Gaps in dietary assessment methodology: meal- vs list-based methods1,2

Lenore Kohlmeier

ABSTRACT The development of new methods and approaches to dietary assessment has not kept pace with advances in related disciplines. One of the main gaps is the lack of innovation in incorporating cognitive psychology and computer game technology in developing useful, less biased tools for epidemiologic research on diet and disease. Another problem is the general lack of validation studies imbedded within research projects to provide data for assessment and correction of measurement error. Conceptual clarity, not only about the dietary components of interest but also about the relevant time frame for the disease under study, is often deficient in the designs of nutritional epidemiologic studies. Inadequate training of nutritional epidemiologists regarding the information requirements of a given study is also a problem, resulting either in “overkill,” i.e., the collection of unnecessarily detailed information at great cost, or in powerless studies, where the information is inadequate for testing the hypothesis of interest. Am J Clin Nutr 1994;59(suppl): 175S–9S.

KEY WORDS Diet history, computer-aided instruction, CAI, measurement error, food frequency

Introduction

If our methods of dietary assessment were satisfactory for our purposes, there would be little interest in convening meetings addressing the topic of dietary assessment methods. If one is truthful, few of us are satisfied or comfortable with the quality of our current methods. Recently, there has been a shift from blindly defending favorite methods to publishing validation papers on methods being used for epidemiologic purposes (1). Little effort has been expended on detailed evaluation for different nutrients and in different groups (2). Without a doubt, given better, stronger methods, nutritional epidemiology would be a stronger science; our results would be more convincing to ourselves and others, and we would need fewer studies and fewer resources to definitively come to conclusions about the relationships between food components and disease development.

Conceptual clarity

What are our true needs for assessment of dietary exposure in nutritional epidemiology? How do we determine whether we are falling short? How much information is enough to determine whether an association exists? What is needed to quantify an effect? When is the information collected with a given method so poor that the experiment should not be or not have been conducted? To address these issues a clear vision of an ideal methodology is needed. With the wealth of foods and nutrients one can study, and the dimensions of time at exposure, frequency of consumption, and amount of food consumed, not to mention preparation, one can easily lose track of the real goals for a given study.

In nutritional epidemiology, the underlying goal is to discover and verify the truth about effects of specific foods on disease development. Enough “methodological power” is needed to recognize or, even better, to quantify the effects of specific foods or food components on risk of disease. Questions of the type, “How much β-carotene is needed daily to halve the risk of breast cancer?” are typical.

With this as the goal, methods that at least can accurately discriminate individuals in regards to their consumption levels of specific goods are the bottom line. Beyond that, good quantitative measures of exposure over the time period of interest are desired. But what is that time period of interest? Conceptual clarity on this is one of the gaps in our science. Does one want to know the daily consumption of carrots during puberty? Does the consumption of 907 g (2 lb) carrots/wk in the summer confer the same protection as the consumption of 227 g (0.5 lb)/wk all year round? Generally, equality of average exposures is assumed, rather than designing studies to differentiate these exposures. Furthermore, the time and duration of exposure are critical aspects of conceptual clarity.

Five questions need to be asked before the study is designed and to be answered as conscientiously as possible. They are “What precisely is the exposure to be studied under the hypothesis being addressed?” “What age period in the life of the subject is of primary interest?” “How many years of exposures does one want to capture?” “How will this exposure be quantified in the modeling of risk?” and “How exact does the information need to be to test the hypotheses with sufficient power?”

Often, the answers are not known the questions are ignored, and the decision is made to capture everything one can afford to. This can result in an unevenness of information that may be a

1 From the Departments of Epidemiology and Nutrition, University of North Carolina at Chapel Hill.
2 Address reprint requests to L Kohlmeier, Departments of Epidemiology and Nutrition, CB# 7400, McGawran-Greenberg Hall, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7400.
fatal flaw in the study of the relationship. Other areas of epidemiology have addressed poorly measured exposures with success. An example is the area of tobacco epidemiology. Both the risks of smoking on carcinogenicity, cardiovascular disease, and respiratory disease, as well as the risks of passive smoking on these conditions have been discovered and quantified through epidemiologic research (3–6). Nutritional epidemiology needs an equivalent of the pack-years used in tobacco epidemiology—for commodities that are not as attractively packaged, not as addictive, not as expensive, and not consumed in isolation.

