Published in final edited form as:

Cancer Causes Control. 2008 August ; 19(6): 595–603. doi:10.1007/s10552-008-9124-y.

Dietary Assessment in the California Teachers Study: Reproducibility and Validity

Pamela L. Horn-Ross^{1,*}, Valerie S. Lee¹, Christine N. Collins¹, Susan L. Stewart², Alison J. Canchola¹, Marion M. Lee³, Peggy Reynolds¹, Christina A. Clarke¹, Leslie Bernstein⁴, and Daniel O. Stram⁴

¹Northern California Cancer Center, Fremont, CA, USA

²Division of General Internal Medicine, Department of Medicine, University of California, San Francisco, CA, USA

³Department of Epidemiology and Biostatistics, School of Medicine, University of California, San Francisco, CA, USA

⁴Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

Abstract

Objective—To evaluate the reproducibility and validity of the food-frequency questionnaire (FFQ) used in the California Teachers Study (CTS) cohort and to use this data to quantify the effects of correcting nutrient-breast cancer relative risks for measurement error.

Methods—195 CTS cohort members participated in a 10-month dietary validation study that included four 24-hour dietary recalls and pre- and post-study FFQs. Shrout-Fleiss intraclass correlations for reproducibility were computed. Under several standard assumptions concerning the correlations of errors in the FFQs and 24-hour recalls, we calculated energy-adjusted deattenuated Pearson correlations for validity and tested for differences in validity according to a number of demographic and other risk factors. For each nutrient, we compared the performance of the FFQ versus the 24-hour recalls, estimating the number of days of recalls that give equivalent information about true intake as does a single FFQ. Finally, the effects of adjustment for measurement error on risk estimates were evaluated in 44,423 postmenopausal cohort members, 1,544 of whom developed breast cancer during seven years of follow-up. Relative risks (RR) and confidence intervals (CI) were calculated using Cox proportional hazards with and without correction for measurement error.

Results—Reproducibility correlations for the nutrients ranged from 0.60 to 0.87. With a few exceptions, validity correlations were reasonably high (range: 0.55–0.85), including $r=0.74$ for alcohol. Performance of the FFQ differed by age for percent of calories from fat and by body mass index and hormone therapy use for alcohol consumption. For most nutrients examined, our FFQ is comparable to two to six recalls for each subject in capturing true intake. In the measurement error-adjusted risk analyses, corrected RRs were within 13% of uncorrected values for all nutrients examined except for linoleic acid. For alcohol consumption the corrected RR (per 20g/1000kcal/d) was 1.36 (95% CI: 1.03–1.51) compared to the uncorrected estimate of 1.25 (95% CI: 1.10–1.42).

Conclusion—The FFQ dietary assessment used in the CTS is reproducible and valid for all nutrients except the unsaturated fatty acids. Correcting relative risk estimates for measurement error resulted in relatively small changes in the associations between the majority of nutrients and the risk of postmenopausal breast cancer.

*Correspondence: Address correspondence to Pamela L. Horn-Ross, Ph.D., Northern California Cancer Center, 2201 Walnut Ave., Suite 300, Fremont, CA 94538; USA; Phone: 510-608-5014; Fax: 510-608-5085; pam@nccc.org.

Keywords

dietary assessment; reproducibility; validity; calibration; breast cancer; alcohol

---

Introduction

The relationship between diet and disease elicits strong public and scientific interest, but it is challenging to measure and quantify usual dietary intake in epidemiologic studies. The most common tool used in these studies is the food-frequency questionnaire (FFQ). The development of any given FFQ is based on the dietary intake of a defined population during a specific period in time. Thus, when these instruments are to be used in populations outside those for which they were specifically developed, it is important to evaluate their reproducibility and validity in the new target population. Additionally, it is important to understand how measurement error can be introduced when nutrient values are assigned to each food group on a FFQ, which are necessarily averages across the foods included in the group, the effects of preparation, seasonal variation, brand differences, and other sources of variation.

While most studies assessing the validity of FFQs generally report good correlations with alternative dietary assessment methods (e.g., 24-hour recalls or diet records), several nutrients often stand out as being measured particularly well or particularly poorly [1–4]. It is important to understand this variation when interpreting the nutrient-disease relationships observed within each population, particularly when these data are being used for purposes of correcting the nutrient-disease estimates for measurement error [1,5,6].

The California Teachers Study (CTS), established in 1995–1996, is a prospective cohort study of 133,479 women who are current, former, or retired public school teachers and administrators [7]. Strengths of this cohort include its wide age range; ethnic, geographic, and socioeconomic diversity; and a comprehensive set of exposure characteristics, including usual diet, that were assessed at baseline and are updated periodically. The baseline dietary assessment used in the CTS was an early version of the Block95 FFQ [8,9]. While various versions of this FFQ have been validated in other populations [10–13], it is critical to describe the reproducibility and validity of this FFQ within the CTS, given the wide range of nutrient-disease investigations being undertaken in this cohort. We also update (with additional follow-up of the cohort) our previously reported results on the association between dietary intake and breast cancer risk among postmenopausal women in the CTS applying the validation results to correct the effect estimates for measurement error [14].

