Published in final edited form as:
*Cancer J*. 2008 ; 14(2): 69–74.

# Methods of Epidemiology: Evaluating the Fat–Breast Cancer Hypothesis – Comparing Dietary Instruments and Other Developments

**Laurence S. Freedman, PhD**[*], **Victor Kipnis, PhD**[†], **Arthur Schatzkin, PhD**[‡], and **Nancy Potischman, PhD**[§]

[*]*Gertner Institute for Epidemiology, Tel Hashomer, Israel*

[†]*Division of Cancer Prevention, National Cancer Institute, Bethesda, MD*

[‡]*Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD*

[§]*Division of Cancer Control and Population Studies, National Cancer Institute, Bethesda, MD*

## Abstract

Results from several large cohort studies that were reported 10 to 20 years ago seemed to indicate that the hypothesized link between dietary fat intake and breast cancer risk was illusory. In this article, we review several strands of more recent evidence that have emerged. These include two studies comparing the performance of dietary instruments used to investigate the dietary fat-breast cancer hypothesis, a large randomized disease prevention trial, a more recent meta-analysis of nutritional cohort studies, and a very large nutritional cohort study. Each of the studies discussed in this article suggests that a modest but real association between fat intake and breast cancer is likely. If the association is causative, it would have important implications for public health strategies in reducing breast cancer incidence. The evidence is not yet conclusive, but additional follow-up in the randomized trial, as well as efforts to improve dietary assessment methodology for cohort studies, may be sufficient to provide a convincing answer.

### Keywords

breast cancer; dietary fat; dietary measurement error; food frequency questionnaire; multiple-day food record

Although a causal has relationship between dietary fat and breast cancer has some biologic plausibility,[1] this relationship has been a topic of controversy for more than 20 years, with positive associations seen in animal studies, international comparisons, and case-control studies,[2] but no association seen in a pooled analysis of several cohort studies, which are free from some of the biases that potentially affect case-control studies.[3] Some years ago, Willett[4] reviewed the evidence for a causal relationship between fat intake and breast cancer risk and concluded “As the findings from large prospective studies have become available, however, support for this relationship has greatly weakened.” In this article we will review several more recently published research results that pertain to this possible relationship.

Reprints: Professor Laurence Freedman, Director, Biostatistics Unit, Gertner Institute for Epidemiology and Health Policy Research, Sheba Medical Center, Tel Hashomer 52161, Israel. E-mail: lsf@actcom.co.il.

A major focus will be the results of 2 studies comparing the performance of dietary instruments used to investigate the dietary fat–breast cancer hypothesis.[5,6] In addition, we will review results of a large randomized disease prevention trial,[7] a more recent meta-analysis of nutritional cohort studies,[8] and the recent results of a very large nutritional cohort study.[9]

## DIETARY ASSESSMENT AND MEASUREMENT ERROR

A major problem besetting studies relating a woman's fat consumption to her risk of breast cancer is dietary measurement error,[10] which, in an observational study, could bias the estimated relationship toward the null and substantially reduce the power to detect the relationship. Case-control and cohort studies use self-reporting techniques for measuring dietary intake. Both study designs reduce (but do not eliminate) the impact of confounders, but the cohort is clearly stronger than the case-control design, effectively eliminating bias arising from differential recall of diet between breast cancer patients and control subjects. However, the cohort design rests heavily on the quality of the dietary reporting by participants, which is strongly influenced by the assessment method used. For reasons of logistics and cost, the reporting instrument most commonly used has been the food frequency questionnaire (FFQ).[11] Little is known about the magnitude and nature of errors in reporting fat intake through an FFQ, and there has been much discussion about whether such errors could have led to the failure of the cohort studies to find a fat–breast cancer association.[12,13]

In 2003, Bingham et al[5] reported the results of a comparison of 2 instruments, a FFQ and a quantitative 7-day diary, both completed by a cohort of 13,070 women living in the United Kingdom, on which the fat-breast cancer association hypothesis was tested. They found a statistically significant positive association (relative risk [RR] 1st vs 5th quintile = 1.79, $P_{trend} = 0.05$) between total fat intake and breast cancer incidence using the 7-day diary, but not using the FFQ (RR 1st vs 5th quintile = 1.31, $P_{trend} = 0.52$). A similar result was found for saturated fat intake. These results suggested that the 7-day diary, being associated with less error than the FFQ for nutrients studied,[14] may be more powerful than the FFQ in detecting this diet-disease relationship. However, the study was based on a relatively small number of breast cancer patients (168).[15]

The results of this study were potentially important because they suggested that the failure of previous nutritional cohort studies to find an association between dietary fat intake and breast cancer may have been due not to the lack of such an association but to the use of an insufficiently precise dietary assessment. This realization motivated conducting a second study to compare the performance of the FFQ relative to a more detailed assessment, in this case a 4-day food record.

