Study design and hypothesis testing: issues in the evaluation of evidence from research in nutritional epidemiology[1,2]

Jo L Freudenheim

ABSTRACT In addressing questions about the relations of dietary factors to disease in human populations, epidemiologic studies must account for the complexity of dietary habits, the intercorrelations among dietary habits, and the correlations of those habits with other behaviors. Furthermore, for studies of chronic disease, relevant dietary exposures may occur over decades. The classic epidemiologic study designs have been used to examine the associations between diet and disease; the strengths and weaknesses of those designs must be considered. Concerns have been raised regarding the validity of the measures of diet, the differential recall of diet by diseased individuals in case-control studies, and confounding by other related factors in both case-control and cohort studies. In clinical trials there may be difficulties in effecting the necessary dietary changes, especially for macronutrients, and there are also concerns about those circumstances in which participants cannot be blinded to their treatment. For case-control and cohort studies and for some clinical trials, intercorrelations among nutrients are a concern in the identification of factors that are important in the etiology of disease. It is important to understand these considerations when interpreting nutritional epidemiologic studies for the purpose of setting public policy. No one study can be considered definitive in the understanding of a diet–disease relation. However, epidemiologic findings from multiple studies taken together can contribute significantly to our understanding of diet in relation to disease in humans. *Am J Clin Nutr* 1999;69(suppl): 1315S–21S.

KEY WORDS Nutritional epidemiology, chronic disease epidemiology, study design, case-control study, cohort study, clinical trial, disease etiology, diet, nutrition, dietary habits, eating habits, humans

INTRODUCTION

Epidemiologic studies address the important but tangled questions concerned with identifying the factors that lead to disease in human populations. In nutritional epidemiology the focus is on dietary factors related to disease while taking into account other, nondietary factors. Such studies must account for the complexity of dietary habits (1, 2), the intercorrelations among dietary factors (3–5), and the correlations between dietary habits and other behaviors that have health consequences (6–10). In the study of chronic diseases, these analyses may need to evaluate exposures spanning decades or even the entire lifetimes of the individuals under study (2, 11, 12). The classic epidemiologic study designs, namely ecologic, case-control, cohort, and clinical trials, have been used to examine the relations of dietary practices to health and disease. These study designs and their strengths and weaknesses are described briefly in this article. A more complete, general discussion of these study designs can be found elsewhere (13, 14). An understanding of the strengths and weaknesses of epidemiologic studies, and of studies about diet in particular, is essential for their interpretation for the purposes of setting public policy.

EPIDEMIOLOGIC STUDY DESIGNS

The study designs used in nutritional epidemiology include ecologic studies, which investigate diet and disease at the population level, and several study designs (cross-sectional, case-control, cohort, and clinical trial) that address such questions at the level of the individual. In ecologic studies the unit of study is a population, which is usually defined by geography; in each population a measure of disease frequency is correlated with a measure of nutrient exposure. Such correlations are particularly useful in the generation of hypotheses about dietary factors that may be associated with variations in disease rates. Ecologic studies often have limitations because the measures of dietary exposure are frequently made for other purposes (eg, food disappearance data collected for economic purposes) and therefore are not necessarily in a form that accurately describes population exposure. More importantly, results from ecologic studies are inherently restricted to an evaluation at the population level and do not provide evidence about whether the individuals in the population who get the disease under study are those who experienced the exposure being measured in the population. Furthermore, in such studies it is difficult to account for other factors correlated with the exposure of interest, which may account for variation among study units in the observed rate of disease.

The epidemiologic study designs that address the relations between dietary exposures and disease risk at the individual level

[1] From the Department of Social and Preventive Medicine, University at Buffalo.

[2] Address reprint requests to JL Freudenheim, Department of Social and Preventive Medicine, 270 Farber Hall, Buffalo, NY 14214. E-mail: Jfreuden@ubmedc.buffalo.edu.

are the cross-sectional study, case-control study, cohort study, and clinical trial. For all of these, there are issues regarding the accurate measurement of dietary exposure, which are addressed in detail in other articles in this supplement. In all but clinical trials, there are also issues regarding the extent to which a study is able to account for confounding factors (factors correlated with both the exposure of interest and the disease outcome). Confounding is a concern when the confounding factor is unmeasured, measured with error, or so closely correlated with the dietary exposure of interest that the dietary factor and the correlated factor cannot be distinguished. Another issue is effect modification, any third factor that alters the association between exposure and disease. Of particular interest in this regard is effect modification by genetic factors, which may cause interindividual differences in the effect of an exposure on disease risk. A lack of knowledge about such factors may result in an effect of importance in a subgroup being masked in that group or, alternatively, an assumption that exposure is related to risk in the entire population when the effect is actually limited to a subgroup. An understanding of such variation in response to exposures may help explain inconsistencies among the findings of different studies that are related to differences in the distribution of genetic factors in the study samples. Recently there has been a proliferation of information on biologic variability related to genetics. These issues can be addressed by classical epidemiologic methodologies with information regarding the appropriate factors to be considered and a sufficient sample size.