Materials and Methods

This validation study has been described in detail previously [15]. Briefly, this substudy of the CTS included a random sample of 386 CTS participants, who were 85 years old or younger at baseline and resided in the greater San Jose, California area. Forty-six (12%) of the women sampled were not contacted to participate in the substudy because they had died, moved out of the substudy area, or could not be located. The 340 women who were contacted were asked to participate in a 10-month study that included four 24-hour dietary recalls. The dietary recalls were spaced at three-month intervals beginning in early 2000. The first dietary recall was conducted in-person and the remaining three via telephone on a random day within the assigned month. Two self-administered FFQs were completed. The first, covering usual dietary intake during 1999, was left with the participant following the first dietary recall and returned to the study office by mail. At the end of the study, a second FFQ, covering usual dietary intake during 2000, was mailed to the participant. The substudy was approved by the Institutional

Review Boards of the Northern California Cancer Center and the California Department of Health Services. Written informed consent covering all phases of the study was obtained prior to the phase I interview. Additional verbal consent was obtained before each telephone recall was conducted (phases II–IV).

Of the 340 women invited to participate, 195 (57%) completed the four dietary recalls, 23 (7%) began the study but dropped out prior to its completion, 108 (32%) declined to participate, and 14 (4%) were not interviewed for other reasons (most notably, illness). Of the 195 women participating in all four dietary recalls, 186 (95%) completed the pre-study FFQ and 183 (94%) completed the post-study FFQ; 178 (91%) women completed all of these assessments.

Dietary assessment

The dietary recalls involved recording all foods, beverages, and vitamin and mineral supplements that were consumed, when they were consumed, how foods and beverages were prepared, and what was added to the foods or beverages (e.g., fats, salt, spices, and condiments) during the 24-hour period directly preceding the interview. During the in-person recall, portion size was estimated using visual aids, such as standard dinner plates, small wood cubes grouped into four portion sizes, glasses, measuring spoons and cups, several models of meat portions, and 2-dimensional images (actual size) of different portions of bananas, pizza, bread, and several other foods. After the in-person recall, interviewers oriented the participants to two-dimensional portion size pictures that were left with the participant and used during the telephone recalls. Nutritionist V software (First DataBank, Inc., San Bruno, CA) was used to assign nutrient values to all dietary recall data.

The FFQ was an early version of the 103 food-item Block95 questionnaire [8,9] with the addition of a few phytoestrogen-rich foods, for a total of 113 food and beverage items/groups. Frequency of consumption (categories from never to once/day or 5+/day depending on the item) and usual portion size (small, medium, or large relative to a given standard medium portion) were assessed as was regular use of vitamin supplements (including multi-vitamins with and without minerals and single supplements of vitamins A, β -carotene, C, E, or calcium; for C and E, dose was also assessed). Ancillary questions addressed cooking and consumption practices and preferences, the addition of fats in cooking and at the table, consumption of low-fat versions of selected foods, and overall estimates of fruit, vegetable, cereal, and milk consumption. The FFQs were self-administered and checked by study staff for completeness. FFQ data were assigned nutrient values based on an updated version of the Block95 nutrient database. These updates included incorporating more recent data obtained from Dr. Block and from the USDA (most notably for carotenoids). Vitamin D data from the USDA and manufacturer data were also added.

Statistical analysis

Chi square and Wilcoxon rank sum tests were used to evaluate differences between the substudy participants and the remainder of the cohort [16,17]. Reproducibility of the nutrient estimates derived from the pre- and post-study FFQs (often referred to as reliability under the assumption that any reported changes in intake during this period are due to reporting errors rather than true changes in eating habits) was estimated using Shrout-Fleiss intraclass correlation coefficients (r_{SF}) [18]. Energy-adjusted (based on using nutrient densities) deattenuated Pearson correlation coefficients (r_p) were used to compare the validity of the FFQ relative to the average of the 24-hour recalls [1,5].

Under the same assumptions (that the errors in each 24 hr recall are independent of each other and independent of the errors in the FFQ) that allow us to provide de-attenuated Pearson correlation coefficients, we also estimated, for each energy-adjusted nutrient the number of

days, n , of recalls that an FFQ is equivalent to in terms of its correlation with true intake. Specifically we solved for n as

$$n = r^2 / (1 - r^2_{p,d}) \text{ Var(one 24hr recall | True Intake) / Var(True Intake)} \quad (1)$$

where r^2 is the square of the corrected correlation coefficient between true intake and the FFQ value and $\text{Var}(\text{True Intake})$, and $\text{Var}(\text{one 24hr recall} | \text{True Intake})$ are both estimated as the between and within person variances in a one way random effects ANOVA model of the repeated 24-hour recalls (see Appendix). Using n days of recalls (with n computed from equation 1) to estimate nutrient intake will then give the same correlation between estimated intake and true intake as does using a single FFQ to estimate the same nutrient. In general, the more recalls the FFQ is equivalent to, the more precise the FFQ-estimated nutrient intake is considered to be.