In 2006, the results of this second study were reported.[6] The study was conducted in a larger cohort than that in Bingham et al's study, comprising the 29,294 women in the control group of the Dietary Modification (DM) arm of the Women's Health Initiative randomized trial.[16,17] To explain some important details of the statistical analysis performed for this study, it is necessary first to describe how women were recruited into this DM trial.

Before the first screening visit, women received several self-administered questionnaires to complete, including an FFQ. At the first visit, the FFQ as well as other questionnaires were reviewed. In particular, those women whose FFQ report showed that less than 32% of their energy intake derived from fat were not considered eligible for the trial. The intention of this requirement was that a group having a relatively high fat intake would be enrolled, thereby increasing the difference in percent energy from fat between women randomly assigned to the dietary intervention and control groups. This screening raised the mean percent energy from fat intake from approximately 32% to approximately 35%. Approximately 42% of those who expressed interest in the DM trial were excluded because of the fat intake criterion.

After this and other exclusion criteria were applied, 48,835 participants entered the DM trial. These women were randomly assigned: 40% (n = 19,541) to a low-fat eating pattern intervention and 60% (n = 29 294) to a control group who were not asked to make dietary changes. Participants included in the study analyses were in the control arm. Staff at the Women's Health Initiative Clinical Coordinating Center frequency matched 2 control subjects for every case patient with invasive breast cancer, with matching on age (50–59, 60–69, and 70–79 years), clinic, and length of follow-up (±12 months), resulting in a sample of 603 case patients and 1,206 control subjects, that together comprise a case-control design nested within the prospective cohort. The median length of follow-up in the study was 83 months at the time of choosing the matched control subjects.

During the second screening visit, participants were given a 4-day food record (FR) booklet[14] and were instructed on how to complete the records, which they did between the second and third screening visits. Before completing the record, they were given instruction through a videotape and a personal training session. After returning the booklet, participants were interviewed for missing details by clinical staff who had reviewed the 4 days of recorded food items consumed. All of the FR booklets were archived at the Clinical Centers and mailed to the Clinical Coordinating Center in 2004 where they were coded to derive daily nutrient intakes.

Unconditional logistic regression adjusted for matching variables was used in the main analyses relating dietary fat intake to breast cancer. The following confounding variables were adjusted for, as they were statistically significantly associated with breast cancer at the conventional 5% level and changed the risk estimates for total fat in the FR or FFQ models by 10% or more: postmenopausal hormone use (current/former or never), family history in a first degree relative (yes/no), and biopsy for benign breast disease (yes/no).

Missing values were present in family history (88 participants) and breast biopsy (201 participants). Of the 201 with missing breast biopsy information, 194 had entered the study early at a time when the baseline questionnaire did not contain this question. In the main analysis the missing values were dealt with by adding an extra "missing" category to the variable in question. An alternative analysis excluded the "early" participants and adjusted for the remaining missing values using the Horvitz-Thompson method,[18] which weights individuals by the inverse probability that they provide full data.

A variety of statistical models relating fat intake to breast cancer were examined. Separate analyses were performed for total, saturated, polyunsaturated, and monounsaturated fats. Each type of fat was examined across quintiles of fat intake, with adjustment for total energy (standard model)[19] or across quintiles of energy-adjusted fat (residual model). Tests for trend across quintiles were conducted using the median values of the quintiles as continuous variables. Versions of these models with continuous intake variables were also considered.

The above analyses are those typically used in nutritional epidemiologic studies. However, as mentioned earlier, participants were selected for the study on the basis of the percent energy from fat reported on their FFQ, with all those reporting less than 32% being excluded. This "truncation" of the sample causes a bias in the estimated RRs, and also a reduction of the variance in reported fat intake, thereby increasing the SEs of the estimated RRs for the 2 instruments. Because selection was based on the FFQ report, these biasing effects were expected to be stronger for the FFQ than for the FR and therefore needed to be adjusted for in any comparison of the 2 instruments.

