Issues in the design and interpretation of studies of fatty acids and cancer in humans$^{1,2}$

Tim Byers and Karen Gieseker

**ABSTRACT** The methods used in nutritional epidemiology to study the relations between fatty acids and cancer risk include ecologic studies, case-control studies, cohort studies, and intervention trials examining either intermediate markers of cancer risk or cancer incidence. Each type of study design has its particular strengths and weaknesses. The inaccuracy of estimates of fatty acid intake from the use of dietary questionnaires linked to nutrient databases is a major limitation in nutritional epidemiology. Information on the concentrations of fatty acids in the circulation or in adipose tissue can complement estimations of dietary intake. Cancer prevention studies now underway are designed to test whole-diet effects on neoplasia and will not be able to separate the effects of specific fatty acids from those of other nutrients in the diet. The development of better intermediate markers of cancer risk could enable the use of experimental methods to assess the relation between specific fatty acids and cancer. Research findings as described in the literature are complicated both by the multiple hypotheses that can be tested when assessing fatty acid effects and by the uncertainties of multivariate adjustment. Although there are substantial obstacles to understanding the relations between fatty acid intakes and cancer risk, there is no better species than humans for inference about diet and cancer risk in people. *Am J Clin Nutr* 1997;66(suppl):1541S–7S.

**KEY WORDS** Cancer, etiology, nutrition, fats, fatty acids, human studies, epidemiology, study design

**INTRODUCTION**

Studies of the relation between the intake of fatty acids and cancer risk in humans are complex in both design and interpretation. Long-term intervention trials, of the type used in experimental research in animals, are rarely practical in human populations. The relations between fatty acid intake and cancer must therefore be inferred from either observational epidemiologic research or from short-term trials examining biological endpoints that are thought to be markers of cancer risk. In this paper we review both the usual challenges in the design and interpretation of such studies in human populations, as well as the special issues in studying the role of fatty acids in the etiology of cancer in humans.

**METHODS USED TO STUDY FATTY ACIDS AS RISK FACTORS FOR CANCER IN HUMANS**

Observational epidemiologic studies include ecologic (correlational) studies, case-control (retrospective) studies, and cohort (prospective) studies. Intervention studies include both small trials examining intermediate markers of cancer risk and large trials examining cancer endpoints. The general approaches to and findings from the investigation of the role of nutrients in human cancers have been reviewed elsewhere (1, 2). In this paper we focus on issues that are particularly relevant to investigating the roles of specific fatty acids in cancer in humans.

**Ecologic studies**

Ecologic studies relate the cancer rates of different populations or of the same population in different time periods to other characteristics that differ between those same populations. In nutritional ecologic studies of cancer, the intent is to look for correlations between dietary intake estimates that might plausibly be thought to be related to cancer etiology and to population-level cancer rates. Correlations between diet and cancer that are thought to have plausible biological relevance can then be tested in more definitive studies. Ecologic studies present a formidable barrier to causal inference because there are usually many other characteristics that differ between populations that might also be correlated with both cancer rates and with the factor (such as diet) under scrutiny. Ecologic studies of fat intake in different countries (3) and of the changes in fat intake among populations after migration between countries (4) are good examples of how hypotheses from ecologic studies may not always be confirmed in studies of individuals. Although international ecologic studies have suggested strong correspondence between dietary fat and breast cancer rates, subsequent studies of the relation between fat intake and breast cancer risk using measures of fat intake of individual women have not shown this association (5). Ecologic studies are often weakened by the nonspecificity of the factor under study. Nonetheless, ecologic studies may be superior in some respects to studies based on measures of diets of individuals because ecologic studies can compare nutrient or food intakes across a wider range than can studies conducted within countries. Moreover, differences in diet between populations at a point in time may also better reflect life-long dietary habits related to cul-

---

$^{1}$ From the Department of Preventive Medicine and Biometrics, University of Colorado School of Medicine, Denver.

