Collection and Analysis of Intake Data from the Integrated Survey

Suzanne P. Murphy

Cancer Research Center of Hawaii, University of Hawaii, Honolulu, HI 96813

ABSTRACT Intake data from the combined CSFII/NHANES survey will be used for many different purposes, each with specific data requirements and appropriate analytic methods. For monitoring and surveillance, the availability of Dietary Reference Intakes will allow estimates of the prevalence of inadequate intakes and the prevalence of intakes with a risk of adverse effects. The accuracy of the nutrient intake estimates will be enhanced by the 5-pass dietary recall methodology, availability of quantified dietary supplement intake data and expanded food and supplement composition data. Food-level dietary monitoring will be improved by using new databases to calculate servings of food groups from the Food Guide Pyramid and intakes of food commodities. Another major strength of the survey is the ability to relate intake data to health measures for individuals. Inferences will continue to be limited by a lack of usual intake for each individual, but the attenuation will be less with 2 d of data than with only 1 d, as in the past. Better data collection and analysis will also lead to more informed nutrition policies and programs. Innovative methods of analyzing the data should be investigated to minimize the effects of underreporting, provide better estimates of usual intake at both the group and individual levels and accurately combine nutrient intakes from foods and supplements. Future modifications to the intake collection methods might be considered to allow larger sample sizes for certain subgroups, more detailed information on supplement use, an expanded food frequency questionnaire, a different number of recall days and incorporation of diet and health knowledge questions. J. Nutr. 133: 585S–589S, 2003.

KEY WORDS: dietary surveys • dietary assessment • nutrition monitoring • Dietary Reference Intakes

National nutrition survey data are used by many types of people for a wide variety of purposes. A detailed list of these uses is given in the overview (1). For simplicity, these are combined into three broad categories: nutrition monitoring and surveillance, relating diet to health and informing nutrition policies and programs. Each of these uses of survey data has somewhat different data requirements and often also uses different analytic approaches to examining the data.

Using survey data for nutrition monitoring and surveillance

Nutrition monitoring and surveillance requires, at a minimum, data on total daily nutrient intakes for all nutrients of public health importance. The dietary adequacy of a group of people is assessed by comparing their intakes with recommended nutrient intakes. The possibility of excessive intakes can also be evaluated for the group by comparing their intakes with recommended maximum intakes. In addition, monitoring may involve evaluation of intakes of foods and food groups by comparing intakes to current food-level guidance such as that offered by the Food Guide Pyramid (2).

Using the DRIs to assess nutrient intakes. The new Dietary Reference Intakes (DRIs) provide the recommended nutrient levels that are used for assessing intakes. DRIs have been set for most of the micronutrients (3–6) and recommendations for macronutrients will be available shortly. A report has also been published on methods for using the DRIs to assess intakes of both individuals and groups (7).

Nutrition monitoring involves assessing intakes of population groups. Three types of DRIs may be used for this purpose. The Estimated Average Requirement (EAR) is used to estimate the prevalence of inadequacy within the group, usually as the percentage of the population below the EAR. For nutrients for which an EAR could not be set, the Adequate Intake (AI) is used as the desirable mean intake of the group—mean intake at or above the AI implies a low prevalence of inadequacy.

1 From the workshop “Future Directions for the Integrated CSFII-NHANES: What We Eat in America—NHANES” held on June 20–21, 2002, in Rockville, MD. This workshop was sponsored by the Office of Dietary Supplements, National Institutes of Health, U.S. Department of Health and Human Services (DHHS) and the Agricultural Research Service, U.S. Department of Agriculture (USDA) and cosponsored by the National Institutes of Child Health and Development, National Institutes of Health, and the National Center for Health Statistics, Centers for Disease Control and Prevention, DHHS, and the Cooperative State Research, Education, and Extension Service and the Economic Research Service, USDA. Guest editors for this workshop were Johanna Dwyer, Agricultural Research Service, USDA; Mary Frances Picciano, Office of Dietary Supplements, National Institutes of Health, DHHS; and Daniel J. Raiten, Office of Prevention Research and International Programs, National Institute of Child Health and Human Development, National Institutes of Health, DHHS.