We also evaluated whether the performance of the FFQ differed by subgroups of participants based on race/ethnicity (white, non-white), age (<50, 50–64, ≥65 years), body mass index (BMI; <25, 25–29, ≥30 kg/m²), average lifetime (to age 54) strenuous physical activity (none, <4, ≥4 hrs/wk), and use of hormone therapy (HT) before or at baseline (never, ever). In these analyses the calibration equation (estimated by regression of the average of the four 24 hour recalls against the FFQ values) included these other covariates (race/ethnicity, age, BMI, physical activity, hormone therapy) as potential modifiers of the calibration slope and intercept parameters. For each nutrient we report the results for those covariates significantly ($p < 0.05$) acting as modifiers of the calibration slope as judged by a test of interaction. In general we interpreted higher slopes as being more favorable since, in those groups (defined by the covariates, race/ethnicity, age, BMI, physical activity, HT) with significantly higher calibration slopes, more of the variance of the 24 hour recalls would tend to be explained by a given change in FFQ value.

Finally, the validation results were used to correct for measurement error when evaluating the association between dietary exposures and disease risk. We previously reported on the association between dietary intake and breast cancer risk in the CTS [14]. Here we update those findings for postmenopausal women (based on additional follow-up of the cohort) and apply the calibration equation to correct the estimates for measurement error.

Included in our main regression model were 44,423 cohort members who were postmenopausal at baseline and had complete data on dietary and all confounder variables. Follow-up was complete through December 31, 2002, during which time 1,544 women developed invasive breast cancer. Cox proportional hazards regression, with age as the time scale, was used to obtain the hazard ratios (relative risks) and 95% confidence intervals for breast cancer associated with a change equivalent to approximately the interquartile range observed for each nutrient of interest, using separate analyses for each nutrient. Each analysis of nutrient density (i.e., nutrient/calories) was adjusted for race/ethnicity/birthplace (white, nonwhite and born in North America, nonwhite and born elsewhere), average lifetime (until age 54) strenuous physical activity (0, <4, ≥4 hours per week), family history of breast cancer in a first degree relative (yes, no), a history of biopsy-diagnosed benign breast disease (yes, no), age at menarche (≤12, ≥13 years), nulliparity/age at first full-term pregnancy (nulliparous, ≤24, 25–29, ≥30 years), a variable representing the joint effects of HT use and BMI (kg/m²) (no HT and BMI <25.8, no HT and BMI ≥25.8, estrogen alone ≤5 years, estrogen alone >5 years, estrogen and progestin ≤5 years, and estrogen and progestin >5 years; women using estrogen alone followed by estrogen and progestin were included in the latter two groups based on the total time HT was used), and average alcohol consumption in the year prior to baseline (none, <20 g/d, ≥20 g/d).

The validation regression used in the calibration analyses included a subsample of 155 women from the validation study who had participated in all four 24-hr dietary recalls, had completed the post-study FFQ, and had complete confounder data (as described above) from the baseline survey. We estimated a calibration equation for each nutrient, which regressed the average of the four 24 hour recalls against the FFQ values, adjusting for all confounders. The slope associated with the FFQ term was considered the validation coefficient.

The corrected hazard ratio (relative risk) for each nutrient was calculated by dividing the parameter estimate by the validation coefficient and then exponentiating this ratio. The 95% confidence intervals were estimated using the method described by Stram et al.¹ This approach is known to correspond to the correction method of Rosner et al. [19] in that it gives the same estimates and nearly the same standard errors of those estimates for small to moderate sized risk estimates. In particular, by our procedure standard errors increase proportionately to the increase in risk estimates so that the significance of a test of the null hypothesis that a nutrient variable is unrelated to risk is not changed by doing the regression calibration.

Results

To evaluate the representativeness of our validation study sample, we compared distributions of selected baseline characteristics for the 195 substudy participants to the remainder of the CTS cohort (age ≤ 85 years) (Table 1; see also ref 15). Substudy participants were similar to the entire cohort in terms of age, body mass index, physical activity, and broad dietary composition (i.e., average daily caloric intake and the contribution of fat and fruits and vegetables to the diet). Menopausal status at baseline was also comparable for substudy participants compared to the remainder of the cohort (50% vs. 53% being postmenopausal; $p=0.61$) [15]. However, substudy participants were more likely than the remainder of the cohort to be Latina (8% vs. 4%) and less likely to be African American (1% vs. 3%) or white (84% vs. 88%), reflecting the racial/ethnic composition of the substudy area compared to the state as a whole.

Table 2 presents the variation in average daily intake of total calories, macronutrients, fiber, selected micronutrients, and alcohol as measured by the pre- and post-study FFQs and the average of the four 24-hour recalls. Usual intakes during the previous year, as measured by FFQ at the beginning and at the end of the 10-month study period, were generally constant, and correlations reflecting the FFQ's reproducibility for these nutrients were generally high (ranging from $r_{SF}=0.60$ for oleic acid to $r_{SF}=0.87$ for alcohol). The average of all reproducibility correlations was 0.71. Correlations quantifying the validity of the FFQ assessment compared to the average of the four 24-hour recalls were low for caloric intake, linoleic acid, and oleic acid ($r_p=0.14$ to 0.44) but moderate to high for all other nutrients (ranging from 0.57 for protein and vitamin A to 0.85 for β -carotene). The average of all validity correlations was 0.63; excluding caloric and oleic and linoleic acid intake, the average correlation was 0.71. Correlations between the recalls and both the pre-study and post-study FFQs were generally similar for all nutrients.