Cross-sectional studies

In cross-sectional studies, exposure and disease are assessed concurrently in individuals selected from a defined population. Such studies frequently include a biologic measurement of disease or of nutrient exposure. An example is the work that correlated calcium intake with blood pressure measurements in healthy populations (15). Cross-sectional studies provide information about disease prevalence and factors associated with that prevalence. Information is collected about dietary exposures for individuals so that in cross-sectional studies, unlike ecologic studies, it is known whether the individuals with the disease are those with the exposure. For example, the calcium and blood pressure data indicated whether the individuals with higher blood pressure were those with lower intakes of calcium. The most important weakness of cross-sectional studies is that information is not available about temporal sequence (ie, whether the dietary exposure as measured is a consequence of the disease rather than a causal factor). In the calcium and blood pressure example, it is not clear from cross-sectional data whether individuals with higher blood pressure had altered their diets and their intake of calcium in response to a previous diagnosis of high blood pressure. Furthermore, cross-sectional studies only identify as diseased those individuals with prevalent disease; factors that affect survival can affect the assessment of causal relations.

Case-control studies

In case-control studies, individuals with recently diagnosed disease (case subjects) are interviewed regarding their past dietary intake. The focus of these interviews may be the period before their disease was diagnosed or may extend back to other periods in their lives. In addition to interview data, measures may be made on biologic materials such as blood, urine, or tissue samples. Data from case subjects are compared with data from a randomly selected sample of individuals (control subjects) from the nondiseased population out of which the case subjects arose. The control subjects may be selected so that their distribution on factors such as age and sex reflects the distribution of those factors in the diseased population.

Case-control studies allow for in-depth inquiry into factors related to risk of disease, with information collected on an individual basis. These studies are relatively efficient and quick, with generally lower costs than either cohort studies or clinical trials. Frequently, case-control studies are population based and findings can be generalized, at least to the population under study and often to a wider population. Case-control studies can focus on dietary exposures in the past, while allowing for the issues regarding measurement of diet in the past (2, 11, 12).

There are several disadvantages to this type of study. One is the difficulty in validating reported exposures, particularly exposures in the distant past. Another is the difficulty in identifying an appropriate group of control subjects, particularly in studies that are not based on geographically defined populations. Ideally, once the appropriate control group has been identified, all eligible case and control subjects would participate. In fact, this is rarely accomplished and there are concerns regarding the representativeness of those who do participate. Furthermore, there is the concern that because the case subjects are sick they will think about and report on their diets differently than will the control subjects, a process called recall bias.

Cohort studies

In a cohort study, a group of individuals is identified and their exposures to dietary factors and other risk factors of interest are measured. Assessment of dietary exposures may include both present and past dietary practices. These individuals are then followed over time to identify those who develop disease; the measured exposures are then used to determine predictors of disease risk. The validity of the study is related to the completeness of the follow-up. Ideally, all individuals in the original sample should be included in later measures of disease status. Most importantly, loss of cohort members to follow-up should not be correlated with exposure.

Studies of this type have the important advantage that the measurement of exposure precedes the identification of disease, thereby avoiding the issue of recall bias and allowing the identification of temporal sequence. Cohort studies require sample sizes on the order of thousands or even tens of thousands, with the exception of studies about diseases with high incidence rates. Additionally, they last longer than do other types of studies, because time is needed for a sufficiently large sample of individuals to develop the disease. Because of the large sample size, the degree of detail obtained by the dietary questionnaire may be more limited than for some case-control studies; both study designs may be hampered by errors in the reporting of dietary intakes. Both case-control and cohort studies provide important information about associations between exposures in a free-living population of humans. Both can address questions regarding long-term exposure previous to the interview, although assessment of past dietary practices is limited. In cohort studies, it is also possible to assess current diet and examine associations with disease over time as the cohort ages. Cohorts have the advantages that there is no concern about recall bias (except for retrospective cohort studies, those that collect at least some information regarding exposures after disease occurrence) and that there is stronger evidence of temporal sequence. For both cohort and case-control

studies, a key issue is uncontrolled confounding. Factors that are associated with both the dietary exposure of interest and the disease outcome and are not accounted for in the design or analysis will confound the results of either of these study types. With uncontrolled confounding, it is not possible to establish correctly the association between the dietary exposure and the disease.