2 E-mail: suzanne@crch.hawaii.edu

3 Abbreviations used: AI, Adequate Intake; CSFII, Continuing Survey of Food Intakes by Individuals; DRI, Dietary Reference Intake; EAR, Estimated Average Requirement; NHANES, National Health and Nutrition Examination Survey; UL, Tolerable Upper Intake Level.
Finally, the Tolerable Upper Intake Level (UL) is used to estimate the prevalence of potentially excessive intakes—those that may be at risk of adverse effects. The prevalence of excessive intakes is estimated as the proportion of the population above the UL. The methods for performing these assessments are covered in more detail in the report on applications in dietary assessment (7). A sample assessment for children ages 4–8 y in the 1994–1996 Continuing Survey of Food Intakes by Individuals (CSFII) is shown in Table 1. The DRIs will allow a variety of new methods for nutrition monitoring and surveillance and in turn will permit more informative interpretations of findings from the combined survey. For example, knowledge of the prevalence of potentially excessive intakes may lead to interventions to reduce very high intakes, whereas information about the prevalence of inadequacy may inform programs to increase intakes of certain nutrients. The availability of prevalence estimates quantifies the magnitude of the concern much better than in the past, when only mean intakes were examined and compared with Recommended Dietary Allowances (8). However, for nutrients with an AI rather than an EAR, prevalence of inadequacy cannot be determined and it is possible to state only whether the mean is desirable (above the AI) or less than desirable (below the AI).

As is covered in more detail by Carriquiry (9), group intakes should be assessed only after the effect of day-to-day variation in intakes (within-person variation) is removed from the intake distribution (10). Without this type of adjustment, the intake distribution is too broad because it reflects not only variation between persons but also variation within persons. Through the use of information obtained from the 2nd d of intake, the true distribution of usual intakes can be estimated for groups within the combined survey. Software has been developed to perform these adjustments (11). After a proper adjustment the tails of the distribution will usually be smaller and the proportion of the group below the EAR and the proportion above the UL will also be smaller. Such an adjustment is not necessary when comparing mean intakes with the AI because the adjustment does not affect the mean intake. Thus, it is not necessary to remove the effect of day-to-day variation from the distribution before making this comparison.

Nutrition monitoring also can involve comparisons of dietary adequacy across subpopulations within a group. These subpopulations might be defined based on, for instance, gender, age, income and use of food programs. For example, it might be of interest to compare intakes of those receiving food stamps with intakes of those not receiving them. An obvious first step is to adjust the intake distribution (to remove day-to-day variation) for each subgroup and then compare the prevalence of inadequate intakes for each. Standard statistical tests can be used to determine whether the prevalences are statistically different. However, it is almost always desirable to adjust for possible confounding variables before undertaking this type of comparison. In the food stamp example, one might wish to adjust for income (because even within a population of families eligible for food stamps, those that choose to obtain them are, on average, at a lower income level than those that do not). To make such an adjustment for confounding variables, it is necessary to perform a multivariate analysis of the prevalence of nutrient inadequacy. An approach to this type of analysis is described in the report on using the DRIs to assess intakes (7). However, this approach has not been tested using national survey data, and further investigations into its strengths and limitations are needed.

### What is needed for accurate nutrient intake estimations for groups?
Several aspects of intake data collection must be considered when ensuring accurate nutrient intake estimates. A primary aspect is to collect unbiased food intakes (i.e., the correct items and the correct portions). The combined survey will use 24-h recalls to collect 2 d of intake data for each participant. The recall method has been associated with substantial underreporting of intakes by many participants as well as some cases of overreporting (12,13). Such biases in the intake data can overestimate the true prevalence of nutrient inadequacies and also inflate the prevalence of potentially excessive intakes. Considerable research has been conducted to increase the accuracy of the recall methodology and has resulted in the new 5-pass approach. In particular, this method appears to reduce the level of underreporting that has been found in previous surveys. A study is now under way to determine how much, if any, underreporting bias remains with this approach.

