Approaches to analysis of dietary data: relationship between planned analyses and choice of methodology[1,2]

George H Beaton

ABSTRACT Dietary intake cannot be estimated without error and probably never will be. The nature and magnitude of the error depends on both the dietary data collection methodology and the subjects studied. The impact of particular types of error depends on the question being asked and the analytical methodology used to address it. Examples of these phenomena are presented in this review paper. The future lies in improved estimation and understanding of the error terms and in the development of analytical and statistical methods of coping with these error terms rather than with "improvements" in dietary methodology per se. *Am J Clin Nutr* 1994;59(suppl):253S–61S.

KEY WORDS Dietary methods, error in food intake estimation, variance, bias

Introduction

In the past decade there has been a great deal published about the errors in dietary data, about the pros and cons associated with particular dietary methods, and about erroneous conclusions drawn in the analysis of dietary data. This is understandable but unfortunate because it can easily leave the impression that dietary data are worthless. Conversely the drive to identify "sources of error" can lead to quite erroneous impressions about where one should invest money and effort to improve dietary information unless this is linked to an evaluation of expected impact in data analysis and interpretation.

We contend that *1*) dietary intake cannot be estimated without error (and never will be), *2*) collection and analysis of dietary data is essential if we are to pursue questions about the relationships between food use and health, and *3*) a very serious limitation at present is not the errors in dietary data but rather our failure to appreciate the nature of these errors, how they differ with choice of methodology of data collection, and what impact they have in specific strategies of data analysis. Thus, we argue that future progress may lie more in the development of an understanding of error structure and analytical methods that recognize and cope with that structure than in the development of new methods of data collection. If ever there was a time for full cooperation and collaboration between those interested in data collection and those charged with data analysis—the "nutrition" and "statistical" communities—the time is now.

I attempt to address two questions: *1*) What is the nature and magnitude of the errors in dietary data? and *2*) What is the impact in analysis and interpretation? I suggest that by appropriate marriage of choice of dietary data collection methodology and choice of analysis methodology, the utility of dietary information can be improved. I do not argue that we are now competent to address all of the issues in this field; however, I suggest that when an issue that affects interpretation cannot be resolved or circumvented, we can still profit by recognizing the sometimes very serious limitation that must be placed on interpretation and move ahead.

Nature and magnitude of error in dietary data

For simplicity I recognize two broad categories of error that may be found in dietary data (or any other data for that matter): bias and "random" error. By bias is meant a systematic under- or overestimation of intake in an individual or group of individuals. It can result from under- or overreporting of intake (intentional or unintentional) as is alleged to be the case with the US Department of Agriculture (USDA) national surveys (1). It can also arise from an error in food-composition databases, as was the case for the iron content of meats in the USDA databases until relatively recent years.

Note however that bias in reporting may be randomly distributed in the study population (some subjects underreport; others overreport) and for many analytical purposes behave like a random error, not a bias. Conversely, bias may associate with a trait in which there is analytical interest, eg, if thin and obese individuals exhibited opposite biases in reporting (or in recording, as might be imposed in methods that assume average portion sizes for all respondents). This bias could appear to be random in the population until one introduced weight status as a variable in the analysis then suddenly the bias covaries with a parameter of analytical interest and very misleading conclusions might be drawn.

Attention to "random" error in dietary data has developed more recently than did interest in bias. Although there is a very long history of interest in random error in the theory of measurements, it was not until the 1980s that it attracted major attention in relation to dietary data (2–16). Perhaps this was because it was only at that time that we began to ask very seriously "What is it we are trying to estimate?" and that vague concept

[1] From the Department of Nutritional Sciences, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada.

[2] Address reprint requests to GH Beaton, Department of Nutritional Sciences, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada M5S 1A8.

of "usual intake" came into vogue. Put simply, with interest in diet and chronic disease we realized that we were really interested in relationships between chronic intake and chronic disease. In the 1980s we began documenting the magnitude of day-to-day variation in intake within the same individuals, now a widely accepted fact of life. However, with this documentation came the sometimes hard-to-understand concept that the intake measured on a particular day was a very poor estimate of the average intake of that individual over a period of several weeks or months, a poor measure of "usual" or "chronic" intake of the individual, even though intake on the study day had been estimated without any error at all. Suddenly the nutrition community had to accept that "error" in the statistical sense did not mean "mistake" in the data collector's sense. With that awareness came the game of "how many days of data are needed to estimate usual intake" and a rash of papers that gave seemingly definitive answers. Only some of these papers (eg, 13, 17–20) were careful to address the key issue. Such a question has to be conditioned: "How many days of data does one need to do what with what precision?" I am afraid that too many users failed to consider the conditioned nature of the answer. Later in this paper some specific examples will be presented.

