Evaluation of a Short Dietary Assessment Instrument for Percentage Energy from Fat in an Intervention Study

Frances E. Thompson, Douglas Midhun, Geoffrey C. Williams, Amy L. Yaroch, Thomas G. Hurley, Ken Resnicow, James R. Hebert, Deborah J. Toobert, Geoffrey W. Greene, Karen Peterson, and Linda Nebeling

Abstract

The need for an inexpensive measure of dietary intake in intervention studies led to evaluation of the National Cancer Institute (NCI) Percentage Energy from Fat short instrument (PFat) in a subgroup of the Behavioral Change Consortium (BCC) intervention sites. The PFat’s performance was evaluated using multiple nonconsecutive 24-h dietary recalls (24HR) as a reference instrument among participants at baseline in 4 demographically diverse intervention sites of the BCC. Mean estimates of percentage energy from fat for 24HR and PFat were within 2.1 percentage points of each other in all but 2 site/gender comparisons. 24HR and PFat estimates were not significantly different (P < 0.05) among men for 2 of 3 sites, and among women for 2 of 4 sites. Deattenuated Pearson correlation coefficients for the PFat and true intake (as estimated from the 24HR using a measurement error model) were significantly different from 0 (P < 0.05) for men and women in all sites, ranging from 0.52 to 0.77 among men and 0.36 to 0.59 among women. Besides gender and site, no other factors examined (age, education, smoking status, and BMI) consistently moderated validity estimates. If accurate assessment of diet at baseline (and presumably at follow-up) is essential, a more detailed instrument such as multiple 24HR may be warranted. The question of whether the PFat adequately measures change in diet is addressed in another article in this supplement. J. Nutr. 138: 193S–199S, 2008.

Introduction

Concerns about the role of total dietary fat and specific categories of fatty acids continue to figure prominently in national health objectives and dietary guidance (1–3). National health objectives for the year 2010 include “increase the proportion of persons aged 2 y and older who consume <10% of calories from saturated fat,” and “increase the proportion of persons aged 2 y and older who consume no more than 30% of calories from total fat” (1). The Dietary Guidelines for Americans 2005 (2) and the Institute of Medicine (3) recommend that for adults total fat intake be between 20 and 35% of energy, with most fats coming from sources of polyunsaturated and monounsaturated fatty acids and limiting intake of fats and oils high in saturated and/or trans fatty acids. It is recognized that a diet that provides recommended levels of fruits and vegetables, 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 Woman’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 To whom correspondence should be addressed. E-mail: thompsof@mail.nih.gov.
fiber, and the wide array of other macro- and micronutrients required must do so within an overall energy constraint and thus cannot include excessive fats. Percentage of energy from fat is a useful indicator of this dimension; the proportion of energy from fat has been found in some studies to be a strong predictor of total caloric intake (4). Dietary interventions have been designed to reduce percentage energy from fat for a variety of conditions in both primary risk factor reduction (5–11) and secondary risk factor reduction (12).

Intervention researchers require precise, reproducible instruments to measure fat intake. The interviewer-administered 24-h dietary recall (24HR) provides the most accurate and complete self-reported information about the individual's diet for a given day (13). However, because this method currently requires highly trained interviewers and imposes a heavy investigator burden with regard to administration, coding, and data processing, it is prohibitively expensive and therefore not feasible for many research applications. In some situations, self-administered FFQ, often optically scannable for inexpensive data entry, are used. However, comprehensive FFQ consisting of >100 items often require up to 1 h to complete and, if self-administered, require a high level of literacy. Furthermore, FFQ have been found to have substantial measurement error (14–17).

Various shorter tools that measure a limited number of dietary factors rather than the entire diet have been developed (13). Although they are more feasible than longer instruments to administer, they are limited in the amount of information they capture. Numerous short instruments have been developed to assess fat intake in U.S. populations. Most rank individuals based on their fat intake but do not attempt to estimate the total absolute amounts of fat consumed (18–24). Furthermore, because dietary guidance is given in terms of percentage energy from fat and saturated fat, instruments that estimate these parameters are desirable. Absolute fat intake and percentage energy from fat measure 2 different constructs. For example, those who eat a lot may have high fat intakes but also relatively low percentage energy from fat. Two instruments have been developed that quantify individual fat intake as a percentage of total energy (25,26). The Behavior Change Consortium (BCC), an NIH-funded set of intervention trials, chose to use the NCI percentage of energy from fat short instrument (PFat) (26).

