Maintaining Time - Related Databases for Dietary Data Collection and Nutrient Calculation

I. Marilyn Buzzard
University of Minnesota
Minneapolis, Minnesota

Introduction

The need for comparability of food and nutrient intakes over time is being increasingly recognized by policy makers at both the local and the national levels, as well as by researchers who conduct long-term dietary intervention trials and other nutrition related research. Time-related databases are required for monitoring progress toward meeting specific nutrition objectives such as the Year 2000 nutrition-related Objectives for the Nation (1). We need to be able to assess the changes in what Americans are eating without the confounding due to using different nutrient databases or different coding schemes for surveys conducted at different periods of time. Similarly, medical researchers must have confidence that changes in nutrient intakes represent actual changes in the subjects' food consumption rather than artifacts of using different versions of a nutrient database.

Investigation of long-term trends in food and nutrient intake requires the use of databases which reflect the constantly changing marketplace and changes in food composition. These databases must also be maintained in a manner that permits recalculation of nutrients at any point in time to take advantage of the availability of improved food composition data. Recalculation of previously collected food intake data also permits the investigator to calculate intakes for other food components that may have been more recently added to the nutrient database.

The Nutrition Coordinating Center (NCC) at the University of Minnesota has developed nutrient database maintenance procedures to accommodate the needs of long-term clinical trials and other nutrition research studies requiring data collection over a period of many years (2-4). These procedures have now been extended to the more complex databases required for interactive collection and automated coding of food intake data (5,6). The NCC has also developed customized database maintenance procedures to meet the special needs of the National Health and Nutrition Examination Survey (NHANES III) (7).

The purpose of this paper is to identify the major factors affecting the comparability of dietary data over time, to describe the NCC databases used for collection and nutrient calculation of food intake data, and to summarize the database maintenance procedures required for the assessment of dietary data over time. These procedures are necessary not only for valid data analysis for long-term studies, but also for comparison of food and nutrient intake data among studies conducted during different time periods.

Time-related factors

There are three general categories of time-related factors pertaining to the long-term collection and nutrient calculation of food intake data. These include: 1) marketplace changes; 2) changes in food preparation practices; and 3) the availability of new or improved food composition data, including both nutrient values and non-nutrient data (such as food densities). Most of the changes in the first two categories reflect actual changes in foods consumed, whereas most of the changes in the third category are related to the availability of better nutrient and non-nutrient data, rather than to actual changes in what people are eating.

Examples of marketplace changes include new products, new serving sizes of existing products, reformulations of existing products, and discontinued products. Current trends toward the use of less sodium and fat in preparing foods are examples of changes in food preparation practices. New analyses of foods and the use of improved analytical methodologies result in the

ongoing availability of new and improved data on food composition. Database maintenance procedures that permit comparison of dietary data over time must account for all of these types of changes.

General approaches to nutrient database maintenance for providing long-term stability of nutrient calculations.

Several different approaches have been used by national nutrition surveys and medical research studies to provide long-term stability for nutrient calculations. One approach is to freeze the nutrient database at the beginning of a study and continue to use that database throughout the study. A number of clinical trials have used this approach in the past. The major problem with this approach is that the study is unable to take advantage of new and better data that become available over the course of the study.

A second approach is to calculate nutrients only at the end of a study, using the most current data available at that time. The benefit of this approach is that the studies are able to take advantage of new data that become available during the period of the study. The United States national nutrition surveys, including the USDA Nationwide Food Consumption Surveys and the DHHS National Health and Nutrition Examination Surveys, have used this approach in the past to provide stability during the period of the survey. The problem with this approach for survey use is that it does not permit comparison with previous surveys that have used different nutrient databases. Another major disadvantage is that nutrients can be calculated only at the end of the study. Therefore, this approach is not an option for most clinical trials and other studies that require ongoing monitoring of differences between study groups or compliance to a dietary intervention protocol.

A third approach is to use the most current nutrient database available whenever nutrient calculations are desired. This approach requires the recalculation of all previously processed data so that all nutrient intakes are calculated from the same nutrient database. With the progressive decline in computer calculation costs and the increasing use of microcomputers, this approach is becoming a viable option for many studies.

