



## Practice of Epidemiology

### Measurement Error of Dietary Self-Report in Intervention Trials

Loki Natarajan, Minya Pu, Juanjuan Fan, Richard A. Levine, Ruth E. Patterson, Cynthia A. Thomson, Cheryl L. Rock, and John P. Pierce\*

\* Correspondence to Dr. John P. Pierce, Cancer Prevention and Control Program, University of California, San Diego, Medical Center, La Jolla, CA 92093 (e-mail: jppierce@ucsd.edu).

Initially submitted April 1, 2010; accepted for publication June 8, 2010.

Dietary intervention trials aim to change dietary patterns of individuals. Participating in such trials could impact dietary self-report in divergent ways: Dietary counseling and training on portion-size estimation could improve self-report accuracy; participant burden could increase systematic error. Such intervention-associated biases could complicate interpretation of trial results. The authors investigated intervention-associated biases in reported total carotenoid intake using data on 3,088 breast cancer survivors recruited between 1995 and 2000 and followed through 2006 in the Women's Healthy Eating and Living Study, a randomized intervention trial. Longitudinal data from 2 self-report methods (24-hour recalls and food frequency questionnaires) and a plasma carotenoid biomarker were collected. A flexible measurement error model was postulated. Parameters were estimated in a Bayesian framework by using Markov chain Monte Carlo methods. Results indicated that the validity (i.e., correlation with "true" intake) of both self-report methods was significantly higher during follow-up for intervention versus non-intervention participants (4-year validity estimates: intervention = 0.57 for food frequency questionnaires and 0.58 for 24-hour recalls; nonintervention = 0.42 for food frequency questionnaires and 0.48 for 24-hour recalls). However, within- and between-instrument error correlations during follow-up were higher among intervention participants, indicating an increase in systematic error. Diet interventions can impact measurement errors of dietary self-report. Appropriate statistical methods should be applied to examine intervention-associated biases when interpreting results of diet trials.

bias (epidemiology); diet; intervention studies; Markov chain Monte Carlo; measurement error; nutrition assessment; reproducibility of results; validity

Abbreviations: CI, credible interval; WHEL, Women's Healthy Eating and Living.

Self-reported dietary intake is commonly assessed via either a general assessment of "usual" foods consumed (i.e., food frequency questionnaires in the recent past (e.g., 3 months)) or a detailed assessment of food intake on a small sample of days (i.e., dietary recalls). Dietary recalls are open-ended and, hence, are more likely to capture food items not listed on food frequency questionnaires. Conversely, foods that are not consumed regularly may be missed on a dietary recall. Both assessment methods are known to have significant random and systematic error (1–5). However, these biases are a concern for randomized dietary trials only if the intervention itself influences the magnitude or direction of the bias. This

intervention-associated bias could occur because participants in a dietary intervention usually receive education on food composition and portion size estimation, which could improve self-report accuracy; alternatively, participants may modify their dietary intake or reporting during monitoring to appear more compliant with intervention goals. Such biases can complicate interpretation of trial results. For instance, significant overreporting of targeted healthy behaviors in the intervention versus nonintervention arm of a null trial could render the results inconclusive, while more accurate self-report in the intervention group could attenuate observed dietary changes between study arms, possibly leading to the erroneous conclusion

that the intervention did not change dietary patterns sufficiently.

When available, biomarkers of dietary intake provide a more objective measure of exposure and can be used to quantify intervention-associated changes (6–8) and to investigate the biases of self-report instruments (1–5, 9–14). By use of biomarkers, differential underreporting by intervention status was noted for sodium intake (6, 7) assessed by 24-hour recalls and for energy and protein intake assessed by food frequency questionnaires (8). Additionally, error correlations (i.e., systematic error) of reported sodium and potassium intakes on 24-hour recalls varied by intervention arm (6).

Little is known about intervention-associated bias in self-reported fruit and vegetable intakes. We investigated such biases using data from the Women's Healthy Eating and Living (WHEL) Study, a randomized intervention trial of 3,088 breast cancer survivors, which examined whether a high vegetable, fruit, fiber, and low-fat diet improved breast cancer-free survival (15, 16). For this analysis, total carotenoid intake was the exposure of interest. Carotenoids, bioactive agents found primarily in fruits and vegetables, reflect adherence to the prescribed WHEL Study intervention. This report compares the validity (i.e., correlation between observed and "true" intake) and within-instrument error correlations (i.e., systematic error) of 2 self-report dietary assessment methods, as well as a plasma carotenoid biomarker, known to correlate with fruit and vegetable intake (17). Higher validity leads to better statistical power and less bias in diet–disease risk estimates. Large systematic error reduces the accuracy of an instrument (1, 3, 18). Thus, quantifying measurement error of assessment methods is key to proper design and analysis of diet studies.

## MATERIALS AND METHODS

### Description of the study sample

Details of the WHEL Study protocol have been reported previously (16). Between 1995 and 2000, the study recruited 3,088 breast cancer survivors (40% stage I, 55% stage II, 5% stage IIIA), averaging 51 years at diagnosis (range: 26–71 years), of whom 85% were non-Hispanic white, 5% Hispanic, 4% African American, and 4% Asian/Pacific Islander. Participants were randomly assigned to 1 of 2 diet groups. The comparison (i.e., nonintervention) arm ( $n = 1,551$ ) received printed materials containing general dietary recommendations for cancer prevention. The intervention arm ( $n = 1,537$ ) was counseled to adopt a micronutrient- and phytochemical-rich diet, including 5 vegetable servings/day, 16 ounces (~473 ml) of vegetable juice/day, and 3 fruit servings/day, all foods rich in *carotenoids*. Hence, the current analysis focused on total carotenoid intake (i.e., sum of  $\alpha$ -carotene,  $\beta$ -carotene, lutein, lycopene, and  $\beta$ -cryptoxanthin) from food sources and supplements, collected at study entry (i.e., preintervention) and at 1 and 4 years post-randomization, thus assessing short- and long-term dietary changes. The institutional review boards of all participating institutions approved study procedures, and written informed consent was obtained from all participants.