At this point it is appropriate to direct attention to what was seen at one time as the salvation—the diet-history or food-frequency methods. These methods were designed to ask "What do you usually eat?" or "What was your pattern of eating over a relatively extended period?" Surely these methods, although potentially susceptible to bias, would avoid the random error of day-to-day variation. They may. But it appears that they may introduce still another type of error. It appears that an error may arise when the respondent, who has one conceptual framework, attempts to use the questionnaire, which potentially has a different framework. A few years ago Beaton (21) attempted to ask about error terms in food-frequency data. He had been impressed by the very high test-retest correlations that many reported as evidence to support the method and wondered whether these correlations might be because subjects repeated the same mistakes when responding rather than because usual intake estimated this way was relatively constant within a subject. In pursuing this curiosity, Beaton was able to rework information from the validation study reported by Willett and colleagues (22–24). The original study design involved two self-administered food-frequency questionnaires administered ≈1 y apart and four 1-wk dietary records collected in the intervening period. By a very indirect analytical approach, Beaton (21) concluded that the food-frequency instruments appeared to have lost some of the real variance between subjects and at the same time introduced new (spurious) between-subject variance. He concluded that the loss of variance might have arisen because it is impossible to ask about all foods consumed and hence that the reductionism and summation needed in responding to the questionnaire of necessity resulted in some underreporting of variation. At the same time, he postulated that the new variance added had arisen from instrument errors that were systematic as far as the particular subject was concerned but that differed between subjects. It was these subject-specific errors, he hypothesized, that gave rise to the very high correlations between repeated applications of the same or similar instruments. Beaton applied this concept of error structure to some other published validation studies and found what he thought was supporting evidence for his concept. The bottom line of this analysis was that the food-frequency method of Willett seemed to have an error structure that might be equivalent to that of ≈3 d of dietary data collected as records or 24-h recalls. There is no reason to believe that the Willett methodology is unique in this regard. This finding should not be taken to imply that the food-frequency and diet-history methods are useless. Willett (personal communication, 1989) pointed out, quite correctly, that even if the food-frequency data were equivalent to only a few days of dietary data, they were data that could feasibly be collected and were certainly better than the 1-d intake estimates that some earlier epidemiologic studies had used. The point then is that the food-frequency method is not a panacea. It has its own error structure that will impact an analysis just as do the errors of other methods. An important distinction is that it is much more difficult to estimate the error structure of food-frequency data than, eg, the errors in repeated 24-h recalls. There has been an unfortunate tendency to assume that if you cannot see it, it is not there. Recently, Freedman et al (25) proposed an analytical approach that might be applied as part of validation studies, to provide estimates of the error terms in food-frequency questionnaires.

In summary, I am aware of no method with which food and nutrient intake can be estimated without some error. It is also apparent that the magnitude of error and the nature of the error varies with the particular method and probably also with the particular culture in which it is administered. We do have methodologies that may better estimate chronic energy expenditure, and thus energy intake. The doubly labeled water method is an excellent example (26). This method is not infallible; however, the magnitude of its error is much smaller than the error associated with food-intake methodologies. Unfortunately, the doubly labeled water methodology remains very expensive and available in only a limited number of research centers. Its main contribution to the evolution of food-intake methodology will lie in its use as a reference standard in examining the validity of food-intake methods and estimating their error components. We need some sort of "gold standard" rather than the process of comparing methods A, B, and C, each of which has its own, often unique, error structure. The doubly labeled water methodology comes close to offering this "gold standard"; however, it examines only total energy.

Analytical strategies and impact of error terms

Again for the sake of simplicity, I address only four broad categories of analytical approaches. As will become apparent, each of these categories addresses different questions; therefore, they are not necessarily alternatives. The objective of this section of the paper is to illustrate that a given type of error (bias or random) has different effects on different analyses and different analytical questions. The major analytical strategies to be discussed are *1*) estimation of group mean intakes or comparison of mean intake across categories of individuals (eg, by anthropometric status, by ethnicity, by income group), *2*) distributional analyses such as are common when the proportion of individuals with inadequate or excessive intakes in nutritional studies or exposure (eg, the 95th percentile of intake) to food components are estimated, *3*) correlation and regression analyses in which one attempts to relate level of intake and some outcome measure, and

4) categorical analyses in which individuals are categorized by intervals of intake and by some health outcome, as is common in epidemiologic studies of diet and cancer or other health conditions.

In all but the second of these analyses, systematic bias is acceptable as long as the only question is "Is there a relationship?" If the question is "What is the relationship?" then bias is a problem. For example, if we ask "does iron intake associate with anemia?" and we determine this by categorizing the population into quartiles or quintiles of iron and into anemic and nonanemic subjects, several statistical techniques can be used to determine whether the proportion of anemics differs across the quintiles of intake. If we increase or decrease intake in every quintile by 5 mg/d, we would still get the same answer to the question "Is there a relationship?" However, because hemoglobin concentrations have a physiologic maximum (and hence a discontinuity is to be expected if intakes are sufficiently high), at what level of iron intake do we cease to see a relationship? Obviously, the addition or removal of 5 mg Fe from each interval could change the answer. This is what happens with systematic bias. Often the epidemiologist is interested in asking "Is there a relationship?", whereas the nutritionist intuitively wants to ask "What is the relationship?"

Exactly the same type of phenomenon can be seen in correlation or regression analyses. Systematic bias has no effect on the regression slope or correlation coefficient; however, it does effect the intercept and hence the plotted regression line. Again we can make the distinction between asking "Is there a relationship?" and "What is the relationship?" This is illustrated in **Figure 1,** based on simulations of the impact of choice of dietary methodology on the apparent relationship between sodium intake and blood pressure (27). In both panels, all regression models were highly significant, although the proportion of variance explained (the adjusted R^2) varied from 2% to 8% depending on the model. All models might have satisfied the question "Is there a relationship?" However, for the nutritionist-biologist, the regression lines appear to describe different relationships. One might feel satisfied if the regression slope were correctly estimated (see bias effects in bottom panel of Figure 1) even though the intercept was wrong and hence the full description of the relationship was erroneous. Few could feel satisfied with the description when both the intercept and slope are wrong (*see* top panel). Interpretation requires some knowledge of the error structure and its likely impact.

Unfortunately, it would be very difficult to establish that bias is systematic. This is the nightmare of all analysts. If one knows that bias exists in a data set, then one must at least suspect that the bias may differ by characteristic of the individuals. In this situation, a comparison of intakes across categories of individuals may really reflect differences in bias rather than differences in intake! This is the real problem of bias: its detection signals the possible presence of the much greater problem of differential bias.