The NCC has developed databases and database maintenance procedures that permit studies to use any of the three approaches described above. The remainder of the paper will describe the two major NCC databases and the database maintenance procedures used to enhance comparability of food and nutrient intakes over time.

Overview of NCC databases

The NCCcMinnesota Nutrition Data System (NDS) is an automated system for interactive data collection and nutrient calculation. NDS requires two NCC databases: the Food Description database (FDdb) and the Nutrient database (Ndb). The FDdb contains the information required for interactive collection and automated coding of food intake data, and the Ndb contains the nutrient values required for nutrient calculation.

Figure 1 shows the three major steps involved in the interactive collection and automated processing of dietary data and the relationships among these steps and the two NCC databases. Interactive data collection is the first of the three major steps involved in this process. Regardless of the method used for collecting the data (eg, 24-hour dietary recall, food records, or diet history), identification of the foods and amounts consumed is captured by the computer in language commonly used by subjects for describing foods. Coding consists of assigning one or more food codes to each food description. These codes are linked to entries in the Ndb. The coding step also involves converting amounts expressed in common household units or food specific units (such as piece, medium, slice, or package) into gram weights for nutrient calculation. The final step involves the calculation of nutrients using the Ndb and calculation software.

The FDdb contains more than ten times as many food descriptions as the Ndb. This is due not only to the fact that many foods in the FDdb are composites of a number of individual ingredients included in the Ndb, but also to the fact that foods having similar nutrient profiles are often grouped together in the Ndb. Thus, the food descriptions listed in the Ndb are primarily generic food descriptions, any one of which may represent many different brand name products or food types. To permit maximum flexibility and specificity for describing foods, all non-nutrient information (such as recipes, densities, yield factors, and other amount conversion data) is maintained for each of the many foods in the FDdb, rather than for the few foods in the Ndb. The Ndb, which is the least complex of the two databases, has been maintained as a time-related database for many years. Therefore, I will address the maintenance of this database first.

Specific guidelines for maintaining the Ndb as a time-related database.

Database maintenance is an ongoing effort by a staff of NCC nutritionists trained in database maintenance procedures. Ongoing maintenance involve

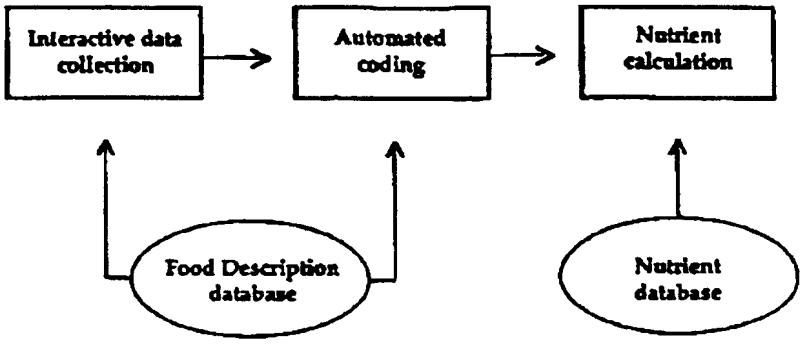

Figure 1. Major steps and databases required for automated collection and processing of dietary data using the Minnesota Nutrition Data System (NDS)

incorporation of new data from the USDA Human Nutrition Information Service and information on commercial products from manufacturers, as well as information from the literature, from foreign food composition tables and from recipe books. A comprehensive listing of the sources of data for the NCC Ndb has been reported by Schakel et al (8). A new version of the Ndb is released every 6 to 12 months; 16 versions of the database have been released over the past decade.