Two self-report methods (repeat 24-hour recalls and the Arizona food frequency questionnaires) and a plasma biomarker (16) were used to assess dietary intake. Repeat 24-hour recalls of dietary intake and supplement use were collected during telephone interviews by trained dietary assessors blinded to the randomization allocation of participants. During these recalls, participants were queried about food intake during the previous 24 hours. Assessments at a particular time point consisted of the average of four 24-hour recalls collected over a 3-week period, including 2 weekdays and 2 weekend days.

The 153-item, semiquantitative Arizona food frequency questionnaires collected information regarding usual foods consumed, frequency of consumption, and food preparation methods over the previous 3 months, by use of age- and gender-specific estimates of portions recorded as small, medium, or large (16). Plasma carotenoids were separated and quantified by a validated high-performance liquid chromatography method (19). For the self-report instruments, total carotenoid intake was derived from reported dietary intake by using standard databases (20, 21). Details of the timing of the assessments were previously reported (16). Briefly, fasting blood samples for plasma carotenoid measurements were collected within 2–4 weeks of obtaining food frequency questionnaires and within 3 months of 24-hour recall assessments.

### The statistical model

Let  $Y_{ijk}$  represent dietary intake measured on participant  $i$  at time  $j$  with instrument  $k$ , where  $j = 1$  represents baseline (i.e., study entry), and  $j > 1$  represents follow-up measurements. Let  $T_{ij}$  be "true intake" for participant  $i$  at time  $j$ . We posit a flexible linear measurement error model:

$$Y_{ijk} = \alpha_k + \beta_k T_{ij} + \epsilon_{ijk} \quad (1)$$

$$T_{ij} = a_i + b_i \times I(j > 1) \times I(\text{group} = \text{intervention}) + \delta_{ij}.$$

Here,  $\alpha_k$  and  $\beta_k$  represent additive and scaling bias of instrument  $k$ ;  $I(\cdot)$  is an indicator function;  $a_i$  and  $b_i$  are individual-specific intakes (random intercept and slope), with mean ( $\mu_a$ ,  $\mu_b$ ) and covariance matrix  $\Sigma_{ab}$ ;  $\delta_{ij}$  and  $\epsilon_{ijk}$  are errors. We assume that 1)  $T_{ij}$  and  $\epsilon_{ijk}$  are independent; 2) errors  $\epsilon_{ijk}$  and  $\epsilon_{rst}$  are independent if  $i \neq r$  but could be correlated otherwise, thus allowing correlated measurement error on repeats of an instrument over time and also between instruments; and 3)  $\delta_{ij}$ s represent random fluctuation around the mean intake for individual  $i$ , are independent over  $i$  and  $j$ , and are also independent of  $a_i$  and  $b_i$ . Note that we model the "true intake" of intervention subjects as baseline level ( $j = 1$ ) plus "average shift" at follow-up ( $j > 1$ ) with random error, whereas we assume each comparison group subject has (on average) a stable diet, fluctuating around her mean intake.

Our objective was to investigate intervention-associated biases of dietary assessment methods. Hence, we actually allowed  $\alpha_k$ ,  $\beta_k$ , and error correlations,  $\text{corr}(\epsilon_{ijk}, \epsilon_{ist})$ , to vary across randomization arms. To keep notation tractable, we do not incorporate these generalizations into model 1, reserving further presentation of our Bayesian hierarchical model (22, 23) to the next section.

We analyzed data on 3,088 WHEL Study participants measured at 3 time points (baseline and 1 and 4 years). We set  $k = 1$  for 24-hour recalls,  $k = 2$  for food frequency questionnaires, and  $k = 3$  for the plasma biomarker. The objective plasma measure was designated as “true intake” plus error (i.e.,  $\alpha_3 = 0, \beta_3 = 1$ ). Errors in 24-hour recalls or food frequency questionnaires were assumed to be independent of plasma errors; that is,  $\text{corr}(\epsilon_{ijk}, \epsilon_{is3}) = 0$  for  $k = 1, 2$ . This assumption is common (2, 5, 9) and defensible as one would not expect self-report and laboratory errors to be correlated. Finally, to render our model identifiable, we assigned fixed correlations ranging from 0 to 0.3 for plasma errors over time (i.e.,  $\text{corr}(\epsilon_{ij3}, \epsilon_{is3})$ , for  $j \neq s$ ) (5). These correlations measure repeated errors in plasma measurements (e.g., due to persistent laboratory errors) and are usually negligible. Hence, allowing error correlations as large as 0.3 represents a worst-case scenario. The aim was to estimate the following:

- $\text{corr}(Y_{ijk}, T_{ij}) = \sqrt{\frac{\beta_k^2 \text{var}(T_{ijk})}{\beta_k^2 \text{var}(T_{ijk}) + \text{var}(\epsilon_{ijk})}}$ , the validity of each

self-report and biomarker over time, thus acknowledging that the plasma marker is an imperfect reference, measured with both random and systematic error, possibly attributable to laboratory error and person-specific factors (“var” denotes “variance”);

- $\text{corr}(\epsilon_{ijk}, \epsilon_{isl})$ , within- and between-instrument error correlations for 24-hour recalls ( $k = 1$ ) and food frequency questionnaires ( $k = 2$ );
- $\alpha_k, \beta_k$  for  $k = 1, 2$ , additive and scaling coefficients for each instrument;
- $\mu_a$ , average true intake at baseline;
- $\mu_b$ , average change achieved by the intervention at follow-up; and
- $\text{corr}(a_i, b_i)$ , the correlation between baseline intake and change achieved for intervention participants.

Previously, we estimated validity and error correlations of these assessment methods among a subsample in the WHEL comparison arm (5). In this analysis, we examine whether the measurement error properties of self-report instruments change after participation in a dietary intervention. Previously (5), we used a *method of moments* technique, an approach that becomes intractable with multiple time points and unbalanced data. Here, we applied a Bayesian approach (22, 23) described in the next section.