In addition to collecting unbiased food intake data, it is also necessary to use an accurate food composition database to calculate nutrient intakes. The U.S. Department of Agriculture (USDA) Nutrient Database for Individual Intake Surveys is maintained for use in national nutrition surveys (14) and will be used with the food intake data from the combined survey. This database is derived in large part from the USDA Nutrient Database for Standard Reference (often referred to as the SR database) (15) but is expanded to contain recipes for mixed dishes and to include imputed values when an analyzed nutrient level is not available for a food. This valuable database is updated to include new analytic data as they become available.

### Table 1

<table>
<thead>
<tr>
<th>Nutrient</th>
<th>Dietary Reference Intake2</th>
<th>Prevalence of inadequacy</th>
<th>Prevalence of excessive intakes</th>
<th>Desirable mean intake</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EAR</td>
<td>AI</td>
<td>UL</td>
<td></td>
</tr>
<tr>
<td>Thiamin</td>
<td>0.5</td>
<td>—</td>
<td>—</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>Riboflavin</td>
<td>0.5</td>
<td>—</td>
<td>—</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>Vitamin C</td>
<td>22</td>
<td>—</td>
<td>650</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>Zinc</td>
<td>4</td>
<td>—</td>
<td>12</td>
<td>&lt; 5%</td>
</tr>
<tr>
<td>Calcium</td>
<td>—</td>
<td>800</td>
<td>2500</td>
<td>N/A</td>
</tr>
</tbody>
</table>

1 Adapted from the Institute of Medicine (7). Before assessment, intake distributions were adjusted to remove the effect of day-to-day variation.  
2 EAR, Estimated Average Requirement; AI, Adequate Intake; UL, Tolerable Upper Intake Level.  
3 N/A, not applicable.
A third component of accurate nutrient intake estimation is collection of unbiased supplement intake data, including the correct items, the right dose and an accurate estimate of frequency of use (daily, weekly, etc.). Proper nutrition monitoring requires measurement of total nutrient intake. Because supplement use is common in the United States, total nutrient intake cannot be captured if supplement intakes are not correctly measured (or are not measured at all). A strength of the combined survey will be the detailed information that will be collected on supplement use in the past month, using methodology similar to that currently used in the National Health and Nutrition Examination Survey (NHANES) surveys. However, very little is known about the accuracy with which people report supplement use. It is possible that a method similar to the 5-pass method used for dietary recalls would be appropriate for collection of supplement intake data. Research to identify and minimize biases during collection of these data is sorely needed.

Just as an accurate food composition table is needed to estimate nutrient intakes from foods, an accurate supplement composition table is needed to estimate nutrient intakes from supplements. Because the number of products available continues to increase at the same time that existing products are being improved, it may become necessary to further increase specificity. For example, a popular drink in Hawaii is Juice Squeeze® by Crystal Geyser. Unlike many similar beverages, this one contains 40% of the daily value for both niacin and vitamin B-12 in a 12-oz (354-mL) bottle. People who drink one or two bottles per day could obtain most of their daily requirement for these B vitamins. Thus, identification of this particular type of juice drink on a database would be important for dietary surveys conducted in regions where this beverage is available and popular.

As noted above, a correct evaluation of both nutrient adequacy and nutrient excess depends on an accurate distribution of total usual nutrient intake. It may be necessary to collect daily supplement use data on the days of the recalls to estimate day-to-day variation in use. This topic is addressed in more detail by Carriquiry (9).

