



Practice of Epidemiology

Validity and Systematic Error in Measuring Carotenoid Consumption with Dietary Self-report Instruments

Loki Natarajan¹, Shirley W. Flatt², Xiaoying Sun¹, Anthony C. Gamst¹, Jacqueline M. Major², Cheryl L. Rock², Wael Al-Delaimy², Cynthia A. Thomson³, Vicky A. Newman², and John P. Pierce² for the Women's Healthy Eating and Living Study Group

¹ Division of Biostatistics, University of California, La Jolla, CA.

² Cancer Center, University of California, La Jolla, CA.

³ University of Arizona, Tucson, AZ.

Received for publication April 3, 2005; accepted for publication August 19, 2005.

Vegetables and fruits are rich in carotenoids, a group of compounds thought to protect against cancer. Studies of diet-disease associations need valid and reliable instruments for measuring dietary intake. The authors present a measurement error model to estimate the validity (defined as correlation between self-reported intake and "true" intake), systematic error, and reliability of two self-report dietary assessment methods. Carotenoid exposure is measured by repeated 24-hour recalls, a food frequency questionnaire (FFQ), and a plasma marker. The model is applied to 1,013 participants assigned between 1995 and 2000 to the nonintervention arm of the Women's Healthy Eating and Living Study, a randomized trial assessing the impact of a low-fat, high-vegetable/fruit/fiber diet on preventing new breast cancer events. Diagnostics including graphs are used to assess the goodness of fit. The validity of the instruments was 0.44 for the 24-hour recalls and 0.39 for the FFQ. Systematic error accounted for over 22% and 50% of measurement error variance for the 24-hour recalls and FFQ, respectively. The use of either self-report method alone in diet-disease studies could lead to substantial bias and error. Multiple methods of dietary assessment may provide more accurate estimates of true dietary intake.

bias (epidemiology); carotenoids; diet; models, statistical; nutrition assessment; reproducibility of results

Abbreviations: FFQ, food frequency questionnaire; WHEL, Women's Healthy Eating and Living.

Dietary patterns emphasizing high intake of fruits and vegetables are recommended for prevention of cancer and heart disease (1). Assessments of dietary intake have generally relied on self-reported consumption from either a census of daily intake (diary or recall interview) or a broad estimate of frequency and quantity of intake of select foods (food frequency questionnaire (FFQ)) from which participants can be ranked on food components of interest. Measurement error is a common concern with self-report instruments (2). FFQs are subject to recall bias because they require individuals to report intake over the previous weeks or months, whereas census approaches

need to sample numerous days to obtain precise estimates of intake.

Errors in dietary intake measures lead to biased estimates of regression coefficients. In the large European Prospective Investigation into Cancer and Nutrition (EPIC) cohort, food diaries indicated that fat intake was associated with breast cancer, whereas the FFQ did not (3). In the New York University Women's Health Study, serum carotenoid measures of vegetable/fruit intake were associated with breast cancer incidence (4), whereas the FFQ measure was not (5). By use of urinary biomarkers as reference instruments (6, 7), studies of potassium, sodium, and absolute protein intake have

Correspondence to Dr. Loki Natarajan, Cancer Center, University of California, San Diego, 3855 Health Sciences Drive, La Jolla, CA 92093-0901 (e-mail: loki@math.ucsd.edu).

demonstrated that the FFQ necessitates substantial correction factors to adjust for measurement error. These studies (6, 7) also report nonnegligible error correlations between FFQs and other self-report instruments (24-hour recalls or 7-day diary) used to calibrate FFQs. Some researchers (8, 9) have argued that these concerns do not invalidate studies whose primary measure is a FFQ, as relative risk estimates based on FFQs can be corrected with calibration methods. However, large validation studies (i.e., those with $\geq 1,000$ participants) are needed to obtain accurate corrections for regression estimates when error correlations are high (10). Furthermore, assessment methods that exhibit large biases may not provide reliable estimates of diet-disease associations, even after adjustment of regression coefficients.

In this investigation, we focus on carotenoids, a group of compounds believed to protect against cancer. A measurement error model is developed and applied to data from the Women's Healthy Eating and Living (WHEL) Study, a randomized trial of 3,088 breast cancer survivors, assessing the impact of a low-fat, high-vegetable/fruit/fiber diet on breast cancer recurrence (11). Carotenoid intake in the WHEL Study was measured by two separate self-report methods: a FFQ and 24-hour recalls. Additionally, plasma carotenoid concentrations, a biomarker of fruit and vegetable intake (12), were also collected.

Blood carotenoid concentrations have been demonstrated to be specifically responsive to self-reported intakes of vegetables and fruit. A major advantage is that they are markers of intake of the presumed active ingredients in the dietary pattern. The degree of responsiveness, however, is known to be influenced by lipid concentrations, body size, and smoking status (12–15). Further, the characteristics of the food can influence the availability of carotenoids for absorption. Nonetheless, even though they are imperfect compared with protein markers, repeated measures of circulating carotenoids, together with self-report measures, allow us to investigate both the validity and the systematic bias of these instruments.