Ideal method

What would our ideal method look like? A clear vision on this would provide something against which the available alternatives can be measured. Certainly, an ideal method for epidemiologic purposes should provide an adequate degree of accuracy about the foods or nutrient consumed by individuals to test the hypothesis of interest in a powerful fashion. (The 24-h recall method is often referred to as being more accurate than other methods. This statement is not sensible without a reference to what it accurately describes. For example, it is well known that a single 24-h recall does not accurately reflect individual intake patterns. Thus, it is not an accurate measure for validating usual dietary behavior. It may, on the other hand, be an accurate measure of usual portion sizes for the foods that were incidentally consumed on that day.)

Further demands on an ideal are governed by logistic constraints. The method, because it needs to be applied to hundreds or tens of thousands of individuals, should be simple to use and foolproof. Furthermore, the subject burden should not be great, or, if so, in a positive cost-benefit relationship to the subject. Currently, this is seen in terms of spending as little time on the assessment as possible. Another alternative to this however could be a long method that the subject enjoys or appreciates. Most importantly, the method should be affordable, preferably inexpensive, when the entire costs of data collection, data processing, coding, and editing are considered (7, 8).

Current reality

Do the available methods represent this ideal? If we divide methods into list- or meal-based, we are separating methods that are administered by asking about frequencies of amounts of food eaten without a meal context (1), which is the general design of the food-frequency method, and those that ask about food consumed during specific meals or eating occasions. Because the subject generally fills out a food-frequency questionnaire without the assistance of an interviewer, the list-based methods are generally inexpensive. They are not necessarily rapid, and often extend to > 100 individual questions, many of which require calculations extending over a multitude of individual foods eaten at separate occasions (9). The degree of accuracy of information collected in this fashion will be addressed later.

Lists of foods are used because it is much simpler to handle categorical responses than to code and document open-ended questions and because it is easier to deal with a set sequence of questions than to allow or encourage wide ranging decision-tree-type responses. Lists are easily printed on paper and it is possible to automate the data entry completely with the use of optic reading devices. Therein lies the strength of list-based methodology. It disregards, however, the psychology of the individual responding to the difficult questions of how often one eats certain foods and in which amounts. How does one respond to the question of how many eggs one eats per week? This information is not on the top of most people’s minds. Normally one attempts to imagine usual meals, add up the eggs over the meals, and, where possible, to take the mixed dishes apart along the way to detect and account for the hidden eggs. Once this is done, we are in the realm of meal-based methods: such as the 24-h recalls, weighed dietary records, and diet histories, which, despite their complexity, according to Nelson (10), have numerous theoretical and practical advantages over prospective methods, 24-h recalls, and questionnaires.

There are cognitive issues to be addressed here in determining the tradeoffs when one compares current methods with the design of a better approach (11). Our interview process needs to be more carefully examined. People are not one-armed bandits. If we drop in a coin and pull the handle, we may truly get random results or biased results. Few of us think of what we eat in terms of a shopping cart full of food. We cannot just look down and see how much milk, yogurt, eggs, or oil we consume. Our eating sessions are much too important to us, experiences associated with emotions of pleasure or pain rather than with an inventory of items consumed.

Responses of individuals are influenced by the form of communication, cognitive processes, and other motivations. We often get the communication wrong in nutrition. Incidents of false conclusions based on language and interpretation are frequent (12, 13), even within countries where the language is constant. Communication is equally wrong if the subject feels compelled to lie, as when the language is not understood because of cultural differences.

Psychology of asking

Three things can happen when a question about diet or anything else is asked: there may be no response, an accurate response, or a wrong answer. Unfortunately, distinguishing between the latter two with dietary questions is particularly difficult. What is really going on here? The unintentional wrong answer may be due to memory (14, 15), concentration level of the subject, understanding of the communication by the respondent, or understanding of the respondent by the interviewer (16). The motivation for intentional wrong answers may be for many reasons, and the risk increases when the individual believes he knows what the interviewer is after or when he believes there is a quality scale involved (17). For example, with many food-frequency questionnaires on paper, categories are lined up under one another, and it may appear that answering towards the left of the categories is more desirable; for other foods, responses more to the right side of the paper are preferable. A better understanding of the motivations may move us towards breakthroughs with these methods.

The psychology of asking should consider that responses are governed by at least four subjective factors:

1) There is the potential threatening quality of the question as perceived by the respondent with concern that they will be judged on the basis of their response.

2) Independent of 1, the desire to conform to the expectations of another can affect responses.
3) There is the belief that verification is impossible. Particularly with historical dietary information, there is a greater freedom to report that which is believed to be desirable when there is no fear of being caught in an exaggeration. It is known that when a spouse is present at the interview, responses are affected (18).