$^{2}$ Address reprint requests to T. Byers, Department of Preventive Medicine and Biometrics, University of Colorado School of Medicine, Box C-245, Denver, CO 80262. E-mail: Tim.Byers@uchsc.edu.
Few ecologic studies of fatty acids and cancer risk have been done (6). Fatty acids would be expected to either covary strongly with total fats in the diet or vary predictably with patterns of the intake of key indicator foods for particular fatty acids, such as red meats, particular cooking oils, or foods such as fish that have unique fatty acid profiles. A potential role for ecologic studies of fatty acids as etiologic factors in cancer might be in the examination of the natural experiments that occasionally result from intentional changes in fatty acids in the food supply. For instance, in reaction to concern about the atherogenicity of saturated fats, Mauritius (an Island nation in the Indian Ocean) changed its import policy in the 1980s to discourage the importation of coconut oils. An immediate reduction of serum cholesterol concentrations of the Mauritian population resulted (7).

Although such policy-related diet changes would be most easily seen in an isolated island nation, there are also other examples of population-level natural experiments that might be amenable to study with the ecologic method. For instance, increasing amounts of trans fatty acids in the food supply resulting from the increased consumption of processed oils in the United States could be examined relative to later cancer rates (8). Such ecologic studies could be useful for estimating the possible bounds of strength of hypothesized etiologic factors in the population. However, inferences that changes in the temporal trend of cancer incidence might correspond with any particular dietary change in the time period are limited by the inherent nonspecificity of the ecologic design.

Case-control studies

Studies based on the assessment of the nutritional status of individuals do not have the same limitations as ecologic studies, but they have their own weaknesses. Case-control studies of diet are based on the contrasts of the reports of recalled diets of people with and without cancer. Usually, the questioning is about the frequency of intake of various foods for the year or so before symptoms leading to the diagnosis. Few studies have attempted to ascertain estimates of diet from the more distant past (9). Many validation studies have shown that diet as reported on questionnaires or in interviews is reliable and that it correlates with diet as measured by other techniques (10–14). Random misclassification due to errors in estimating diet will tend to bias effect estimates toward the null value.

Of more concern has been the question of possible differential recall of diet by case subjects compared with control subjects. Even a small effect of the recent experience of cancer on diet recall would create a systematically different estimate of past diet and hence a bias that would lead to artifactual associations between diet and cancer risk. Several investigators have looked for differences between case and control subjects in the correspondence between recalled diet and diet as previously measured. Investigators in the Nurses Health Study found evidence for recall bias by the subjects with breast cancer (15), but other studies of similar design have not found bias related to cancer status (16–19). These different findings may support the idea that prior knowledge of the hypothesis under investigation can accentuate recall bias by case subjects. The health professional women in the Nurses Health Study (15) were likely more aware of the major dietary hypothesis related to breast cancer than were those in the other populations studied (16–19).

Cohort studies

Cohort studies have an important advantage over case-control studies because the assessment of diet is made before diagnosis and hence is unbiased by the cancer experience. Cohort studies also have their own limitations, however. Because of the large number of subjects that must be followed in cohort studies, diet is usually assessed by a self-administered questionnaire. This method of diet estimation is not as detailed and has less quality control than does a structured diet interview. Cohort studies are also expensive, they usually require several years to accrue enough cases for analysis, and diet changes subsequent to the baseline assessment are often not considered. Cohort studies are nonetheless important in nutritional epidemiology because the diet measures are unbiased by the outcomes under investigation. Several large cohort studies that have included validated dietary measures are underway in the United States and several other countries. Many of these studies have also collected and stored serum for later analyses as nested case-control studies.

Intervention trials

Intervention trials offer several advantages to observational studies in investigating the relations between diet and disease. In human populations, however, there are important limitations in our ability to implement intervention studies of diet and cancer (20). Ethical considerations prohibit the study of extreme intakes for prolonged periods. Cancer endpoints are difficult to study because the studies take many years, they require large numbers of subjects, and ensuring adherence to the intervention is costly. Cost, compliance, and ethics are therefore all serious limitations to nutritional intervention studies for cancer.