MATERIALS AND METHODS

Study population

The data used in the current investigation were obtained from 1,551 women assigned randomly to the comparison (nonintervention) group of the WHEL Study. Details regarding the WHEL Study protocol are reported elsewhere (11). Briefly, 3,088 women were assigned at random to one of two diet groups (1,537 to the intervention group; 1,551 to the comparison group), stratified by stage of disease, age, and clinic site. The comparison group was advised to consume a diet consistent with current National Cancer Institute dietary recommendations for cancer prevention, while the intervention group was counseled to eat fruits and vegetables rich in phytochemicals and micronutrients. Fasting blood samples were collected during clinic visits at baseline and at specified times thereafter. The institutional review boards of all participating institutions approved the procedures for this study, and written, informed consent was obtained from all study participants.

Pierce et al. (16) reported that WHEL Study intervention and comparison groups had similar baseline dietary pat-

terns. The comparison group had not changed significantly at 1 year, whereas the intervention group showed a marked increase in the intakes of fruits and vegetables. The current investigation focused on quantifying error associated with estimating habitual diet by self-report methodologies. Hence, this analysis is restricted to women in the WHEL Study comparison group, who had (on average) stable diets (16), thus avoiding the influence of dietary intervention on measurement error properties. Of the 1,551 women assigned randomly to the comparison group, 1,013 of these had complete dietary carotenoid data at baseline and at 1 year on all three measures of interest, namely, FFQ, 24-hour recalls, and plasma concentrations. The measurement error analysis was conducted on this sample of 1,013 women.

Fifty-seven percent of the participants at baseline and 59 percent at 1 year reported consuming supplements containing β -carotene on 24-hour recalls; the corresponding lutein usage was 6 percent at study entry and 12 percent at 1 year. On the FFQ, 62 percent reported consuming supplements containing β -carotene at both times, while 7 percent at study entry and 14 percent at 1 year were taking supplements containing lutein. Use of supplements containing other carotenoids was not reported. Total carotenoid intake was calculated by summing dietary and supplement intakes for each self-report instrument.

Plasma carotenoid and cholesterol measurement

Plasma carotenoids, including α -carotene, β -carotene, lutein, lycopene, and β -cryptoxanthin, were separated and quantified with high-performance liquid chromatography (17). Zeaxanthin and lutein elute together with this method, so the values presented as lutein are assumed to be lutein plus zeaxanthin. The total plasma carotenoid concentration is calculated as the sum of α -carotene, β -carotene, lutein, lycopene, and β -cryptoxanthin concentrations. The laboratory participates in the National Institute of Standards and Technology Branch of Quality Assurance Round Robin studies.

Total plasma cholesterol concentrations were determined with the Kodak Ektachem Analyzer system (Johnson & Johnson Clinical Diagnostics, Rochester, New York) and used to interpret plasma carotenoid data. Standard reference materials were used to validate the analytical precision of these procedures. The laboratory participates in the American College of Pathologists quality assurance program to monitor the precision and reliability of lipid measures.

Dietary information

Carotenoid intake (mg/day), defined as dietary plus supplement values, and energy (kcal/day) were assessed by use of 24-hour dietary recalls and the Arizona FFQ (11, 18, 19). Each participant provided four 24-hour dietary recalls including 2 weekdays and 2 weekend days over a 3-week period. For each point in time (baseline and 1 year), the four recall measures were averaged, and this average was used in the analysis. Trained dietary assessors, blinded to the randomization assignment of participants, collected these data during telephone interviews. Nutrient calculations were performed by use of Nutrition Data System for Research software, developed by the Nutrition Coordinating Center, University of Minnesota, Minneapolis, Minnesota (Food and Nutrient Database

31, version 4.03, released in November 2000). Participants were queried regarding actual dietary supplement use on recall days during their telephone interview. Supplement content was quantified by methods described previously (20).

The 153-item, semiquantitative, scannable Arizona FFQ (18) was used to collect information regarding the usual foods consumed and the frequency of consumption over the previous 3 months, using age- and gender-specific estimates of portions recorded as small, medium, or large. The database used to quantify nutrient intake from this FFQ was derived from the US Department of Agriculture Food Composition Database and the Nutrient Database for Standard Reference (versions 11–13), as well as from the Continuing Survey of Food Intake by Individuals 1994–1996, 1998. Carotenoid data in the FFQ-linked database were updated from the 1998 US Department of Agriculture-Nutrition Coordinating Center Carotenoid Database for US Foods. Supplement data were based on an inventory of products reported as consumed over the previous month, during interviews of the participants at baseline and follow-up clinic visits.