4) Responses may be a function of the self-presentation strategies of the individual, which tend to be assertive if the interviewer is unknown to the subject and defensive if the interviewer or recipient of the responses is known to the subject (H Hofhuis, unpublished observation, 1992). This knowledge can be better used for improving the validity of responses to dietary questioning (19).

Improvement can be made in all dimensions of dietary-data collection: what was usually eaten, how often was it eaten, and how much was generally consumed. Usually we examine only the similarities between amounts reported using various methods. Smith (20) has made us aware of the extent of reports of eating foods that were known not to have been eaten. These phantom foods, or "intrusions," are very frequent in recall memory from 0 to 6 wk after the occasion, as shown in Figure 1. Four- to six-wk records of 170 subjects showed that 30–40% of the foods reported were phantom foods, and that this intrusion rate increased with the time elapsed since the end of the recorded dietary phase.

The question in nutritional epidemiology is how this effect can be reduced, adjusted for, or eliminated. Can we prevent people from reporting figments of their imagination as consumed foods? To prevent this, we need to understand the reasons for this behavior. Are the subjects trying to please the interviewer or perhaps fool themselves? Can this be minimized through the threat of verification, perhaps through the use of biomarkers of validity (21). Do we need to eliminate this effect or is it adequate to recognize the effect and adjust for it? Clues to minimizing the effect, which lend support to the belief in the superiority of meal-based methods over food-frequency questionnaires, can be found in a further experiment by Smith (20). He used different sets of cues to jog the memories of the subjects. A comparison of the use of food group–oriented cues and meal-oriented cues are presented in Figure 2. It was seen that, across the board, independent of time since the diet in question, intrusion rates were consis-

![FIG 1](https://example.com/fig1.png)

**FIG 1.** Mean intrusion rates. The mean intrusion rate is the number of foods reported as eaten but not recorded divided by the total number of food items reported multiplied by 100. The weeks since recording are reported. (Adapted from ref 20.)

![FIG 2](https://example.com/fig2.png)

**FIG 2.** Mean intrusion rates by method. The mean intrusion rate is the number of foods reported as eaten that were not recorded divided by the total number of food items reported, multiplied by 100. The meal and food group methods were sets of cues used to prompt memory of eating occasions. The weeks since the recording are presented. (Adapted from ref 20.)

tently lower when meal-based cues were used. This underlines the need for more extensive research on the cognitive aspects of our questioning methods. Response behavior is different with different approaches, but research on the extent of the differences has barely begun.

**Validation/calibration**

Validation is carried out to measure the extent to which a method actually measures that aspect of the diet it was designed to measure with the group being measured (22). There are many components of measurement error, some of which are characteristic of any measurement, others specific to dietary-assessment methods, still others specific to some methods but not observed with others, which may effect the validity of a method for measuring specific dimensions of diet. These five sources of error include random error, the error associated with assuming usual long-term historical intakes as based on current behaviors, error in portion size estimates, method-related bias in which consistent over- or underrepresentations occur, and differential misclassification. This occurs when different subjects within a study react differently to the method used. An example of differential misclassification is when obese individuals underestimate or underreport their dietary intakes whereas lean individuals report accurately or overreport. Another instance would be differential recall of prior diet of cases and control subjects such that one group tends to distort reports of prior behavior and the other does not.

Each source of error calls for a different method of correction or adjustment. One of the needless gaps in the current use of dietary assessment is the underuse of the statistical and mathematical procedures available to improve estimation of diet. The first three sources of error described above can be adjusted for if errors are not correlated between the methods used (23, 24). Differential misclassification is much more difficult to detect and address because the risk of the investigator falsely manipulating the findings is great (25).

Although valid dietary-assessment results are the fundamental basis of findings in nutritional epidemiology, dietary assessment...
methods cannot be validated currently for their ability to measure long-term intakes. This is because of the lack of a gold standard against which the method can be measured. Currently gold standards exist only for short-term energy intake as measured by doubly labeled water (26) or para-aminobenzoic acid (PABA) (27) for short-term protein intake. Other biomarkers of intake, such as serum concentrations of vitamins or fatty acid composition of adipose tissue, are at best indicators that can provide a basis for categorical comparisons of individuals, but not at this time for intraindividual validation (21).

Interest is, however, generally directed at validating methods on their ability to provide stable and accurate information on energy intakes, fat intakes, and the intakes of specific vitamins and other nutrients, for which no markers can quantitatively provide information on historical intakes (23).

Publications validating dietary methods generally, instead of validating studies, compare two data sets for precision but not accuracy. Generally, the two approaches are compared for correlational strength of association, and the authors conclude that the described method is adequate and valid. Correlations are good measures of neither accuracy nor precision. They are dependent on the breadth of variation within the group studied. They do not adequately inform the reader of the extent of misclassification of exposure histories to be expected under the method of measurement being validated.