To ensure that intake data collected in the past can be recalculated on any previous or future version of the Ndb, the following guidelines are strictly adhered to by the database nutritionists:

1. All new foods and new formulations are added to all previous versions. This allows studies using a "frozen" database to keep current with the marketplace.
2. No foods are ever deleted from the Ndb. Products no longer on the market continue to be updated when the database is expanded to include additional nutrients or other non-nutrient food components.
3. Modifications to nutrient values are added only to the current version of the database. Even though better analytic data become available, any modifications to a database already in use by a study would confound the interpretation of study results.
4. For products that are reformulated, multiple date-related nutrient values are maintained in the database. The computer accesses the appropriate nutrients based on the date the food was consumed.
5. When an entry in the database is expanded into two or more entries, the entry that is the most representative of past intakes retains the original code number.

Adherence to these guidelines ensures that dietary data collected at any point in time can be recalculated to take advantage of all updates to the Ndb. These guidelines have been in place at the NCC since 1983, and many US and Canadian studies have taken advantage of the ability to compare nutrient data over time by recalculating previously coded data using an updated version of the Ndb.

Description and maintenance of the NCC Food Description database (FDdb).

The FDdb maintained at the NCC has been described in detail elsewhere (5,6). Briefly, the FDdb consists of two components. One component is a hierarchical structure of food descriptions that drives the interactive food identification process. This hierarchy of food descriptions is based on the way ordinary people describe foods. A series of menu-type screens prompt the user for as much detail as is necessary to permit the computer to automatically assign the appropriate code numbers and gram amounts. The level of food description becomes progressively more detailed as the hierarchy is traversed for a given food item. These detailed descriptions may include brand names as well as methods of food preparation.

The second component of the FDdb is the informa-

tion required to automatically assign code numbers and gram weights to the foods identified through the interactive prompting described above. Such information includes all of the common synonyms for describing a particular food, the recipes or formulations for describing combination foods, designation of ingredients that require further description (eg, the type of fat used to prepare a food), and the defaults to be assigned if the necessary level of detail cannot be provided by the subject. In addition, the FDdb must be able to provide adequate information for converting amounts as described by the subject into gram weights. Amount conversions may require densities, raw to cooked yields, edible portion factors, geometric shape conversions, or weights of food specific units.

The more traditional non-automated procedures for processing dietary data result in substantial loss of descriptive detail at the coding level. For example, most brand name products are assigned code numbers corresponding to generically described foods in the Ndb. Thus, brand identification is lost at the time of coding, since only the generic code numbers and amounts are entered into the computer. If a product had been assigned an incorrect code number, the only way to correct the error would be to manually recode the data. Because manual coding is very labor intensive, recoding is not a feasible option unless the coding process is automated. Automated coding is possible only if the computer is able to capture all of the food description detail required for coding. Thus, the FDdb described above incorporates the features needed for automated recoding at a later time. However, automated recoding is possible only if successive versions of the FDdb are compatible. (Procedures to ensure compatibility between successive versions of the FDdb are discussed in the next section.)

All FDdb modifications are made to the current working version of the database. Updates include additions of new products that enter the market, modifications for reformulated products, and changes reflecting new or improved data. Foods no longer on the market are deleted only after a reasonable period that takes into account the potential time lag between purchase and consumption of the product.

To maintain currency with the marketplace and changing food consumption patterns, ongoing studies are encouraged to collect data using only the most recent version of the FDdb. However, all previous versions used by the study must be retained to permit editing of food intake records collected in the past, since each record is tied to that version of the FDdb on which it was originally collected.

Development and maintenance of a Multi-Version Food Description database.

To enhance the comparability of dietary data collected over time, the NCC has developed a Multi-Version Food Description database (MVdb). Use of the MVdb eliminates the logistical problems of maintaining many different versions of the FDdb for editing food records collected in the past. All existing versions of the FDdb are collapsed into a single database, thus eliminating the large amount of redundancy among the separate databases. The MVdb automatically accesses the appropriate version each time a dietary intake record is edited. Most importantly, the MVdb permits coding changes to be retroactive to previous versions of the FDdb. Thus, food intake records collected in the past are able to take advantage of subsequent corrections and additions to the FDdb.