The Women’s Health Initiative is a multicenter trial now underway designed to test several risk factors for chronic disease in US women (21). Among those factors is the intake of fats, which has been hypothesized to be related to both coronary atherosclerosis and breast cancer. Total fat intake is targeted to be reduced to <25% of energy intake in the Women’s Health Initiative dietary intervention. This degree of reduction in fat intake is necessarily accompanied by many other changes in diet, including increased intakes of fiber, fruit, and vegetables. The Women’s Health Initiative dietary intervention will therefore be interpretable only as a total diet effect, so effects of specific fatty acids will not be determinable. The Polyp Prevention Trial is a dietary intervention trial now underway to test the effects on colonic adenoma recurrence of substantial increases in fruit and vegetable intake along with reduced fat intake (22). The Polyp Prevention Trial is another example of a total diet intervention that will have limited ability to identify effects of particular types of fats. Changes in particular fats or fatty acids in either the Women’s Health Initiative or the Polyp Prevention Trial will not be specifically attributable to changes that might be observed in neoplasia risk. There are no trials underway that are designed to test the effects of changes in particular fatty acids as related to cancer endpoints.

Much more common types of intervention studies are those that assess the effects of dietary manipulations on short-term
physiologic measures that are thought to be related to eventual disease risk. Such studies are limited in two ways. First, short-term changes may not be maintained in the long term because homeostatic mechanisms often take effect. Second, and more importantly, because we do not fully understand the sequence of events that leads to cancer, the relevance of any particular short-term physiologic change to eventual cancer risk is uncertain. Studies of intermediate markers for coronary artery disease risk have been much more common than for cancer because the relevance of circulating lipid markers to heart disease risk is much better understood than is the relevance of intermediate markers to cancer risk. Most of the feeding trials of fatty acids have therefore examined fatty acid effects on plasma lipids (23). Nonetheless, studies documenting dietary effects on plausibly relevant intermediate physiologic markers of cancer risk, such as on circulating estrogen concentrations in women (24), could strengthen the argument for the etiologic relevance of fatty acids and could lead the way to studies on other neoplasia risk markers that are more proximal to cancer.

METHODS FOR ESTIMATING FATTY ACID INTAKE

Dietary estimation methods

Fatty acids are not directly measured, of course, in epidemiologic studies of dietary intakes, but estimates of intake are derived from measures of food intake. In case-control and cohort studies, food intake is usually measured by using food-frequency questionnaires, in which respondents report their usual frequency of intake of each food or food group from a specified list (25). Some food-frequency instruments include portion sizes in the food question whereas others ask respondents to report separately their usual portion size for each item. Estimates of nutrients can then be generated from these food reports by summing the products of the reported frequency and nutrient content of each food included in the questionnaire. The limitations to dietary assessment methods used in nutritional epidemiologic studies apply to all nutrients, but there are also specific limitations for estimating fatty acid intakes. Many validation studies have been done in which food-frequency reports have been compared with other methods of estimating diet, such as multiple 24-h recalls or diet diaries. Such studies have typically shown correlations ranging from 0.4 to 0.6 for most nutrients (10–14). More important than the correlations, however, is the extent of serious misclassification at the extremes of intake, as many analyses use categorical methods to compare risk across ordinal categories of intake. In general, the degree of misclassification at the extreme categories is small. The correspondence between methods for specific types of fats is similar to that for other nutrients.

There are particular methodologic limitations in estimating fatty acid intake in epidemiologic studies. The nutrient contents of foods are specified in nutrient databases, which are catalogs of the nutrient amounts found in various foods, either as assayed or as estimated by analogy. The US Department of Agriculture (USDA) maintains and updates a nutrient database of > 5200 foods for many nutrients that includes information on the amounts of 19 specific fatty acids (26). Other nutrient databases are also used widely in nutritional epidemiology, such as that maintained at the Nutrition Coordinating Center at the University of Minnesota (25). The strength of the Nutrition Coordinating Center database is its detailed information on fats and fatty acids, as obtained from multiple sources, including the USDA nutrient database, published studies, and manufacturers’ brand-specific information. Brand specificity is important for many nutrients because the fatty acid content can vary considerably by brand for a particular food, such as cookies or pastries. Even within a brand, however, the fatty acids used in the recipe can change over time because of variations in the cost and availability of fats and oils in the marketplace. Many nutrient databases are often incomplete for specific fatty acids, however. For example, the amounts of trans fats in foods is not well documented in any of the nutrient databases.