**Monitoring of group intakes may occur at the food level.** Some types of monitoring and surveillance use food-level intake data rather than nutrient-level data. For example, it is often of interest to compare intakes with food guidance such as the recommendations of the Food Guide Pyramid (2). Comparisons of this type have been facilitated by the availability of a Pyramid Servings Database developed by the USDA (17). This important new database allows direct estimation of the number of Food Guide Pyramid servings for each participant in the 1994–1996 and 1998 CSFII. An essential feature of the database is the disaggregation of mixtures before accumulating the number of servings of each food. Thus, mixtures such as chicken pot pie are not directly assigned to a food group but instead the flour in the crust is considered a grain product; the chicken is assigned to the meat group; and peas, carrots and onions are assigned to the vegetable group. An example of the results for CSFII participants is shown in Table 2 (18).

In some ways, evaluation of food intakes parallels evaluation of nutrient intakes: intakes are compared with recommendations to estimate the quality of diets. Although prevalence of nutrient inadequacy cannot be directly estimated from food-level data, it is possible to evaluate the prevalence of intakes that are below (or above) the recommendations, as shown in Table 2. When undertaking this type of evaluation, day-to-day variation in food intakes should first be removed, as was necessary when nutrient intakes were evaluated. The prevalence of usual intakes below the recommendation will be distorted if day-to-day variation is not considered. However, the methodology for this type of adjustment at the food level is not as well developed as that for nutrient intake adjustment. More research is needed on this point.

A second use of food-level intake data is to estimate exposures to contaminants, toxins, pathogens and other undesirable components of the food supply. Another important new database, the Food Commodity Database, will facilitate these types of analyses (19). This database was developed by the USDA in collaboration with the Environmental Protection Agency and provides estimates of the intake of food commodities (e.g., apples, oranges, potatoes) by each participant in the 1994–1996 and 1998 CSFII. Like the Pyramid Servings Database, all food mixtures are disaggregated into their component ingredients before the commodity intakes are accumulated. Thus, for example, intake of apples would include apples consumed alone as well as apples in pies, salads and other mixtures. Although the database was developed specifically for estimates of pesticide exposure, it is likely to have other uses as well. If it is to be used to monitor intakes (e.g., to look at the proportion of the population consuming specific types of

---

**TABLE 2**

Assessment of intakes of Food Guide Pyramid food groups for all individuals aged 2 y and over, CSFII 1994–1996

<table>
<thead>
<tr>
<th>Food group</th>
<th>Minimum recommended svg/d</th>
<th>Mean intake/d</th>
<th>Percentage consuming minimum svg/d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grain products, svg</td>
<td>6</td>
<td>6.7</td>
<td>52</td>
</tr>
<tr>
<td>Vegetables, svg</td>
<td>3</td>
<td>3.3</td>
<td>49</td>
</tr>
<tr>
<td>Fruits, svg</td>
<td>2</td>
<td>1.5</td>
<td>28</td>
</tr>
<tr>
<td>Dairy products, svg</td>
<td>2</td>
<td>1.5</td>
<td>27</td>
</tr>
<tr>
<td>Meat and alternates, svg</td>
<td>2</td>
<td>1.9</td>
<td>37</td>
</tr>
<tr>
<td>Discretionary fat, g</td>
<td>None</td>
<td>56.1</td>
<td></td>
</tr>
<tr>
<td>Added sugar, tsp</td>
<td>None</td>
<td>20.0</td>
<td></td>
</tr>
</tbody>
</table>

1 Adapted from U.S. Department of Agriculture, Agricultural Research Service (19).
2 svg, serving.
foods), then analytic issues related to the distribution of usual intakes need to be addressed.

In summary, the combined survey will offer the best opportunities to date for national nutrition monitoring and surveillance. However, work on analysis methods for population groups is still needed, in particular more information on the magnitude of underreporting of intake, methods of multivariate analyses of prevalence of inadequacies, how to combine nutrient intakes from foods and supplements, ensuring the accuracy and appropriate specificity of nutrient composition databases and adjusting food-level data to remove day-to-day variation in intakes.

