Guidelines for use of dietary intake data

Sue Ann Anderson, Ph.D., R.D.
Senior Staff Scientist, Life Sciences Research Office,
Federation of American Societies for Experimental Biology,
Bethesda, Maryland 20814

Abstract  This article presents the guidelines from the report of the ad hoc Expert Panel on Guidelines for Use of Dietary Intake Data. The report was prepared under terms of a contract with the Life Sciences Research Office of the Federation of American Societies for Experimental Biology to provide the Food and Drug Administration with a working reference for evaluating dietary intake data. The guidelines represent the Expert Panel's conclusions on appropriate ways to interpret dietary data relating to (a) prevalence of consumption of particular levels of foods or food components, (b) comparisons of intake of different groups within the U.S. population, (c) time trends in consumption of foods or food components, and (d) relationships of intake of a food or food component to a given health outcome. The guidelines address general issues concerning dietary data collection and statistical concerns as well as specific issues important in interpretation of dietary data. J Am Diet Assoc 88:1258,1988.

Estimation of dietary intake and/or food available for consumption is a vital component of scientific inquiry in disciplines including nutrition, toxicology, and epidemiology. Dietary intake data provide information that can be used in combination with information on food components (for example, nutrients, food additives, or contaminants such as radionuclides, pesticides, or industrial chemicals) to estimate intake of or exposure to such substances in the food supply. Dietary data are also used extensively in epidemiological studies to relate exposure to food components to health status or disease incidence. Findings from those types of studies are important not only for their scientific value but also for their role in determining policies for food safety and public health.

This article presents guidelines from the report of the ad hoc Expert Panel on Guidelines for Use of Dietary Intake Data. The ad hoc Expert Panel was convened by the Life Sciences Research Office (LSRO) of the Federation of American Societies for Experimental Biology in response to a request from the Center for Food Safety and Applied Nutrition of the Food and Drug Administration to provide a working reference for evaluating data on dietary intake. The report conveys the Expert Panel's conclusions about appropriate ways to interpret dietary data, with an orientation to the regulatory needs of the Food and Drug Administration. In particular, the report addresses interpretation of dietary data in the context of (a) prevalence of consumption of particular levels of foods or food components, (b) comparison of intakes of different groups within the U.S. population, (c) time trends in consumption of foods or food components, and (d) relationship of intake of a food or food component to a given health outcome. The ad hoc Expert Panel considered guidelines for design of de novo studies with defined purposes and interpretation of secondary analyses made on existing databases.

The text of the report details the scientific rationale that the Expert Panel considered important in analysis and interpretation of dietary data for the areas of study stated above. The panel members recognized that their rationale and guidelines may change as further scientific progress emerges. The following brief synoptic statements are the guidelines on methodological, statistical, and biological issues addressed in the final report of the Expert Panel.

General guidelines relevant to all areas of study

Definition of problem
- The attribute or event of interest should be characterized as completely as possible.
- The target population should be identified as definitively as possible.
- The purposes for which the data are to be used and the precision needed should be determined.
- The adequacy of the dietary database for the food component(s) should be determined.

2Members of the Expert Panel were George H. Beaton, Ph.D., University of Toronto, Toronto; David R. Jacobs, Jr., Ph.D., University of Minnesota, Minneapolis; Nancy E. Johnson, Ph.D., University of Hawaii at Manoa, Honolulu; Kathryn M. Kolasa, Ph.D., East Carolina University, Greenville, NC; Frances A. Larkin, Ph.D., University of Michigan, Ann Arbor; Liang Liu, Ph.D., Northwestern University Medical School, Chicago; Karen J. Morgan, Ph.D., University of Missouri, Columbia (current address: Nabisco Brands, Inc., East Hanover, NJ); Milton Z. Nichaman, M.D., D.Sc., University of Texas School of Public Health, Houston; and Helen Smiciklas-Wright, Ph.D., Pennsylvania State University, University Park.
Sampling
- The sample should represent the target population as closely as possible. A probability sample of the target population is best if the results are to be used for making inferences about the larger population.
- The critical elements of sampling are the design and the sample size. Those elements will determine the precision of the estimates and the power of the statistical tests.
- Estimates based on small samples or weak experimental designs may have very wide confidence intervals, which will probably limit their utility. Statistical tests based on small samples or weak experimental design will have reduced power.
- Sample size is related to the variance associated with the estimate of the mean. Increasing sample size by increasing the number of subjects or the number of days of dietary data collection provides an approach for improving the precision of the estimate of the mean.
- If a study design uses a probability sample, then the results are generalizable to the target population. If the sample is flawed or is not a probability sample, the results may still be generalizable but on the basis of expert judgment rather than on mathematical theory. In the latter situation, the presence of biases is a concern.
- Independent corroboration of results is desirable if biases are or may be present. Independent corroboration of results is required if the generalizability of the results to the target population is based on expert judgment.