Measurement error model

We introduce some notation. Let Z_i denote “true” (unobservable) average carotenoid intake for the i th participant. For n participants, the Z_i , $i = 1, 2, \dots, n$ are assumed to be independently distributed random variables with first and second moments, $EZ_i = \mu_Z$ and $\text{var}(Z_i) = \sigma_Z^2$. The observed self-reported intake or biomarker value for the i th participant is the vector

$$\vec{Y}_i^T = (Y_{i11}, Y_{i21}, Y_{i12}, Y_{i22}, Y_{i13}, Y_{i23}),$$

where Y_{ijk} refers to the observed carotenoid exposure of participant i at time j measured by method k , with $j = 1, 2$ denoting baseline and year 1 values, and $k = 1, 2, 3$ representing 24-hour recalls, FFQ, and plasma concentrations, respectively. The model is

$$Y_{ijk} = \alpha_k + \beta_k Z_i + \epsilon_{ijk}, \quad (1)$$

where α_k and β_k are the intercept and scale parameters to be estimated. We assume that the “true” intake Z_i is independent of “measurement error” ϵ_{ijk} . The error is assumed to have a mean of 0 ($E\epsilon_{ijk} = 0$) and covariance matrix

$$\text{cov}(\vec{\epsilon}_i) = \text{cov} \begin{pmatrix} \epsilon_{i11} \\ \epsilon_{i21} \\ \epsilon_{i12} \\ \epsilon_{i22} \\ \epsilon_{i13} \\ \epsilon_{i23} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho_{11}\sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{12}\sigma_1\sigma_2 & 0 & 0 \\ \cdot & \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{12}\sigma_1\sigma_2 & 0 & 0 \\ \cdot & \cdot & \sigma_2^2 & \sigma_2^2\rho_{22} & 0 & 0 \\ \cdot & \cdot & \cdot & \sigma_2^2 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \sigma_3^2 & \sigma_3^2\rho_{33} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \sigma_3^2 \end{pmatrix}.$$

Thus, σ_k^2 denotes the variance of ϵ_{ijk} , while correlation between ϵ_{ijk} and ϵ_{isl} is ρ_{kl} . In order to make our model identifiable, we will assume that ρ_{33} is 0. This assumption of uncorrelated errors in plasma is unlikely to hold in practice. Hence, to assess the sensitivity of our results to departures from this assumption, we repeated the analysis for values of ρ_{33} set equal to 0.2 and 0.4.

The covariance equation is

$$\text{cov}(\vec{Y}_i) = \sigma_Z^2 \vec{\beta} \vec{\beta}^T + \text{cov}(\vec{\epsilon}_i) \quad (2)$$

with

$$\vec{\beta}^T = (\beta_1, \beta_1, \beta_2, \beta_2, \beta_3, \beta_3).$$

Our model belongs to the class of models considered by Plummer and Clayton (21). It is similar to models considered by others (6–8, 13, 22), except that we included slope terms in all equations relating observed exposure to “true” intake. Also, although we do not include separate subject-specific bias terms in the model (i.e., subject-specific random effect), these are estimable in terms of correlated error in repeated administrations of the same instrument (Appendix).

Estimation of parameters

A *method of moments* approach was used to estimate the parameters in equation 2. With our model, we can estimate $\beta_k^2 \sigma_Z^2 / (\beta_k^2 \sigma_Z^2 + \sigma_k^2)$, the proportion of variance in each assessment method attributable to true intake variance. The square root of this proportion is the *attenuation factor* for the standardized regression coefficient (21), which is the multiplicative factor by which the true diet-disease relation is “attenuated” because of measurement error. This square root is also the correlation between true and measured intake, that is, the validity of the instrument. We also calculate the within-instrument (across time) and across-instrument error correlations (ρ_{kl}).

Although it might be useful to estimate the variance of true intake and the regression parameters, α_k and β_k , problems with identifiability make this an untenable objective, unless restrictions are imposed on the parameters. However, we can estimate the product $\beta_k \sigma_Z$ and, hence, the ratios of the β_k s by our methods.

Standard errors and 95 percent confidence intervals for the estimated parameters were computed by bootstrap resampling (23). All calculations utilize the software packages, S-PLUS, version 5 (Insightful Corporation, Seattle, Washington) and Statistical Analysis System, version 8.01 (SAS Institute, Inc., Cary, North Carolina).

Diagnostics

Model fit was assessed to detect deviations from modeling assumptions. In linear regression, fit is investigated by plotting residuals against independent variables. In the measurement error setting, the independent variable is true

TABLE 1. Summary statistics expressed as the mean and standard deviation for 1,013 participants randomly assigned between 1995 and 2000 to the nonintervention arm of the Women's Healthy Eating and Living Study

Variable	Baseline	1 year
Total carotenoids (dietary + supplement)		
24-hour recalls (mg/day)	20.21 (34.67)	18.78 (27.35)
Food frequency questionnaire (mg/day)	24.88 (27.97)	25.31 (32.50)
Plasma carotenoids (mmol/liter)	2.40 (1.56)	2.34 (1.40)
Total energy intake		
24-hour recalls (kcal/day)	1,742.09 (410.86)	1,618.54 (394.24)
Food frequency questionnaire (kcal/day)	1,924.65 (789.32)	1,820.50 (796.43)
Plasma cholesterol (mg/dl)	196.02 (38.69)	194.88 (38.14)
Body mass index (kg/m ²)	27.16 (6.10)	27.46 (6.20)

intake, which is “unobservable.” Nevertheless, the best linear unbiased estimator of true intake for each individual can be estimated, and this quantity replaces the independent variable in residual plots. True intake is estimated by conditioning on observed data and assuming joint normality of observed data and true intake, that is,

$$E(Z_i|\vec{Y}_i) = \mu_Z + \Sigma_{Z\vec{Y}}\Sigma_{\vec{Y}\vec{Y}}^{-1}(\vec{Y}_i - \mu_Y), \quad (3)$$

where Z_i and \vec{Y}_i were defined earlier, μ_Z is the population mean of true intake, and μ_Y is the vector of population means of observed data; $\Sigma_{Z\vec{Y}}$ is the covariance of true intake with observed data, which is computed from estimated model parameters; and $\Sigma_{\vec{Y}\vec{Y}}$ is the covariance matrix for observed data, which is estimated using sample covariances.

