-derived percentage of energy from fat in the BCC sample (n = 228)1

<table>
<thead>
<tr>
<th>Subject Category</th>
<th>Baseline measures</th>
<th>Follow-up measures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>$b^0$</td>
</tr>
<tr>
<td>Males</td>
<td>48</td>
<td>0.01</td>
</tr>
<tr>
<td>Females</td>
<td>180</td>
<td>-0.56</td>
</tr>
<tr>
<td>High Education6</td>
<td>74</td>
<td>-0.62</td>
</tr>
<tr>
<td>Low Education6</td>
<td>102</td>
<td>-0.61</td>
</tr>
</tbody>
</table>

1 Values shown are based on the ME model; bias is estimated by comparing the percentage of energy as fat derived from the PFat instrument with the percentage of energy as fat derived from the 24HR.

2 Slope of the regression line ($b$ as the computed $b$ coefficient) obtained by regressing baseline screener-derived value on 24HR-derived value.

3 Standard error of the corresponding $b$ coefficient.

4 P-value for the test $H_0$: $b = 0$.

5 Slope of the regression line ($b$ as the computed $b$ coefficient) obtained by regressing follow-up screener-derived value on 24HR-derived value.

6 Educational level is defined as <college or ≥college; 4 women were missing educational level.

TABLE 4 Effect of SD bias in FVS-derived intake of fruits and vegetables (servings/wk) in the BCC sample (n = 231)1

<table>
<thead>
<tr>
<th>Subject Category</th>
<th>Baseline measures</th>
<th>Follow-up measures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>$b^0$</td>
</tr>
<tr>
<td>Males</td>
<td>53</td>
<td>0.01</td>
</tr>
<tr>
<td>High education6</td>
<td>24</td>
<td>-0.05</td>
</tr>
<tr>
<td>Low education6</td>
<td>27</td>
<td>0.05</td>
</tr>
<tr>
<td>Females</td>
<td>178</td>
<td>-0.01</td>
</tr>
<tr>
<td>High education6</td>
<td>71</td>
<td>0.12</td>
</tr>
<tr>
<td>Low education6</td>
<td>103</td>
<td>-0.11</td>
</tr>
</tbody>
</table>

1 Values shown are based on the ME model; bias is estimated by comparing number of fruit and vegetable servings derived from the FVS with servings derived from the 24HR.

2 Slope of the regression line ($b$ as the computed $b$ coefficient) obtained by regressing baseline screener-derived value on 24HR-derived value.

3 Standard error of the corresponding $b$ coefficient.

4 P-value for the test $H_0$: $b = 0$.

5 Slope of the regression line ($b$ as the computed $b$ coefficient) obtained by regressing follow-up screener-derived value on 24HR-derived value.

6 Educational level is defined as <college or ≥college; 2 men and 4 women were missing data on educational level.
improving methods for assessing intake of specific foods or food categories is of interest to those studies as well (14).

The role of fat intake in health has been a matter of debate and controversy for well over 2 decades, in part because of the problems associated with measurement of fat and the inconsistency of study results (82–90). This has been a particularly contentious area of research for a variety of reasons including the ubiquitous nature of the exposure (88), the association between caloric density of the diet and the rise in prevalence of obesity (91–94), continuing problems with assessment methodologies (88,95–97), and the fact that food production and marketing is a multi-trillion-dollar industry (98,99). So, as for fruit and vegetable intake, understanding potential biases in the report of fat intake has considerable importance.

Because of the generally greater value attached to slimmer and low total dietary intake in women (as compared with men) (100,101), we hypothesized that dietary self-reports of fat in women would more strongly reflect a tendency to choose socially desirable responses (rather than more objectively accurate ones) in an effort to present themselves in a positive light and avoid criticism (102). Though biases have been observed earlier, those studies have been conducted almost exclusively in primarily middle-class and well-educated individuals (28,29). Our recent work conducted in multiethnic female health center employees indicates that the SD bias is concentrated in more highly educated women, i.e., those with college or more education (33). Unlike the previous studies, the results of the current study are not consistent with formal education influencing SD bias except in the 1-item being associated with an SD-related overestimate in less-educated men at baseline and in less-educated women at follow-up. However, they do confirm the results of other studies indicating that fat and total energy intakes (9,29–31,33,103) are consistently associated with an SD-related downward biasing of data derived from structured questionnaires.

In analyzing the BCC data to determine whether there were differences by intervention status, we found that FVS-derived intake was associated with an SD-related overestimate in men in the intervention condition at follow-up but not at baseline. This result is consistent with that found by Kristal et al. (104) showing that an intervention-related bias can occur in dietary assessment instruments. Concentration of an upward bias in the 1-item in all men at baseline and in less-educated intervention women at follow-up calls attention to the sensitivity of the effect modification of the measures by SD in relation to gender, education, and whether or not someone was randomized to intervention. A possible explanation for the differential effect of SD by intervention category might have been that there was a real intervention effect that was related to the acquiescent personality type associated with high scores on the SD scale (105). The absence of an apparent intervention effect (50) argues against this. It is much more likely that the observed effect is caused by bias in self-report and not by an indirect relation between SD (e.g., acting as a proxy for a psychological predisposition related to acquiescent personality type leading to a successful intervention effect) and true change in fruit and vegetable consumption. Clearly, differences in the level of bias according to intervention condition have important implications for both measuring phenomena under study and for developing methods that might be used to control for such biases. Although results observed in this study need to be reproduced, our findings do suggest that outcome data for intervention studies could be biased solely based on the proportion of subjects with high SD, potentially distorting the outcome of a study away from the null. A logical consequence of the drive to design questions that are short and simple is that they are therefore very face valid. Indeed, it is this face validity that may drive the bias by revealing to the respondent a very obvious “right” answer (28–30,32,106,107).