Multiple 24-hour dietary recalls can be used as a gold standard of dietary assessment, under the assumption that each gives an independent and unbiased estimate of true nutrient exposure. For individual nutrients, the number of recalls needed to achieve the same degree of measurement precision represents a useful metric for evaluating a FFQ; the higher the number of recalls a FFQ represents, the greater its precision for estimating the intake of a given nutrient. For most nutrients, estimates obtained from our FFQ corresponded to approximately two to six recalls (Table 2). Dietary fiber, saturated fat, vitamin A and vitamin C were measured particularly well by our FFQ as they were comparable to more than four recalls. As reflected

by the low validation coefficients, our FFQ captured caloric and linoleic and oleic acid intake less well. For each, the FFQ assessment was worth less than one recall.

To understand whether the performance of the FFQ differed by selected characteristics of the participants (i.e., race/ethnicity, age, BMI, physical activity, and HT use), we examined the slope of the calibration equation for each dietary factor (Table 3). There was a significant relationship between age and the slope for percent of calories from fat, however, the trend was not monotonic. For white women with a BMI<25 who did not exercise or use HT, the slopes for ages <50, 50–64, and ≥65 years were 0.97 ($= 0.78 + 0.19$, i.e., as shown in Table 3, the baseline slope estimate + the estimate for women age <50), 0.36, and 0.78, respectively, evidence that the FFQ captured intake less well for women age 50–64 compared to either younger or older women. Similarly, the slope for β -carotene differed by participant age; BMI modified performance of the FFQ for fiber, vitamins C and D, and alcohol; physical activity influenced the estimates for saturated fat (data not shown), vitamin D, and folate; and the slopes for fiber and alcohol differed by HT use.

Table 4 presents the uncorrected and corrected relative risks for the association between dietary intake and postmenopausal breast cancer risk in the study cohort. For all nutrients except linoleic acid, the corrected risk estimates were within 13% of the uncorrected estimates. The only dietary factor significantly associated with breast cancer risk was alcohol (uncorrected RR=1.25 per 20g/1000kcal/d, 95% CI: 1.10–1.42). Correcting for measurement error provided a relative risk of 1.36 per 20g/1000kcal/d (95% CI: 1.03–1.51) for this association. When simultaneous adjustment for measurement error in caloric intake, alcohol consumption, and each nutrient of interest was made using the methods described by Rosner et al.[19] and Spiegelman et al.[6] corrected relative risks of similar magnitude were observed (data not shown).

Discussion

These data suggest that the FFQ assessment and its associated nutrient database used in the CTS produce reproducible and valid estimates of nutrient intake. Saturated fat, fiber, β -carotene, vitamins C and E, and alcohol were captured particularly well when expressed as nutrient densities while neither the unsaturated fatty acids nor absolute caloric intake were well measured.

Repeated self-administration of our FFQ 10-months apart resulted in correlations ranging from 0.60 to 0.87 for most nutrients, with an average correlation of 0.71. This is similar to the average correlation observed over a 6-month period in the Iowa Women's Health Study (IWHS) cohort and higher than reproducibility in the IWHS and the Nurse's Health Study (NHS) over periods of two and four years, respectively [2] [4]. Validity correlations, r_p , based on comparing estimates of intake from our FFQ to those from the alternative dietary assessment method (i.e., the average of four 24-hour dietary recalls over a 10-month period), ranged from 0.55 to 0.85 for most of the nutrients examined. These estimates are consistent with validity estimates for FFQs used in other large cohort studies [1,2,4,20]. In all these studies, as well as ours, validity estimates for absolute caloric intake are relatively poor while the relative composition of the diet can be generally well estimated using nutrient densities. This property of FFQs was elegantly demonstrated by the Observing Protein and Energy Nutrition (OPEN) study which performed an extraordinarily careful characterization of total energy, protein, and protein density using biomarkers and found that FFQ correlations were poor for absolute measure of protein and energy, but substantially better for protein density [21,22].

We have presented correlations between our dietary recalls and two FFQs. The timing of the two FFQs have complementary strengths and limitations [5]. The pre-study FFQ was not likely

to be influenced by the increased awareness of diet stemming from study participation and refers to an earlier time period. Responses to the post-study FFQ may have been somewhat biased by the awareness resulting from study participation but the time period covered by this FFQ and the dietary recalls is the same. However, the similarity between the two sets of validity correlations (i.e., the FFQs with the recall data) and the reproducibility coefficients (i.e., comparing estimates from the two FFQs) suggest only minimal biases in this regard.

Our validity coefficients reflect not only the correspondence between the two assessment measures but also between the two nutrient databases used. Caloric and nutrient intake from the 24-hour recalls was based on the brand- and preparation-specific data available in Nutritionist V database while the corresponding measures for the FFQs were based on averaged data for foods or food groups. Database differences may have accounted for the higher median β -carotene estimates from the FFQ compared to the dietary recalls and for the relatively poor validity correlations observed for the unsaturated fatty acids. In addition, foods high in these nutrients may be poorly measured by the FFQ or infrequently consumed resulting in inaccurate estimation from only four 24-hour recalls. However, the relatively high correlations for the majority of the nutrients examined suggest that our FFQ database is accurate and robust for measuring nutrient intake.