Even if we had perfect knowledge of the fatty acid content of all foods, uncertainties would still be introduced because of the nonspecificity of many of the food items ascertained in questionnaires. For instance, the Health Habits and History Questionnaire asks the frequency of intake of beef, steaks, or roasts and of doughnuts, cookies, cake, or pastries (25, 27). Clearly, there are many different types and cuts of beef, steaks, and roasts and many different types of doughnuts, cookies, cakes, and pastries in the USDA nutrient database. To convert the particular frequency response to such questions into nutrients, such as fats or fatty acids, the investigator must define a specific nutrient content for that questionnaire item. In the Health Habits and History Questionnaire this has been done by assigning to the group the nutrient values of the one specific food from that group that is most commonly consumed in the population, as found in a national survey. For other questionnaires, the decision has been to assign to the group either the average nutrient value of all foods in the group or an average that is weighted by the relative frequencies of consumption of the foods in the group. Regardless of the method used to assign nutrient values to food groups in questionnaires, the assumptions that must be made about nutrient content for foods measured in groups necessarily adds additional error to the estimation of nutrient intake of individuals.

The estimation of fats and specific fatty acids is more difficult than the estimation of other nutrients in the diet for several reasons. Because fats are present in so many different foods, diet questionnaires need to be lengthy to capture the various food sources of fats (28, 29). For most uses, a full-length questionnaire of > 100 food items will be needed to estimate energy intake and to produce meaningful variations in intake of fats and fatty acids. Short questionnaires have been developed for crude classification of specific foods or nutrients (25). In some settings, some foods will be important indicators of particular fatty acids. For example, in the Mediterranean diet, olive oil is the predominant determinant of oleic acid (18:1) intake (30). However, for most nutrients and in most settings, there are no clear indicator foods. For example, in the US diet, 18:1 intake is determined not so much by the use of olive oil as by a wide variety of foods, including meats.

There are many similarities in the food sources for the different types of fats and fatty acids. The leading food sources for selected fats and fatty acids in the US diet are shown in Table 1 (29). Despite some differences in the food sources for the different fats, many of the foods are the same key indicator foods for total fat, saturated fat, and monounsaturated fat. Therefore, we would expect to see considerable collinearity between total energy, fats, and the particular types of fats from food-frequency questionnaires.
TABLE 1
Foods that were the leading sources of total fat, saturated fat, oleic acid, and linoleic acid in the US diet, 1976–1980: rank order

<table>
<thead>
<tr>
<th>Food</th>
<th>Total fats</th>
<th>Saturated fats</th>
<th>Oleic acid</th>
<th>Linoleic acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hamburgers, meatloaf</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td>Hot dogs, ham, lunch meats</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Whole-milk beverages</td>
<td>3</td>
<td>2</td>
<td>7</td>
<td>23</td>
</tr>
<tr>
<td>Doughnuts, cookies, cake</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Beef steaks, roasts</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>28</td>
</tr>
<tr>
<td>White bread, rolls, crackers</td>
<td>6</td>
<td>10</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Eggs</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>Cheeses</td>
<td>8</td>
<td>3</td>
<td>10</td>
<td>29</td>
</tr>
<tr>
<td>Margarine</td>
<td>9</td>
<td>12</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Mayonnaise, salad dressing</td>
<td>10</td>
<td>14</td>
<td>12</td>
<td>1</td>
</tr>
</tbody>
</table>

¹ From reference 29.

Furthermore, cooking methods influence the fat content of foods. Indication of the method of cooking, such as broiling or frying or the addition of cooking oils, can add important information about the types and amounts of fat retained in the foods as eaten. In many cases, however, study subjects are unaware of the type of cooking oils and the cooking methods used to prepare many of the foods they consume. There are also substantial amounts of hidden fats in many foods, such as pastries, about which subjects often have little knowledge.