The purpose of this article is to evaluate the performance of the PFat in a set of intervention studies at baseline. We compare estimates from the PFat instrument to those from multiple 24HR, controlling for within-person variability in the 24HR. In addition, we examine factors that may moderate the screener’s performance. Results of analyses evaluating the ability of the screener to measure change caused by the various interventions are presented in another article in this supplement (27).

Subjects and Methods

Study description. Individual sites participating in the BCC Nutrition Working Group are described in detail in Yaroch et al. (28). Four sites administered both a PFat and multiple 24HR and are included in these analyses: University of Rhode Island (URI); Harvard School of Public Health (HSPH); Emory University; and University of Rochester (ROC). These 4 sites reflect 4 distinct populations. URI participants were all at least 60 y old and predominantly white. HSPH participants were postpartum women attending WIC clinics, mostly Hispanic and under 40 y of age. Emory participants were predominantly African American and mostly women. ROC participants were all smokers and mainly white and under 60 y of age. Distributions of demographic and lifestyle factors by site are listed in Table 1.

In 3 of the 4 sites, the baseline PFat was administered after enrollment into the intervention study and before randomization to usual care vs. treatment; at Emory, the PFat screener was administered after randomization but before intervention. The PFat was self-administered in 2 sites (Emory, ROC) and interviewer-administered in 2 sites (URI, HSPH). Up to 3 nonconsecutive unannounced 24HR, including 2 weekdays and 1 weekend day, were administered by telephone to each participant. These 24HR were conducted after administration of the screener except for Emory, where the sequence was reversed. The PFat consists of 16 questions that ask about usual consumption of foods over the past year (28). The foods on the screener were identified as those that best explained variability in percentage energy from fat in a nationally representative sample of adults. Foods could be positive predictors of percentage energy from fat, e.g., bacon, French fries, or could be negative predictors of percentage energy from fat, e.g., fruit, cold cereal, skim milk. Some major sources of fat in the U.S. population, e.g., beef, fried chicken, are not major predictors of variability in percentage energy from fat and so are not included on the screener. Estimates of percentage of energy from fat are computed from respondent-reported frequency responses, assigned externally derived gender- and age-specific portion sizes, and gender-specific regression coefficients using USDA's 1994–1996 Continuing Survey of Food Intakes by Individuals. Details of the tool's development, scoring, and testing are given by Thompson et al. (26). The PFat is available electronically (29).

The total baseline sample consisted of 1474 participants. Of these, 524 (36%) completed both the screener and at least 1 recall, 855 (58%) completed only the PFat, and 95 (6%) completed only 24HR. In this article, we include for analysis and presentation data from only those 524 participants who completed both the PFat and 24HR. Of these, 74% had 3 24HR; 19% had 2 24HR; and 7% had a single 24HR.

Analytical procedures. Validity is defined as the concurrence between measured and true exposure. Although true usual dietary intake in free-living populations is impossible to measure (30), its distribution in the population can be estimated, as can the relation between true and screener-reported intake, by use of an appropriate reference instrument and statistical methods. The reference instrument used in the BCC study is multiple 24HR and is analyzed using a latent variable measurement error model, described by Freedman et al. (31).

Measurement error is defined as the difference between reported and true exposure and is conceptualized as being composed of systematic bias and within-person random error. The systematic bias may be related to true intake or to other characteristics such as age or BMI, whereas the within-person error has a mean of 0 and is unrelated to any other measurements. True intake is modeled as a latent variable using information from repeated measures of the reference instrument. The screener is assumed to have systematic bias and within-person random error, whereas the reference instrument is assumed to have within-person random error but no systematic bias. The model can incorporate covariates such as age or BMI to allow for systematic biases other than those related to true intake.