Clinical trials

In clinical trials, confounding can be controlled for completely. Participants are randomly assigned to receive the dietary treatment under study or to receive a control diet or placebo. Participants are then followed for the development of the disease or the outcome of interest. If the sample size is sufficient and the assignment to groups is done correctly, it can be assumed that confounding characteristics will be evenly distributed between groups. For this study design there is no problem of recall bias in that assessment of exposure is made before the occurrence of disease. In addition, there is certainty that the exposure preceded the outcome. If the possibility that chance could explain the findings is ruled out, and a difference between the treated and untreated groups in the rate of outcome is observed, it is possible to assume that the difference is an effect of the treatment and that there is causality.

Clinical trials are often costly and difficult to conduct, and some questions may be difficult or impossible to address with a clinical trial because the exposure period is too long. Because a clinical trial entails changes in participants' lives, there are often ethical considerations that limit the questions that can be addressed. It is not ethical to administer substances for which there is evidence of deleterious effects, nor is it ethical to conduct a trial of a treatment and withhold it from the untreated group when there is sufficient evidence that the treatment is beneficial.

STUDY DESIGNS AND ISSUES OF INTERPRETATION OF FINDINGS

In the interpretation of epidemiologic studies, no one study can be considered definitive. Rather, information from several epidemiologic studies and from human metabolic, animal, and in vitro studies needs to be considered. For analytic epidemiology, the issue of consistency of findings is important because of the possibility of confounding by one or more unknown variable; different populations will likely have differences in the correlations of the factors of interest with the confounders. Nonetheless, when several studies are assessed to arrive at nutritional recommendations, some studies will carry more weight than others.

Because of the efficiency of case-control studies, there is frequently a much larger number of such studies available for assessment of an issue; this allows for comparisons of findings between different populations. Case-control studies may provide significant information because they often include more detailed interview data. In terms of diet, it is generally more feasible in case-control than in cohort studies to assess a larger number of foods and to include aspects of diet such as seasonality of consumption, cooking methods used, and detailed information about portion size. Although data collection with this degree of detail is technically possible in a cohort study, it is often not practical because of the large number of individuals involved.

Regardless of the study design, the relative validity of the dietary measure should be evaluated. For case-control studies, several factors need to be considered when evaluating the internal and external validity of the findings. One concern, as noted above, is the participation rate. In many populations there is difficulty in obtaining the cooperation of randomly selected individuals, which leads to concern about the representativeness of the study sample. The absolute participation rate in a case-control study may be less important than whether participation is correlated with factors of interest in the study (ie, that there is selection bias). With a participation rate close to 100% this issue is not a concern, but with a moderate or low participation rate, selection bias limits the extent to which the conclusions of the study are generalizable. One solution to this problem is to assess, to the extent possible, characteristics of both nonparticipants and participants and to compare those characteristics. In one small study of nonparticipation, for a crude measure of diet there were no differences in reported intakes of several categories of foods but there were differences in smoking habits between participants and nonparticipants (16, 17). Such considerations may need to be a more standard feature of case-control studies.

A second concern, as noted above, is the question of recall bias. For hypotheses that are not well known or for factors that are not considered to be related to health by the general population, this source of bias may be of less importance. That is, there may be a greater tendency for the memory of past exposure to be influenced by disease status when there is general knowledge that the exposure is related to disease risk. However, in most populations there is a general sense that diet is related to health, and thus recall bias may be of concern for all case-control studies regarding diet and disease. Some studies examined recall bias in case-control studies nested within cohorts (18–22). In these studies, a sample of participants in a cohort (both case subjects and appropriate control subjects), all with predisease measures of dietary intake, were subsequently reinterviewed about their diet in the past. The measured associations between diet and disease from the prospectively collected dietary data were compared with those from retrospectively collected dietary data for the same individuals. In 2 such studies from Canada and Sweden, there was little, if any, evidence of recall bias (18–20). In another analysis conducted in the Nurses' Health Study, there was evidence of some effect of recall bias on the assessment of the association between dietary fat and breast cancer (21). In that study, the estimate of risk for total fat intake in the highest quintile went from 0.87 for the prospective analysis to 1.43 for the retrospective analysis. This may represent the upper limit of recall bias, because this was an analysis of a well-known hypothesis studied in a presumably well-informed group of nurses. There was no evidence of recall bias in the assessment of the association between alcohol intake and breast cancer risk in that same cohort (22). These studies of recall bias were conducted in highly selected populations and may not completely address the issue as it occurs in a general population sample. Although the issue of recall bias is important in the interpretation of case-control studies, such studies appear to provide valid information about the relation of human diet to disease.