Let us focus on random error. Here we must accept that the random error we see is a mixture of methodologic error (eg, random bias in reporting by the same individual on different days) and true variation in intake from day to day. These cannot be separated easily. If we are comparing group mean intakes, then random error has no impact except that it increases variance and decreases statistical power. There is no a priori reason to think

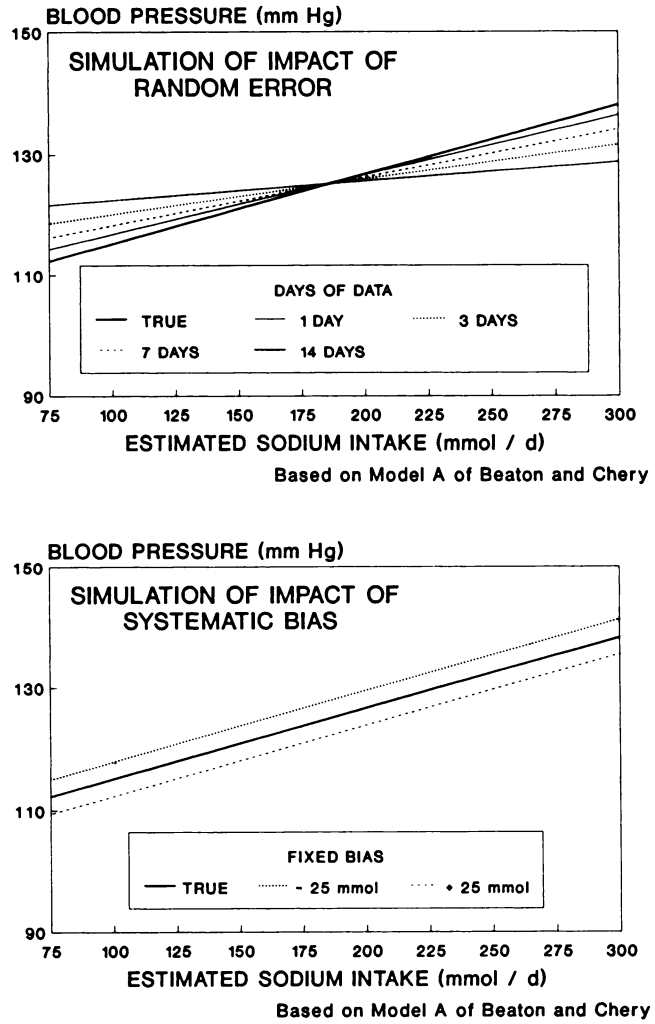

FIG 1. Simulation of impact of random error (top) and bias (bottom) on regression analyses. Random error was manipulated in accord with expectations of pooling multiple days of data for each individual. Bias was introduced as the removal or addition of 25 mmol Na to all intakes. For details of the simulation model itself, *see* Beaton and Chery (27).

that random error will bias the estimated group mean intake. In this type of analysis, then, random error is not a serious problem, and one can, for example, compare intakes across ethnic groups or across income strata with only 1 d of data if there is a sufficient number of subjects to provide statistical power. If power is low it can be improved by either collecting more data or by adding more subjects, the choice often being based on logistical considerations (3) as well as sampling issues (*see* below).

Distributional analysis concerns the prevalence of inadequate or excessive intakes lying beyond set cutoffs applied to the observed distribution of intakes. With this type of analysis the effect of random error can be serious, as is illustrated in **Figure 2,** which portrays hypothetical distributions of nutrient intakes. The mean intake had been set at 100 arbitrary units. The true variation in usual intake among individuals was assumed to have a CV of 25%, about what would be expected for energy or macronutrient intakes in free-living subjects. The within-person variation (day-to-day variation) was also assumed to have a CV of 25%, again comparable to that estimated for energy and macronutrient in-

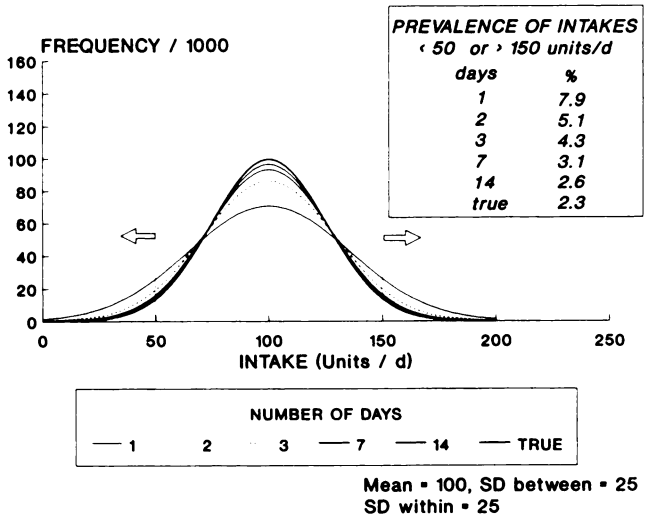

FIG 2. Simulation of impact of random error on observed intake distributions. The assumptions are that both between- and within-person variations have a CV of 25% and that both are normally distributed. Also shown is the impact that days of data collected would have on the proportions of individuals with low (< 50 units) or high (> 150 units) intakes.

takes. Figure 2 portrays the distributions of intake that would be expected if observations comprised single days or the average of 2, 3, 7, or 14 d of data for each subject. The impact of within-individual random variation is to inflate the observed distribution (note contraction of distribution as number of days increases). This then increases the proportion of individuals with intakes beyond the selected cutoff (here shown as the proportion with intakes < 50 units or > 150 units; in this model the proportions are identical in each tail). That is, the presence of random error in the intake estimate biases the estimated prevalence of inadequate or excessive intakes. As was discussed in reports from the National Research Council (28) and the Life Sciences Research Office (29), this can lead to very serious errors. It is an all-too-frequent problem in published reports of nutrition studies. The conventional approach to controlling this is also shown in Figure 2: collect more days worth of data and pool them for each individual, thereby reducing the within-individual random error term (obtaining a better and better estimate of usual intake of the individual). For many nutrients, increasing the number of days of data beyond three does not materially contract the distribution (the contraction is a function of the square root of the number of days). In an earlier paper (30), I made the rash statement that 3 d of data were enough. The statement referred to the distributional type of analysis, but it still comes back to haunt me as a generic statement, "Beaton said 3 d were enough." As was illustrated in Figure 1, 3 d of data are not adequate to resolve problems in correlation and regression analyses. We have, then, a relatively clear example of the pitfalls of asking "How many days of data do I need?" without also specifying what is to be done with the data.