The selection was adjusted for so as to produce unbiased estimates of RRs (selection-adjusted RRs) for breast cancer on the basis of either of the 2 dietary instruments, enabling a proper evaluation of the fat-breast cancer association. It was necessary also to calculate specially

adjusted unbiased estimates of standardized log RRs for the 2 instruments to provide a fair comparison of the power of the 2 dietary instruments to detect a fat-breast cancer association in a nontruncated study.

Estimated RRs adjusted for selection into the study are shown in Table 1, where missing data are handled using the missing category method. In all cases the estimated RR in the highest quintile was higher when based on the FR than when based on the FFQ. Statistically significant trends were seen for total, polyunsaturated, and monounsaturated fat when using the FR. Some trends were seen also when using the FFQ, but these were not statistically significant. However, note that the confidence limits for the RRs were wider when the FFQ rather than the FR was used, and this was partly due to the selection bias.

When the Horvitz-Thompson method was used for handling missing data, the trends were similar to those discussed above but stronger. For each type of fat, the adjusted RRs were stronger for the FR and weaker for the FFQ. For total fat, the 4th and 5th quintile RRs adjusted for total energy and corrected for selection were for the FR, 1.86 and 2.54 ($P_{trend} = 0.006$) and for the FFQ, 1.06 and 1.24, respectively ($P_{trend} = 0.41$). For saturated fat, these adjusted RRs for the 4th and 5th quintiles were 1.33 and 1.79 for the FR ($P_{trend} = 0.06$) and 1.01 and 0.85 for the FFQ ($P_{trend} = 0.49$).

Direct comparison of the instruments' power to detect a fat–breast cancer association was made through an adjusted standardized log RR. These values are shown together with the resultant projected statistical power for various continuous models in Table 2, according to the 2 different ways of adjusting for missing values. The power was higher for the FR than for the FFQ for all types of fat. The differences in power reached statistical significance ($P < 0.05$) under the Horvitz-Thompson method and were close to statistically significant under the missing category method. Note that for the FR, the statistical powers for total, polyunsaturated, and monounsaturated fats were >0.8, whereas for the FFQ they were all <0.5.

In summary, the data in this study showed a clear positive association between dietary fat intake and breast cancer incidence, for total fat and several subtypes, based on the dietary assessments from the FR. The association was far less clear when the data from the FFQ were used.

In addition, the analysis indicated that the FFQ carried less statistical power than the FR for detecting this association. These results provide direct evidence that the FR is a more powerful instrument than the FFQ for detecting fat–breast cancer associations, although one should not consider the matter proven without corroborating evidence from other large studies that use both types of instrument. Together with the results of Bingham et al,[5] they suggest a possible reason for the failure to detect such associations in previous cohort studies, given that those studies used FFQs. Day et al[14] in a comparison of a 7-day diary with a FFQ reported finding less error with the diary, although this finding was disputed.[20] The large Observing Protein and Energy Nutrition (OPEN) study, in which doubly labeled water and urinary nitrogen were used as unbiased biomarkers of energy and protein intake, has shown that absolute intakes reported on a FFQ have weaker correlations with true intake than the corresponding intakes reported on 24-hour recalls.[21] Thus, it seems plausible that the difference in results seen when the FR and FFQ are used is not a chance event, but one due to the different properties of the instruments. Additional evaluation of dietary assessment strategy in large studies is clearly warranted, as alternative self-report instruments (dietary records or multiple 24-hour recalls) may be preferable to the FFQ.

## A RANDOMIZED DIETARY INTERVENTION TRIAL

The strongest design to investigate the effects of dietary intake on breast cancer incidence is acknowledged to be the randomized controlled trial (RCT). However RCTs are expensive and

time-consuming, and can be used for only a few select questions. The FFQ-FR comparison study described above was conducted in the control group of the Women's Health Initiative DM trial of a low-fat eating pattern. Principal results of the randomized comparison in this trial were reported recently.[7] The intervention group reported (on a FFQ) consuming an average 25.5% calories from fat per day over the follow-up period compared with 35% in the comparison group. The comparison of breast cancer incidence showed a reduction in breast cancer risk in the intervention group that did not quite attain conventional statistical significance. Over a median follow-up period of 8.1 year, the estimated RR (hazard ratio [HR]) between the intervention and comparison group was 0.91 ($P = 0.07$) with a 95% confidence interval (CI) of 0.83–1.01.