**Estimation**

A Markov chain Monte Carlo approach using the software WinBUGS (24) was applied. The model presents itself as a hierarchy starting from an assumed multivariate normal likelihood on the dietary intake measurements (observed and true intake)  $Y_{ijk}$  and  $T_{ij}$ , characterized by parameters  $\alpha, \beta, a, b$ , and various covariance matrices on which prior distributions are placed. In particular, for participants  $i = 1, \dots, n$ , time points  $j = 1, 2, 3$ , and instruments  $k = 1, 2, 3$ , with  $\mathbf{I}_p$  denoting a  $p \times p$  identity matrix, assume the following:

$$\begin{pmatrix} Y_{i.1} \\ Y_{i.2} \\ Y_{i.3} \end{pmatrix} \sim \text{normal} \left\{ \begin{pmatrix} \mu_{24h} \\ \mu_{FFQ} \\ \mu_{lab} \end{pmatrix}, \begin{pmatrix} \Sigma_{RQ} & \mathbf{0} \\ \mathbf{0} & \Sigma_L \end{pmatrix} \right\}, \quad (2)$$

where  $Y_{i.k} = (Y_{i1k}, Y_{i2k}, Y_{i3k})$  is the observed dietary intake vector at baseline and 1 and 4 years for individual  $i$  on instrument  $k$ ,

$$\mu_{24h} = (\alpha_1 + \beta_1 T_{i1}, \alpha_1 + \beta_1 T_{i2}, \alpha_1 + \beta_1 T_{i3}),$$

$$\mu_{FFQ} = (\alpha_2 + \beta_2 T_{i1}, \alpha_2 + \beta_2 T_{i2}, \alpha_2 + \beta_2 T_{i3}),$$

$$\mu_{lab} = (T_{i1}, T_{i2}, T_{i3}),$$

$$\Sigma_{RQ} = \begin{pmatrix} \Sigma_{24h} & \Sigma_{24h,FFQ} \\ \Sigma_{24h,FFQ} & \Sigma_{FFQ} \end{pmatrix},$$

and

$$\Sigma_L = \begin{pmatrix} \sigma_{lab_1}^{\frac{1}{2}} & 0 & 0 \\ 0 & \sigma_{lab_2}^{\frac{1}{2}} & 0 \\ 0 & 0 & \sigma_{lab_3}^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{pmatrix} \begin{pmatrix} \sigma_{lab_1}^{\frac{1}{2}} & 0 & 0 \\ 0 & \sigma_{lab_2}^{\frac{1}{2}} & 0 \\ 0 & 0 & \sigma_{lab_3}^{\frac{1}{2}} \end{pmatrix}.$$

$\Sigma_L$  denotes the error covariance matrix over time for the plasma marker. The  $3 \times 3$  matrices  $\Sigma_{24h}$  (24 hours) and  $\Sigma_{FFQ}$  (food frequency questionnaires) represent error covariances over time within instrument and the  $3 \times 3$  matrix  $\Sigma_{24h, FFQ}$  represents error correlations between instruments. As mentioned previously,  $\alpha$ 's,  $\beta$ 's, and  $\Sigma_{RQ}$  were allowed to differ by randomization arm.

True dietary intakes are modeled in the second level of the hierarchy as

$$T_{ij} \sim \text{normal}(\mu_{T_{ij}}, \sigma_{T_j}^{-1}), j = 1, 2, 3, \quad (3)$$

where  $\mu_{T_{i1}} = a_i$  and  $\mu_{T_{i2}} = \mu_{T_{i3}} = a_i + b_i \times I(\text{group} = \text{intervention})$ .

Conjugate prior distributions are placed on the parameters in expressions 2 and 3 as follows:

$$\begin{aligned} \alpha_k &\sim \text{normal}(\mu_{\alpha_k}, \sigma_{\alpha_k}^{-1}), k = 1, 2 \\ \beta_k &\sim \text{normal}(\mu_{\beta_k}, \sigma_{\beta_k}^{-1}), k = 1, 2 \\ \begin{pmatrix} a_i \\ b_i \end{pmatrix} &\sim \text{normal} \left\{ \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix}, \Sigma_{ab} \right\} \\ \Sigma_{RQ}^{-1} &\sim \text{wishart}(\mathbf{I}_6, \nu_{RQ}) \\ \sigma_{lab_j}^{-1} &\sim \text{gamma}(L_a, L_b), j = 1, 2, 3 \end{aligned}$$

$$\sigma_{T_j} \text{ iid gamma}(T_a, T_b), \quad (4)$$

**Table 1.** Dietary Intake and Sample Characteristics for Key Covariates by Randomization Arm and Follow-up Year in a WHEL Sample of 3,088 Breast Cancer Survivors Recruited Between 1995 and 2000 and Followed Through 2006

Covariate, Assessment Method, and Years	Randomization Group, mean (SD)	
	Comparison (n = 1,551)	Intervention (n = 1,537)
<b>Carotenoids<sup>a</sup></b>		
24-Hour recalls, mg/day		
Baseline	19.61 (34.93)	18.76 (15.46)
1 year	19.83 (48.61)	59.44 (32.63)
4 years	18.16 (49.71)	45.41 (31.01)
Arizona food frequency questionnaire, mg/day		
Baseline	24.32 (25.11)	23.98 (19.22)
1 year	26.43 (52.58)	48.16 (26.85)
4 years	22.38 (26.90)	36.25 (22.14)
Plasma, μmol/L		
Baseline	2.32 (1.46)	2.26 (1.28)
1 year	2.33 (1.40)	3.70 (2.41)
4 years	2.21 (1.19)	3.15 (2.05)
<b>Energy intake</b>		
24-Hour recalls, kcal/day		
Baseline	1,717.0 (414.6)	1,719.0 (400.7)
1 year	1,605.0 (391.4)	1,603.0 (351.1)
4 years	1,574.0 (392.7)	1,552.0 (358.1)
Arizona food frequency questionnaire, kcal/day		
Baseline	1,919.0 (808.0)	1,912.0 (905.5)
1 year	1,819.0 (781.2)	1,910.0 (738.3)
4 years	1,670.0 (734.0)	1,755.0 (702.3)
<b>Cholesterol</b>		
Plasma, mg/dL		
Baseline	196.5 (39.7)	195.7 (40.4)
1 year	195.2 (39.1)	192.8 (39.2)
4 years	197.8 (36.0)	198.8 (38.3)
<b>Body mass index, kg/m<sup>2</sup></b>		
Baseline	27.2 (6.1)	27.2 (5.9)
1 year	27.5 (6.1)	27.1 (5.6)
4 years	27.6 (6.0)	27.5 (5.8)

Abbreviations: SD, standard deviation; WHEL, Women’s Healthy Eating and Living.