There are statistical techniques by which the magnitude of the random error can be estimated and the distribution adjusted to at least reduce if not eliminate the effect. Thus, if ≥ 2 d of data on a large sample of subjects are used, it may be possible to obtain a good estimate of the true distribution of usual intakes. The National Research Council report (28) described a method that is applicable to distributions that approximate normality. Since then, Aickin and Ritenbaugh (31) described a technique applicable to the highly skewed distribution of vitamin A intakes, and are still evolving their nonparametric approach (32). Independently a group at Iowa State University has developed another approach to estimating the "true" distribution of intakes (33). Very recently, still another report looking at possible adjustment methods has appeared (34). It appears that for distribution analyses there are, or very shortly will be, acceptable methods of dealing with the problem of within-subject variation, at least in large data sets. Any such method will still require that at least a sample of replicated observations be available so that the within-subject variance can be estimated. It requires also that the "population" be initially stratified into relatively homogeneous, unimodal groups, conveniently by age and sex stratification but perhaps also by ethnicity or any other factor that is expected to have a major influence on food-selection behavior and hence the partitioning of variance. Although I often refer to "population," in this paper, I mean "population group," implying that such stratifications have been performed.

Note that the error discussed here is the error in estimation of usual intake. This may not be the real interest in every analysis. Consider for example the situation in which interest lies in acute exposure to a potentially harmful substance in food (29). One might want to estimate the distribution of single-day intakes, not the distribution of usual or chronic intakes. The relative importance of components of error shifts again.

In correlation and regression analyses, random error presents very serious problems. It alters the detected relationship as was illustrated in Figure 1. The observed correlation or regression coefficient was less than that which would characterize the true relationship between usual intake and the other variable. In simple linear regressions and correlations, the attenuation effect can be predicted from a knowledge of the error terms, and it is theoretically possible to either adjust the observed coefficients or at least state what they might have been without error (35). Rosner and Willett (23) have described the derivation of confidence intervals for the correlation coefficient, given the effect of random error in the dietary data. Further, in simple linear-regression analyses, the slope is biased if the random error term is in the independent variable but not if it is in the dependent variable (the R^2 of the regression is affected to same extent in either mode). In multiple-regression analyses the impact of error terms is extremely difficult to estimate (36, 37). Depending on the structure of errors, coefficients in multivariate analyses may be under- or overestimated. It seems unlikely that there will be simple statistical methods that can address the analytical problems that arise. We are compelled to look at data collection methods that reduce the random error component (usually implying the collection of more days of data).

The final type of analysis mentioned above involved categorizing people by their intake and then asking whether the incidence of some health outcome differed across intervals of intake. In one way or another this is a frequent approach in epidemiologic studies. Here random error means that a particular individual may be placed in the wrong interval. In turn that means that relationships between interval of intake and the outcome measure are muddied. Like the regression analysis, we again find an at-

tenuation of apparent relationships and encounter the danger of the false-negative conclusion. The impact of random error in these types of analyses has been discussed by Freudenheim and colleagues (38–40). There is a trick that is sometimes used. If the outcome can be measured with substantially less error than the intake, then one can reverse the question and ask whether subjects with different outcomes have different intakes. Now the random error in intake data is merely a nuisance, decreasing statistical power but not leading to attenuation of the estimated association. However, the question asked has been changed and many find that unacceptable. Nevertheless, reversing the question this way is an interesting way of testing to see whether a negative answer might have been attributable to major random error in the intake data.

Sampling of subjects and of time

Any discussion of errors in dietary data and their impact in analysis would be seriously incomplete if it omitted discussion of sampling. In dietary analyses there are two important aspects of sampling: *1*) sampling of people and *2*) sampling of time. One can think of two fundamental questions that must always be asked: "Who do you wish to examine?" and "What is the time frame you wish to examine?" Much has been written about sampling theory in the context of survey design, and it will not be reviewed here. One need only warn, again, that despite the adequacy of sampling plans, if planned respondents refuse to participate there must be very serious concern that the collected data no longer represent the intended group because of response bias or self-selection bias. [This was discussed with reference to a particular USDA survey (41). Low response rates have been reported in many large surveys. In the Nutrition Canada national survey (42), the overall response rate was estimated to be only ≈46%; there were major differences in the rate across strata of the population (eg, by sex and urbanization). Low response rates must always be a matter of concern particularly if the characteristics of nonrespondents are not known.] Those who respond are quite likely to be different from those that do not respond. Clearly that uncertainty must detract from the utility of the data for many (but not necessarily all) purposes.

Note that, in general, the impact of within-person variation can be reduced if the sample of subjects provides a rich between-person variation. It is often the ratio of within- to between-person variances that predicts the analytical effect. It follows that the most severe problems are found when analyses are performed in very homogeneous population groups.