Objective biomarker measures are available for energy and protein intake (14–17) but not for fat and percentage energy from fat (32). Self-report measures of diet are associated with error from a variety of sources. The 24HR is considered one of the better self-report methods available, not only because of the detailed description of the diet obtained but also because lack of literacy among potential respondents does not adversely impact the quality of information obtained, and because, if obtained without prior notice, there is no reactive effect (13). However, even with its advantages, the 24HR has been found in biomarker studies to contain individual-level bias for some nutrients, with bias toward underreporting of energy (14,16,17,33,34). In the BCC, we used multiple 24HR as the criterion gold standard, albeit an “alloyed” gold standard.

Our analytical objectives in this intervention context were several. First, we evaluated the ability of the PFat to estimate the mean and

---

12 Abbreviations used: 24HR, 24-h dietary recall; BCC, Behavioral Change Consortium; HSPH, Harvard School of Public Health; NCI, National Cancer Institute; PFat, Percentage Energy from Fat short instrument; ROC, University of Rochester; URI, University of Rhode Island.
distribution of percentage energy from fat intake in the population and
to adequately characterize the baseline intake for comparison to
postintervention intake. Second, we evaluated the ability of the PFat
to estimate intake of an individual and to rank an individual’s intake within
the population, useful to determine eligibility for entry into the study
and/or later assignment to the appropriate treatment group. Third, we
examined whether other factors moderated the relation between the
 screener and true intake. Because preliminary analyses revealed signif-
icant site and gender differences in all parameters, we performed all
analyses presented here stratified by site and gender.

We assessed the ability of the PFat to estimate mean intake in the
population by comparing means of individuals’ percentage energy from
fat values from multiple recalls to that from their screener and used a
paired t-test to test for differences, using $P < 0.05$. To see how well
the PFat estimates other characteristics of the distribution of intake, we
compared prevalence estimates from the screener to true prevalence as
estimated from multiple 24HR in the measurement error model.
Prevalence is the proportion of individuals in the population above or
below a specified level of intake. The deattenuated Pearson correlation
coefficient and its standard error also are estimated from the model
and use external variance adjustment factors to adjust prevalence estimates
and have poor sensitivity. Thompson et al. (35) suggest multiplying the
screener by a variance adjustment factor so that its variance more closely
approaches the variance of true intake in the population. We applied these external variance adjustment factors to
the PFat screener in the BCC samples when estimating prevalence and
positive predictive values were estimated using the measurement error model.

The scoring algorithm for the screener was developed in an external
data set (Continuing Survey of Food Intakes by Individuals, 1994–1996)
and uses a regression model to estimate the conditional expectation of
true intake given the screener responses. Because the screener estimates
derives from a conditional expectation, mathematically, the screener estimate should have about the same mean as, but smaller variance than,
true intake (35) and so, without any adjustment, will provide poor estimates of prevalence and have poor sensitivity. Thompson et al. (35)
suggest multiplying the screener by a variance adjustment factor so that
its variance more closely approximates the variance of true intake in the
population. The formula for the variance-adjusted screener is:

\[
\text{variance-adjusted screener} = (\text{variance adjustment factor}) \\
\times (\text{unadjusted screener} - \text{mean}_{\text{adjusted}}) + \text{mean}_{\text{adjusted}}
\]

In the NIH-AARP Diet and Health Study, the estimated variance
adjustment factor for the PFat screener was 2.0 for men and 1.7 for
women (26). We applied these external variance adjustment factors to
the PFat screener in the BCC samples when estimating prevalence and
positive predictive value.

We examined agreement between the screener and the 24HR by
treatment group (control, fat as secondary intervention, fat as primary
intervention) to confirm effective randomization of participants to
treatment group. In addition, we examined whether other factors
moderated the agreement between the screener and the 24HR. These
included individual demographic variables: years of age (18–39, 40–59,
60 or older), educational status (less than high school, high school only,
more than high school); and standard defined categories of BMI [weight
(kg)/height (m)$^2$—normal weight (<18.5–24.9), overweight (25–29.9),
and obese (≥30)] (36). We created 2 variables to characterize the level
of agreement in mean estimates between the 24HR and screener: the
difference score, defined as 24HR – screener, and the ratio score, defined
as 24HR/screener. The difference score would characterize the amount
than the true prevalence in the population. Positive and negative predictive values were estimated using the measurement error model.