Separation between the incidence curves in the 2 groups began to diverge approximately 4 years after entry to the trial. Subset analyses, although not providing definitive evidence, accorded well with results that would be expected if the intervention effect were real. Thus, the reduction in risk was greatest (RR = 0.78) among women who reported consuming 36.8% calories from fat at baseline (the highest quartile); these were the women able to make the greater reductions in their fat intake. Reduction in breast cancer risk was greater among those women who adhered to the protocol intervention. Thus, although the first definitive analysis of this trial did not result in a statistically significant reduction in risk, the overall pattern of results certainly suggested that the low-fat dietary pattern would reduce breast cancer incidence. It should be noted that the median follow-up of 8.1 years was somewhat shorter and the difference between the fat intakes of the 2 groups was less than that originally planned, so that the power to detect a reduction in breast cancer fell below the designed level. Longer, planned follow-up may yield a more conclusive result.

## COMPARISON OF THE FR-FFQ STUDY AND THE RCT

The results of the FR-FFQ comparison and of the RCT analyses can be shown to be quite consistent with each other. The RR found in the FR-FFQ study, using the FR data, can be applied to the reported difference in fat intakes of the 2 groups (25.5% vs 35% calories from fat) to yield a predicted RR between the 2 groups of 0.78,[6] a 22% reduction. However, the FR-FFQ analysis was based on baseline reported dietary intakes thought to reflect long-term intake, whereas the RCT analysis estimated the effect of short-term dietary change and was designed under the assumption that breast cancer risk would decline linearly over time, with the intervention achieving its maximum effect after 10 years. If the estimate of 22% reduction is taken as that maximum effect, then one would expect an average risk reduction over the first 8 years of the trial to be 22% × 0.4 = 8.8%, an estimate very close to the observed 9.1% in the RCT.

Willet and Hu[22] claimed that if this prediction had been adjusted for the attenuation resulting from the use of just 4 days of diet records in the FR-FFQ analysis, then it would have been seen to be much higher than the 8.8% and consequently inconsistent with the trial results. However, the amount of attenuation that occurs for reported fat intake is uncertain and is offset by the difference in underreporting between the intervention and control groups reported by Neuhouser et al.[23] Such differential misreporting between the trial groups would exaggerate the reported gap in fat intake, upon which the predicted reduction of 8.8% is based, in which case this prediction would have to be adjusted downward. In summary, the results of the FR-FFQ and RCT analyses do appear to be quite consistent, within the bounds of our current knowledge.

## A RECENT META-ANALYSIS OF OBSERVATIONAL STUDIES

Other more recent evidence for an association of dietary fat and breast cancer comes from a meta-analysis of the published observational studies up to July 2003.[8] The design of this meta-analysis differed from that of the pooled cohort analysis of Smith-Warner et al.[3] It was based on reported results in the literature rather than on source data (a potential disadvantage), but it also included more studies (a potential advantage). In fact, the authors included 45 studies for analysis of total fat (14 cohort studies and 31 case-control studies), with 33 studies for types of fat and 36 studies with information on the 3 most common food groups studied in relation to breast cancer: meat, milk, and cheese. This compares with the 8 cohort studies reported by Smith-Warner et al.[3]

Most of the studies in the report of Boyd et al[8] had used FFQs and analyzed the data by comparing the lowest to highest intake categories. The authors used the RR estimate that had been adjusted for energy and all of the established breast cancer risk factors in each of the studies. In addition, for each study a quality score was generated on the basis of predetermined methodologic criteria, such as histologic confirmation of the breast cancer and details of the study population. These scores were used in subanalyses to divide the studies into more and less reliable studies to see whether the observed associations held in the 2 groups. Premenopausal and postmenopausal breast cancer estimates were merged and not evaluated separately.