The issue of sampling of time is perhaps less obvious to many of us. This paper has already addressed the matter of day-to-day variation and the need to collect a sufficient number of days to permit estimation of the error term. However, in that discussion it was presented as a "random" error. We know it is not! Virtually everyone accepts that weekends are different from weekdays; this presumably derives from the fact that food intake is influenced by social patterns and these tend to follow systematic cycles within individuals. Recently Tarasuk and Beaton (43) examined energy and macronutrient intake patterns of the 29 subjects who each contributed 365 d of intake data in the Beltsville 1-y dietary-intake study (44). They found definite day-of-the-week intake patterns in most of the subjects, but these patterns were not in phase across subjects. Each subject had his or her own pattern (43). This should not be surprising because it simply says that lifestyles and social customs are not identical in all individuals. The implication is clear. In many dietary studies investigators purposefully sample on weekdays and weekends and then attempt some sort of weighting to adjust intakes. They assume that this division of days represents a valid categorization of behavioral effects. Unquestionably it is better than ignoring weekend and weekday differences; however, on the basis of the work of Tarasuk and Beaton (43), it falls short of providing a correct representation for all subjects. Randomized sampling of all days of the week would be necessary to give true representation.

Those who have worked in developing countries are familiar with another issue of time sampling. In some settings there may be very pronounced cyclic shifts in energy intake, and perhaps energy expenditure, over time (45). In these situations it becomes very obvious that the investigator must ask what time frame a particular data set has represented.

In North America we still have seasonal patterns in the usage of particular foods, notably in the seasonal foods, but when food intake is converted to estimated energy and nutrient intakes, true seasonal effects are hard to find. They may be there but they are now relatively small. That does not mean that there are no long-term patterns in intakes of individuals. Again, from the Beltsville data by Tarasuk and Beaton (43, 46), there was evidence that many of the individual subjects did exhibit definable patterns in their energy and macronutrient intakes. One such subject is portrayed in **Figure 3.** This particular subject showed a pattern that could be modeled as a cubic regression explaining ≈16% of the total day-to-day variation in that subject. What is important to recognize is that the patterns detected were not the same across subjects and that for the group as a whole, no meaningful general pattern could be established. This was not a seasonal effect. As Tarasuk and Beaton (46) pointed out, within-person variation is highly individualistic both in magnitude and in pattern. The implication is that there is not a clear strategy for limited sampling that would serve to capture the correct variance estimate and hence correctly estimate an individual's intake. To estimate the

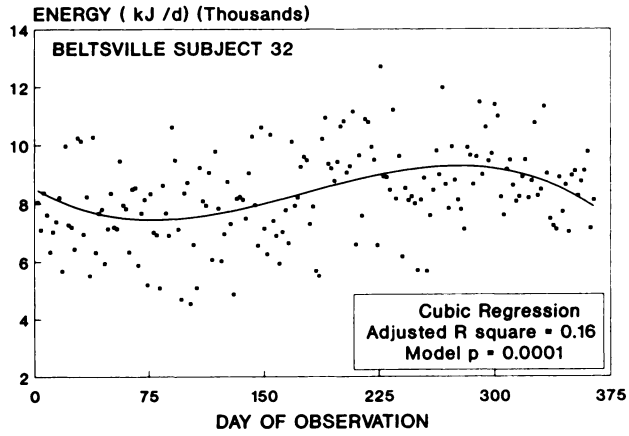

FIG 3. Distribution of estimated energy intakes as reported by one subject over 365 successive days. Only alternate days are plotted. The regression line was fitted by using all 365 d. [Redrawn from Tarasuk and Beaton (43).]

intakes of the Beltsville subjects and to correctly estimate the between- and within-person variances, one would have to randomly sample the 365 d for each subject or perhaps even better do a stratified random sampling such that you were confident that each of the days of the week was represented as well as the overall interval (47). The correct, but generally impossible, time sampling strategy is a random sampling of days across the time period of interest.

The message emerging here seems clear. At the level of study design, if one has knowledge of the likely error structure in dietary data, there are many things that can be done in sampling that will serve to reduce potential bias in estimates. These strategies do not eliminate variation. Rather they are intended to make the variation as close to a random phenomenon as possible. We have a better chance of dealing with random error in analysis than with an error structure that may relate to other variables of analytic interest.

Summation: impact of error terms in analysis and possible solutions

Table 1 attempts a summation of the discussion of error types and their impact in analysis. The symbols + to +++++ are meant to imply relative impact, not direction of the impact. When these are in parentheses, eg, (++), the meaning is that the analytical result may be wrong but that does not matter for many purposes.

The three types of error identified are systematic bias (all subjects showing the same error), random bias (direction and magnitude of bias varying between subjects), and random error (random variation within subjects over time). As has been pointed out these categories are gross simplifications. Hopefully they are useful for an overview even though true error structures are much more complex than this. Another type of analytical problem that may exist in dietary data is an association between the magnitude of within-person variance and the subject's mean intake, commonly contributing to the skew in data sets. Logarithmic transformation is often used to break these associations as well as to normalize the distributions.

The portrayal in Table 1 makes another critical simplification. It assumes that only the dietary variable has error, ie, it assumes that when the relationship between intake and some health outcome is examined, the outcome measure has been estimated without error or with a very small error component.