Using survey data to study relationships between diet and health

A major use of national nutrition surveys, in particular the NHANES survey, has been to study associations between dietary intake and a variety of health outcomes as measured in the mobile examination center. Such cross-sectional analyses are possible because the intake and health measures are available for the same individual. Furthermore, the potential exists to collect data (particularly morbidity and mortality data) over time for the participants, creating the opportunity to perform even more powerful longitudinal analyses. Dietary intake variables at the time of the baseline survey can then be related to subsequent health outcomes, thus avoiding the confounding that occurs when a health condition affects concurrent intakes. The possibility of longitudinal collection of disease outcomes is another major strength of the combined survey.

To examine associations between diet and health, it is usually desirable to rank or classify each participant’s intake for foods or nutrients of interest. Therefore, the analyses are performed on individual survey participants rather than on the group as a whole. Unfortunately, the statistical methods that are valid for removing the effect of day-to-day variation in intakes for a group are not valid at the individual level. As a result there is usually an attenuation of associations with intake variables. However, with 2 d of intake data, there will be less attenuation than in the past when only 1 d of data was available for most of the NHANES participants. Even though 2 d of intake does not accurately capture true usual intake, it does so better than is possible with just a single day.

If the primary goal is to rank individuals, does underreporting of intake matter? If the bias is uniform across all individuals (e.g., everyone underreports energy and nutrient intakes by the same amount), then the rankings would not change. This is seldom the case, of course: under- and overreporting vary depending on characteristics such as body size and literacy level (20). Thus, for example, it is difficult to study relationships between energy intake and body mass index because heavier participants underreport energy intake more than lighter-weight participants. Because the bias in reporting may well vary with the health outcomes of interest, it becomes crucial to reduce the bias as much as possible. Thus, minimizing intake reporting bias with the new 5-pass method is as important for individual-level analyses as it is for group-level analyses.

Another parallel between the appropriate methods for analyzing survey data for monitoring and for studying diet and health is the importance of accurately measuring and combining nutrient intakes from foods and from supplements. However, at the level of the individual, it is necessary to approximate usual intake for each participant rather than for the group as a whole. Currently, estimates of usual supplement intakes are calculated from the frequency of use over the past month, averaged to reflect daily intake (e.g., if a participant reports taking 500 mg of vitamin C once/wk, daily intake would be calculated as $\frac{500}{7} = 71$ mg/d). Intakes from supplements are then added to the 2-d average intake from foods to give an approximation of total usual intake. As noted earlier, there is little information on the validity of the supplement use data collection methodology or on the accuracy of the supplement composition table that is currently being used. An alternative is to collect supplement use data only for the 2 d of the recall. However, if participants can report intermittent use of supplements (e.g., weekly) with reasonable accuracy, then the monthly use data may be a more valid measure of usual supplement intake than data collected only on the 2 d of the recalls. Thus, the current method of combining the monthly supplement use data and the 2-d food intake data may be preferable if estimating usual nutrient intake for an individual is the goal. Various options for combining food and supplement intake data should be evaluated for their validity and respondent burden.

A variety of food, nutrient and supplement intake variables can be used when data from the combined survey are used to study diet and health relationships. In some analyses, researchers may wish to study just the effect of nutrients from foods or nutrients from supplements as well as the total nutrient intake. Thus, it would be useful to continue to provide nutrients from foods and nutrients from supplements as separate variables. In addition, studies of relationships between foods and health outcomes may be as revealing as studies of nutrients and health outcomes. Interactions among nutrients and non-nutrients in foods cannot always be predicted and thus food-level associates of various health or disease outcomes can often provide additional insights.

In summary, the combined survey should be the best to date for relating diet to health. However, work on analytic methods is still needed, and most of the needs discussed under monitoring and surveillance apply equally here. In addition, analyses could be undertaken to evaluate the attenuation that occurs when 2 d of dietary data, collected with the 5-pass method, are used in analyses of diet and health; determine whether there are valid ways to adjust an individual’s intake data to remove biases due to under- or overreporting (perhaps considering individual characteristics such as body mass index and literacy level); investigate ways to adjust individual-level dietary data for the effect of day-to-day variation in intakes; and further evaluate the usefulness of adjusting macro- and micronutrient intakes from foods for energy intakes.