Our model assumed that true intake Z_i was independent of measurement error ϵ_{ijk} . We checked this assumption by plotting “residuals” $v_{ijk} = Y_{ij3} - (\beta_3/\beta_k)Y_{ijk} = \epsilon_{ij3} - (\beta_3/\beta_k)\epsilon_{ijk}$ against estimated Z_i from equation 3. The 24-hour recalls and FFQ residuals were obtained by setting $k = 1$ or 2, respectively. If the assumption that true intake is independent of measurement error holds, then the v_{ijk} and Z_i should be uncorrelated and, hence, the plot should show no particular pattern.

Other statistical issues

All dietary measures were log transformed to stabilize variance and improve normality. The self-reported carotenoid values were adjusted for body mass index and energy intake (kcal/day) because intake is associated with these variables, and because energy adjustment is known to reduce the impact of measurement error (7, 10). These adjustments were achieved by performing a linear regression of carotenoid intake on body mass index and energy measured by each self-report instrument and by using residuals of these models in the subsequent measurement error analysis. Similarly, plasma carotenoid concentrations were regressed on plasma cholesterol and body mass index (12), and the analysis was conducted using residuals from this regression. Although

smoking status influences plasma concentrations (12), fewer than 5 percent of WHEL Study women were current smokers and, hence, no adjustment was made for smoking.

RESULTS

Sample means (table 1) show that, on average, the total carotenoids estimated by the three methods (self-report or biomarker) and body mass index, total energy intake, and plasma cholesterol concentrations were stable from study entry to 1 year, for our sample. The lack of significant change in dietary pattern of the WHEL Study comparison group has been noted previously (16). Interestingly, the FFQ posted significantly higher mean carotenoid values than did 24-hour recalls at baseline and 1 year ($p < 0.001$ by paired t tests).

Pearson's correlations of (log-transformed) total carotenoids (table 2) between the self-report instruments ranged from 0.36 to 0.48. The correlation between plasma concentrations and carotenoid intake from 24-hour recalls ranged from 0.36 to 0.45, and that between the plasma marker and FFQ ranged from 0.28 to 0.35. Additionally, correlations over time within each method were 0.45, 0.64, and 0.80, respectively, for 24-hour recalls, FFQ, and plasma marker. Spearman's correlations between carotenoid measures were similar (data not shown) and comparable to those of a previous publication (19).

The model-based estimate of correlation between either self-report instrument and true carotenoid intake (i.e., the validity) was moderate (0.44 for 24-hour recalls; 0.39 for FFQ) (table 3) when ρ_{33} was assumed to equal 0. The correlation between true intake and plasma concentrations was the highest at 0.86. Sensitivity analyses conducted for $\rho_{33} = 0.2$ or 0.4 were not qualitatively different, whereby the validities of 24-hour recalls and FFQ increased to 0.51 and 0.44, respectively, while the validity of plasma decreased to 0.75 (table 3).

The validity of the self-report methods and plasma was also estimated separately for each of the five component carotenoids. For 24-hour recalls, the correlations with true

TABLE 2. Pearson's correlations of total carotenoid measures for 1,013 participants randomly assigned between 1995 and 2000 to the nonintervention arm of the Women's Healthy Eating and Living Study*

	24-hour recall		Food frequency questionnaire		Plasma	
	Baseline	1 year	Baseline	1 year	Baseline	1 year
24-hour recall						
Baseline	1.00	0.45	0.48	0.38	0.45	0.37
1 year	0.45	1.00	0.36	0.41	0.36	0.40
Food frequency questionnaire						
Baseline	0.48	0.36	1.00	0.64	0.35	0.28
1 year	0.38	0.41	0.64	1.00	0.29	0.31
Plasma						
Baseline	0.45	0.36	0.35	0.29	1.00	0.80
1 year	0.37	0.40	0.28	0.31	0.80	1.00

* All correlations are based on log-transformed carotenoid values.

intake for α -carotene, β -carotene, β -cryptoxanthin, lutein, and lycopene were, respectively, 0.45, 0.52, 0.33, 0.29, and 0.27. The corresponding FFQ correlations were 0.45, 0.47, 0.36, 0.21, and 0.24, and for plasma they were 0.85, 0.87, 0.83, 0.86, and 0.76. Thus, although the self-report instruments appear to have the highest validity for α - and β -carotene and the lowest validity for lutein and lycopene, the results for the component carotenoids were not qualitatively different from the results for total carotenoids.

Identifiability constraints preclude the estimation of the scaling parameters β_k . However, the product $\beta_k\sigma_Z$ is estimable in our model and was 0.28, 0.20, and 0.41 for 24-hour recalls, FFQ, and plasma, respectively. The ratio of the β_k s for 24-hour recalls to plasma was 0.68, the ratio for FFQ to plasma was 0.49, and the ratio for 24-hour recalls to FFQ

was 1.38. The differences in these ratios provide additional evidence of the lack of concordance in the self-report instruments when measuring carotenoid intake.