Findings from well-conducted cohort studies with good follow-up necessarily carry considerable weight in the assessment of epidemiologic and other scientific literature regarding the relation of diet to disease. Because of the advantages of predisease dietary intake assessment and the large number of individuals involved, they provide significant information about the relation of dietary practices to disease risk in free-living individuals. However, for the assessment of findings in order to set public policy, generalizability must be considered. Participants in cohort studies of dietary fac-

tors may be highly selected. Of those invited to participate, as few as 25% or fewer may agree to do so. There are some cohorts that are population-based with higher rates of participation [eg, the Beaver Dam cohort (23) and the first National Health and Nutrition Examination Survey (NHANES I) Epidemiologic Follow-up Study (24–26)]. In other studies, all participants may be volunteers [eg, the American Cancer Society cohort (27) and the Women's Health Initiative Observational Study (28)]. For many factors and study questions, it is appropriate to assume that such selection is not relevant, the underlying biologic mechanisms are the same, and findings from cohort studies can be generalized to the general population. However, there may be important differences in the confounding factors affecting observed relations and therefore affecting generalizability. For example, the Nurses' Health Study (29) includes women who are registered nurses; for the most part they are or have been employed and are therefore older on average at first pregnancy than the general population. People willing to participate in a cohort study may be more interested in health and health practices than the average person and therefore may have different dietary practices from the general population. Practices of the general population may be more difficult to study. Conversely, the special characteristics of the cohort may allow for the examination of factors that could not otherwise be examined. For example, in the Health Professionals Follow-up Study (30), the distribution of fat intake is lower and that of fiber intake is higher than in the general population. These characteristics allow for analysis of the effects of low fat and high fiber intakes.

Findings from clinical trials carry a great deal of weight but have some limitations, particularly regarding their usefulness for nutritional epidemiologic research. In any clinical trial there are always concerns about the way the trial was conducted, including issues of study participant compliance, drop-out rate, and adherence to the study protocol. With dietary interventions there may be additional concerns. For many kinds of dietary change, especially changes in macronutrient intake, it is difficult for participants to be blinded regarding their assigned group. This lack of blinding among participants and investigators may lead to some bias in the assessment of study outcomes. For changes in micronutrient intake, it is more feasible to give a treatment in pill form and therefore to blind participants to their treatment. Even then there may be difficulty in controlling participants' exposure to the nutrient of interest from other dietary sources; this is of particular concern if the dietary change under study is within the range of intakes found in the diet. In many trials, however, the treatment dose is far greater than the usual intake from dietary sources and thus variation in intake from the diet would not have an important effect on the study results.

Another important limitation of clinical trials is that interpretation of the findings, particularly negative findings, is necessarily limited. When there is no difference in an intervention study between the treatment group and the control group in terms of the measured outcome, it is not clear whether there would have been an effect of the treatment if the study conditions (eg, study duration or participant age) had been somewhat different. Because of the difficulties involved in maintaining individuals in a trial for a long period of time, many trials are limited in duration. Chronic disease may result from decades of dietary exposure, and a change in risk may not be evident within the period of a trial. This difficulty is often addressed by using an intermediary marker of disease risk. However, intermediary markers of risk are limited by the extent to which a factor that increases risk of the marker is also a risk factor for disease. Furthermore, a dietary factor that has an effect in another part of the pathway that is not related to the marker being measured may erroneously appear to be unrelated to the disease.

An additional limitation is the interpretation of findings regarding the dose of the treatment. If the findings are negative there is always the question of whether the findings would have been different if the dose had been higher or lower. Therefore, positive findings from trials carry considerable weight; negative findings need to be considered in light of other epidemiologic and scientific evidence before a lack of effect can be concluded.

Multicollinearity

One factor to consider when developing policy recommendations based on findings from nutritional epidemiologic studies is the analytic approach used in the studies. Intakes of nutrients tend to be highly correlated; that is, there is considerable *multicollinearity* for nutrients and other food components. Conclusions about the association of a single nutrient or small number of nutrients with disease risk may be flawed because of correlations that may exist among nutrients (3–5). Numerous other food components that are not necessarily nutrients may also affect disease risk. For some of these food components, complete food composition data may not be available. Carotenoids and flavonoids are examples of food components that are currently thought to influence disease risk.

In the analysis of data from both case-control and cohort studies, a true relation between a group of food components and a disease may be erroneously attributed to just one food component. Thus dietary recommendations based on such studies are more likely to be correct if they focus on dietary patterns rather than intakes of single nutrients. The nature of patterns of food intake is not well understood. Of concern is the pattern of foods consumed as well as the variation in pattern of intake over time (eg, by the day of the week) (1). Because there is considerable interaction among nutrients, both within dietary sources and in their effects on the human system, it may sometimes be more appropriate to examine the effects of changes in the pattern of food intake rather than changes in a single nutrient. There may be effects of diet on disease risk that can be explained by examining pattern of consumption but that cannot be explained by analyzing single nutrients in the diet (31).