<sup>a</sup> Self-reported carotenoid data comprise dietary plus supplement intake.

where *iid* denotes variables independent and identically distributed.

The next level of the hierarchy specifies further conjugate distributions on the hyperparameters characterizing the prior distributions in expression 4 as follows:

$$\begin{aligned} \mu_{\alpha_1}, \mu_{\alpha_2} & \quad iid \text{ normal}(0, \sigma_{\alpha}) \\ \mu_{\beta_1}, \mu_{\beta_2} & \quad iid \text{ normal}(0, \sigma_{\beta}) \\ \sigma_{\alpha_1}, \sigma_{\alpha_2}, \sigma_{\beta_1}, \sigma_{\beta_2} & \quad iid \text{ gamma}(\tau_a, \tau_b) \\ \mu_a, \mu_b & \quad iid \text{ normal}(0, \sigma_{ab}) \end{aligned}$$

$$\Sigma_{ab}^{-1} \sim \text{wishart}(\mathbf{I}_2, \nu_{ab}). \tag{5}$$

All further parameters in the distributions of expressions 4 and 5 are specified and known. Noninformative prior distributions were assumed: Gamma prior distribution parameters were set to 0.01, normal distribution parameters were characterized by large dispersions  $\sigma_{\alpha} = \sigma_{\beta} = \sigma_{ab} = 1,000$ , and “wishart” distribution degrees of freedom were specified as  $\nu_{ab} = 2$  and  $\nu_{RQ} = 6$ . In sensitivity analysis, different covariance structures, namely, compound symmetry (i.e.,  $r_{jt} = r$  fixed) or first-order autoregressive (i.e.,  $r_{jt} = r^{|j-t|}$ ) were specified for  $\Sigma_L$ , the plasma error covariance matrix, with  $r$  varied between 0 and 0.3.

These models were fit to WHEL data. Self-reported total carotenoid intakes were adjusted for body mass index and total energy intake, while plasma carotenoid concentrations were adjusted for body mass index and total plasma cholesterol (17). Fewer than 5% of WHEL women were current smokers; hence, this factor was not included in the models (17). Similar to other reports (3), these adjusted outcome variables were derived as marginal residuals from linear mixed-effects models (25), which included a random intercept. All variables were log transformed. A total of 40,000 iterations were run, the first “burn-in” 5,000 were discarded, and the remaining 35,000 were used for inference. Trace, density, and autocorrelation plots were obtained to assess model diagnostics, and they indicated that model fit was adequate. Posterior medians with 95% credible intervals (of measurement error parameters derived from model 1) were estimated. Additionally, summary statistics for unadjusted, untransformed carotenoid levels were calculated.

## RESULTS

As previously reported, the WHEL intervention achieved large changes in dietary pattern at 1 and 4 years of follow-up (26–28). Carotenoid intake increased markedly in the intervention arm (26, 28, 29) irrespective of assessment method, with minimal changes in the comparison group (Table 1). Similar results hold for the model-based true intake distribution,  $T_{ij}$ , represented in model 1 by 2 subject-specific effects,  $a_i$  (representing carotenoid intake at baseline), and  $b_i$  (parameterizing changes in carotenoid levels among intervention group participants). The estimated median true intake at baseline was  $\mu_a = -0.12$  (95% credible interval (CI):  $-0.13, -0.10$ ), the negative values likely due to log transformation and various covariate adjustments. The median change for intervention subjects was  $\mu_b = 0.42$  (95% CI: 0.40, 0.45), corroborating the significant increases in observed carotenoid intake achieved by the WHEL intervention (Table 1). The median correlation between the 2 subject-specific effects was 0.10 (95% CI:  $-0.01, 0.23$ ),

**Table 2.** Validity<sup>a</sup> of the Self-Report Instruments and Plasma Marker for Total Carotenoids in a WHEL Sample of 3,088 Breast Cancer Survivors Recruited Between 1995 and 2000 and Followed Through 2006

Assessment Method <sup>b</sup> and Years	Error Correlation in Repeated Plasma Measurements ( <i>r</i> )							
	<i>r</i> = 0				<i>r</i> = 0.3 <sup>c</sup>			
	Comparison ( <i>n</i> = 1,551)		Intervention ( <i>n</i> = 1,537)		Comparison ( <i>n</i> = 1,551)		Intervention ( <i>n</i> = 1,537)	
	Validity Estimate	95% Credible Interval	Validity Estimate	95% Credible Interval	Validity Estimate	95% Credible Interval	Validity Estimate	95% Credible Interval
24-Hour recalls								
Baseline	0.44	0.40, 0.48	0.44	0.40, 0.48	0.44	0.40, 0.47	0.44	0.40, 0.47
1 year	0.46	0.42, 0.49	0.56	0.52, 0.60	0.45	0.41, 0.48	0.55	0.51, 0.59
4 years	0.48	0.44, 0.52	0.58	0.54, 0.62	0.49	0.45, 0.53	0.57	0.54, 0.61
Arizona food frequency questionnaire								
Baseline	0.39	0.35, 0.43	0.39	0.35, 0.43	0.38	0.34, 0.42	0.38	0.34, 0.42
1 year	0.42	0.38, 0.46	0.53	0.49, 0.57	0.41	0.37, 0.45	0.52	0.48, 0.56
4 years	0.42	0.38, 0.46	0.57	0.53, 0.61	0.42	0.38, 0.46	0.57	0.53, 0.61
Plasma								
Baseline	0.85	0.82, 0.86	0.85	0.82, 0.86	0.87	0.84, 0.90	0.87	0.84, 0.90
1 year	0.90	0.88, 0.93	0.93	0.91, 0.95	0.90	0.87, 0.93	0.93	0.91, 0.96
4 years	0.87	0.85, 0.90	0.91	0.89, 0.93	0.87	0.84, 0.89	0.91	0.89, 0.93

Abbreviation: WHEL, Women's Healthy Eating and Living.