Given that a gold standard for validation is not available, the best one can do is to calibrate one method against another believed to be more accurate for the purpose at hand. Ideally, the slope of the regression line between the two should penetrate the y-axis at the 0 point and show a slope of 1. Calibration will help recognize the difference between both methods, will allow adjustment of one to the levels of the other, and allow adjustment over time of any time-, season-, or measurement-related differences.

Calibration studies can, if the method being used as a basis for calibration contains errors that are independent of the method applied, be used to quantify and adjust for measurement error in the study. Calibration studies can also be used to apply a less demanding or less expensive method in a large study and, on the basis of the calibration results, adjust these findings to the predicted outcomes with the more demanding method. More on this can be found elsewhere (28). Recommendations of the commission of the Federal Health Office on nutritional epidemiology on the planning of studies to account for measurement error follow.

Innovation

Although technology has made great strides in the last 20 y, there has been little change in the development of better dietary assessment methods. Better methods are those that more accurately reflect eating behavior at the time of interest, are inexpensive, are simple to conduct, and minimize subject burden. Minimizing subject burden can be accomplished in one of two ways: by reducing the time it takes to ask boring or bothersome questions or by designing the interview and conducting it in a fashion that is attractive, even pleasurable to the subjects.

Modern computers have the capabilities of bringing sophisticated interviews, which are entirely under the control of the scientist, into all possible environments. Until now this capability has been largely in the field of computer-assisted telephone interviews (29). However, computer game development has engineered the complex and intriguing use of graphics and animation to liven and support the game process, methods that could also be built into interviews. Beyond this, computer-interactive interviewing allows immediate range checks to ensure that reports of extreme intakes are not due to recording error and can provide the subject with an informative report of their dietary intakes, as compared with recommended intakes, directly upon completion of the interview.

Currently, computers are being used almost always for food-to-nutrient conversion. They are often used for coding and entering information on dietary intakes (30, 31). They are seldom used instead of an interviewer for asking the questions in a given sequence and to a level of probing specified by the scientist. This latter case is the form of dietary interview applied in the national health examination survey conducted in 1991–1992 in the states of the former East Germany.

The automated diet history method

This computerized diet history was designed under the concept that a complete in-depth diet history of an individual can be obtained in a precise and standard form only if the multitude of possible questions are programmed into a complex decision tree so that any given response triggers the same follow-up questions on type of food, method of preparation, and amount.

Starting with taped interviews of dietitians conducting retrospective dietary assessments, a set of primary questions was developed. Their follow-up questions were determined based on the level of detail needed to completely define a food and determine its nutrient value.

Portable notebook or laptop computers and a set of household measures (plates, bowls, cups, glasses, and spoons) are carried by the interviewer to the scene of the examination. Subjects are introduced to the aim of the program (to capture usual dietary behavior) and are asked to report to the best of their knowledge the frequency of intakes of the foods that they eat at least once a month during all meals of the day (32).

The program leads the subject through a detailed set of questions about all foods eaten regularly during the day, meal by meal, beginning with foods (at the group level) usually eaten during breakfast. Then, for each food group reported, information on the specific item, the frequency of consumption, and, where appropriate, its preparation form, is collected. Finally, the amount eaten is asked, with amounts being entered in grams, or determined with household measures on hand at the interview.

As each meal is entered, a window at the bottom of the screen reminds the subject which meal and level is currently being asked. Beyond that, upon completion of each meal, the results are presented to the subject for review and editing.

Upon completion of responses regarding foods usually consumed during all of the eating occasions during the day, the subject is shown graphically how these usual intakes relate to recommended amounts. An eight-page printout provides high and low intakes, provides information on the main dietary sources of important nutrients in the subject’s diet, and summarizes the information presented during the interview.

This method has been shown to work in the field in large representative surveys of adults and in the clinical situation in case control studies. Further developments along these lines, which
integrate questions that have a solid base in cognitive science with modern advances in computer games, may be important developments in improving our measurement of dietary exposures.

Summary

Five gaps in our science have been identified: the need for more conceptual clarity on the design of nutritional epidemiology studies; improved standards of assessing and handling measurement error on analyses of nutritional epidemiologic studies; the regular inclusion of calibration studies within long-term nutritional studies; improved training of nutritional epidemiologists with the skills to design and analyze studies with consideration of their sources of measurement error; and finally, the need for great innovative advances in the development of new, better, more valid, and less burdensome methods of dietary assessment.

References