Error correlations from baseline to 1 year (table 3) were higher for the FFQ (i.e., 0.51–0.53) compared with 24-hour recalls (i.e., 0.22–0.28). These error correlations are related to systematic error (7, 24). In fact, the parameter ρ_{kk} for across-time error correlations in our model can be interpreted as the ratio of systematic error variance to total measurement error variance (Appendix). Thus, systematic error in the FFQ accounted for more than 50 percent of the total error variance, whereas for the 24-hour recalls the corresponding value was 22–28 percent (table 3).

The error correlations between 24-hour recalls and FFQ were 0.31 when ρ_{33} was assumed to be 0. A sensitivity

TABLE 3. Estimated measurement error parameters for 1,013 participants randomly assigned between 1995 and 2000 to the nonintervention arm of the Women's Healthy Eating and Living Study

Parameter and method	Plasma error correlation					
	$\rho_{33} = 0$		$\rho_{33} = 0.2$		$\rho_{33} = 0.4$	
	Estimate	95% confidence interval*	Estimate	95% confidence interval*	Estimate	95% confidence interval*
Correlation between measured carotenoid exposure and true carotenoid intake						
24-hour recall	0.44	0.39, 0.50	0.46	0.41, 0.52	0.51	0.45, 0.57
Food frequency questionnaire	0.39	0.33, 0.44	0.40	0.35, 0.46	0.44	0.38, 0.51
Plasma	0.86	0.83, 0.89	0.82	0.78, 0.85	0.75	0.69, 0.80
Error correlations (from baseline to 1 year)						
24-hour recall	0.28	0.20, 0.35	0.26	0.18, 0.34	0.22	0.12, 0.31
Food frequency questionnaire	0.53	0.47, 0.59	0.52	0.46, 0.58	0.51	0.43, 0.57
Error correlations						
Food frequency questionnaire and 24-hour recall	0.31	0.25, 0.37	0.30	0.24, 0.36	0.27	0.20, 0.34

* Bootstrap percentile interval (1,000 replicate samples).