<sup>a</sup> "Validity" is defined as the correlation between observed and true dietary intake, reflecting the accuracy of the assessment method.

<sup>b</sup> Self-reported dietary intakes adjusted for body mass index and total caloric intake; plasma markers adjusted for plasma cholesterol and body mass index.

<sup>c</sup> Assumed first-order autoregressive correlation structure for plasma error matrix.

implying that the changes achieved by the intervention were minimally influenced by a participant's carotenoid intake at baseline.

### Validity of dietary assessment methods

Validity is the correlation between observed and "true" intake. Higher validity indicates better accuracy at ranking individuals by dietary intake level (18). Clearly, a validity value of 1 would be ideal, but values  $\leq 0.5$  are common for dietary self-report (3). For the WHEL sample, the plasma marker exhibited the highest validity ( $\text{corr}(Y_{ijk}, T_{ij})$ ) irrespective of intervention status, although, as expected, these values were less than the ideal validity coefficient of 1, reflecting the realistic scenario of laboratory and person-specific errors for this marker (Table 2). The validity of 24-hour recalls ranged from 0.44 to 0.48 at different time points within the comparison arm and from 0.44 to 0.58 for intervention subjects, assuming independent plasma errors (i.e.,  $r = 0$ ) (Table 2). Interestingly, 24-hour recalls exhibited significantly higher validity at years 1 and 4 for intervention subjects, as evidenced by nonoverlapping 95% credible intervals between intervention and comparison group estimates. A similar nonoverlapping of credible intervals was observed for food frequency questionnaires, with comparison group validity ranging from 0.39 to 0.42 and intervention group validity between 0.39 and 0.57. Thus, participating in the intervention improved self-report accuracy.

Comparing between instruments, 24-hour recalls exhibited numerically higher validity than food frequency questionnaires at each time point, although all credible intervals were overlapping. Importantly, both methods had similar validity at 1- and 4-year assessments in the intervention arm, indicating comparable accuracy for foods consumed regularly (e.g., fruit and vegetable intakes which were a major focus of the WHEL intervention).

Results were robust to varying assumptions regarding the plasma error covariance structure  $\Sigma_L$  (plasma error correlation  $r = 0.3$  presented in Table 2; data not shown for  $r = 0.1, 0.2$ ).

### Error correlations within and between self-report methods

*Within-instrument error correlations.* Error correlations between repeated assessments (Table 3) represent the systematic error of each instrument and equal the proportion of error variance due to person-specific bias (5). For comparison group subjects, error correlations of baseline with follow-up 24-hour recalls were similar: 0.32 for baseline-1 year and 0.27 for baseline-4 year correlations. The 1 year-4 year error correlations = 0.42, however, were significantly higher than the baseline-follow-up correlations, suggesting a higher systematic error on postrandomization measures. In the intervention group, baseline-follow-up error correlations were much lower (0.02 for baseline-1 year and 0.09 for baseline-4 year). These negligible error

**Table 3.** Error Correlations Within<sup>a</sup> and Between<sup>b</sup> Self-Report Instruments for Total Carotenoid Intake in a WHEL Sample of 3,088 Breast Cancer Survivors Recruited Between 1995 and 2000 and Followed Through 2006

Assessment Method <sup>c</sup> and Year	Error Correlation in Repeated Plasma Measures ( <i>r</i> )							
	<i>r</i> = 0				<i>r</i> = 0.3 <sup>d</sup>			
	Comparison ( <i>n</i> = 1,551)		Intervention ( <i>n</i> = 1,537)		Comparison ( <i>n</i> = 1,551)		Intervention ( <i>n</i> = 1,537)	
	Error Correlation Estimate	95% Credible Interval	Error Correlation Estimate	95% Credible Interval	Error Correlation Estimate	95% Credible Interval	Error Correlation Estimate	95% Credible Interval
Within 24-hour recalls								
Baseline, 1 year	0.32	0.27, 0.37	0.02	-0.06, 0.10	0.32	0.27, 0.38	0.00	-0.07, 0.08
Baseline, 4 years	0.27	0.21, 0.33	0.09	0.02, 0.16	0.27	0.21, 0.33	0.08	0.01, 0.15
1 year, 4 years	0.42	0.37, 0.48	0.59	0.53, 0.63	0.43	0.37, 0.48	0.59	0.54, 0.64
Within Arizona food frequency questionnaire								
Baseline, 1 year	0.55	0.50, 0.59	0.19	0.12, 0.26	0.55	0.51, 0.59	0.17	0.10, 0.24
Baseline, 4 years	0.39	0.33, 0.44	0.23	0.17, 0.30	0.39	0.34, 0.44	0.23	0.16, 0.29
1 year, 4 years	0.51	0.46, 0.56	0.58	0.52, 0.63	0.51	0.46, 0.56	0.58	0.53, 0.63
Between Arizona food frequency questionnaire and 24-hour recalls								
Baseline	0.40	0.36, 0.45	0.40	0.36, 0.45	0.41	0.36, 0.45	0.41	0.36, 0.45
1 year	0.42	0.36, 0.47	0.68	0.65, 0.72	0.42	0.36, 0.47	0.69	0.65, 0.72
4 years	0.39	0.34, 0.44	0.51	0.45, 0.56	0.39	0.34, 0.44	0.51	0.46, 0.56

Abbreviation: WHEL, Women's Healthy Eating and Living.

<sup>a</sup> Within-instrument error correlations represent systematic error and can be interpreted as the proportion of error variance due to subject-specific bias.