This issue of analysis and interpretation of findings in terms of focusing on specific nutrients or on patterns of food consumption is also relevant when choosing questions to be addressed in clinical trials. Intervention studies of supplementation with one or more nutrients are far easier to implement than are studies of the effects of changes in dietary patterns. Nonetheless, there are some important strengths of studies that examine the effect of a change in dietary pattern. Such studies allow for the examination of a wider group of dietary factors simultaneously and sometimes may be important in identifying the responsible factor or factors. There can be considerable difficulties in the implementation of dietary changes in trials that focus on individuals; generally this entails in-depth training of study participants. Results from the Polyp Prevention Trial (32) and the Women's Health Initiative (28) should provide some indication of appropriate methodology and of the utility of focusing on dietary pattern in such trials. Smaller trials conducted for shorter periods of time may bypass the difficulties of educating study participants by providing food to participants and obtaining their compliance in eating only foods provided by the trial or other approved foods.

Another strategy that has been used is to focus on whole communities. Community-based trials have examined community-wide interventions regarding health promotion, including dietary changes. Such studies include the Minnesota Heart Health Program (33), the Pawtucket Heart Health Program (34), and the Stanford Five-City Project (35). Although there continues to be a place for the evaluation of individual nutrients or groups of nutrients in clinical trials, such evaluations may be best conducted after positive findings are obtained from a trial examining changes in dietary patterns.

Energy adjustment

Of the issues involved in identifying the correct model for analysis of multicollinear dietary data, particular concern has focused on controlling for energy intake. Willett and Stampfer (36) argued that diet and disease associations may in some cases result from effects on risk of differences in body size, efficiency of energy metabolism, or physical activity, all factors that are correlated with energy intake. Because intakes of most nutrients are correlated with energy intake, it is important to control for energy intake so that the true relations can be identified in these cases. Furthermore, adjusting for energy intake allows differentiation between the effects of total intake of a nutrient and the effects of diet composition. The correct methodology for energy adjustment has been discussed (37, 38), as has the effect of that methodology in identifying the correct model (38, 39). The correlation between the intake of some nutrients as estimated from a food frequency questionnaire and the estimate from biochemical markers of intake was shown to improve with adjustment for energy consumption (40).

As noted by Willett and Stampfer (36), the decision to adjust for energy intake depends on the model under consideration and is not necessarily appropriate for all analyses of nutritional data. In many epidemiologic studies, in which the underlying mechanism is not known, it may be most appropriate to examine the association of nutrients with risk both with and without adjustment for energy. Furthermore, for nutrients that are highly correlated with energy intake, particularly the macronutrients, the use of energy adjustment assumes that all shared variation should be attributed to energy and that only the variation that is independent of energy should be attributed to the second nutrient. The energy-adjusted variable would be most influenced by unusual diets (eg, high in energy and low in fat or low in energy and high in fat), which might lead to inappropriate conclusions.

For nutrients that are highly correlated with energy intake, it may not be possible to distinguish between the effects of energy intake and the effects of the other nutrient in observational epidemiologic studies. It may be necessary to perform studies in other populations in which the nutrient of interest is less strongly correlated with energy or to employ the controlled conditions of a trial or metabolic study to distinguish between the effects of the correlated substances. In some cases, such distinctions are not important in terms of public health in that recommendations can be made regarding the intake of foods. However, in circumstances in which individuals are likely to try to change their intake of a single nutrient, the lack of clarity about the nutrient of importance would be significant.

Multiple hypothesis testing

Because of the nature of epidemiologic studies, multiple questions can be addressed within the context of a single interview or collection of biologic materials. The cost and difficulty of conducting these studies make it most efficient to address more than a single question. In case-control studies, several exposures may be assessed in relation to a single disease. In cohort studies, both multiple diseases and multiple exposures may be examined. However, this practice of addressing multiple hypotheses has raised the question of whether findings that seem statistically significant are possibly the result of chance.

In response to this concern, Rothman (41) argued that the classical statistical approach to multiple hypothesis testing is not without trade-offs in that it results in a reduction in study power. Furthermore, the assumption underlying the concern about multiple hypothesis testing is that chance is the best first explanation for observations. Rothman argued that such an assumption is incorrect and that given the empirical finding of order in biologic phenomena, chance as an explanation is refuted with regularity. Instead, findings need to be examined critically and a scientific evaluation needs to be made as to their validity. Savitz (42) also addressed this issue and reached conclusions similar to those of Rothman. He also pointed out that it is difficult to assess the number of comparisons that are being made. Does one count the number of comparisons made in one study, in one part of a study, or in one investigator's career? In general the discussion of multiple comparisons is based on the use of statistical testing for decision-making, a process that is inherently arbitrary and incomplete (43). Because evaluation of epidemiologic findings to develop recommendations necessarily is based on replication in other studies, this issue is of less importance. When an unexpected finding is reported, the replication of that finding in several studies would indicate that chance is an unlikely explanation.