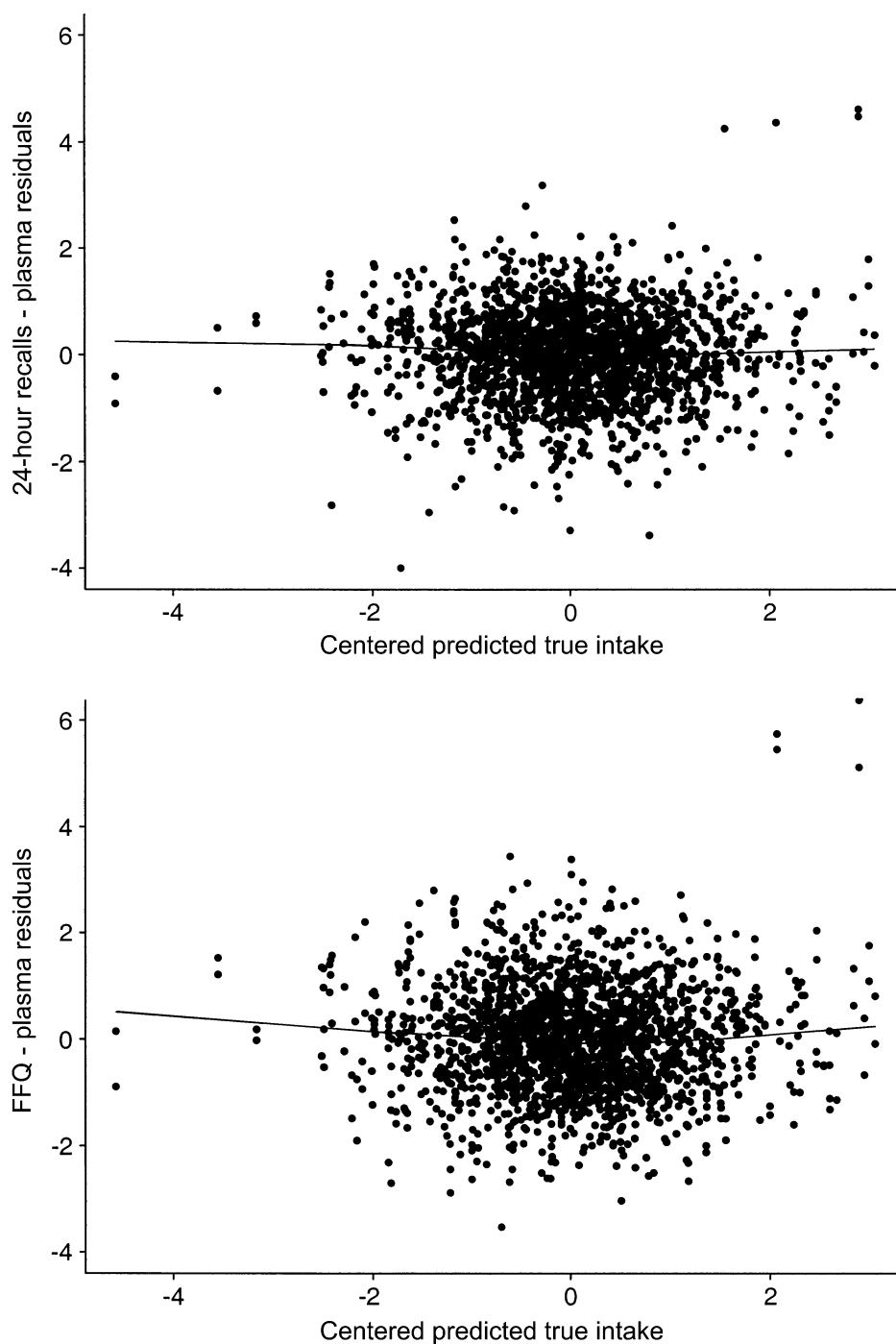


FIGURE 1. Model diagnostics for 1,013 participants randomly assigned between 1995 and 2000 to the nonintervention arm of the Women's Healthy Eating and Living Study. Top, 24-hour recalls; bottom, food frequency questionnaire (FFQ).

analysis was conducted to allow for possibly different error correlations when the two instruments were administered contemporaneously versus a year apart. As expected, the error correlation between the FFQ and 24-hour recalls when both assessments were conducted contemporaneously was higher (i.e., 0.37) compared with when one of the instruments

was collected at baseline and the other at 1 year (i.e., 0.31). These error correlations did not change substantially when the analysis was repeated for values of $\rho_{33} = 0.2$ or 0.4.

We assessed the fit of the model (Diagnostics section above). The sample correlations between estimated true intake (equation 3) and the residuals for each of the

carotenoids were negligible ($<1.85 \times 10^{-16}$). We plotted the residuals for 24-hour recalls and FFQ against centered true intake scaled to have unit variance (figure 1). A smoothed LOESS fit is also included in each graph. The fit of the model appears adequate, and the assumption that true intake, Z , is uncorrelated with error, ϵ , appears to hold.

DISCUSSION

In this study, we estimated the ability of a FFQ, repeat 24-hour recalls, and a plasma marker to measure carotenoid exposure. Plasma carotenoids represent the internal dose of exposure and are therefore closer to the true exposure of interest. The plasma marker had high validity with correlation to true intake ranging between 0.75 and 0.86. The self-report methods exhibited modest correlations to true carotenoid intake, ranging from 0.39 to 0.44 for FFQs and from 0.44 to 0.51 for 24-hour recalls, suggesting that the use of either method alone in diet-disease studies would require large corrections for regression dilution. It is possible that increasing the number of items on the FFQ would improve its validity. Additionally, the FFQ is usually designed to measure diet in general population studies and, hence, may not perform as well when the focus is on specific food compounds (e.g., carotenoids). Using multiple methods to assess dietary intake could alleviate these problems.

To our knowledge, this is the largest study investigating the validity and biases of FFQs and 24-hour recalls in measuring carotenoid intake over time. Other studies (25–27) have relied on smaller samples and have not estimated the systematic error in these instruments. These studies (26, 27) applied the method of “triads” (13) and concluded that the FFQs had similar or higher validity than did blood carotenoid markers. Differences in their cohorts versus WHEL Study participants and the fact that the method of triads assumes random errors in self-report instruments to be uncorrelated are possible explanations for the discrepancy between their results and ours. Notably, the validity coefficient of the FFQ in our study is similar to values reported for protein density (7).

The availability of the plasma marker permitted calculation of the error correlation between 24-hour recalls and the FFQ, which was estimated to be approximately 0.3 (table 3). This implies that the errors of these two instruments are not independent. Thus, as previously noted (7, 21), using 24-hour recalls to calibrate FFQs would underestimate attenuation factors and overestimate validity.

The FFQ was prone to systematic error, with a moderately high error correlation of more than 0.5 over time (table 3). Large systematic error components imply that repeated administration of the instrument will not improve validity. The 24-hour recalls had a lower error correlation over time of 0.22–0.28 (table 3).

Our findings suggest that both self-report instruments were subject to biases and error. An earlier WHEL Study investigation (28) found that these self-report methods were independent predictors of the plasma marker and had comparable prediction errors. Because different time periods of intake (4 days vs. 3 months) were reflected in these self-report instruments, it is not surprising that they *both* contributed to pre-

dicting plasma values. These results (28) in conjunction with the current analysis suggest that the 24-hour recalls, FFQ, and plasma marker could complement each other in capturing dietary intake. Statistical methods for combining validation and main study results have been developed for logistic regression models (9), and more research in this area is needed.

There are potential limitations to our methods. Plasma carotenoid concentrations are not a true “gold standard” because they are influenced by nondietary factors, such as body size, plasma cholesterol, and smoking status (12–15). Nevertheless, these biomarkers represent an objective approach to characterizing intake and would not be subject to recall bias, a common problem with self-report methods. More importantly, the availability of plasma markers enables us to apply identical modeling assumptions to the two self-report instruments.