We assessed the ability of the PFat to estimate individual intake by
estimating the screener’s positive and negative predictive values at intake
levels at or <30% energy from fat [recommended in Healthy People 2010 (1)] and <35% energy from fat [recommended by Dietary
Guidelines for Americans (2) and Institute of Medicine (3)]. The positive
predictive value of a screening tool represents the proportion of subjects
selected into the study who are truly eligible, whereas the negative
predictive value represents the proportion of subjects excluded from the
study who were truly ineligible. For an instrument to be useful as a
screening tool, the positive predictive value should be substantially larger
and direction of bias, whereas the ratio score would characterize the proportional bias. Because both scores appeared to be approximately normally distributed, analyses of variance methods were used to test whether mean scores differed among different subgroups using P < 0.05. Extreme scores, as defined below, were excluded from these analyses. Because of cross-site differences in the distribution of potential moderating variables, sample sizes were not sufficient to test all variables in all sites. (Small sample size and consequently unstable model estimates also precluded examination of differences in correlation coefficients across site/genders.)

In our analysis, the measurement error model we used assumes all variables are normally distributed. Because percentage fat was in this study approximately normally distributed for both the PFat and 24HR, no transformation to normality was needed. Before analysis, we excluded extreme values of percentage energy from fat from each instrument to avoid their undue influence. For each gender, values > 3 interquartile ranges below quartile 1 or 3 interquartile ranges above quartile 3 of that variable’s distribution were excluded (for each of 3 d of 24HR, percentage energy from fat values were −10.3 to −17.3 and 73.4 to 80.2 for males and −15.4 to −17.9 and 80.0 to 80.8 for females; for PFat, percentage energy from fat values were 15.2 and 45.8 for males and 9.6 and 50.9 for females). This procedure was followed for each day of dietary recall and for the screener. Under these criteria, no values were excluded for the 24HR; 1 value (57.4) was excluded for the screener. Subsequent to fitting the measurement error model, diagnostic statistics were used to identify influential observations. Three observations (2 women in Emory and 1 woman in ROC) were identified that had a significant impact on the estimate of the correlation coefficient. Thus, deattenuated correlations are presented excluding these 3 subjects.

Results

Validity measures by site and gender. Estimates of percentage energy from fat ranged from 5 to 55% from 24HR (for each individual, sum of energy from fat divided by sum of energy) and 18 to 47% from the PFat. Agreement in estimates of population mean intakes between the multiple 24HR and screener is shown in Table 2. The PFat screener and 24HR estimates of mean intakes were not statistically significantly different among men at URI and men and women at ROC. Estimates from the 2 instruments were significantly (P < 0.05) or borderline significant (0.05 < P < 0.10) different for the remaining 4 gender/site groups. Differences in the estimates between the 2 instruments ranged from 0.3 percentage points (men at URI) to 3.8 percentage points (men at Emory) and 5.2 percentage points (women at Emory).

Estimated deattenuated correlation coefficients between the screener and true intake are shown also in Table 2. Among men, deattenuated correlation coefficients were 0.52 (URI) and 0.77 (ROC), and coefficients of determination (R²) were 27% and 59%, respectively. Among women, deattenuated correlations ranged from 0.36 (Emory) to 0.59 (URI), and coefficients of determination ranged accordingly from 13 to 35%.

Estimated prevalence of being at or below 30% energy from fat and <35% energy from fat for true intake and screener are presented in Table 3. Prevalence estimates at or below 30% energy from fat were identical for true intake and screener among URI men but were quite different among Emory women and ROC men. The positive predictive values were at least 10% larger than the true prevalence in the population for all sites and at least 24% larger for all sites except Emory. The negative predictive values were 0.61 or higher for all sites. At 35% energy from fat, prevalence estimates between true intake and screener were similar among URI men and women and ROC men but differed greatly for women from HSPH, Emory, and ROC. Also presented in Table 3 are estimates of the percentage energy from fat values where sensitivity and specificity are approximately equal by site and gender. These estimates of optimal screening level vary across sites and genders.