For total fat, there were 25,015 case patients and 580,000 control subjects for evaluation from studies conducted in Europe, North America, Asia, Australia, and Uruguay. FFQs were used in 32 studies, with varying numbers of food items and method of administration (interviewer vs self-administered). Results were, for the most part, similar between the cohort and case-control studies. Focusing on the results from the cohort studies, for total fat, the overall RR estimates for high compared with low intake were 1.13 (95% CI: 1.04–1.23). For other types of fat, the RR estimates were very similar to those of total fat for the cohort studies, 1.15 (95% CI: 1.02–1.30) for saturated fat, 1.10 (95% CI: 0.83–1.44) for monounsaturated fat, and 1.11 (95% CI: 1.00–1.22) for polyunsaturated fat. Analyses of the association of total fat and breast cancer risk by quality of the studies revealed stronger associations among the studies meeting 80% of the quality standards than among those meeting fewer of the standards. Results for foods showed increased risk for high intake of meat (RR = 1.17; 95% CI: 1.06–1.29) and nonsignificantly elevated risks for high intakes of milk (RR = 1.12; 95% CI: 0.88–1.43) and cheese (RR = 1.26; 95% CI: 0.96–1.66).

## THE NATIONAL INSTITUTES OF HEALTH (NIH)-AARP COHORT

Aside from dietary measurement error, a second methodologic problem is that many previous epidemiologic studies were conducted in homogeneous populations with a relatively narrow range of fat intakes. This could add to the difficulties of detecting an association between fat intake and breast cancer risk.[24] A narrow range of exposure may also increase attenuation of estimated RRs, even when the same dietary assessment instrument is used, because the smaller is the ratio of true variation in intake to measurement error variation the greater is the attenuation.[25] Therefore, it is important to study cohorts with a wide range of fat intake.

A report on the association of dietary fat intake with risk of postmenopausal breast cancer in the National Institutes of Health (NIH)–AARP (formerly American Association of Retired Persons) Diet and Health Study, a very large prospective cohort of more than 500,000 US men and women among whom dietary fat intakes varied substantially was recently published.[9] Details of the study are described elsewhere.[26] The analysis included 188,736 postmenopausal women aged 50–71 years at baseline (in 1995–1996), when they completed a 124-item FFQ. The FFQ was calibrated against 2 24-hour dietary recalls that were administered to a subset of

2,053 NIH–AARP participants by telephone an average of 25 days apart.[26] Over an average follow-up of 4.4 years, 3,501 cases of invasive breast cancer were observed.

Hazard ratios (HR), which may be thought of as age-specific RRs averaged over time, were estimated using the Cox proportional hazards regression models with age as the primary time variable. Analyses were performed with fat intake as either a continuous or a categorical variable. The HRs for fat increase on the continuous scale were calculated for a 2-fold increase, eg, from 20% to 40% of energy from total fat, and in categorical analyses, the HRs were calculated between quintiles of fat intake. Tests for linear trend in the categorical analysis were performed by using the median intake level in each quintile.

Several potential confounders were adjusted for in a multivariate model, including alcohol consumption (continuous), smoking history (ever versus never), age at birth of first child and number of children combined, age at menopause (<50, 50–54, or ≥55 years), menopausal hormone therapy use (current, never, or former user), and body mass index (<25, 25–<30 kg/$m^2$, or ≥30 kg/$m^2$), as well as nonalcohol energy intake.

The HR of breast cancer for the highest (median intake, 40.1% energy from total fat) versus the lowest (median intake, 20.3% energy from total fat) quintile of total fat intake was 1.11 (95% CI: 1.00–1.24; $P_{trend}$ = 0.017). The corresponding HR for a 2-fold increase in percent energy from total fat on the continuous scale was 1.15 (95% CI: 1.05–1.26). Positive associations were also found for subtypes of fat (HR for a 2-fold increase in percent energy from saturated fat = 1.13; 95% CI: 1.05–1.22; from monounsaturated fat, HR = 1.12, 95% CI: 1.03–1.21; and from polyunsaturated fat, HR = 1.10, 95% CI: 1.01–1.20).