<sup>b</sup> Between-instrument error correlations represent shared systematic error (e.g., recall bias) between instruments.

<sup>c</sup> Self-reported dietary intakes adjusted for body mass index and total caloric intake; plasma markers adjusted for plasma cholesterol and body mass index.

<sup>d</sup> Assumed first-order autoregressive correlation structure for plasma error matrix.

correlations possibly reflect changes in reporting practices accompanying the changes in dietary intake among intervention participants. However, 1 year-4 year error correlations for 24-hour recalls were 0.59 for intervention participants, significantly higher than for nonintervention subjects. Thus, for 24-hour recalls, the proportion of error variance attributable to subject-specific factors was higher for postrandomization assessments versus baseline-follow-up assessments for both study arms, with a significantly larger impact for intervention subjects.

Error correlations for repeated food frequency questionnaires among comparison subjects (Table 3) were higher than for 24-hour recalls (with nonoverlapping credible intervals for baseline-1 year and baseline-4 year values). Interestingly, for food frequency questionnaires, baseline-1 year and 1 year-4 year error correlations were similar (0.55 vs. 0.51), indicating that the error structure of food frequency questionnaires was less affected at follow-up compared with 24-hour recalls among comparison group women. For intervention subjects, baseline-follow-up error correlations for food frequency questionnaires were significantly lower than the corresponding comparison arm values, but they were significantly higher than 24-hour recall estimates (with nonoverlapping credible intervals) for intervention participants, indicating more consistent systematic error in food frequency questionnaires even after commencement of a dietary intervention. For food fre-

quency questionnaires, 1 year-4 year intervention group error correlations were 0.58, numerically higher than comparison group values, and similar to corresponding 24-hour recall intervention group estimates. Results were essentially unchanged when plasma error correlations were varied between 0 and 0.3.

**Between-instrument error correlations.** Between-instrument error correlations (Table 3), representing shared systematic error (e.g., similar recall biases) between instruments, were consistent over time, equaling  $\approx 0.40$  for the comparison group, but were significantly higher at 1 and 4 years for the intervention arm (0.68 at 1 year and 0.51 at 4 years). Thus, using one self-report instrument to validate another could overstate the accuracy of the instruments because of shared errors between them (2, 3, 5). More importantly, our results suggest that participating in a dietary intervention actually increased error correlation between self-report instruments.

#### Additive and scaling coefficients of self-report instruments

Additive ( $\alpha_k$ ) and scaling ( $\beta_k$ ) coefficient values (model 1) for an unbiased instrument would be  $\alpha_k = 0$  and  $\beta_k = 1$ , yielding observed intake equal to "true intake" plus "random error." Given the different measurement scales between self-report (mg/day) and plasma carotenoids

**Table 4.** Additive ( $\alpha_k$ ) and Scaling ( $\beta_k$ ) Bias for Each Self-Report Instrument for Total Carotenoid Intake in a WHEL Sample of 3,088 Breast Cancer Survivors Recruited Between 1995 and 2000 and Followed Through 2006

Assessment Method <sup>a</sup>	Error Correlation in Repeated Plasma Measures ( <i>r</i> )							
	<i>r</i> = 0				<i>r</i> = 0.3 <sup>b</sup>			
	Comparison ( <i>n</i> = 1,551)		Intervention ( <i>n</i> = 1,537)		Comparison ( <i>n</i> = 1,551)		Intervention ( <i>n</i> = 1,537)	
	Bias Estimate	95% Credible Interval	Bias Estimate	95% Credible Interval	Bias Estimate	95% Credible Interval	Bias Estimate	95% Credible Interval
24-Hour recalls								
$\alpha_1$	-0.24	-0.27, -0.22	0.06	0.02, 0.10	-0.25	-0.27, -0.22	0.07	0.03, 0.11
$\beta_1$	0.77	0.70, 0.84	1.10	1.01, 1.19	0.74	0.67, 0.81	1.04	0.96, 1.13
Arizona food frequency questionnaire								
$\alpha_2$	-0.11	-0.13, -0.09	0.02	-0.01, 0.05	-0.11	-0.13, -0.09	0.03	-0.01, 0.06
$\beta_2$	0.55	0.49, 0.61	0.70	0.64, 0.76	0.53	0.47, 0.58	0.66	0.61, 0.72

Abbreviation: WHEL, Women's Healthy Eating and Living.

<sup>a</sup> Self-reported dietary intakes adjusted for body mass index and total caloric intake; plasma markers adjusted for plasma cholesterol and body mass index.

<sup>b</sup> Assumed first-order autoregressive correlation structure for plasma error matrix.

( $\mu\text{mol/L}$ ) and the various transformations and covariate adjustments applied to the observed data ( $Y_{ijk}$ ), neither self-report method is expected to exhibit these ideal coefficient values. Nevertheless, comparing coefficients between instruments and randomization arms permits further illustration of different biases.

The scaling factors ( $\beta_k$ 's) were significantly higher (closer to the ideal value of 1) for 24-hour recalls versus food frequency questionnaires, irrespective of randomization arm (Table 4). For food frequency questionnaires,  $\beta$  estimates were closer to unity for the intervention versus comparison group, but the 95% credible intervals did not include unity for either group. Interestingly, for 24-hour recalls,  $\beta_k$  switched from being below unity in the comparison arm to above unity in the intervention arm, again indicating profound intervention effects on biases of 24-hour recalls. Additive factors ( $\alpha_k$ ) were close to zero for both instruments in the intervention group, but they were significantly different from zero for the comparison group. Further, in the comparison arm, 24-hour recalls displayed significantly larger additive bias than did food frequency questionnaires.