When the problems associated with random error are addressed, the conventional approach has been to increase the number of days of data collected. By pooling days, the residual random error component is reduced. However, for some of the relationships in which we are interested, the number of days of data required to reduce attenuation to an acceptable level and make detection of a relationship a reasonable expectation is prohibitive. For example, Liu et al (2) estimated that it might require ≈28 d of data to draw a conclusion, with reasonable confidence, about the presence or absence of a relationship between dietary cholesterol intake and serum cholesterol concentration. Although this remains the correct approach, particularly when it is coupled with random sampling of days in the time period of interest, seldom is it logistically feasible. What is emerging are statistical approaches that estimate the error components and use that information to adjust the analyses. At present there are satisfactory approaches for the adjustment of distributions that are not highly skewed (28, 35). More refined new approaches for application also to skewed distributions are evolving (32–34). These have been designed for application in the estimation of prevalence of inadequate or excessive intakes. There is at least hope that the recent evolutions in this field may hold promise for application in other analytical strategies in the future. It is in anticipation of such developments that the current National Health and Nutrition Examination Survey (NHANES) has taken steps to obtain replicated intake estimates, thereby permitting estimation of the random error term in the data set.

It is critically important that the nutrition community recognize that in dealing with the particular individual (as opposed to a

TABLE 1
Summary portrayal of impact of dietary data error terms in data analyses

Class of analysis	Type of error: Systematic bias*	Random bias*	Random*	Comments
Diet as independent variable				Systematic bias in intake estimates does not affect the ability to detect a relationship but it will lead to an erroneous description of the relationship
Simple regression or correlation	(+)†	++++†	++++†	
Categorical anlyses	(+)†	++++†	++++†	
Multivariate analyses	(+++)‡	+++++‡	+++++‡	
Diet as dependent variable				Random error reduces statistical power but does not distort the analytical result
Regression, correlation, categorical analyses	++	+	+	
Assess inadequate or excessive intakes				For group examinations there are techniques to adjust distributions if random error can be estimated.
Group prevalence	++++‡	(++)‡	(++)‡	
Assess particular individual	+++++	+++++	+++++	There is no recognized way to estimate or adjust for errors in the isolated individual. Valid assessment probably logistically impossible

* Very serious analytical problems arise if the errors relate to other analytical variables (continuous or categorical variables).

† If error terms are known, statistical adjustments are possible. *See* text for explanation of + and ( ) symbols.

‡ In multiple regression or comparable analyses, random error can attenuate coefficients as occurs in simple regression, or it can lead to overestimation of coefficients depending on correlation of errors across variables [*see* Liu (36, 37)].

group of individuals), there is no known way to estimate error terms and use them to adjust intake estimates before assessment. This means, in turn, that there really is no way of making dietary assessments of particular individuals more reliable except by collecting more days of data. Even then, serious problems remain. As already noted, the diet history and food-frequency methods probably do not accomplish more than would a few days of data and, for individual assessment, a few days of data is not adequate to permit reliable assessment. It follows that although we are continually improving our ability to deal with data from groups, we really have made very little progress in addressing the particular individual, and I see no promise of startling developments on the horizon. All we can do is recognize the magnitude of the error we face and be very cautious in our interpretations of data at the level of specific individuals.

One major development in the dietary assessment field has been the evolution of the "probability approach" to assessment (28). This gives recognition to the fact that nutrient requirements vary among seemingly similar individuals. Unlike the more common use of fixed cutoffs, this approach does not attempt to categorize intakes as adequate or inadequate in a binary manner. Rather it assigns a probability that a given level of intake is or is not adequate. For reasons given above, it cannot be reliably applied to particular individuals. Very often the error associated with the estimated intake imposes confidence intervals on the estimated intake that are broader than the distribution of requirements. One might end up with an assertion that the probability of inadequacy lies somewhere between 0 and 1 (95% CI). This is not a very useful statement! Furthermore, to apply the probability approach to the particular individual requires a relatively precise knowledge of the distribution of requirements. These constraints do not apply in application of the approach to groups as in the estimation of prevalence of apparently inadequate intakes.

Accompanying these developments in assessment approach, has been an accompanying evolution of the description of nutrient requirement estimates at least at the international level (48, 49). We now find an attempt to describe "requirements for what" as well as "requirements for whom." We find also increased attention to the estimation of mean requirement and the distribution rather than the previous focus on the "recommended intake" as a point in the upper tail of the distribution of requirements and the most recent development is the expression of suitable/safe/advocated group mean intakes rather than individual intakes [FAO/WHO/IAEA (50)].

The discussions in this paper have emphasized the statistical properties of dietary data and their implications for analyses. The reader must be aware of other, easily forgotten issues. One arises from our practice of reducing information on food use to estimated intakes of nutrients based on the variables that happen to have been included in food-composition tables. We must periodically stop and recognize that the generated proxy variables may or may not have captured the important characteristics of the original information gathered. Le Marchand et al (51), in a population control study of diet and cancer in Hawaii, had shown an apparent positive association of β-carotene intake with prostatic cancer risk in older men. Because of the implications of this finding, they chose to rerun the analyses, this time looking at the reported intakes of fruits and vegetables that were major sources of carotene in their subjects' diets. These new analyses showed that Papaya consumption had a strong positive association with prostatic cancer in men older than 70 y, whereas other yellow and orange fruits and vegetables, tomatoes, dark green leafy vegetables, and cruciferous vegetables did not show the association. They concluded that the originally observed β-carotene effect was in reality a papaya effect and did not involve intakes of β-carotene, lutein, indophenol, or other recognized phytochemicals. Clearly then, these additional analyses, based on food usage rather than derived carotene intake, reversed the conclusion previously drawn. The proxy variable, β-carotene, was flawed.

In dietary data there are many covariates among the computed nutrient intakes as well as with the food sources as represented in the example from Le Marchand et al (51). It is not easy to assess separately the associations with individual nutrients in the presence of strong covariation (whether that be covariation of true intakes or correlation of errors in estimated intake). We have long been warned about this problem in analyses of diet-disease relationships. Recently, Kushi et al (52) have shown, through comparative analyses of the same data set, that four reported methods of controlling for energy intake during analyses yielded different pictures of the association between fat intake and breast cancer. To establish which is the most correct analytical strategy requires a detailed consideration of the biological as well as statistical concepts involved and undoubtedly also requires consideration of the variance structure of the food-frequency methodology used to collect data.