A related issue concerns findings that appear to be significant but that were not based on a priori hypotheses. This question of exactly what constitutes an a priori hypothesis can become muddied within the context of epidemiologic studies. Many studies are specifically designed to allow for the testing of hypotheses that are not even formulated at the time that the study is conducted (eg, biologic specimens may be stored to allow new questions to be addressed without incurring the cost of a new study). A related issue concerns unexpected findings in a trial; in a recent trial of the effect of selenium supplementation on skin cancer risk, there was no effect of the supplement on risk of skin cancer but there was reduced risk of prostate cancer in the treatment group (44). It would be imprudent to discount such a finding because there was no a priori hypothesis. For all of these situations, an assessment would need to be made about whether the study would have been conducted differently to correctly address the new hypothesis. For example, would additional information have been collected? Would the biologic samples have been handled differently? Would different efforts have been made to ascertain the additional outcome? Would these differences have affected the validity of the study findings? It would be appropriate to discuss which hypotheses were considered and how the ones presented were chosen from among those considered. Again, as for the case of multiple hypothesis testing, replication of findings in additional studies would render chance the less likely explanation for a finding. Findings from epidemiologic studies are useful in setting nutritional recommendations if the study was valid regardless of whether there was an a priori hypothesis.

ASSESSMENT AND SUMMARIZATION OF FINDINGS FROM MULTIPLE STUDIES

One tool for assessing the findings of epidemiologic and other studies is meta-analysis. Meta-analyses are generally of 2 sorts: those that quantitate findings based on the published literature and those that are based on reanalysis of pooled data sets. Meta-

analysis does not eliminate problems of confounding and bias in the original data. As with any evaluation of the published literature, to the extent that there is publication bias (ie, a tendency toward not publishing findings that are not statistically significant or are new or unpopular), that bias will affect quantitative and qualitative assessment of that literature. Conclusions may also be biased if a meta-analysis excludes some studies.

Some concerns about this tool have been raised (45–47). One concern regards the way that heterogeneity is handled; excluding studies that are heterogeneous from the analysis could produce misleading results. In some meta-analyses, study findings are weighted on the basis of the investigators' assessment of study quality. Such an assessment can be flawed and may not be appropriate (45, 48). Assessments themselves need to be evaluated and the effect of the assessment on the results of the meta-analysis needs to be indicated (49).

Like other statistical tools, meta-analysis is useful but limited. It is important that findings from such analyses not be given more weight simply because they are apparently more quantitative than other literature reviews. Furthermore, the precision of the estimate of overall effect needs to be evaluated carefully. In an examination of the concordance between meta-analyses and large randomized clinical trials regarding the same questions, findings from the clinical trials and the meta-analyses generally agreed about the direction of the effect. There was less agreement about the size of the effect (50). One opinion is that efforts to reduce study findings to a single value are generally flawed (47). In any case, meta-analyses are not necessarily more objective than other reviews; they are a resource for critically evaluating several studies (48, 49).

Meta-analysis does not address the issue of the biologic plausibility of the findings. Assessment of causality may be based on findings from human metabolic, animal, and in vitro studies in addition to epidemiologic findings. Metabolic studies of humans under controlled circumstances can provide insight into human biologic mechanisms, although the findings are necessarily limited by the length of the study and thus they do not generally address chronic effects of dietary factors on biochemical markers of risk. The animal literature is a source of insight about biologic mechanisms toward understanding human disease. There are frequently study questions that need to be addressed under the controlled conditions possible in animal studies that cannot be addressed ethically in human populations. Studies in animals may be limited by interspecies differences, which need to be considered when evaluating findings. However, when experimental findings are replicated in several species, particularly those most closely related to humans, they can be extrapolated to humans with more confidence.

Results from animal experiments may depend on certain characteristics of the experiment that limit generalizablity. For example, it is frequently cited that in rodents fat intake, particularly polyunsaturated fat intake, is positively related to mammary tumor development (51). However, it now appears that the effect of fat intake on mammary tumors is affected by the particular circumstances of the experiment. The observed effect is limited to experiments in which the rodents are virgin, are given a semipurified diet, and are fed ad libitum. When parous rodents are exposed to fat in a diet that includes natural ingredients or when there is restriction of energy intake, the effect is decreased or eliminated (52). This difference might indicate that the apparent lack of concordance between animal studies, in which there has been a consistent relation between fat intake and mammary tumors, and epidemiologic studies, in which the association between fat and breast cancer has been less consistent (53), may result from differing circumstances in the human and animal studies.

Another source of information for the assessment of diet in relation to disease is cell culture and other in vitro experiments. Such experiments allow for identification of disease mechanisms under highly controlled conditions. However, there is often considerable disparity among the findings from different cell lines, making it difficult to draw conclusions. It is also of concern how well the findings regarding mechanism translate to more chronic conditions in whole animals.