We also note that both self-report measures could be strongly biased by the inherent errors in the nutrient databases’ ability to estimate true carotenoid intake. The higher validity of both instruments for assessing β -carotene intake compared with lycopene and lutein is likely a reflection of higher quality food content data for provitamin A compounds, such as β -carotene, compared with non-provitamin A carotenoids. These limitations are well known and are beyond the scope of this paper.

This investigation focused on measurement error in estimating habitual intake of carotenoids. We did not incorporate the possibility of dietary change, such as might be promoted by a successful intervention. It is conceivable that errors in estimating intakes in diet intervention studies might exhibit different biases. However, a rigorous investigation of this possibility would involve modeling changes in “true” dietary intake over time and would require multiple measurements (i.e., >2) per individual.

In summary, our methods illustrate the difficulties in accurately estimating dietary intakes with a single self-report dietary assessment method. A strength of our study is the availability of multiple measures of dietary intake over time on a large sample of breast cancer survivors. The analytical tools used here stem from classic measurement error theory (29) applied to nutritional epidemiology (6–9, 13, 21, 22). We provide a simple method of graphics for checking model fit. The use of multiple dietary assessment methods, rather than reliance on a single measure, may provide more accurate estimates of diet-disease associations.

Note added in proof: We would like to reference a recent article by Spiegelman et al. (30) that shows how one might combine data from the food frequency questionnaire, recalls (or diet records), and biomarkers to estimate the attenuation factor.

ACKNOWLEDGMENTS

The Women’s Healthy Eating and Living Study Group: WHEL Study Coordinating Center: University of California, San Diego, Cancer Prevention and Control Program, San Diego, California: Dr. John P. Pierce (Principal Investigator), Dr. Cheryl L. Rock, Susan Faerber, Vicky

Newman, Shirley W. Flatt, Sheila Kealey, Dr. Barbara Parker, Dr. Loki Natarajan, Jacqueline Major. *WHEL Study clinic sites: Center for Health Research, Portland, Oregon: Dr. Cheryl Ritenbaugh, Dr. Mark Rarick; Kaiser Permanente Northern California, Oakland, California: Dr. Bette J. Caan, Dr. Lou Ferenbacher; Northern California Cancer Center, Palo Alto, California: Dr. Marcia L. Stefanick, Dr. Robert Carlson; University of Arizona, Tucson and Phoenix, Arizona: Dr. Cynthia Thomson, Dr. James Warnecke; University of California, Davis, California: Dr. Ellen B. Gold, Dr. Sidney Scudder; University of California, San Diego Cancer Center, San Diego, California: Dr. Linda Wasserman, Dr. Kathryn A. Hollenbach; University of Texas M. D. Anderson Cancer Center, Houston, Texas: Dr. Lovell A. Jones, Dr. Richard Theriault.*

Conflict of interest: none declared.