## DISCUSSION

Participation in a dietary intervention trial can affect the accuracy of self-reported intake (6–8). Our results show that the validity of 24-hour recalls and food frequency questionnaires for capturing carotenoid intake improved substantially for participants randomized to the WHEL intervention arm. WHEL Study participants were counseled to consume carotenoid-rich fruits and vegetables and trained to better estimate portion sizes. These factors could have resulted in more accurate recall of intake, as well as real increases in usual intake of fruits and vegetables. However, this improvement in validity among intervention subjects was accompanied by a substantial increase in within-instrument error correlations for 1 year–4 year (i.e., postrandomization) assessments, especially for 24-hour recalls. Participant burden

could have induced subjects to reduce 24-hour recall interview time by either repeatedly under- or overreporting intake in certain set ways, thereby increasing systematic error. Additionally, since 24-hour recalls were prescheduled in the WHEL Study, participants may have altered (i.e., increased) fruit and vegetable intake on the scheduled recall day, in order to appear more compliant with the intervention goals or simply because of the continual intervention emphasis on self-monitoring of food intake (16). This effect could potentially be mitigated by having random, unscheduled 24-hour recall interviews. However, unscheduled recalls reduce response rates. Hence, the WHEL Study elected to conduct prescheduled 24-hour recalls in order to minimize missing data.

Reconciling these apparently conflicting results among WHEL Study intervention participants is important: Improved validity likely resulted from participant training and habituation to a diet rich in fruits and vegetables, thereby improving self-report accuracy. Psychosocial factors and participant burden possibly led to increased systematic error.

Our findings for the comparison group confirm previous reports (2–4, 10–13) that self-report methods have moderate validity and high systematic and between-instrument errors. These similarities are striking given the different populations (healthy individuals vs. cancer survivors) and dietary components investigated across studies. Interestingly, within 24-hour recall error correlations were higher for 1 year–4 year versus baseline-follow-up measures, indicating an increase in postrandomization systematic error among comparison group women. Volunteering for a dietary intervention trial and participant burden (i.e., a response-set bias from repeated assessments) may have systematically biased reporting among WHEL Study comparison group women.

This study has many strengths and limitations. The WHEL Study offers a large sample with dietary self-report and an objective biomarker assessed at multiple time points. However, there is no recovery biomarker for fruit and vegetable consumption. Hence, the absolute extent and degree

of bias in fruit and vegetable self-report cannot be determined. Our results are derived from a predominantly non-Hispanic white sample of breast cancer survivors participating in a clinical trial, a population that is likely to be better educated and more motivated than the general population. We cannot compare changes in reporting practices between an intervention and true control group, namely, an observational cohort followed for the same period, which did not receive any dietary advice. Our Bayesian approach (22, 23) permits flexible model development, can handle missing data, can incorporate data from multiple sources, and appropriately assesses uncertainty in parameter estimates, but it can be computationally challenging.

There are important lessons to be learned. These data confirm that both self-report methods have moderate validity and substantial subject-specific bias (2, 3). Strategies to improve reporting accuracy and to reduce participant burden may increase validity and lower within-instrument error correlations. From a statistical perspective, longitudinal studies should use flexible approaches for data analysis, such as mixed models (25) or Bayesian methods (22, 23), which can incorporate person-specific bias terms and model flexible error covariance structures. Improper modeling could yield biased regression coefficients and standard error estimates, as well as incorrect *P* values, leading to potentially erroneous interpretation of intervention effects.

Accurate assessment of dietary intake is crucial for obtaining unbiased diet-disease risk estimates (1). Our results suggest that participating in an intervention might improve self-report validity, thus reducing biases in diet-disease risk estimates. However, repeated dietary assessment increases the systematic error of self-report. Biomarkers, albeit expensive, are useful for quantifying biases and calibrating self-report methods (4, 9, 14). Analytical methods that carefully model the various biases and error structures should be implemented. Incorporating and planning for these factors in study design and analysis are a key first step toward uncovering the role of diet in disease prevention.

## ACKNOWLEDGMENTS

Author affiliations: Rebecca and John Moores UCSD Cancer Center, School of Medicine, University of California, San Diego, La Jolla, California (Loki Natarajan, Minya Pu, Ruth E. Patterson, Cheryl L. Rock, John P. Pierce); Arizona Cancer Center, University of Arizona, Tucson, Arizona (Cynthia A. Thomson); and Department of Mathematics and Statistics, San Diego State University, San Diego, California (Juanjuan Fan, Richard A. Levine).

This work was supported by the National Cancer Institute (grant CA 69375 to L. N., M. P., R. P., C. R., C. T., J. P. and grant CA117292-02 to L. N.) and by the National Eye Institute (grant R21EY018698-02 to J. F., R. L.), both at the US National Institutes of Health. The Women's Healthy Eating and Living Study was initiated with the support of the Walton Family Foundation and was continued with funding from the National Cancer Institute (grant CA 69375). Some of the data were collected from the Gen-

eral Clinical Research Centers under National Institutes of Health grants M01-RR00070, M01-RR00079, and M01-RR00827. Partial support was also provided by National Institutes of Health grants CA117292-02 and R21EY018698-02.

*WHEL Study Coordinating Center: Cancer Prevention and Control Program, Rebecca and John Moores UCSD Cancer Center, School of Medicine, University of California, San Diego, La Jolla, California* (Dr. John P. Pierce, Susan Faerber, Dr. Barbara A. Parker, Dr. Loki Natarajan, Dr. Cheryl L. Rock, Vicky A. Newman, Shirley W. Flatt, Sheila Kealey, Dr. Linda Wasserman, Dr. Wayne A. Bardwell, Dr. Lisa Madlensky, and Dr. Wael Al-Delaimy); *WHEL Study clinical sites: Center for Health Research Portland, Portland, Oregon* (Dr. Njeri Karanja and Dr. Mark U. Rarick); *Kaiser Permanente Northern California, Oakland, California* (Dr. Bette J. Caan and Dr. Lou Fehrenbacher); *Stanford Prevention Research Center, Department of Medicine, School of Medicine, Stanford University, Stanford, California* (Dr. Marcia L. Stefanick and Dr. Robert Carlson); *Arizona Cancer Center, University of Arizona Health Sciences Center, Tucson, Arizona* (Dr. Cynthia Thomson, Dr. James Warneke, and Dr. Cheryl Ritenbaugh); *Department of Public Health Sciences, School of Medicine, University of California, Davis, Davis, California* (Dr. Ellen B. Gold and Dr. Sidney Scudder); *Rebecca and John Moores UCSD Cancer Center, School of Medicine, University of California, San Diego, La Jolla, California* (Dr. Kathryn A. Hollenbach and Dr. Vicky Jones); and *M. D. Anderson Cancer Center, University of Texas, Houston, Texas* (Dr. Lovell A. Jones, Dr. Richard Hajek, and Dr. Richard Theriault).