Impact of potential moderators on validity. As expected, the treatment condition at baseline (experimental vs. standard/control) was unrelated to the level of agreement in mean estimated intakes between the screener and the 24HR using either the difference score or the ratio score. Within site and gender groups, few of the potential moderators we examined appeared to affect validity. Similarly, no relations were found between validity and age, educational status, and smoking status (data not shown). However, age and smoking status, like race/ethnicity and gender, were related to the level of agreement in mean intake estimates, with smokers having significantly higher correlations (and consequently larger correlation coefficients) than non-smokers. As expected, the deattenuated correlation coefficients were 0.52 (URI) and 0.77 (ROC), and coefficients of determination (R²) were 27% and 59%, respectively. Among women, deattenuated correlations ranged from 0.36 (Emory) to 0.59 (URI), and coefficients of determination ranged accordingly from 13 to 35%.

TABLE 2 | Estimated mean percentage energy from fat (95% CI) from 24HR and screener and deattenuated Pearson correlation coefficient (and SE) between true intake1 and screener, by site and gender: BCC, 2001–2004

<table>
<thead>
<tr>
<th>Site/gender</th>
<th>n</th>
<th>Mean (95%CI)</th>
<th>Difference²</th>
<th>Deattenuated Pearson correlation coefficient (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>URI</td>
<td></td>
<td>24HR</td>
<td>Screener</td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>46</td>
<td>29.8 (27.7–31.9)</td>
<td>29.5 (28.7–30.3)</td>
<td>0.3 (−1.6–2.2)</td>
</tr>
<tr>
<td>Women</td>
<td>124</td>
<td>29.4 (28.1–30.7)</td>
<td>27.3 (26.7–28.0)</td>
<td>2.1 (0.9–3.2)</td>
</tr>
<tr>
<td>HSPH</td>
<td></td>
<td>30.4 (28.6–30.7)</td>
<td>31.9 (30.9–32.9)</td>
<td>−1.5 (−3.2–0.3)</td>
</tr>
<tr>
<td>Emory</td>
<td></td>
<td>35.0 (31.8–38.3)</td>
<td>31.2 (30.0–32.4)</td>
<td>3.8³ (0.4–7.2)</td>
</tr>
<tr>
<td>Men</td>
<td>18</td>
<td>36.1 (34.9–37.3)</td>
<td>30.9 (30.3–31.6)</td>
<td>5.2² (3.9–6.5)</td>
</tr>
<tr>
<td>Women</td>
<td>147</td>
<td>34.4 (32.2–36.7)</td>
<td>32.9 (31.5–34.4)</td>
<td>1.5 (−0.4–3.4)</td>
</tr>
<tr>
<td>ROC</td>
<td></td>
<td>31.4 (29.4–33.4)</td>
<td>33.1 (31.7–34.4)</td>
<td>−1.7 (−3.7–0.4)</td>
</tr>
</tbody>
</table>

1 True intake is estimated from multiple nonconsecutive 24HR in a measurement error model. Pearson correlation coefficients are deattenuated.
2 24HR-derived value minus Screener-derived value.
3 Marginally significantly different from estimated true intake, P < 0.05.
4 Significantly different from estimated true intake, P < 0.05.
—, Sample size too small for stable estimate.

Note: Two Emory women and 1 ROC woman were excluded in the estimation of the deattenuated Pearson correlation coefficient because their data exerted undue influence on the model results.
ethnicity, vary much more between than within sites, so that the power to detect any moderating effect of these variables is limited.

Table 4 shows the mean difference scores for the BMI subgroups. (Results for the ratio score are qualitatively similar and are not presented.) Agreement between 24HR and PFat was significantly \( (P < 0.05) \) higher in normal-weight than in overweight and obese categories in URI women, and although not statistically significant, followed the same pattern in Emory women and ROC men.

Discussion

The PFat screener evaluated in this study is 1 of 2 short instruments available to estimate percentage energy from fat. The other short instrument, the “Block” screener, is composed of 2 subscales: a 15-item meat/snack section and a 7-item fruit/vegetable section (25). Little evaluative work has been published on either instrument. In 1 study, the (Spearman) correlation between the Block screener and the 1995 Block 100-item FFQ estimates of percentage energy from fat was 0.63 (25). This correlation is most likely an overestimate of the correlation between screener and true percentage energy from fat because both instruments are essentially FFQ, and it would be expected that similar instruments would have errors that are highly correlated (34,37). In a second study comparing the Block screener to 5 nonconsecutive 24HR in 88 medical students, mean estimated percentage energy from fat was significantly lower \( (P < 0.01) \) for the screener than for the recalls (28.4 vs. 33.8%); the deattenuated Pearson correlation between the 2 instruments (0.36) was statistically significant (38).