Using survey data to inform nutrition policies and programs

A third important use of survey data is to guide nutrition policies and programs at the federal, state or local level. Woteki (21) covered this subject in more detail, but it is useful to recognize that the two uses of survey data discussed above (monitoring/surveillance and diet-health evaluations) ultimately can lead to decisions about policies and programs. Some of the current uses of survey data for this purpose include establishing nutrient standards, such as the DRIs; developing food intake recommendations, such as the Dietary Guidelines and the Food Guide Pyramid; evaluating program effectiveness (e.g., Women, Infants, and Children’s Supplemental Food Program; food stamp program; and school meal program); developing new programs to reach specific targets (such as the nutrition education program for food stamp recipients); and guiding regulatory policies, including those used for food and supplement labeling.
The combined survey will offer many opportunities to refine and update nutrition policies and programs. However, some types of additional data also would be helpful. For example, to better target nutritionally vulnerable groups, it may be necessary to provide larger sample sizes (e.g., for some specific ethnic groups). Some states are not covered by the national nutrition surveys; residents of Hawaii and Alaska are not covered at all, and they represent several unique cultural and ethnic populations with a variety of dietary patterns and consume foods not routinely reported in the 48 coterminous states.

Programs designed to provide nutrition education can benefit from data on lifestyle and behavioral variables related to food choices. For example, it might be useful to know about food expenditures and food shopping behaviors and how these are related to dietary quality. Likewise, information on knowledge, attitudes and beliefs related to foods and nutrients, such as that collected in the past as part of the Diet and Health Knowledge Survey (22), would be helpful. A subsample that receives a health examination; reports dietary data; and also completes a questionnaire on food-related knowledge, attitudes, beliefs and behaviors should be considered. The ability to look at interrelationships among all these measures could significantly improve our ability to appropriately design and target interventions, such as nutrition education programs.

What might be added to or deleted from the combined survey in the future?

Almost any user of national nutrition survey data could compile a list of desirable additions to the current surveys. Some of the more obvious items on most such lists include a larger sample size to permit more analyses of vulnerable subgroups; a 3rd d of recalls to permit better characterization of an individual’s usual intake; a more detailed food frequency questionnaire to allow better quantification of infrequently consumed foods (although there are concerns about the validity of many of the currently used food frequency questionnaires, a combination of recalls and a food frequency questionnaire would provide more accurate information than either alone); a questionnaire regarding food-related behaviors, knowledge, attitudes and beliefs; and information on supplement use on the day of the dietary recalls.

Of course, funding would be needed to implement any of these additions unless current components could be deleted from the combined survey. One trade-off might be to recruit a larger sample for groups that are considered nutritionally vulnerable and select a smaller sample for groups that are less vulnerable. Another possible saving would be to replace all but the 1st d of dietary recall with a food frequency questionnaire if one with acceptable validity could be identified or developed. At least 1 d of recall would probably still be needed to characterize meal patterns and specific types of foods consumed. Savings might also be realized by collecting less detailed information on supplement use, particularly on dosages, if descriptive analyses show that more generic data would not significantly reduce the accuracy of estimated intakes.

Summary

The combined survey will provide data that will allow unprecedented analyses and applications of national nutrition survey data. Monitoring and surveillance will be enhanced by the more accurate 5-pass method of collecting 24-h recall data and by new methods for determining the prevalence of dietary inadequacies and excesses. The availability of new or expanded databases for food composition, supplement composition, Food Guide Pyramid servings and food commodities will provide a wealth of new variables of public health importance. Analyses of relationships between diet and health will also be improved by these new methods and databases as well as by a 2nd d of dietary recalls. Better information on nutrition monitoring and on relationships between diet and health will, in turn, result in more effective nutrition policies and programs.