REFERENCES

1. Healthy People 2010. Conference ed. Vols. 1 and 2. Washington, DC: US Department of Health and Human Services, 2000.
2. Black AE, Cole TJ. Biased over- or under-reporting is characteristic of individuals whether over time or by different assessment methods. *J Am Diet Assoc* 2001;101:70–80.
3. Bingham SA, Luben R, Welch A, et al. Are imprecise methods obscuring a relation between fat and breast cancer? *Lancet* 2003;362:212–14.
4. Toniolo P, Van Kappel AL, Akhmedkhanov A, et al. Serum carotenoids and breast cancer. *Am J Epidemiol* 2001;153:1142–7.
5. Smith-Warner SA, Spiegelman D, Yaun SS, et al. Intake of fruits and vegetables and risk of breast cancer: a pooled analysis of cohort studies. *JAMA* 2001;285:769–76.
6. Day NE, McKeown N, Wong MY, et al. Epidemiological assessment of diet: a comparison of a 7-day diary with a food frequency questionnaire using urinary markers of nitrogen, potassium and sodium. *Int J Epidemiol* 2001;30:309–17.
7. Kipnis V, Subar AF, Midthune D, et al. Structure of dietary measurement error: results of the OPEN Biomarker Study. *Am J Epidemiol* 2003;158:14–21.
8. Spiegelman D, Schneeweiss S, McDermott A. Measurement error correction for logistic regression models with an “alloyed gold standard.” *Am J Epidemiol* 1997;145:184–96.
9. Spiegelman D, Carroll RJ, Kipnis V. Efficient regression calibration for logistic regression in main study/internal validation study designs with an imperfect reference instrument. *Stat Med* 2001;20:139–60.
10. Spiegelman D. Commentary: correlated errors and energy adjustment—where are the data? *Int J Epidemiol* 2004;33:1387–8.
11. Pierce JP, Faerber S, Wright F, et al. A randomized trial of the effect of a plant based dietary pattern on breast cancer recurrence: the Women’s Healthy Eating and Living (WHEL) Study. *Control Clin Trials* 2002;23:728–56.
12. Rock CL. Carotenoids: biology and treatment. *Pharmacol Ther* 1997;75:185–97.
13. Kaaks RJ. Biochemical markers as additional measurements in studies of the accuracy of dietary questionnaire measurements: conceptual issues. *Am J Clin Nutr* 1997;65(suppl):1232S–9S.
14. Willett W. *Nutritional epidemiology*. New York, NY: Oxford University Press, 1990.
15. Rock CL, Thornquist MD, Kristal AR, et al. Demographic, dietary and life-style factors differentially explain variability in serum carotenoids and fat-soluble vitamins: baseline results from the sentinel site of the Olestra Post-Marketing Surveillance Study. *J Nutr* 1999;129:855–64.
16. Pierce JP, Newman VI, Flatt SW, et al. Telephone counseling intervention significantly increases intakes of micronutrient- and phytochemical-rich vegetables, fruit and fiber in breast cancer survivors. *J Nutr* 2004;134:452–8.
17. Gamboa-Pinto AJ, Rock CL, Ferruzzi MG, et al. Cervical tissue and plasma concentrations of α -carotene and β -carotene are correlated. *J Nutr* 1998;128:1933–6.
18. Martinez ME, Marshall JR, Graver E, et al. Reliability and validity of a self-administered food frequency questionnaire in a chemoprevention trial of adenoma recurrence. *Cancer Epidemiol Biomarkers Prev* 1999;8:941–6.
19. Thomson CA, Giuliano A, Rock CL, et al. Measuring dietary change in a diet intervention trial: comparing food frequency questionnaires and dietary recalls. *Am J Epidemiol* 2003;157:754–62.
20. Newman V, Rock L, Faerber S, et al. Dietary supplement use by women at risk for breast cancer recurrence. The Women’s Healthy Eating and Living Study Group. *J Am Diet Assoc* 1998;98:285–92.
21. Plummer M, Clayton D. Measurement error in dietary assessment: an investigation using covariance structure models. Parts I and II. *Stat Med* 1993;12:925–48.
22. Tosteson TD, Buonaccorsi JP, Demidenko E. Covariate measurement error and the estimation of random-effect parameters in a mixed model for longitudinal data. *Stat Med* 1998;17:1959–71.
23. Efron B, Tibshirani RJ. An introduction to the bootstrap. In: *Monographs on statistics and applied probability*. Vol 57. London, United Kingdom: Chapman & Hall, 1993.
24. Prentice R. Measurement error and results from analytic epidemiology: dietary fat and breast cancer. *J Natl Cancer Inst* 1996;88:1738–47.
25. McNaughton SA, Marks GC, Gaffney P, et al. Validation of a food-frequency questionnaire assessment of carotenoid and vitamin E intake using weighed food records and plasma biomarkers: the method of triads model. *Eur J Clin Nutr* 2005;59:211–18.
26. Kabagambe EK, Baylin A, Allan DA, et al. Application of the method of triads to evaluate the performance of food frequency questionnaires and biomarkers as indicators of long-term dietary intake. *Am J Epidemiol* 2001;154:1126–35.
27. Shai I, Rosner BA, Shahar DR, et al. Dietary evaluation and attenuation of relative risk: multiple comparisons between blood and urinary biomarkers, food frequency, and 24-hour recall questionnaires: the DEARR Study. *J Nutr* 2005;135:573–9.
28. Natarajan L, Rock CL, Major J, et al. On the importance of using multiple methods of dietary assessment. *Epidemiology* 2004;15:738–45.
29. Fuller WA. *Measurement error models*. New York, NY: John Wiley & Sons, 1987.
30. Spiegelman D, Zhao B, Kim J. Correlated errors in biased surrogates: study designs and methods for measurement error correction. *Stat Med* 2005;24:1657–82.

APPENDIX

We did not include person-specific bias terms explicitly in our model. However, by reformulation of the model, the

error variance can be decomposed into “systematic” and “pure” noise components. By use of earlier notation, our model can be rewritten as

$$Y_{ijk} = \alpha_k + u_{ik} + \beta_k Z_i + \tilde{\epsilon}_{ijk}, \quad (\text{A1})$$

where α_k and β_k are intercept and scale parameters, $\tilde{\epsilon}$ s are mean 0 errors with variance $\tilde{\sigma}_k^2$, and $\text{corr}(\tilde{\epsilon}_{ijk}, \tilde{\epsilon}_{isk}) = 1$ if $j = s$ and 0 otherwise. The term u_{ik} is a subject-specific random effect with mean 0 and variance $\sigma_{u_k}^2$. We assume that Z_i , u_{ik} , and $\tilde{\epsilon}_{ijk}$ are mutually independent.

The covariance across time of Y_{ijk} is

$$\begin{aligned} \text{cov}(Y_{ijk}, Y_{isk}) &= \beta_k^2 \sigma_Z^2 + \rho_{kk} \sigma_k^2 \quad (\text{by equation 1}) \\ &= \beta_k^2 \sigma_Z^2 + \sigma_{u_k}^2 \quad (\text{by equation A1}). \end{aligned}$$

Equating the right sides of the above equations and solving for ρ_{kk} gives $\rho_{kk} = \frac{\sigma_{u_k}^2}{\sigma_k^2}$, indicating that the error correlation over time is the proportion of error variance attributable to systematic error. Thus, we can decompose the error variance σ_k^2 as $\rho_{kk} \sigma_k^2$, “systematic error,” and $(1 - \rho_{kk}) \sigma_k^2$, “pure noise” components.