Just as the results of epidemiologic studies cannot be considered final because of issues of uncontrolled confounding, findings in animals are limited by the possibility of interspecies differences, and in vitro findings are limited by the very nature of the controlled experiments. When there is agreement among the findings reported in the in vitro, animal, and epidemiologic literature, it gives credence to the findings and to the conclusion that there is a relation of importance. Disparities between findings indicate that there needs to be an evaluation of the different kinds of evidence based on the strengths and weaknesses of the data. Lack of concordance may also indicate that there is missing information driving the differences and that further examination of the question is required to fully understand the dietary factor as it relates to the disease under study.

REFERENCES

1. Tarasuk V, Beaton GH. The nature and individuality of within-subject variation in energy intake. Am J Clin Nutr 1991;54:464–70.
2. Willett W. Nutritional epidemiology. 2nd ed. New York: Oxford University Press, 1998.
3. Gordon T, Fisher M, Rifkind BM. Some difficulties inherent in the interpretation of dietary data from free-living populations. Am J Clin Nutr 1984;39:152–6.
4. McGee D, Reed D, Yano K. The results of logistic analyses when the variables are highly correlated: an empirical example using diet and CHD incidence. J Chronic Dis 1984;37:713–9.
5. Mendola P, Marshall J, Graham S, Laughlin RH, Freudenheim JL. Dietary correlates of fat intake. Nutr Cancer 1995;23:161–9.
6. Hebert JR, Kabat GC. Differences in dietary intake associated with smoking status. Eur J Clin Nutr 1989;44:185–93.
7. Colditz GA, Giovannucci E, Rimm EB, et al. Alcohol intake in relation to diet and obesity in women and men. Am J Clin Nutr 1991;54:49–55.
8. Le Marchand L, Kolonel LN, Hankin JH, Yoshizawa CN. Relationship of alcohol consumption to diet: a population-based study in Hawaii. Am J Clin Nutr 1989;49:567–72.
9. Puccio EM, McPhillips JB, Barrett-Connor E, Ganiats TG. Clustering of atherogenic behaviors in coffee drinkers. Am J Public Health 1990;80:1310–3.
10. Subar AF, Harlan LC, Mattson ME. Food and nutrient intake differences between smokers and non-smokers in the US. Am J Public Health 1990;80:1323–9.
11. Byers TE, Rosenthal RI, Marshall JR, Rzepka TF, Cummings KM, Graham S. Dietary history from the distant past: a methodological study. Nutr Cancer 1983;5:69–77.
12. Rohan TE, Potter JD. Retrospective assessment of dietary intake. Am J Epidemiol 1984;120:876–87.
13. Hennekens CH, Buring JE. Epidemiology in medicine. Boston: Little, Brown and Company, 1987.
14. Langseth L. Nutritional epidemiology: possibilities and limitations. Washington, DC: ILSI Press, 1996.