The papers by Le Marchand et al (51) and Kushi et al (52) were cited merely as examples of issues that remain in the analysis and interpretation of dietary data. All of these developments mentioned in this paper mark the gradual blending of statistical and biological principles in the collection, analysis, and interpretation of dietary data. That, I suggest, is where the future lies. Unless and until we merge concepts between the nutrition and statistical fields, we (in both fields) will continue to make horrendous misinterpretations of dietary data and will continue to discredit the utility of dietary data.

Summary and conclusions

Some very general conclusions can be drawn. *1*) Dietary intake cannot now be estimated without error; it never will be! *2*) The nature and magnitude of the error depends on both the dietary data collection methodology and the subjects studied. *3*) The importance (impact) of error depends on the intended analysis and analytical question being asked. *4*) False-negative and false-positive answers can both be generated. *5*) There is no perfect dietary methodology but there are preferred methodologies for defined purposes. *6*) It is not meaningful to ask "How many days of data do I need?" without stipulating "To do what with what level of required confidence." *7*) There is little prospect for major development in dietary methodology. *8*) There is real hope that we will gain a better understanding of errors in dietary data and that we will improve our approaches to analysis and interpretation. We must do so because dietary analysis remains the keystone in understanding relationships between this major aspect of our environment and human health.

References

1. Mertz W, Judd JT, Reiser S, et al. What are people really eating? The relation between energy intake obtained from estimated dietary

records and intake determined to maintain body weight. Am J Clin Nutr 1991;54:291–5.
2. Liu K, Stamler J, Dyer A, McKeever J, McKeever P. Statistical methods to assess and minimize the role of intraindividual variability in obscuring the relationship between dietary lipids and serum cholesterol. J Chronic Dis 1978;31:399–418.
3. Beaton GH, Milner J, Corey P, et al. Sources of variance in 24-hour dietary recall data: implications for nutrition study design and data interpretation. Am J Clin Nutr 1979;32:2546–59.
4. Beaton GH, Milner J, McGuire V, Feather T, Little JA. Sources of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. Carbohydrate sources, vitamins and minerals. Am J Clin Nutr 1983;37:986–95.
5. McGee D, Rhoads G, Hankin J, Yano K, Tillotson J. Within-person variability of nutrient intake in a group of Hawaiian men of Japanese ancestry. Am J Clin Nutr 1982;36:657–63.
6. Rush D, Kristal AR. Methodological studies during pregnancy: the reliability of the 24 hour dietary recall. Am J Clin Nutr 1982;35:1259–68.
7. Black AE, Cole TJ, Wiles SJ, Winter F. Daily variation in food intake in infants from 2 to 18 months. Hum Nutr Appl Nutr 1983;37A:448–58.
8. Hunt WC, Leonard AG, Garry PJ, Goodwin JS. Components of variance in dietary data for an elderly population. Nutr Res 1983;3:433–44.
9. Jeor S, Guthrie H, Jones M. Variability in nutrient intake in a 28-day period. J Am Diet Assoc 1983;83:155–62.
10. Todd KS, Hudes M, Calloway DS. Food intake measurement: problems and approaches. Am J Clin Nutr 1983;37:139–46.
11. Gibson RS, Gibson IL, Kitching J. A study of inter- and intrasubject variation in seven-day weighed dietary intakes with particular emphasis on trace elements. Biol Trace Elem Res 1985;8:79–81.
12. McAvay G, Rodin J. Interindividual and intraindividual variation in repeated measures of 24-hour dietary recall in the elderly. Appetite 1988;11:97–110.
13. Nelson M, Black AE, Morris JA, Cole TJ. Between- and within-subject variation in nutrient intake from infancy to old age: estimating the number of days required to rank dietary intakes with desired precision. Am J Clin Nutr 1989;50:155–67.
14. Neuhaus JM, Murphy SP, Davis MA. Age and sex differences in variation of nutrient intakes among U.S. adults. Epidemiology 1991;2:447–50.
15. Sempos CT, Johnson NE, Smith EL, Giligan C. A two year study of middle aged women: repeated dietary records as a measure of usual intake. J Am Diet Assoc 1984;84:1008–13.
16. Sempos CT, Johnson NE, Smith EL, Gilligan C. Effects of intraindividual and interindividual variation in repeated dietary records. Am J Epidemiol 1985;121:120–30.
17. Borrelli R, Cole TJ, Di Biase G, Contaldo F. Some statistical considerations on dietary assessment methods. Eur J Clin Nutr 1989;43:453–63.
18. Marr JW, Heady JA. Within and between person variation in dietary surveys: number of days needed to classify individuals. Hum Nutr Appl Nutr 1986;40A:347–64.
19. Basiotis PB, Welsh SO, Cronin FJ, Kelsay JL, Mertz W. Number of days of food intake records required to estimate individual and group nutrient intakes with defined confidence. J Nutr 1987;117:1638–41.
20. Hartman AM, Brown CC, Palmgren J, et al. Variability in nutrient and food intakes among older middle-aged men: implications for design of epidemiologic and validation studies using food recording. Am J Epidemiol 1990;132:999–1012.
21. Beaton GH. Interpretation of results from diet history studies. In: Kohlmeier L, ed. The diet history method. Proceedings of the Second International Meeting on Nutritional Epidemiology. London: Smith-Gordon, 1989:15–36.
22. Willett WC, Sampson L, Stampfer MJ, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. Am J Epidemiol 1985;122:51–65.
23. Rosner B, Willett WC. Interval estimates for the correlation coefficient corrected for within person variation: implications for study design and hypothesis testing. Am J Epidemiol 1988; 127:377–86.
24. Willett W. Nutritional epidemiology. Monographs in epidemiology and statistics. Vol 15. New York: Oxford University Press, 1990.
25. Freedman LS, Carroll RJ, Wax Y. Estimating the relation between dietary intake obtained from a food frequency questionnaire and true average intake. Am J Epidemiol 1991;134:310–20.
26. Prentice AM, ed. The doubly-labeled water method for measuring energy expenditure: technical recommendations for use in humans. A consensus report by the IDECG Working Group. International Dietary Energy Consultancy Group. NAHRES-4, Vienna: International Atomic Energy Agency, 1991.
27. Beaton GH, Chery A. Evaluation of methods used to assess dietary intake: simulation analyses. Can J Physiol Pharmacol 1986;64:772–80.
28. National Research Council Subcommittee on Criteria for Dietary Evaluation. Nutrient adequacy: assessment using food consumption surveys. Washington, DC: National Academy Press, 1986.
29. Life Sciences Research Office. Estimation of exposure to substances in the food supply. Bethesda, MD: Federation of American Societies for Experimental Biology, 1988.
30. Beaton GH. What do we think we are estimating? In: Beal VA, Laus MJ, eds. Proceedings of a symposium on dietary data collection, analysis and significance. Massachusetts Agriculture Experiment Station, College of Food and Natural Resources, Amherst: University of Massachusetts at Amherst, 1982:36–48. (Research bulletin 675).
31. Aickin M, Ritenbaugh C. Estimation of the true distribution of vitamin A intake by the unmixing algorithm. Commun Stat Simulation 1991;20:255–80.
32. Aickin M, Ritenbaugh C, Woolf K, et al. A general method for adjusting dietary intake to obtain stable estimates of usual intake. Am J Clin Nutr 1994;59(suppl):303S (abstr).
33. Carriquiry AL, Jensen HH, Fuller WA, Guenther P. Methods for estimating usual intake distributions. Am J Clin Nutr 1994; 59(suppl):305S (abstr)
34. McConnell S. The robustness of estimates of the distribution of usual dietary intake to alternate sample sizes. Washington, DC: Mathematica Policy Research, Inc., 1992. (Report to the USDA Food and Nutrition Service, Office of Analysis and Evaluation.)
35. Life Sciences Research Office. Guidelines for use of dietary data. Bethesda, MD: Federation of American Societies for Experimental Biology, 1986.
36. Liu K. Measurement error and its impact on partial correlation coefficient and multiple linear regression analyses. Am J Epidemiol 1988;12:864–74.
37. Liu K. Consideration of and compensation for intraindividual variability in nutrient intake. In: Kohlmeier L, Helsing E, eds. Epidemiology, nutrition and health. London: Smith-Gordon, 1989.
38. Freudenheim JL, Johnson NE, Wardrop RL. Misclassification of nutrient intake of individuals and groups using one-, two-, three- and seven-day food records. Am J Epidemiol 1987;126:703–13.
39. Freudenheim JL, Johnson NE, Wardrop RL. The effect of controlling for kilocalories on classification to quintiles of nutrient intake: one- to seven-day food records as estimators of usual intake. Nutr Res 1988;8:745–53.
40. Freudenheim JL, Johnson NE, Wardrop RL. Nutrient misclassification: bias in the odds ratio and loss of power in the Mantel test for trend. Int J Epidemiol 1989;18:232–8.