One limitation of this study, which is shared with most in this area, is the absence of a true criterion measure of exposure. Another limitation includes the number of days of 24HR available. Although the 24HR is a relatively robust method for estimating macronutrient and total energy intake for groups (42,63,64,108), additional days of intake representing a longer period of time would have provided a more precise estimate of individual usual intake. Currently, the expense of this method precludes having very many days of observation. Increases in variability with fewer days of 24HR (note that >80% of subject had 3 d) would tend to obscure real correlations in the data (i.e., increased type II error). Relatively few men were represented in the analyses, and this was a severe limitation for PFat. To have robust overall estimates of bias, especially in educational strata, it would be necessary to have much larger numbers with which to work. Although 4 sites contributed data to these analyses, regional representation of subjects in the study was limited. There were virtually no individuals included outside of the Northeast, the Chicago area, and Metropolitan Atlanta. This meant that special, high-risk, groups such as the 55% of all African Americans who live in the rural Southeast were completely unrepresented. As noted, each site recruited a special demographic group (36). Therefore, there was a strong correlation between site (and, therefore, region of the country) and demographic group (i.e., age, gender, race/ethnicity). Thus, our finding of no significant effect (or modification of other effects) when BMI, age, and race/ethnicity were included in the models must be interpreted cautiously. The implication for future studies is that they will need to recruit sufficient numbers of each particular demographic subgroup to deepen our understanding of how these biases express themselves among people with different backgrounds and attitudes about eating. Finally, participants were excluded if they did not complete all items on the screeners. This may lead to biases in factors that could determine reporting accuracies on self-report instruments of the type we examined.

### TABLE 5 Effect of SD bias in estimated intake of fruits and vegetables (servings/wk) in BCC sample (n = 225) 1-item question vs. 24HR

<table>
<thead>
<tr>
<th>Subject category</th>
<th>n</th>
<th>( b^1 )</th>
<th>SE</th>
<th>P-value ( ^1 )</th>
<th>( b^5 )</th>
<th>SE</th>
<th>P-value ( ^1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>51</td>
<td>0.14</td>
<td>0.06</td>
<td>0.02</td>
<td>0.00</td>
<td>0.05</td>
<td>0.99</td>
</tr>
<tr>
<td>High education</td>
<td>24</td>
<td>-0.01</td>
<td>0.07</td>
<td>0.92</td>
<td>-0.08</td>
<td>0.14</td>
<td>0.57</td>
</tr>
<tr>
<td>Low education</td>
<td>27</td>
<td>0.23</td>
<td>0.08</td>
<td>0.01</td>
<td>0.04</td>
<td>0.05</td>
<td>0.47</td>
</tr>
<tr>
<td>Females</td>
<td>174</td>
<td>0.05</td>
<td>0.05</td>
<td>0.54</td>
<td>0.09</td>
<td>0.05</td>
<td>0.09</td>
</tr>
<tr>
<td>High education</td>
<td>71</td>
<td>0.08</td>
<td>0.09</td>
<td>0.33</td>
<td>0.04</td>
<td>0.09</td>
<td>0.69</td>
</tr>
<tr>
<td>Low education</td>
<td>108</td>
<td>-0.01</td>
<td>0.07</td>
<td>0.92</td>
<td>0.13</td>
<td>0.05</td>
<td>0.02</td>
</tr>
</tbody>
</table>

1 Values shown are based on the ME model; bias is estimated by comparing number of fruit and vegetable servings per week derived from the 1-item fruit and vegetable question with servings derived from the 24HR.
2 Slope of the regression line (\( \beta \) as the computed \( \beta \) coefficient) obtained by regressing baseline screener-derived value on 24HR-derived value.
3 Standard error of the corresponding \( \beta \) coefficient.
4 \( P \)-value for the test \( H_0: \beta = 0 \).
5 Slope of the regression line (\( \beta \) as the computed \( \beta \) coefficient) obtained by regressing follow-up screener-derived value on 24HR-derived value.
6 Educational level is defined as <college or ≥college; 2 men and 4 women were missing educational level.

Social desirability and dietary self-report 231S
In summary, despite its limitations, the results of this study were broadly consistent with findings from previous work showing that estimations of fat intake based on structured questionnaires are biased for women according to SD. Unlike other studies, however, this bias was found to be generalized across education levels rather than being limited only to highly educated women. One surprising result is the finding of bias in the 1-item estimator of fruit and vegetable intake that was most pronounced in subjects with low educational attainment and appeared to be affected differentially over time by gender. If investigators wish to use this method, careful monitoring of bias should be put in place. The magnitudes of the biases observed in this study are large enough to be of concern with respect to distorting estimates of effect in epidemiologic studies of diet and health. Further studies in which larger numbers of groups typically underrepresented in this area of research are recommended. Furthermore, these findings suggest that methodology could be developed to control for SD in scoring these measures or in analytic control of statistical analyses, or both.

Literature Cited


69. McCory MA, Fuss PJ, McCallum JE, Yao M, Vinken AG, Hays NP, Roberts SB. Dietary variety within food groups: association with Social desirability and dietary self-report. 233S