The investigators also estimated the HR for a 2-fold increase in fat intake using measurement error adjustment methods (regression calibration),[27] with data from the calibration study to make the adjustment. The adjustment gave an estimated HR for total fat of 1.32 (95%CI: 1.11–1.58). Although the use of 24-hour recalls to correct for measurement error may not fully adjust for the error in the FFQ, evidence from the OPEN study suggests that the adjustment is in the right direction.[25]

## CONCLUSIONS

Results from several large cohort studies that were reported 10–20 years ago appeared to indicate that the hypothesized link between dietary fat intake and breast cancer risk was illusory. In this article we have reviewed several strands of more recent evidence that have emerged and together point in a different direction. Each of the studies discussed in this article suggests that a modest but real association between fat intake and breast cancer is likely. If, as the randomized trial result indicates, the association is causative, it would have important implications for public health strategies in reducing breast cancer incidence. The evidence is not yet conclusive, but further follow-up in the randomized trial, as well as efforts to improve dietary assessment methodology for cohort studies, may be sufficient to provide a convincing answer.

### Acknowledgements

We acknowledge the contribution of our coauthors of the article comparing 2 dietary instruments,[6] Doug Midthune, Fran Thompson, Ross Prentice, Ruth Patterson, Raymond Carroll and Amy Subar, which is featured in this review article.

## References

1. Rothman, KJ.; Greenland, S., editors. Modern Epidemiology. Philadelphia, PA: Lippincott Williams & Wilkins; 1998. p. 24-28.

2. Howe GR, Hirohata T, Hislop TG, et al. Dietary factors and risk of breast cancer: combined analysis of 12 case-control studies. J Natl Cancer Inst 1990;82:561–569. [PubMed: 2156081]
3. Smith-Warner SA, Spiegelman D, Adami HO, et al. Types of dietary fat and breast cancer: a pooled analysis of cohort studies. Int J Cancer 2001;92:767–774. [PubMed: 11340585]
4. Willett, WC. Cancer prevention: diet and chemopreventive agents: Section 1. Dietary fat. In: DeVita, VT.; Hellmann, S.; Rosenberg, S., editors. Cancer: Principles and Practice of Oncology. 7. Philadelphia, PA: Lippincott Williams & Wilkins; 2004.
5. Bingham SA, Luben R, Welch A, et al. Are imprecise methods obscuring a relation between fat and breast cancer? Lancet 2003;362:212–214. [PubMed: 12885485]
6. Freedman LS, Potischman N, Kipnis V, et al. A comparison of two dietary instruments for evaluating the fat-breast cancer relationship. Int J Epidemiol 2006;35:1011–1021. [PubMed: 16672309]
7. Prentice RM, Caan B, Chlebowski RT, et al. Low fat dietary pattern and risk of invasive breast cancer: The Women's Health Initiative randomized controlled dietary modification trial. JAMA 2006;295:629–642. [PubMed: 16467232]
8. Boyd NF, Stone J, Vogt KN, et al. Dietary fat and breast cancer risk revisited: a meta-analysis of the published literature. Br J Cancer 2003;89:1672–1685. [PubMed: 14583769]
9. Thiébaut AC, Kipnis V, Chang SC, et al. Dietary fat and postmenopausal invasive breast cancer in the National Institutes of Health-AARP Diet and Health Study cohort. J Natl Cancer Inst 2007;99:451–462. [PubMed: 17374835]
10. Freudenheim JL, Marshall JR. The problem of profound mismeasurement and the power of epidemiologic studies of diet and cancer. Nutr Cancer 1988;11:243–250. [PubMed: 3217262]
11. Willett, WC. Nutritional Epidemiology. New York, NY: Oxford University Press; 1990. p. 69-91.
12. Prentice R. Measurement error and results from analytic epidemiology: dietary fat and breast cancer. J Natl Cancer Inst 1996;88:1738–1747. [PubMed: 8944004]
13. Kipnis V, Midthune D, Freedman LS, et al. Empirical evidence of correlated biases in dietary assessment instruments and its implications. Am J Epidemiol 2001;153:394–403. [PubMed: 11207158]
14. 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;30:309–317. [PubMed: 11369735]
15. Prentice RL. Dietary assessment and the reliability of nutritional epidemiology reports. Lancet 2003;362:182–183. [PubMed: 12885475]
16. Women's Health Initiative Study Group. Design of the Women's Health Initiative Clinical Trial and Observational Study. Control Clin Trials 1998;19:61–109. [PubMed: 9492970]
17. Hays J, Hunt JR, Hubbell A, et al. The Women's Health Initiative recruitment methods and results. Ann Epidemiol 2003;13:S18–S77. [PubMed: 14575939]
18. Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. J Am Stat Assoc 1952;47:663–685.
19. Kipnis V, Freedman LS, Brown CC, et al. Interpretation of energy adjustment models for nutritional epidemiology. Am J Epidemiol 1993;137:1376–1380. [PubMed: 8333419]
20. Willett W. Commentary: Dietary diaries versus food frequency questionnaires—a case of undigestible data. Int J Epidemiol 2001;30:317–319. [PubMed: 11369736]
21. 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 OPEN study. Int J Epidemiol 2003;32:1054–1062. [PubMed: 14681273]
22. Willett WC, Hu FB. The food frequency questionnaire [Letter to the editors]. Cancer Epidemiol Biomarkers Prev 2007;16:182–183. [PubMed: 17220351]
23. Neuhouser, M.; Tinker, L.; Schoeller, D., et al. Use of nutritional biomarkers to describe participant-related measurement error from dietary self-report in the Women's Health Initiative. In: Heitmann, BL.; Lissner, L.; Winkvist, A., editors. Sixth International Conference on Dietary Assessment Methods: Complementary Advances in Diet and Physical Activity Assessment Methodologies. Copenhagen, Denmark: Diet Research Foundation, The Danish Network of Nutritional Epidemiologists, Swedish Network for Nutritional Epidemiology; 2006. p. SY01-01.