While the number of dietary recalls we chose to administer as part of this study was based on a balance between statistical precision and practicality, there is always concern that a relatively small number of recalls can affect the accuracy of estimated intake, especially when intake of the nutrients of interest varies substantially within individuals [5]. To minimize possible misclassification related to the number of recalls, we did not tell participants about the specific nature of the recalls until the first in-person meeting and then conducted the second, third, and fourth dietary recalls on random days (including weekends and holidays) within the specified month. In addition, we used statistical techniques to adjust for the number of recalls which were conducted and to account for the resulting within- and between-person variation [1].

Unlike prior studies [1,23,24], we found no racial/ethnic differences in the performance of our FFQ. This may relate to the highly-educated nature of our cohort where the vast majority of teachers hold a bachelor's degree or higher. However, we did observe differences for several nutrients by age, BMI, physical activity level, or HT use. For percent of calories from fat and vitamin C we can compare our results directly with those found in the MEC cohort [1]. Among women participating in the MEC cohort young age was associated with better reporting of vitamin C, but not fat, on the FFQ; in our cohort the opposite was observed, i.e., young women reported fat intake better than older women whereas there were no age differences in reporting vitamin C intake. BMI did not affect the FFQ performance in MEC for either of these nutrients but in our cohort, the FFQ characterized vitamin C intake better in thinner women (BMI <25 kg/m²) than those who were overweight (BMI $\geq$ 30 kg/m²).

A major reason to conduct validation/calibration studies is to obtain the data needed for correction of measurement error in exposure-disease association estimates. We examined both uncorrected and corrected estimates for postmenopausal breast cancer risk, finding that the correction usually resulted in generally similar point estimates and, as expected, wider confidence intervals.

In summary, our dietary assessment appears to be both reproducible and valid with the observed correlations being similar to those reported by other cohort studies. Correcting relative risk estimates of diet-breast cancer associations for measurement error using the data from this validation/calibration study resulted in relatively small changes (<13%) for most nutrients. Our analyses also confirmed the high measurement quality underlying the observation of increased

breast cancer risk associated with greater alcohol consumption, which is the sole dietary factor associated with overall breast cancer risk in this study.

Acknowledgments

The authors would like to thank Dr. Esther John for her work in compiling the vitamin D nutrient data and the members of the CTS Steering Committee, who are responsible for the formation and maintenance of the cohort within which this study was conducted, but are not authors on the current paper: Hoda Anton-Culver, Rosemary Cress, Dennis Deapen, David Peel, Rich Pinder, Ronald K. Ross (deceased), Dee W. West, William E. Wright, Giske Ursin, and Argyrios Ziogas.

Funding Acknowledgement: This research was supported by grant R01 CA77398 from the National Cancer Institute and contract 97–10500 from the California Breast Cancer Research Fund. The funding sources did not contribute to the design or conduct of the study, nor to the writing or submission of this manuscript. The collection of cancer incidence data used in this study was supported by the California Department of Health Services as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885; the National Cancer Institute's Surveillance, Epidemiology and End Results Program under contract N01-PC-35136 awarded to the Northern California Cancer Center, contract N01-PC-35139 awarded to the University of Southern California, and contract N02-PC-15105 awarded to the Public Health Institute; and the Centers for Disease Control and Prevention's National Program of Cancer Registries, under agreement #U55/CCR921930-02 awarded to the Public Health Institute. The ideas and opinions expressed herein are those of the authors and endorsement by the State of California, Department of Health Services, the National Cancer Institute, and the Centers for Disease Control and Prevention or their contractors and subcontractors is not intended nor should be inferred.

References

1. Stram DO, Hankin JH, Wilkens LR, et al. Calibration of the dietary questionnaire for a multiethnic cohort in Hawaii and Los Angeles. *Am J Epidemiol* 2000;151:358–370. [PubMed: 10695594]
2. Munger RG, Folsom AR, Kushi LH, Kaye SA, Sellers TA. Dietary assessment of older Iowa women with a food frequency questionnaire: nutrient intake, reproducibility, and comparison with 24-hour dietary recall interviews. *Am J Epidemiol* 1992;136:192–200. [PubMed: 1415141]
3. Smith-Warner S, Spiegelman D, Ritz J, et al. Methods for pooling of results of epidemiologic studies: the pooling project of prospective studies of diet and cancer. *Am J Epidemiol* 2006;163:1053–1064. [PubMed: 16624970]
4. Willett WC, Sampson L, Browne ML, et al. The use of a self-administered questionnaire to assess diet four years in the past. *Am J Epidemiol* 1988;127:188–199. [PubMed: 3337073]
5. Willett, WC. *Nutritional Epidemiology*. New York: Oxford University Press; 1998.
6. Spiegelman D, McDermott A, Rosner B. Regression calibration method for correcting measurement-error bias in nutritional epidemiology. *Am J Clin Nutr* 1997;65:1179S–1186S. [PubMed: 9094918]
7. Bernstein L, Anton-Culver H, Deapen D, et al. High breast cancer rates among California teachers: Results from the California Teachers Study Cohort. *Cancer Causes Control* 2002;13:625–635. [PubMed: 12296510]
8. Block G, Hartman AM, Dresser CM, Carroll MD, Gannon J, Gardner L. A data-based approach to diet questionnaire design and testing. *Am J Epidemiol* 1986;124:453–469. [PubMed: 3740045]
9. Block G, Woods M, Potosky A, Clifford C. Validation of a self-administered diet history questionnaire using multiple diet records. *J Clin Epidemiol* 1990;43:1327–1335. [PubMed: 2254769]
10. Block G, Thompson FE, Hartman AM, Larkin FA, Guire KE. Comparison of two dietary questionnaires validated against multiple dietary records collected during a 1-year period. *J Am Diet Assoc* 1992;92:686–693. [PubMed: 1607564]
11. Kumanyika SK, Mauger D, Mitchell DC, Phillips B, Smiciklas-Wright H, Palmer JR. Relative validity of food frequency questionnaire nutrient estimates in the Black Women's Health Study. *Ann Epidemiol* 2003;13:111–118. [PubMed: 12559670]
12. Potischman N, Carroll RJ, Iturria SJ, et al. Comparison of the 60- and 100-item NCI-block questionnaires with validation data. *Nutr Cancer* 1999;34:70–75. [PubMed: 10453444]
13. Subar AF, Thompson FE, Kipnis V, et al. Comparative validation of the Block, Willett, and National Cancer Institute food frequency questionnaires: the Eating at America's Table Study. *Am J Epidemiol* 2001;154:1089–1099. [PubMed: 11744511]