Maintenance of the MVdb requires that every update to the database is assigned one of three _troactivity codes.y One code is used to indicate that the change is retroactive to all previous versions of the FDdb; a second code is used to indicate that the change is not retroactive to any previous versions; and a third code, representing a specific version of the FDdb, is used to indicate that the change is retroactive back to the designated version. NCC database nutritionists use the following guidelines for selection of retroactivity codes:

1. Improved numeric data are always retroactive to all versions of the FDdb. For example, a more accurate value for a food density or for a raw to cooked conversion factor needs to be reflected in the recalculated data whenever food intake records collected with a previous version of the database are recalculated on an updated version.
2. Changes in the food supply are retroactive to the version reflecting the time the change occurred. For example, the addition of a new commercial product is retroactive to that version of the database most closely matching the date the product was introduced into the marketplace. This allows the editing of past food intake records for which the new product was reported and temporarily assigned a missing food status because the product had not yet been added to the FDdb.
3. Deletions are never retroactive. Even though a product is no longer on the market, the product must remain in all previous databases to permit editing of food intake records collected in the past for which the product may have been reported. However,

deletions are permitted in a newer version of the database reflecting the foods available in the current food supply.

Adherence to these guidelines ensures currency for data collection and comparability of coding and nutrient calculation for dietary data collected at any point in time.

Summary

Time-related databases that permit comparison of food and nutrient intakes over time are essential for nutrition monitoring, as well as for long-term studies investigating diet-health relationships. It is possible to maintain nutrient and coding databases in a manner that will permit routine updating to reflect the changing marketplace and the availability of improved data while simultaneously providing data analysis stability for long-term studies and trend analysis of food consumption data.

Comparison of food and nutrient intake data over a long time period is possible only if certain requirements are met. These requirements include the following: 1) the database used for food identification must be frequently updated to maintain currency with the dynamic marketplace and changing food preparation and food consumption patterns; 2) detailed food descriptions, including any relevant information on brand names or food preparation methods, must be captured; 3) coding procedures must be automated so that the raw data (ie, the detailed food descriptions) collected at different times can be recoded using the same coding procedures; and 4) all nutrient intakes must be calculated using the same nutrient database. Compliance with these requirements will ensure comparability of food intake data collected at any point in time, permitting trend analysis of food and nutrient intakes with minimal confounding due to differences in coding procedures or nutrient databases.

References

1. Healthy People 2000, National Health Promotion and Disease Prevention Objectives. DHHS Publication No. (PHS) 91-50212, Supt of Documents, US Government Printing Office, Washington DC, 1991.

2. Sievert, YA, Schakel, SF, and Buzzard, IM: Development and maintenance of a nutrient database for clinical trials. Controlled Clinical Trials, 10:416-425, 1989.

3. Dennis B, Ernst N, Hjortland M, Tillotson J, Grambsch V: The NHLBI Nutrition Data System. J Am Dietet Assoc 77:641-647, 1980.

4. Tillotson JL, Gorder DD, Kassim N: Nutrition data collection in the Multiple Risk Factor Intervention Trial (MRFIT). J Am Dietet Assoc 78:235-240. 1981.

5. Feskanich D, Buzzard IM, Welch BT, Asp EH, Dieleman LS, Chong KR Bartsch GE: Comparison of a computerized and a manual method of food coding for nutrient intake studies. J Am Dietet Assoc 88:1263-7, 1988.

6. Feskanich,D, Sielaff BH, Chong K, and Buzzard IM: Computerized collection and analysis of dietary intake information. Computer Methods and Programs in Biomedicine 30:47-57 1989.

7. McDowell M, Briefel RR, Warren RA, Buzzard IM, Feskanich D, Gardner SH: The Dietary Data Collection System: an automated interview and coding system for NHANES III. In: Proceedings of the 14th National Nutrient Databank Conference, June 19-21, 1989, Iowa City, IA; Stumbo PJ, ed; The CBORD Group, Inc, Ithaca, NY; 1990. pp125-131.

8. Schakel SF, Sievert YA, and Buzzard IM: Sources of data for developing and maintaining a nutrient database. J Am Dietet Assoc 88:1268-71, 1988.