The NCI PFat instrument used in this BCC Nutrition Working Group validation study has been evaluated in 1 other study. In that study, the PFat was compared with 2 nonconsecutive 24HR in a subsample of 401 men and women participating in the NIH-AARP Diet and Health Study, a prospective cohort study (26). In the study, mean estimated percentage energy from fat intake was significantly lower for the screener than for 24HR in women (28.4 vs. 31.3) but not in men (26). Deattenuated Pearson correlations between the 2 instruments were 0.64 for men and 0.58 for women. The study sample was comprised of older adults (ages 50–71 y), members of AARP (formerly the American Association of Retired Persons); 91% were white, and 72% had higher than high school education. To our knowledge, this BCC study constitutes the first evaluation of the screener in an intervention study and among a diverse group of study samples.

In this study, performance of the PFat varied across sites and genders. There was only limited agreement between screener and 24HR in mean estimated intakes—means for the 2 instruments were statistically significantly different \( (P < 0.05) \) or nearly so to 5 nonconsecutive 24HR in 88 medical students, mean estimated percentage energy from fat was significantly lower (28.4 vs. 33.8%); the deattenuated Pearson correlation between the 2 instruments (0.36) was statistically significant (38).

Table 4

<table>
<thead>
<tr>
<th>Gender and BMI status(^1)</th>
<th>n</th>
<th>Mean</th>
<th>P-value</th>
<th>Gender and BMI status(^1)</th>
<th>n</th>
<th>Mean</th>
<th>P-value</th>
<th>Gender and BMI status(^1)</th>
<th>n</th>
<th>Mean</th>
<th>P-value</th>
<th>Gender and BMI status(^1)</th>
<th>n</th>
<th>Mean</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td></td>
<td></td>
<td></td>
<td>Women</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>14</td>
<td>-0.88</td>
<td>0.41</td>
<td>Normal</td>
<td>37</td>
<td>0.11</td>
<td>0.02</td>
<td>Normal</td>
<td>37</td>
<td>3.97</td>
<td>0.52</td>
<td>Normal</td>
<td>40</td>
<td>1.51</td>
<td>0.64</td>
</tr>
<tr>
<td>Overweight</td>
<td>19</td>
<td>1.87</td>
<td></td>
<td>Overweight</td>
<td>33</td>
<td>-2.22</td>
<td>0.92</td>
<td>Overweight</td>
<td>33</td>
<td>5.27</td>
<td></td>
<td>Overweight</td>
<td>46</td>
<td>-1.46</td>
<td></td>
</tr>
<tr>
<td>Obese</td>
<td>13</td>
<td>-0.77</td>
<td></td>
<td>Obese</td>
<td>22</td>
<td>3.97</td>
<td>0.52</td>
<td>Obese</td>
<td>22</td>
<td>6.04</td>
<td></td>
<td>Obese</td>
<td>37</td>
<td>-4.17</td>
<td>0.41</td>
</tr>
</tbody>
</table>

1 Normal, BMI = 18.5–24.9 kg/m\(^2\); Overweight, BMI = 25–29.9 kg/m\(^2\); Obese, BMI = ≥30 kg/m\(^2\).

2 — No sample, or every cell n < 10.
(0.05 < P < 0.10) from true intake for 4 of 7 site-gender subgroups. Except for men and women at Emory, the magnitude of these statistically significant differences was 2.1 percentage points or less. In the context of an intervention study, the level of error in a screener at baseline is less important than the error in estimating change over time among the same individuals, the area addressed by Williams et al. (27).

Short instruments may be used in dietary intervention studies as screening devices for selection into, or exclusion from, a study. The aim of such screening is to recruit a sample of subjects having, say, true percentage energy from fat >35%; because of misclassification, however, not all subjects selected into the study will be truly eligible. The BCC study indicates that the PFat screener would be only moderately useful in increasing the proportion of truly eligible subjects in a study, at least if one screens at the same level as the desired true level, as we did in this analysis. One could improve the proportion of subjects with true percentage energy from fat >35% by screening at a higher level, say 37%. This would lead to a lower negative predictive value, however, and as a result would require one to screen a larger pool of potential subjects to achieve recruitment goals.