Methods for collection of dietary intake data
- The data collection method should reliably measure the variable that the investigator is attempting to measure. If the method used results in collection of data that are imprecise, incomplete, insufficient, or inaccurate, then use of the data will be compromised and will probably lead to misinterpretation.
- With appropriate quality control, quantitative daily consumption methods (replicated for 2 or more days to provide estimates of variance ratios from within the study data), semiquantitative food frequency methods, or even specific questions may provide useful data.
- Use of quantitative daily consumption methods is preferred when information is needed about actual intake. For estimates of usual intake, large intra-individual variability in day-to-day intake is the most serious concern with use of these methods, particularly when data are collected for only 1 day. Replicated collection of data with appropriate corrections may lower the total variance and lessen bias. Measures to control variability and bias are most effective if planned prior to data collection and should be sufficient to attain the necessary precision of the estimate.
- Data collected by use of food frequency methods may provide a less accurate estimate of actual recent consumption than data collected by quantitative daily consumption methods, but intra-individual variation is a much less serious problem. Such data may be subject to a systematic bias of subject estimation of intake. The estimate may be either higher or lower than actual intake.

Analysis of data
- Analysis of intake of a food component requires an adequate database for content of the substance in foods and a survey design and dietary method appropriate for estimating intake of foods containing the component.
- The precision of the estimate is determined by the study design, the sample size, and the extent of variability in the data.
- Analysis should take into account the inherent biases and limitations of the study design and data collection methods. Adjustments and corrections should be appropriate for the database.
- Even after adjustments and corrections, results must be assessed for adequacy of individual assumptions and for any remaining biases before an analysis is accepted.
- In studies utilizing data collected for a single day, external estimates of variance ratios should be applied very cautiously, at least until more information is available on sources of intra-individual variation.
- Analysis of data collected in studies based on experimental designs employing complex sampling procedures requires the use of appropriate sample weighting factors.
- Cautious interpretation is required for data based on a single day's data collection or a food composition database judged to be inadequate.
- Interpretation of results should be based on biological and statistical significance rather than on statistical significance alone.

Guidelines specific for particular areas of study

Prevalence of consumption of particular levels of foods or food components
- Use of either fixed cutoff points or a probability approach for making prevalence estimates should have a sound biological basis.
- When prevalence estimates are made, generalization to the total population is often required; therefore, probability sampling is the method of choice.
- Estimation of centiles may be biased by the enlarged variance resulting from a large amount of intra-individual variation. Higher centiles (greater than the median) and lower centiles (less than the median) are overestimated and underestimated, respectively.
- A systematic bias in the data resulting in overestimation of intake shifts the entire distribution curve to the right. Prevalence will be overestimated from the upper tail of the curve and underestimated from the lower tail of the curve. Conversely, a systematic bias in the data resulting in underestimation of intake shifts the entire distribution curve to the left, producing an underestimation or overestimation of prevalence, depending on which tail of the distribution is of interest.
- Per capita food availability data cannot be used for making prevalence estimates because distributions of intakes of individuals cannot be derived from such data.