15. Hamet P. The evaluation of the scientific evidence for a relationship between calcium and hypertension. J Nutr 1995;125(suppl):311S–400S.
16. Freudenheim JL, Marshall JR, Vena JE, et al. Premenopausal breast cancer risk and intake of vegetables, fruits and related nutrients. J Natl Cancer Inst 1996;88:340–8.
17. Graham S, Hellmann R, Marshall J, et al. Nutritional epidemiology of postmenopausal breast cancer in western New York. Am J Epidemiol 1991;134:552–66.
18. Friedenreich CM, Howe GR, Miller AB. Recall bias in the association of micronutrient intake and breast cancer. J Clin Epidemiol 1993;46:1009–17.
19. Friedenreich CM, Howe GR, Miller AB. The effect of recall bias on the association of calorie-providing nutrients and breast cancer. Epidemiology 1991;2:424–9.
20. Holmberg L, Ohlander EM, Byers T, et al. A search for recall bias in a case-control study of diet and breast cancer. Int J Epidemiol 1996;25:235–44.
21. Giovannucci E, Stampfer MJ, Colditz GA, et al. A comparison of prospective and retrospective assessments of diet in the study of breast cancer. Am J Epidemiol 1993;137:502–11.
22. Giovannucci E, Stampfer MJ, Colditz GA, et al. Recall and selection bias in reporting past alcohol consumption among breast cancer cases. Cancer Causes Control 1993;4:441–8.
23. Mares-Perlman JA, Klein BEK, Klein R, Ritter LL, Freudenheim JL, Luby MH. Nutrient supplements contribute to the dietary intake of middle- and older-aged adult residents of Beaver Dam, Wisconsin. J Nutr 1993;123:176–88.
24. National Center for Health Statistics. Plan and operation of the first National Health and Nutrition Examination Survey Epidemiologic Follow-up Study, 1982–84. Washington, DC: US Government Printing Office, 1987. [Series 1, 22. DHHS publication (PHS) 87 1324.]
25. National Center for Health Statistics. Plan and operation of the first National Health and Nutrition Examination Survey Epidemiologic Follow-up Study, 1986. Washington, DC: US Government Printing Office, 1990. [Series 1, 25. DHHS publication (PHS) 90 1307.]
26. National Center for Health Statistics. Plan and operation of the first National Health and Nutrition Examination Survey Epidemiologic Follow-up Study, 1987. Washington, DC: US Government Printing Office, 1992. [Series 1, 27. DHHS publication (PHS) 92 1303.]
27. Stellman SD, Garfinkel L. Smoking and tar levels in a new American Cancer Society prospective study of 1.2 million men and women. J Natl Cancer Inst 1986;76:1057–63.
28. The 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.
29. Hu FB, Stampfer MJ, Manson JE, et al. Dietary fat intake and the risk of coronary heart disease in women. N Engl J Med 1997;337:1491–9.
30. Giovannucci E, Rimm EB, Stampfer MJ, Colditz GA, Ascherio A, Willett WC. Intake of fat, meat, and fiber in relation to risk of colon cancer in men. Cancer Res 1994;54:2390–7.
31. Randall E, Marshall JR, Brasure J, Graham S. Dietary patterns and colon cancer in western New York. Nutr Cancer 1992;18:265–76.
32. Lanza E, Schatzkin A, Ballard-Barbash R, et al. The Polyp Prevention Trial II: dietary intervention program and participant baseline dietary characteristics. Cancer Epidemiol Biomarkers Prev 1996;5:385–92.
33. Mittelmark MB, Luepker RV, Jacobs DR, et al. Community-wide prevention of cardiovascular disease: education strategies of the Minnesota Heart Health Program. Prev Med 1985;15:1–17.
34. Carlton RA, Lasater TM, Assaf AR, et al. The Pawtucket Heart Health Program: community changes in cardiovascular risk factors and projected disease risk. Am J Public Health 1995;85:777–85.
35. Farquhar JW, Fortmann SP, Flora JA, et al. Effects of community-wide education on cardiovascular disease risk factors. The Stanford Five-City Project. JAMA 1990;264:359–65.
36. Willett W, Stampfer MJ. Total energy intake: implications for epidemiologic analyses. Am J Epidemiol 1986;124:17–27.
37. Howe GR, Miller AB, Jain M. Re: total energy intake: implications for epidemiologic analyses. Am J Epidemiol 1986;124:156–7.
38. Kushi LH, Sellers TA, Potter JD, et al. Dietary fat and postmenopausal breast cancer. J Natl Cancer Inst 1992;84:1092–9.
39. Beaton GH, Burema J, Ritenbaugh C. Errors in the interpretation of dietary assessments. Am J Clin Nutr 1997;65(suppl):1100S–7S.
40. Bingham SA, Day NE. Using biochemical markers to assess the validity of prospective dietary assessment methods and the effect of energy adjustment. Am J Clin Nutr 1997;65(suppl):1130S–7S.
41. Rothman KJ. No adjustments are needed for multiple comparisons. Epidemiology 1990;1:43–6.
42. Savitz DA, Olshan AF. Multiple comparisons and related issues in the interpretation of epidemiologic data. Am J Epidemiol 1995;142:904–8.
43. Poole C. Multiple comparisons? No problem! Epidemiology 1991;2:241–3.
44. Clark LC, Combs GF Jr, Turnbull BW, et al. Effects of selenium supplementation for cancer prevention in patients with carcinoma of the skin. A randomized controlled trial. Nutritional Prevention of Cancer Study Group. JAMA 1996;276:1957–63.
45. Shapiro S. Is meta-analysis a valid approach to the evaluation of small effects in observational studies? J Clin Epidemiol 1997;50:223–9.
46. Feinstein AR. Meta-analysis: statistical alchemy for the 21st century. J Clin Epidemiol 1995;48:71–9.
47. Bailar JC. The promise and problems of meta-analysis. N Engl J Med 1997;337:559–61.
48. Weed DL. Meta-analysis under the microscope. J Natl Cancer Inst 1997;89:904–5.
49. Thacker SH. Meta-analysis. A quantitative approach to research integration. JAMA 1988;259:1685–9.
50. LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large, randomized, controlled trials. N Engl J Med 1997;337:536–42.
51. Rogers AE. Diet and breast cancer: studies in laboratory animals. J Nutr 1997;127:9335–55.
52. Ip C. Controversial issues of dietary fat and experimental mammary carcinogenesis. Prev Med 1993;22:728–37.
53. Hunter DJ, Willett WC. Diet, body size and breast cancer. Epidemiol Rev 1993;15:110–32.