Biomarkers of fatty acids

Directly measuring the concentrations of fatty acids in tissues eliminates the uncertainties of both self-reports of diet and the inadequacies of nutrient databases. Fatty acids can be measured in plasma lipoproteins, cell membranes, and adipose tissue (31–34). There are no reliable biomarkers of total fat intake, only of the relative proportions of various fatty acids in the diet. The concentrations and proportions of the various fatty acids in circulating cholesterol esters and phospholipids reflect intake in the preceding several days to weeks. Concentrations of fatty acids in the circulation can therefore be useful markers for defining the cross-sectional fatty acid intakes of a population and classifying individuals by their usual intakes in recent weeks. Fatty acids in the circulation correlate with both multiple diet recalls and with food-frequency estimates of dietary intakes of fatty acids (34). Concentrations of fatty acids in cholesterol esters may have a higher degree of reliability (ie, less within-person variation) than do concentrations of fatty acids in plasma phospholipids (35).

Fatty acids incorporated into cell membranes indicate nutrient intake in a time between the times indicated by plasma lipid and adipose tissue concentrations (32). Circulating concentrations have been used for validation studies of diet assessment methods (34) as well as for direct estimations of dietary patterns in etiologic studies (36).

Adipose tissue concentrations reflect the relative intakes of various fatty acids in the diet over longer periods of time. Adipose tissues, which can be easily taken from subcutaneous sites, have fatty acid ratios that correlate with dietary estimations (37–42). In one study the correlation between 19 diet recalls over 1 year and linoleic acid (18:2n–6) concentrations in gluteal adipose biopsies was 0.77 (38). The correlations between fatty acids in gluteal adipose biopsies and estimates of intake from less quantitatively precise diet measures, such as food-frequency questionnaires, is somewhat lower: 0.48 for men (41) and 0.37 for women (40).

Correlations between diet and both adipose tissue and plasma concentrations is greatest for the polyunsaturated fatty acids, intermediate for the saturated fatty acids, and lowest (virtually uncorrelated) for the monounsaturated fatty acids. This is likely due to the greater proportion of monounsaturated than polyunsaturated fatty acids in the total body pool; monounsaturated fatty acids are endogenously produced whereas polyunsaturated fatty acids are almost entirely diet-derived. Correlations between polyunsaturated fatty acids in adipose tissues and in circulating plasma lipids change predictably when fatty acids in the diet are intentionally altered (23, 42).

Interestingly, all adipose depots may not concentrate fatty acids in the same way. Saturated fatty acids have been found in higher concentrations in the subcutaneous abdominal adipose depots, whereas polyunsaturated fatty acids have been found in higher concentrations in the subcutaneous depots of the buttocks and thighs (43–45). This is an intriguing finding in view of the apparent importance of both central adiposity and saturated fat intake for heart disease and some cancers.

In case-control studies, the possible effects of recent illness and weight loss may be a barrier to the interpretation of plasma fatty acid values. For example, fatty acid profiles can change with acute weight loss, even in adipose tissues, because of the selective mobilization of α-linolenic acid (18:3n–3) during weight loss (46–48). A case-control study of fatty acids in subcutaneous adipose tissue from biopsies of the buttocks showed similar fatty acid profiles for breast cancer patients and control subjects, suggesting that differences in dietary fatty acid ratios are not associated with breast cancer (49). Future case-control studies using adipose tissue biopsies will need to carefully assess the possible confounding effects of weight loss on fatty acid concentrations. Other factors have also been shown to be related to both cancer risk and circulating fatty acid concentrations (34). These factors include obesity and tobacco use, which will also need to be carefully accounted for in future case-control and cohort studies examining fatty acid concentrations in biological samples as related to cancer risk.