14. Horn-Ross PL, Hoggatt KJ, West DW, et al. Recent diet and breast cancer risk: the California Teachers Study (USA). *Cancer Causes Control* 2002;13:407–415. [PubMed: 12146845]
15. Horn-Ross PL, Barnes S, Lee VS, et al. Reliability and validity of an assessment of usual phytoestrogen consumption (United States). *Cancer Causes Control* 2006;17:85–93. [PubMed: 16411057]
16. Kelsey J, Thompson W, Evans A. *Methods in observational epidemiology*. 1986
17. Breslow, NE.; Day, NE. *The Analysis of Case-Control Studies*. Vol. Volume I. Lyon: International Agency for Research on Cancer; 1980. *Statistical Methods in Cancer Research*.
18. Shrout P, Fleiss J. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86:420–428. [PubMed: 18839484]
19. Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. *Am J Epidemiol* 1990;132:734–745. [PubMed: 2403114]
20. Kroke A, Klipstein-Grobusch K, Voss S, et al. Validation of a self-administered food-frequency questionnaire administered in the European Prospective Investigation into Cancer and Nutrition (EPIC) Study: comparison of energy, protein, and macronutrient intakes estimated with the doubly labeled water, urinary nitrogen, and repeated 24-h dietary recall methods. *Am J Clin Nutr* 1999;70:439–447. [PubMed: 10500011]
21. 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. [PubMed: 12835281]
22. Schatzkin A, Kipnis V, Carroll RJ, et al. A comparison of a food frequency questionnaire with a 24-hour recall for use in an epidemiological cohort study: results from the biomarker-based Observing Protein and Energy Nutrition (OPEN) study. *Int J Epidemiol* 2003;32:1054–1062. [PubMed: 14681273]
23. Coates R, Monteilh C. Assessments of food-frequency questionnaires in minority populations. *Am J Clin Nutr* 1997;65:S1108–S1115.
24. Kristal AR, Feng Z, Coates RJ, Oberman A, George V. Associations of race/ethnicity, education, and dietary intervention with the validity and reliability of a food frequency questionnaire. The Women's Health Trial Feasibility Study in Minority Populations. *Am J Epidemiol* 1997;146:856–869. [PubMed: 9384206]

Appendix

Appendix

Estimation of the number of days, n, of 24-hour recalls a FFQ report is worth

We define n by equating the correlation, r, between the true nutrient intake and the FFQ with the correlation between true nutrient intake and n 24-hour recalls (which increases as n increases). Note that this is equivalent to equating $\text{Var}(\text{True Intake} \mid \text{FFQ value})$ to $\text{Var}(\text{True Intake} \mid n \text{ 24-hour recalls})$ since $1-r^2$ is equal to $\text{Var}(\text{True} \mid \text{Estimated}) / \text{Var}(\text{True})$, i.e., it is the “unexplained” portion of the variance of True not captured by the estimate. If we assume that all variables are distributed normally then we can compute both $\text{Var}(\text{True Intake} \mid \text{FFQ value})$ as well as $\text{Var}(\text{True Intake} \mid n \text{ 24-hour recalls})$ by a standard variance components analysis.

Let X_i be true intake of the nutrient, then by fitting a random intercept model (as in SAS proc mixed) to the 4 recalls ($R_{ij}, j=1,4, i=1,N$) per subject, then R_{ij} is estimated by

$$R_{ij} = X_i + \varepsilon_{ij}$$

where ε_{ij} is the random error in our measurement. We can estimate $\text{Var}(X)$ as the variance of the random intercept and $\text{Var}(\varepsilon)$ as the residual variance (i.e. the variance of one 24 hr recall

given true intake). By an elementary application of conditional probability for normal random variables, and defining $\bar{R}_i = R_{ij}/n$, we have

$$\begin{aligned}\text{Var}(X|\bar{R}_i) &= \text{Var}(X) - \text{Cov}(X, \bar{R}_i)^2 / \text{Var}(\bar{R}_i) \\ &= \text{Var}(X) - \text{Var}(X)^2 / (\text{Var}(X) + \text{Var}(\varepsilon)/n)\end{aligned}$$

