Activities of the National Nutrient Databank Conference Organizing Committee on Data Quality

by Suzanne P. Murphy, Chair, Department of Nutritional Sciences, University of California, Berkeley

A. Background

Attendees at the 19th National Nutrient Databank Conference (held in May, 1994) asked that a committee be formed to address issues surrounding nutrient data quality. The importance of this topic was emphasized in a session on data quality at the 19th conference, which focused in part on a recent report from the General Accounting Office (1).
In response to these discussions, the Steering Committee of the National Nutrient Databank Conference formed an Organizing Committee on Data Quality, with myself as chair. Over the past year we have formed five subcommittees to address specific aspects of nutrient data quality, each with two co-chairs:

Quality Codes (Joanne Holden and Loretta Hoover)
Clearinghouse for Industry Data (Jean Pennington and Jack Smith)
Letters to Legislators (Roberta Markel and Judi Douglass)
Letters to Editors (Phyllis Stumbo and Jean Hankin)
Setting Priorities for Analysis (Margaret McDowell and Sue Gebhardt)

A Committee meeting was held, and a brief presentation on the Committee organization and objectives was made at the Dietary Assessment Conference in Boston (1/95).
The American Institute of Nutrition has developed a white paper on food composition data needs. A copy of this paper was distributed to conference attendees (see appendix).

B. General Needs Identified at the 1994 Conference

- Although the GAO report made many suggestions, it overlooked a number of issues of importance to nutrient database users; these should not be overlooked when setting priorities for improving data quality.
- Conference participants would like to find ways to work with USDA to facilitate improvement in nutrient data quality.
- The staff at USDA would like a mechanism to solicit feedback from users.

C. Specific Needs Identified at the 1994 Conference

- Data quality is not a binary variable; users would prefer to have an estimate of the quality rather than have "poorer" quality data omitted from national databases.
- The food processing industry has a wealth of information on food composition, but ways to compile these data centrally have not been identified.
- USDA's Nutrient Data Laboratory (formerly HNIS) and Food Composition Laboratory must be adequately funded if the U.S. is to have a high quality national nutrient database.

•Since funding will never allow comprehensive analyses of all nutrients in all foods, priorities for analyses should be set based on: nutrients which are of public health importance, foods which are good sources of these nutrients, and gaps in our knowledge.

D. Subcommittee Reports

Quality Codes

Co-chairs: Joanne Holden (USDA/NDL) and Loretta Hoover (Univ of Missouri).

The Quality Codes Subcommittee convened for an initial meeting on June 13, 1995. The volunteer and invited participants and guests at the meeting were: Alison Eldridge, Bonnie Sherr, Bonita Hoverson, Margaret McDowell, and Lena Bergstrom. A variety of issues were identified and subsequently grouped into the following seven categories:

- Rationale

How will quality codes be used?
What do they represent? (e.g., quality of data vs confidence in data)
How do source codes, data derivation codes and quality codes differ?

- Criteria

Constructed to minimize bias introduced when assigned
Flexible, especially relative to state-of-the-art considerations

- Specification

What should they be? (e.g., numeric, alphanumeric, etc.)
Should they be nutrient/component specific?
What process/procedure will be used to determine quality codes?
What are the major sources of variation that should be reflected in quality codes?

- Implementation

Should assignment of the quality codes be centralized?
How will quality codes be assigned for non-analytical data? (e.g., imputed values)
How will quality codes be determined for mixed dishes calculated from recipes?
How will quality codes be manipulated? (probably will require statistical consultation)
How will missing values for quality codes be handled?

- Priorities

Which foods should be addressed initially?
Which nutrients/components should be addressed initially?

- Storage/Maintenance

What data structure will be used for transfer and storage of quality codes?

What data elements are necessary to permit automated recalculation of quality codes when better information or better data become available? (i.e., quality codes are dynamic, perishable, and subject to information available at a given point in time.)
What strategy will be used to provide a trend analysis or history of changes in quality codes? (i.e., which quality code should be used at a specific point in time?)
How will users demonstrate their accountability in maintaining a nutrient database in an up-to-date status relative to quality codes?

- Education
  How will developers of data bases be educated relative to the incorporation and maintenance of quality codes in their products?
  How will end users be educated to interpret and make appropriate use of quality codes?

International Data Quality Committee Meeting

The minutes from the NDBC Quality Codes Subcommittee Meeting were submitted to the International Data Quality Committee Meeting convened by the International Union of Nutrition Scientists (IUNS) in Riverdale, Maryland during June 14-15, 1995. Representatives included Gary Beecher, (chairperson)(USDA), Joanne Holden, (USDA), Barbara Burlingame (INFOODS), Prapasiri Puwastien (ASEAN FOODS), Lilia Masson (Latin Foods), and Lillian Marovatsanga (AFRO Foods). The Eurofoods representative was not available to attend the Riverdale meeting. The discussion focused on the needs for the assessment of data quality, the relevant parameters, the types of food composition data requiring quality indicators, and the possible formats of indicators. Representatives agreed that data quality indicators for analytical and derived (calculated and selected impute) values would be desirable. Representatives accepted in concept, the following parameters for assessing analytical data quality: analytical method, analytical quality control, sampling plan, sample handling, and statistical characteristics of the data. A rating scale pertaining to individual data sources and parameters was discussed. The need for a summary indicator for a single food and component was considered. The discussants recognized the importance of having good quality analytical data as the basis of calculated or imputed values. Algorithms and factors (e.g. 6.25 for N to protein conversion) used to determine estimates for food components should be reviewed for chemical and biochemical validity and updated, as necessary.