ISSUES IN THE ANALYSIS OF DIETARY FATTY ACID DATA

Several different systems have been used to classify dietary fatty acids. Some investigators have reported only total dietary fats, with no details about food sources or types of fatty acids.
Others have classified fats according to animal or vegetable sources. Perhaps the most common practice is to classify fats according to their degree of saturation (saturated, monounsaturated, or polyunsaturated). In recent years investigators have begun to present findings for specific fatty acids. Of course, many of the fatty acids will be highly collinear with the classifications based on degree of saturation. This is particularly true for fatty acids such as 18:1, which accounts for most of the monounsaturated fatty acids, and 18:2n−6, which accounts for most of the polyunsaturated fatty acids.

The problem of multiple hypothesis testing is widespread in nutritional epidemiology but is particularly apparent in presentations of findings for specific fatty acids. From reports of a set of foods, many different nutrients can be computed from a nutrient database, including many different fatty acids. In addition, nutrients can be cross-classified with food groups to examine separately the nutrient intake relations as supplied by particular types of foods. For example, saturated fats can be analyzed separately by their food group source, such as dairy, or meat, and monounsaturated fats can be analyzed separately by animal or vegetable sources.

Separating the effects of a particular nutrient from possible confounding or modifying effects of other nutrients or foods is a challenging task in data analysis. There has been considerable discussion about the relative merits of alternative methods to account for energy intake in epidemiologic studies. The energy density method, in which fats (either total, by class, or for specific fatty acids) are expressed as a proportion of all energy, is intuitively simple and has been used in animal laboratory research for years. However, the expression of exposure as a ratio of two variables complicates the interpretation of effects. Alternatively, analysis can be done by using multivariate adjustment, in which the effects of a nutrient can be statistically isolated from the effects of other nutrients as well as from total energy intake. Often regression residuals are used to define nutrient estimates that are independent of energy intake; these residual values are then used in multivariate models that often include other nutrients to control for their possible confounding. Although multivariate-analysis programs can rapidly perform such adjustment procedures, the high degree of collinearity between nutrients complicates the interpretation of results from such modeling.

To demonstrate the potential for problems of collinearity between specific fats and fatty acids, we examined dietary data from subjects taking part in the San Luis Valley Diabetes Study. Twenty-four-hour recalls for 1340 people were completed from 1984 to 1992. Nutrient intakes from those recalls were estimated by using the Nutrition Coordinating Center software. The correlations between energy, fats, and selected types of fats and fatty acids as well as their partial correlations, accounting for energy intake, are shown in Table 2. Although some of the correlation between fats and fatty acids is accounted for by the mutual collinearity of fats and fatty acids with energy, much of it is not. This is to be expected both because many of the fats are present in the same foods and because they are also indicative of similar types of diets. Multivariate analyses that include adjustments for both total energy and other fats when assessing the associations between particular fatty acids and cancer risk might therefore be difficult to interpret. Presenting the correlations between the various nutrients and presenting both the adjusted and the unadjusted effect estimates helps readers interpret studies featuring these types of analyses.

### CONCLUSIONS

Studies of dietary fatty acids in human populations are needed to make inferences about cancer risk as related to the intake of fatty acids in people. Various study designs can be used, but each has both advantages and limitations. Ecologic studies will be of limited value in making causal inference because many factors can covary with fatty acid intake between populations and over time. Case-control and cohort studies are both limited by the uncertainties of methods currently used to estimate dietary intakes of specific fatty acids. Continued improvements in nutrient databases for fatty acids are needed, as is continued improvement in dietary assessment methods. The use of biomarkers as primary indicators of dietary intake is promising. Plasma fatty acid concentrations will be informative, especially from samples stored and analyzed as nested studies within cohorts, and adipose tissue samples from subjects in case-control studies can be informative. In all studies using biomarkers, though, careful control for the confounding effects of obesity, weight loss, and tobacco use is needed. Finally, the presentation of findings for specific fatty acids in the literature is clouded by the problems of the many hypotheses that can be tested for fatty acids within any dietary data set and by the uncertainties of inferring independent effects of specific nutrients from multivariate analysis.

![Table 2](https://academic.oup.com/ajcn/article-abstract/66/6/15415/4656014/1340)
REFERENCES

45. Calder P, Harvey D, Pond C, Newsholme E. Site-specific differences