41. US Department of Agriculture. Nutrition monitoring: mismanagement of nutrition survey has resulted in questionable data. Washington, DC: US General Accounting Office, 1991. (Report RCED91-117.)
42. Sabry ZI. Nutrition Canada National Survey: a report by Nutrition Canada to the Department of National Health and Welfare. Ottawa: Information Canada, 1973.
43. Tarasuk V, Beaton GH. The nature and individuality of with-subject variation in energy intake. Am J Clin Nutr 1991;54:464–70.
44. Mertz W, Kelsay JL. Rationale and design of the Beltsville one-year dietary intake study. Am J Clin Nutr 1984;40(suppl):1323–6.
45. Ferro-Luzzi A, Pastore G, Sette S. Seasonality in energy metabolism. In: Scurch B, Scrimshaw NS, eds. Chronic energy deficiency: consequences and related issues. Lausanne, Switzerland: Nestle Foundation, 1987:37–58.
46. Tarasuk V, Beaton GH. Day-to-day variation in energy and nutrient intake: evidence of individuality in eating behaviour. Appetite 1991;18:43–54.
47. Tarasuk V, Beaton GH. Statistical estimation of dietary parameters: implications of patterns in within-subject variation—a case study of sampling strategies. Am J Clin Nutr 1992;55:22–7.
48. Beaton GH. Human nutrient requirement estimates: derivation, interpretation and application in evolutionary perspective. Rome: Food and Agricultural Organization, 1992. (Food Nutrition and Agriculture 1:No.2/3:3–15.)
49. Beaton GH. Criteria of an adequate diet. Vol 2. In: Shills ME, Olson JA, Shike M, eds. Modern nutrition in health and disease. 8th ed. Philadelphia: Lea & Febiger, 1993:1491–506.
50. FAO/WHO/IAEA. Trace elements in human nutrition. Geneva: WHO (in press).
51. Le Marchand L, Hankin JH, Kolonel LN, Wilkens LR. Vegetable and fruit consumption in relation to prostate cancer risk in Hawaii: a reevaluation of the effect of dietary β-carotene. Am J Epidemiol 1991;133:215–9.
52. Kushi LH, Sellers TA, Potter JD, et al. Dietary fat and postmenopausal breast cancer. J Natl Cancer Inst 1992;84:1092–9.