Conflict of interest: none declared.

## REFERENCES

1. Carroll RJ, Ruppert D, Stefanski LA. Measurement error in nonlinear models. In: *Monographs on Statistics and Applied Probability # 63*. Boca Raton, FL: Chapman & Hall/CRC; 1995.
2. Day N, McKeown N, Wong M, 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(2):309–317.
3. Kipnis V, Subar AF, Midthune D, et al. Structure of dietary measurement error: results of the OPEN biomarker study. *Am J Epidemiol*. 2003;158(1):14–21.
4. Freedman LS, Midthune D, Carroll RJ, et al. Adjustments to improve the estimation of usual dietary intake distributions in the population. *J Nutr*. 2004;134(7):1836–1843.
5. Natarajan L, Flatt SW, Sun X, et al. Validity and systematic error in measuring carotenoid consumption with dietary self-report instruments. *Am J Epidemiol*. 2006;163(8):770–778.
6. Espeland MA, Kumanyika S, Wilson AC, et al. Lifestyle interventions influence relative errors in self-reported diet intake of sodium and potassium. *Ann Epidemiol*. 2001;11(2):85–93.
7. Forster JL, Jeffery RW, VanNatta M, et al. Hypertension prevention trial: do 24-h food records capture usual eating behavior in a dietary change study? *Am J Clin Nutr*. 1990;51(2):253–257.



8. Neuhouser ML, Tinker L, Shaw PA, et al. Use of recovery biomarkers to calibrate nutrient consumption self-reports in the Women's Health Initiative. *Am J Epidemiol*. 2008;167(10):1247–1259.
9. Kaaks RJ. Biochemical markers as additional measurements in studies of the accuracy of dietary questionnaire measurements: conceptual issues. *Am J Clin Nutr*. 1997;65(4 suppl):1232S–1239S.
10. Subar AF, Kipnis V, Troiano RP, et al. Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN Study. *Am J Epidemiol*. 2003;158(1):1–13.
11. Daurès JP, Gerber M, Scali J, et al. Validation of a food-frequency questionnaire using multiple-day records and biochemical markers: application of the triads method. *J Epidemiol Biostat*. 2000;5(2):109–115.
12. 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(12):1126–1135.
13. 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(2):211–218.
14. Prentice RL, Shaw PA, Bingham SA, et al. Biomarker-calibrated energy and protein consumption and increased cancer risk among postmenopausal women. *Am J Epidemiol*. 2009;169(8):977–989.
15. Pierce JP, Natarajan L, Caan BJ, et al. Influence of a diet very high in vegetables, fruit, and fiber and low in fat on prognosis following treatment for breast cancer: the Women's Healthy Eating and Living (WHEL) randomized trial. *JAMA*. 2007;298(3):289–298.
16. Pierce JP, Faerber S, Wright FA, et al. A randomized trial of the effect of a plant-based dietary pattern on additional breast cancer events and survival: the Women's Healthy Eating and Living (WHEL) Study. *Control Clin Trials*. 2002;23(6):728–756.
17. Rock CL. Carotenoids: biology and treatment. *Pharmacol Ther*. 1997;75(3):185–197.
18. Kaaks R, Riboli E, Estève J. Estimating the accuracy of dietary questionnaire assessments: validation in terms of structural equation models. *Stat Med*. 1994;13(2):127–142.
19. Gamboa-Pinto AJ, Rock CL, Ferruzzi MG, et al. Cervical tissue and plasma concentrations of  $\alpha$ -carotene and  $\beta$ -carotene in women are correlated. *J Nutr*. 1998;128(11):1933–1936.
20. Food and nutrient database. Nutrition Data System for Research (NDS-R) software 31, version 4.03. Minneapolis, MN: Nutrition Coordinating Center, University of Minnesota, 2000.
21. US Department of Agriculture. *Continuous Survey of Food Intakes by Individuals (CSFII), 1994–1996, 1998*. Beltsville, MD: Beltsville Human Nutrition Research Center; 2000.
22. Richardson S, Gilks WR. Conditional independence models for epidemiological studies with covariate measurement error. *Stat Med*. 1993;12(18):1703–1722.
23. Gustafson P, Le Nhu D. Comparing the effects of continuous and discrete covariate mismeasurement, with emphasis on the dichotomization of mismeasured predictors. *Biometrics*. 2002;58(4):878–887.
24. WinBUGS software, version 1.4. Cambridge, United Kingdom: MRC Biostatistics Unit, Institute of Public Health, University Forvie, 2004. (<http://www.winbugs-development.org.uk/>). (Accessed March 26, 2010).
25. Diggle PJ, Liang KY, Zeger SL. *Analysis of Longitudinal Data*. New York, NY: Oxford Science Publications; 1996.
26. Thomson CA, Giuliano A, Rock CL, et al. Measuring dietary change in a diet intervention trial: comparing food frequency questionnaire and dietary recalls. *Am J Epidemiol*. 2003;157(8):754–762.
27. Pierce JP, Newman VA, Flatt SW, et al. Telephone counseling intervention increases intakes of micronutrient- and phytochemical-rich vegetables, fruit and fiber in breast cancer survivors. *J Nutr*. 2004;134(2):452–458.
28. Pierce JP, Newman VA, Natarajan L, et al. Telephone counseling helps maintain long-term adherence to a high-vegetable dietary pattern. *J Nutr*. 2007;137(10):2291–2296.
29. Pierce JP, Natarajan L, Sun S, et al. Increases in plasma carotenoid concentrations in response to a major dietary change in the Women's Healthy Eating and Living Study. *Cancer Epidemiol Biomarkers Prev*. 2006;15(10):1886–1892.