24. Riboli E, Kaaks R. Invited commentary: The challenge of multi-center cohort studies in the search for diet and cancer links. Am J Epidemiol 2000;151:371–374. [PubMed: 10695595]

25. Kipnis V, Subar AF, Midthune D, et al. The structure of dietary measurement error: results of the OPEN biomarker study. Am J Epidemiol 2003;158:14–21. [PubMed: 12835281]

26. Schatzkin A, Subar AF, Thompson FE, et al. Design and serendipity in establishing a large cohort with wide dietary intake distributions: the National Institutes of Health-American Association of Retired Persons Diet and Health Study. Am J Epidemiol 2001;154:1119–1125. [PubMed: 11744517]

27. Carroll, RJ.; Ruppert, D.; Stefansk, LA., et al. Measurement Error in Nonlinear Models: A Modern Perspective. 2. Boca Raton, FL: Chapman and Hall CRC Press; 2006.

**TABLE 1**

RRs* (95% CI) for Breast Cancer in Quintiles of Fat Intake, Estimated According to Standard and Residual Models, Adjusted for Selection Criteria

| Type of Fat and Model | Q1 | Q2 | Q3 | Q4 | Q5 | $P_{trend}$ |
|---|---|---|---|---|---|---|
| Total | | | | | | |
| FR fat + log energy | 1.00 | 1.48 (0.94–2.33) | 1.64 (1.06–2.54) | 1.96 (1.22–3.15) | 2.09 (1.21–3.61) | 0.008 |
| FFQ fat + log energy | 1.00 | 1.16 (0.72–1.85) | 1.15 (0.65–2.02) | 1.56 (0.79–3.08) | 1.71 (0.70–4.18) | 0.18 |
| FR residual + log energy | 1.00 | 1.11 (0.72–1.72) | 1.23 (0.79–1.92) | 1.30 (0.88–1.94) | 1.58 (1.08–2.33) | 0.007 |
| FFQ residual + log energy | 1.00 | 1.26 (0.28–5.69) | 1.22 (0.28–5.33) | 1.16 (0.27–4.93) | 1.43 (0.33–6.25) | 0.30 |
| Saturated | | | | | | |
| FR fat+ log energy | 1.00 | 1.21 (0.79–1.86) | 1.29 (0.85–1.95) | 1.07 (0.69–1.66) | 1.51 (0.94–2.43) | 0.20 |
| FFQ fat + log energy | 1.00 | 1.12 (0.70–1.80) | 1.09 (0.66–1.81) | 1.19 (0.67–2.11) | 1.00 (0.49–2.02) | 0.95 |
| FR residual + log energy | 1.00 | 0.82 (0.55–1.24) | 1.24 (0.88–1.74) | 1.14 (0.81–1.61) | 1.17 (0.85–1.62) | 0.11 |
| FFQ residual + log energy | 1.00 | 0.95 (0.62–1.45) | 1.02 (0.68–1.53) | 0.94 (0.63–1.39) | 0.99 (0.68–1.44) | 0.99 |
| Polyunsaturated | | | | | | |
| FR fat + log energy | 1.00 | 1.00 (0.67–1.51) | 1.43 (0.95–2.14) | 1.27 (0.83–1.95) | 1.74 (1.06–2.84) | 0.01 |
| FFQ fat + log energy | 1.00 | 0.89 (0.60–1.32) | 0.90 (0.60–1.36) | 0.94 (0.59–1.52) | 1.02 (0.57–1.83) | 0.79 |
| FR residual + log energy | 1.00 | 1.37 (0.92–2.03) | 1.19 (0.81–1.74) | 1.40 (0.96–2.04) | 1.56 (1.05–2.33) | 0.03 |
| FFQ residual + log energy | 1.00 | 0.95 (0.61–1.49) | 0.89 (0.58–1.36) | 0.90 (0.57–1.42) | 1.03 (0.66–1.61) | 0.76 |
| Monounsaturated | | | | | | |
| FR fat + log energy | 1.00 | 1.27 (0.78–2.07) | 1.63 (1.09–2.45) | 1.58 (0.99–2.51) | 1.96 (1.11–3.45) | 0.02 |
| FFQ fat + log energy | 1.00 | 1.06 (0.68–1.64) | 1.03 (0.62–1.72) | 1.41 (0.77–2.59) | 1.39 (0.64–3.01) | 0.25 |
| FR residual + log energy | 1.00 | 1.11 (0.70–1.76) | 1.69 (1.16–2.47) | 1.66 (1.16–2.38) | 1.46 (1.01–2.11) | 0.03 |
| FFQ residual + log energy | 1.00 | 0.95 (0.45–2.03) | 1.16 (0.57–2.40) | 1.08 (0.52–2.23) | 1.27 (0.62–2.63) | 0.13 |