Note also that by fitting a second random intercept model

$$R_{ij} = a + b\text{FFQ}_i + W_i + \varepsilon_{ij}$$

and equating $X_i = a + b\text{FFQ}_i + W_i$ (so that the random intercept W is the portion of true X unexplained by the regression on the FFQ value), we can estimate $\text{Var}(X|\text{FFQ}) \equiv \text{Var}(W)$ as the variance of the random intercept term. Thus to estimate the number of days of 24-hr recalls that one FFQ is “worth” we equate $\text{Var}(W) = \text{Var}(X|\bar{R}_i) = \text{Var}(X) - \text{Var}(X)^2 / [\text{Var}(X) + \text{Var}(\varepsilon)/n]$ and solve for n as

$$n = \text{Var}(\varepsilon) (\text{Var}(X) - \text{Var}(W)) / (\text{Var}(X) * \text{Var}(W))$$

Since the square of the corrected correlation between FFQ and the 24-hour recalls can be written as

$$r^2 = (\text{Var}(X) - \text{Var}(W)) / \text{Var}(X), \text{ so that } \text{Var}(W) = (1 - r^2) \text{Var}(X)$$

(this follows since $X - W$ is the portion of X that is “explained” by the FFQ and r^2 is always the ratio of the variance of the “explained” portion of X to the total variance of X) we have

$$n = r^2 \text{Var}(\varepsilon) / (1 - r^2) \text{Var}(X)$$

which is identical to equation (1).

Selected baseline characteristics of participants in the validation substudy and the California Teachers Study cohort.

Characteristic	Median (Interquartile range)			p-value
	Substudy participants (n=195)	CTS cohort ^d (n=131,045)		
Age	51 (42–61)	52 (44–64)		0.15
Body mass index (kg/m ²)	23.4 (21.3–26.7)	23.7 (21.3–27.3)		0.54
Physical activity (hours/week) ^b	1.5 (0.3–2.9)	1.3 (0.3–3.0)		0.90
Caloric intake (kcal/day)	1,525 (1,194–1,824)	1,491 (1,175–1,864)		0.94
Percent of daily calories from fat	33.0 (28.1–37.8)	33.6 (27.9–38.9)		0.26
Fruits & vegetables (servings/day)	2.3 (1.5–3.5)	2.1 (1.3–3.5)		0.15

^aExcluding substudy participants and women over age 85.
^bAverage lifetime (to age 54 years) strenuous exercise activity.

Table 2

Reproducibility, validity and precision of the food frequency questionnaire (FFQ) dietary assessment used in the California Teachers Study.

Nutrient	Median intake				Correlations ^a		
	Pre-study FFQ (n=186)	Post-study FFQ (n=183)	24-hr recalls (n=195)	Pre- and post-study FFQs (n=178)	Recalls w/ pre-study FFQ (n=186)	Recalls w/ post-study FFQ (n=183)	Number of recalls represented by the FFQ
Calories (kcal/d)	1,526	1,532	1,713	0.71	0.33	0.44	0.5
Macronutrients (g/d)							
Total Fat	56.3	55.0	58.4	0.67	0.61	0.62	1.9
Saturated fat	17.8	17.6	18.5	0.71	0.78	0.78	4.3
Linoleic acid	10.4	10.0	5.8	0.63	0.19	0.14	0.5
Oleic acid	20.6	19.8	11.1	0.60	0.28	0.23	0.9
Carbohydrates	199.4	173.0	220.9	0.69	0.70	0.67	1.9
Protein	58.5	59.8	65.8	0.65	0.61	0.57	1.8
Fiber (g/d)	14.1	13.8	17.7	0.74	0.79	0.79	5.8
Micronutrients							
Vitamin A (IU/d) ^b	8,636	8,479	9,370	0.73	0.55	0.57	5.7
β-carotene (μg/d) ^b	2,733	2,652	582	0.71	0.78	0.85	2.0
Vitamin C (mg/d) ^b	174	192	217	0.69	0.70	0.77	4.9
Vitamin D (μg/d) ^c	375	425	222	0.77	0.64	0.78	1.9
Vitamin E (mg/d) ^b	28.8	29.4	20.6	0.63	0.75	0.74	1.1
Folate (μg/d) ^c	323	307	412	0.74	0.45	0.58	1.3
Calcium (mg/d) ^b	971	985	1,095	0.74	0.67	0.77	1.8
Alcohol (g/d)	0.9	1.6	1.4	0.87	0.76	0.74	1.6

^a Coefficients comparing pre- and post-study FFQs are based on Shout-Fleiss intraclass correlations and coefficients comparing the recalls with the FFQs are calorie-adjusted (using the residual method) deattenuated Pearson correlations.

^b Including multi-vitamin and single vitamin supplements.

^c Including multi-vitamin supplements only.

Table 3

Expanded calibration model for dietary components by participant demographic and lifestyle factors; the estimates represent the amount of the change in baseline slope (and the standard error).