Representatives of the IUNS committee will continue to interact with their colleagues at the Regional and National level to identify and discuss data quality issue and to provide input to the IUNS Committee. The USDA Nutrient Data Laboratory has begun to define appropriate data quality indicators to be included in the USDA Nutrient Database for Standard Reference and will continue to develop the process and system. Representatives to the NDBC subcommittee on Data Quality Indicators will be asked to continue to play an important role in this development during the next year.

Joanne Holden, Suzanne Murphy, Gary Beecher, and other U.S. and Canada representatives attended the 2nd International Food Database Conference in Lahti, Finland during August 28-31, 1995. Both Joanne Holden and Barbara Burlingame reported on the discussion and outcome of the IUNS Committee meeting. Joanne Holden spoke on the Determination of Food Composition Data Quality. Comments from participants indicate support for the determination of food composition data quality and the need for data quality indicators which are available to data users. The Eurofoods COST 99 project recently funded by the European Union Conducted three workshops at the Lahti Meeting for COST 99 participants and invited expert observers. One of those workshops address the topic of Food Composition Data Quality and will continue to address this topic during the next three years.

Clearinghouse for Industry Data

Co-chairs: Jack Smith (Univ of Delaware) and Jean Pennington (FDA).

The Subcommittee has had relatively little activity during the past year. The reasons for this include the organizational changes within USDA, the Nutrient Data Lab's moving and the activity of IFDA in creating the Data Transfer Standard.
There are several challenges for the Subcommittee in the coming year. These include:

- Establishment of a Partner Relationship between NNDBC, USDA/NDL and industry.
- Enhancing the dialog between these partners and other interested parties.
- Define the function, contributions and context of a clearinghouse.
- Address the real and perceived concerns of all parties related to the sharing of data.
- Evaluate acceptable mechanisms for imputing data not available from analysis.

The Subcommittee needs a number of volunteers from data providers, data users and those maintaining the data.

Letters to Legislators

Co-chairs: Roberta Markel (DINE) and Judi Douglass (TAS).

The Subcommittee has drafted several letters to legislators which were distributed to attendees of the conference and are included here as an appendix. The availability of these letters also was posted on both the food-comp and nutepi listservs, and subscribers were invited to request electronic copies.

Letters to Editors

Co-Chairs: Phyllis Stumbo (Univ of Iowa) and Jean Hankin (Univ of Hawaii).

The Subcommittee identified factors critical to maintaining the high quality of USDA nutrient data. Since some nutrient data is gleaned from published literature, independent investigators could lessen criticisms of the published tables by using internal standards in their analytical work, describing their quality control procedures when they publish and indicating measures of

variability when reporting nutrient data. Thus when data are taken from the published literature they would include the quality control information considered essential by the GAO report (1). These issues, along with pleas for stronger funding for nutrient composition activities, are recommended content for letters to the editors.

Setting Priorities for Analyses

Co-chairs: Margaret McDowell (CDC/NCHS) and Sue Gebhardt (USDA/NDL).

1. Background from NCHS

NCHS has used the USDA Survey Nutrient Data Base (SNDB) to code foods and estimate nutrient intakes for HANES surveys since 1982, beginning with Hispanic HANES, 1982-84 and most recently with NHANES III, 1988-94.
Hispanic population subgroups were specifically oversampled in recent surveys; ARS analyzed a number of ethnic foods, primarily Hispanic and Asian foods, at NCHS's request during NHANES III.

2. Background from USDA/ARS

Maintaining food composition data bases has been an important activity at USDA for many years. In recent years, this responsibility has become increasingly complicated and challenging. The reasons have been mentioned many times during this Conference.

- The food supply has become more complex. The number of food products is ever increasing. Although some items are discontinued there are many more that are being introduced. Foods are being modified to address public health concerns, providing varieties of foods such as low fat, high fiber, low sodium, and fat replacements.
- The population is becoming more diverse, introducing new foods and combinations of foods that are being consumed.
- Data for more nutrients and food components are being requested by researchers, for example carbohydrate fractions, selenium and vitamin K.

Because the cost for food analyses is very high and resources are not available to analyze all foods for all nutrients, establishing criteria for setting priorities for food analysis is essential.

3. Looking to the Future

The goal of this Subcommittee is to recommend procedures for identifying those foods that are of highest priority for nutrient analysis. To do this, it is necessary to identify key foods as described by Pamela Pehrsson in the session on "Trends and Changes in the Food Supply" at this conference. This concept has been evolving at the Nutrient Data Lab at USDA to help identify those foods that are the main contributors of nutrients in the US diet and therefore the ones for which the quality of the data would have the greatest impact. The Subcommittee will be using this paper and list of foods as our starting point.

It is also important to anticipate changes in the composition of key foods over time, and to have the capability to identify these foods and reanalyze them as necessary. Examples include flour products and breakfast cereals.

Finally, we need to anticipate an increase in the need for information about ethnic foods and be able to identify priorities for analysis of key foods consumed by ethnic subgroups.

Requests for analyses of key foods may require funding from data users.

The following individuals have signed up to be on the Subcommittee:

Shirley Gerrior, Center for Nutrition Policy and Promotion, USDA
Bonita Hoverson, Grand Forks Human Nutrition Center, USDA
Li-Ching Lyu, Cancer Research Center of Hawaii
Betty Perloff, Survey Systems/Food Consumption Lab, USDA
Laura Sampson, Harvard School of Public Health

E. Closing Comments

Attendees of the 20th Conference were urged to become involved in these issues, and were invited to sign up to participate in the work of one or more of the subcommittees. The need for additional subcommittees was discussed and a subcommittee on analytic methodology was suggested.
As a users group with representatives from government, industry, and academia, we can be an effective voice in promoting the development of high quality nutrient data.

Reference

1. General Accounting Office. Better Guidance Needed to Improve Reliability of USDA's Food Composition Data. October, 1993.