*Adjusted for the following baseline variables: duration of follow-up, age at entry into study (in 5-year age groups), clinical center region (Northeast, South, Midwest, or West), hormone use (never or ever), family history (missing, no, or yes), and breast biopsy (missing, no, or yes).

**TABLE 2**

Estimated Standardized log RR Parameters* and Resulting Statistical Power for FR and FFQ, According to 2 Methods of Adjusting for Missing Data, Adjusted for Selection Criteria, in a Continuous Intake Model

| Model and Instrument | Missing Data Category | | Horvitz-Thompson Method | |
|---|---|---|---|---|
| | Standardized Log RR Adjusted for Selection† | Statistical Power | Standardized Log RR Adjusted for Selection† | Statistical Power |
| Log total fat and log energy | | | | |
| FR | 3.32 (1.16) | 0.91 | 3.70 (1.26) | 0.96 |
| FFQ | 1.24 (1.33) | 0.24 | 0.71 (1.37) | 0.11 |
| Difference | $P = 0.08$ | | $P = 0.01$ | |
| Log saturated fat and log energy | | | | |
| FR | 1.68 (1.06) | 0.39 | 2.33 (1.14) | 0.64 |
| FFQ | 0.14 (1.17) | 0.03 | −0.12 (1.11) | 0.02 |
| Difference | $P = 0.17$ | | $P = 0.02$ | |
| Log polyunsaturated fat and log energy | | | | |
| FR | 3.02 (1.08) | 0.86 | 2.95 (1.10) | 0.84 |
| FFQ | 0.77 (1.29) | 0.12 | 0.42 (1.18) | 0.06 |
| Difference | $P = 0.07$ | | $P = 0.03$ | |
| Log monounsaturated fat and log energy | | | | |
| FR | 3.47 (1.14) | 0.93 | 3.92 (1.20) | 0.98 |
| FFQ | 1.82 (1.30) | 0.44 | 1.16 (1.36) | 0.21 |
| Difference | $P = 0.16$ | | $P = 0.02$ | |

SEs (in parentheses) and *P* values were estimated by the bootstrap method.

*Adjusted for the following baseline variables: duration of follow-up, age at entry into study (in 5-year age groups), clinical center region (Northeast, South, Midwest, or West), hormone use (never or ever), family history (missing, no, or yes), and breast biopsy (missing, no, or yes).

†Estimates of what the standardized log RR would have been if the selection criterion had not been used to select participants into the study: (standardized log RR in a study with no selection) = (log RR adjusted for selection)/(SE log RR in a study with no selection).