Nutrient	Baseline slope ^a	Non-white	Age			BMI	Physical activity (hrs/wk)		Ever used HT
			<50	50-64	25-29		<4	≥4	
Total fat (% kcal)	0.78	-0.18 (0.35)	0.19 (0.27)	-0.42 ^b (0.19)	-0.05 (0.20)	-0.39 (0.28)	-0.11 (0.20)	-0.56 (0.29)	0.03 (0.19)
Carbohydrates (% kcal)	0.88	0.05 (0.36)	0.03 (0.33)	-0.36 (0.24)	0.12 (0.24)	-0.21 (0.32)	-0.24 (0.25)	-0.71 (0.37)	0.08 (0.22)
Protein (% kcal)	0.29	0.35 (0.25)	0.11 (0.46)	-0.16 (0.19)	-0.07 (0.21)	-0.37 (0.20)	0.19 (0.24)	-0.14 (0.31)	0.23 (0.20)
Fiber	0.62	0.07 (0.22)	-0.09 (0.23)	-0.31 (0.21)	0.22 (0.18)	-0.44 ^b (0.21)	0.19 (0.24)	0.21 (0.31)	0.32 ^c (0.18)
B-carotene	0.67	0.00 (0.07)	-0.44 (0.06)	-0.35 ^b (0.07)	0.04 (0.05)	-0.04 (0.08)	-0.15 (0.09)	-0.27 (0.19)	-0.11 (0.07)
Vitamin C	0.39	-0.34 (0.30)	0.10 (0.21)	-0.05 (0.11)	-0.15 (0.13)	-0.33 ^c (0.14)	0.24 (0.32)	0.50 (0.23)	-0.11 (0.12)
Vitamin D	0.19	0.10 (0.22)	0.01 (0.22)	-0.05 (0.11)	0.23 (0.12)	0.31 ^c (0.15)	-0.05 (0.14)	0.80 ^b (0.28)	-0.07 (0.12)
Vitamin E	0.36	-0.28 (0.31)	0.16 (0.16)	-0.03 (0.13)	0.06 (0.13)	-0.13 (0.17)	0.07 (0.16)	0.47 (0.32)	0.15 (0.14)
Folate	0.20	0.20 (0.22)	-0.04 (0.22)	0.10 (0.13)	-0.12 (0.13)	0.15 (0.17)	-0.01 (0.17)	0.82 ^b (0.25)	0.08 (0.13)
Calcium	0.69	0.51 (0.40)	0.44 (0.43)	-0.33 (0.19)	0.42 (0.21)	0.21 (0.25)	-0.03 (0.20)	-0.45 (0.34)	-0.14 (0.18)
Alcohol	0.88	0.68 (0.48)	-0.12 (0.31)	0.04 (0.16)	0.40 (0.17)	-0.38 ^b (0.30)	0.17 (0.17)	0.14 (0.57)	-0.67 ^b (0.20)

^aThe baseline slope represents the slope for a white woman, age ≥65, with a BMI of <25, who is physically inactive and never used HT (as of baseline).

^bGlobal p-value for the test of homogeneity of the slope estimates across all levels of a given subgroup ≤0.05.

^c p-value >0.05 and ≤0.10.

Table 4

Uncorrected and corrected^a relative risks for the relationship between specific nutrients^b and postmenopausal breast cancer risk in the California Teachers Study cohort.

Nutrient (per 1000kcal/d)	Uncorrected		Validation coefficient	Corrected	
	RR	95% CI		RR	95% CI
Calories	1.02	0.92–1.13	0.35	1.06	0.76–1.36
Macronutrients					
Total Fat (per 10g)	1.00	0.95–1.06	0.44	1.01	0.88–1.14
Saturated fat (per 5g)	1.05	0.97–1.13	0.56	1.09	0.92–1.20
Linoleic acid (per 5g)	0.95	0.85–1.05	0.06	0.41	0.04–22.7
Oleic acid (per 5g)	0.98	0.91–1.05	0.13	0.87	0.51–1.89
Carbohydrates (per 30g)	0.96	0.89–1.04	0.50	0.92	0.82–1.12
Protein (per 10g)	1.04	0.97–1.11	0.42	1.10	0.88–1.23
Fiber (per 5 g)	0.96	0.89–1.04	0.72	0.94	0.86–1.07
Micronutrients					
Vitamin A (per 5000 IU)	0.98	0.90–1.06	0.68	0.97	0.87–1.10
β-carotene (per 2000 μg)	0.99	0.97–1.01	0.38	0.97	0.93–1.06
Vitamin C (per 400 mg)	1.00	0.95–1.04	0.43	0.99	0.90–1.10
Vitamin D (per 300 IU)	0.98	0.90–1.06	0.25	0.91	0.69–1.39
Vitamin E (per 150 α-TE)	0.99	0.94–1.04	0.41	0.97	0.87–1.12
Folate (per 300 μg)	0.95	0.87–1.04	0.32	0.85	0.71–1.27
Calcium (per 300 mg)	1.00	0.94–1.07	0.67	1.00	0.91–1.10
Alcohol (per 20 g)	1.25	1.10–1.42	0.72	1.36	1.03–1.51

^aCorrected for estimated measurement error in the FFQ estimates of nutrient intake.

^bNutrient densities per 1000 kcal adjusted for race/ethnicity/birthplace, physical activity, family history of breast cancer, biopsy-diagnosed benign breast disease, age at menarche, nulliparity/age at first full-term pregnancy, the joint effect of hormone therapy use and body mass index, and as appropriate, alcohol consumption; age as the time-scale.