In this study, site- and gender-specific deattenuated correlations between the PFat and 24HR ranged widely, from 0.36 to 0.77. All correlations were positive and statistically significantly different from 0, and the proportion of variance in true percentage energy from fat explained by the screener ranged from 13 to 59%. In the context of an intervention study, it is not enough to show that the screener is correlated with true intake at baseline. One must also establish similar levels of correlation at follow-up and at postintervention, so that any changes in reported intake would be due to actual dietary changes, and not to changes in the validity of the instrument that might occur over time or with the intervention, or both.

If factors that moderate validity could be identified, these could be used to explain differences seen between sites, and could possibly be included with the screener in analytical models to improve overall predictive value. In the BCC, in general, the PFat did not perform as well in women as in men. Some studies have shown higher underreporting of energy and fat among overweight and obese than among normal-weight individuals (39,40). Although there was some indication of such an effect among URI women and possibly ROC men and Emory women, the effect was not evident for all groups examined. Differences in performance were not associated with differences in age, educational status, or smoking status. The lack of moderating effects on validity for a wide range of variables is somewhat surprising. Several factors may have contributed to this. First, the effective sample size for examining these types of moderating effects was quite limited, as these variables did not vary generally within each site, and thus statistical power was low. Second, because a screener instrument is so limited, it may not be sensitive enough to detect small differences in performance among various subgroups. Furthermore, in our analyses of potential effect moderators, we were not able to disentangle design variables such as mode of screener administration that were specific to each site from effects of site itself.

Overall, the PFat did not perform as well in the baseline BCC data presented in the current study as in the NIH-AARP validation study. An important distinction is that this BCC study occurred in the context of an intervention setting, whereas the previous NIH-AARP validation study occurred within a prospective observational study. Participation in an intervention study requires a higher level of commitment than that required in an observational study, which could result in a sample that differs from the overall population in ways that affect measurement of the validity of the instrument in that sample (sometimes called selection bias). For example, intervention participants may be particularly inclined to “talk a good diet”; it is thought that this type of response bias may be more problematic in food frequency-type instruments than in 24HR. However, comparison of the NIH-AARP validation study results to those from URI, the sample most similar demographically to the NIH-AARP sample, reveals very similar agreement in means and correlations. Thus, although the potential for selection bias still exists, it may not be the most important reason for the poorer overall performance of the PFat in the BCC. A more important reason may stem from the diversity of study sites.

The PFat performed somewhat more poorly among Emory men and women than among those in other sites. Although the results for Emory men might result from the small sample size (n = 18), the results for Emory women were based on a relatively large sample (n = 147), thus minimizing the possibility of unstable estimates. The Emory screener was self-administered, which may have led to poorer quality data compared with interviewer-administered instruments. Alternatively, it may be that the relationships between the foods assessed on the screener and percentage energy from fat differ for various subgroups of the population and in particular are different for African Americans in the Emory site. Thus, it may be necessary to develop separate scoring algorithms for specific subpopulations defined by, for example, race/ethnicity or BMI status. One might even need to include particular foods important to estimating percentage energy from fat in specific subpopulations. More research is needed in large representative population samples to model population-specific algorithms. Testing of such population-specific algorithms should be done in both cross-sectional and intervention studies.

In summary, the ability of the PFat to estimate usual intake of percentage energy from fat at baseline in the BCC intervention studies varied across genders and sites. Generally, performance was better among men than women and was somewhat poorer in the site composed of Southern African-Americans. Targeted food lists and/or scoring algorithms may be necessary for certain subgroups of the population. If accurate assessment of diet at baseline (and presumably at follow-up) is essential, a more detailed instrument such as the 24HR, administered multiple times for each participant, is warranted. A self-administered automated 24HR is being developed currently (41) for public use at minimal charge, and it promises to offer researchers a more precise yet affordable option in the near future. If less precise assessment is acceptable, the PFat may be appropriate but should be pretested rigorously in the target population.

Literature Cited