Comparison of intakes of different groups within the U.S. population
- Probability sampling is desirable for estimating differences in absolute levels of intake for groups. However, when it can be assumed that the difference under investigation is independent of sample selection criteria, differences among groups may be representative of differences in the total population. Even when a sample

---

3Survey or probability sampling is the discipline concerned with selection, observation, and analysis of samples from a population in order to make inferences about the whole population.
is not representative of the total population (i.e., sampling only certain socioeconomic groups), differences among groups may still be important.

- Estimates of the means must be sufficiently precise to detect differences among the groups compared. Sample sizes must be large enough to ensure adequate precision in estimating means.
- If sample sizes are small, variances of sample means will be large. This may preclude detection of differences among means of groups.
- Single-day dietary intake data may be adequate for comparisons of groups if the samples are sufficiently large.
- Sampling and data collection methods should be equivalent for the groups being compared.
- If instruments are equivalent, food frequency data for dietary intake may be useful for comparison of group means because systematic biases will probably cancel out. However, instruments for all methods for collection of dietary intake data may be subject to different interpretation by different cultural groups. This type of bias will not cancel out in comparison of intakes among different cultural groups. Food frequency instruments are more subject to this problem than are quantitative daily consumption instruments.
- Per capita food availability data cannot be used for comparison of intakes of groups because distributions of intakes of individuals within the groups cannot be derived from such data.

Time trends in consumption of foods or food components

- The conceptual basis for variables should be constant over time.
- Sampling procedures should be equivalent across all time points studied.
- Methods should be equivalent across all time points studied.
- Values in food composition tables may change over time because analytical methods have changed or because the composition of certain foods actually has changed. Food composition databases used for data analysis should accurately represent the actual composition of foods available at each time point.
- Changes over time may occur because the target population has changed over time.
- Time trends observed over a short time interval should be interpreted cautiously until confirmed for a longer time span. Differences smaller than methodological error may not be detectable, particularly if a time trend is based on only two time points.
- Per capita food availability may be useful for estimating time trends in consumption of foods or food components.

- The statistical considerations pertaining to comparison of groups and analysis of relationships also apply to intervention trials and clinical studies.

Classification

- When dietary intake is used as the classification variable, the large amount of intra-individual variation in intake observed with use of quantitative daily consumption methods may result in a large likelihood of misclassification. Similarly, estimates made by use of frequency methods are sufficiently imprecise that classification by health outcome may result in fewer misclassifications. Therefore, caution should be exercised if intake of dietary components is to be used as an independent variable.

Bivariate analysis

- In bivariate analyses (e.g., ANOVA, simple correlation, regression), misclassification of subjects attenuates associations.
- Systematic bias has little effect on bivariate analyses. Systematic bias is unlikely to attenuate the correlation coefficients in bivariate analyses, but it will yield an error in the intercepts in linear regressions.
- In bivariate analyses, the impact of intra-individual variation and methodological error varies among analyses.
- Statistical correction of correlations and regressions is possible but can be misleading. Therefore, it must be done with caution.

Multivariate analysis

- Intra-individual variation results in serious distortion for multivariate analyses. Statistical correction is generally not feasible, and an increased number of replicates is the only way to reduce the distortion.
- Systematic bias has little effect on multivariate analyses. It will not preclude detection of relationships, although the description of the relationship (e.g., intercept in multiple linear regression) will be in error.
- Methodological error in frequency methods can result in serious attenuation for multivariate analyses. Neither statistical corrections nor increasing the number of measurements reduces the error in the measures of association.

Factor analysis

- Although factor analysis potentially provides a means to deal with multicollinearity in intake data, its value as an analytic tool for use with dietary intake data remains to be determined.

Applications

The ad hoc Expert Panel on Use of Dietary Intake Data prepared these guidelines as an aid to policymakers in using and interpreting dietary data for regulatory or legislative purposes in the contexts specified in the introduction to this article. Additionally, the Expert Panel anticipates that the guidelines will prove useful to investigators for the design of new studies and for secondary analyses of dietary data gathered for other primary purposes.