Capstone Presentation – 22nd National Nutrient Databank Conference: EMERGING ISSUES FOR THE NEXT GENERATION OF DATABASES

Jean A.T. Pennington, PhD, RD. Division of Nutrition Research Coordination, National Institutes of Health, Bethesda, MD

It is encouraging to look back over the past 23 years and consider how much progress has been made in our knowledge and understanding of the many components of foods and their associations with health and disease. The computer systems we have today have certainly improved since 1975 with increased storage capacity and speed of response. The food values have also improved, but at a slower rate, one food at a time and one food component at a time. It is difficult to "see" the improvements in food composition databases because the documentation of the analytical methods and sampling designs and the treatment of the resulting data on many analytical samples are not apparent in the data tables or computer disks that hold the summarized information. We often note that trust in the validity and representative nature of the data seem be related to one's knowledge of how data are collected, compiled, and summarized. Those who know little about food composition data tend to have more faith in the data, while those who understand the difficulties in this area of research are more conservative in their assessments. Those who use the data appropriately, knowing the potential weaknesses of the data are more likely to draw reasonable conclusions from analysis of dietary intakes.

This 22nd Conference has helped us to think about a number of different issues regarding food composition databases. I would like to highlight six of them:

The **National Food and Nutrient Analysis Program**, addressed by Joanne Holden and developed by Joanne and by Dr. Abby Ershow of the National Institutes of Health, has reminded us of the importance of routine analysis for maintaining and updating our national and other databases. The composition of our foods changes and evolves over time as new foods or formulations become available; analytical techniques and methods to determine the levels of food components become more accurate and reliable; sampling schemes become better designed to ensure representative samples; and we learn more about food components that are important to human health. Information on food composition is the basis for all research work on the relationships between diet, health, and disease. As stated in previous Databank Conferences, the National Nutrient Database is a US treasure and it should be treated as such. Perhaps one of the outcomes of this and future Databank Conferences might be to find ways to encourage Congress to provide the needed funding for this most important work.

The **harmonizing of databases in North America**, as discussed by representatives from Mexico, Canada, and the Caribbean, is an exciting and intriguing venture. How enriching it would be to combine the resources and expertise of Canada, Mexico, the Caribbean, and the US to develop a single database or perhaps four databases that

are compatible, i.e., that use same format, computer software, food names and descriptors, nutrient names, and nutrient data. Food trade flows easily among these four countries. Much US produce comes from Mexico. The cuisine of Southwestern States is based on many Mexican or Southwestern foods such as red and green chilies, posole, tortillas, fry bread, burritos, enchiladas, flan, and mole. Many of these foods, although widely consumed in the US, are not in the National Nutrient Database. Foods consumed in Canada are basically the same as those in the US, and there is much food trade between these two countries. We need to find better ways to exchange, share, and document our food composition data among North American countries. If the US, Mexico, Canada, and the Caribbean could find ways to share knowledge and resources regarding food composition databases, perhaps the government funding agencies in these countries would be more inclined to increase resources to develop and maintain nutrient databases.

The work on **folic acid databases,** presented by Dr. Jesse Gregory, Dr. Jeanne Rader, and Sue Gebhardt, gives us hope for revised estimates of the dietary intake of this vitamin. Data on the folic acid content of foods is confounded by the labile nature of this vitamin (i.e., its sensitivity to oxidation); the different forms of the vitamin with different biological activities; its easy destruction during food storage, processing, and cooking; its destruction during laboratory analysis; and the difficulties of the microbiological assay used to determine it presence in foods. With the FDA's 1998 requirement for fortification of grain products with folic acid, there are additional challenges for food composition database users and developers. Considering all the grain and grain-based products that are to be affected by folic acid fortification, we must resolve to update the folic acid values for these products in our databases. It is exciting to consider the potential health effects of folic acid fortification on birth outcomes. This will be an important area of research, and it will necessitate that accurate information on dietary intake of folic acid be available.

The topic of **databases for dietary supplements**, presented by Margaret McDowell, has long loomed over the heads of database developers and users. The enormity of the task has caused us to largely ignore it, i.e., to calculate nutrient intakes from foods and not to include nutrients from supplements. It is time for us to face up to the contribution of dietary supplements to the total nutrient intakes of population groups. Supplement databases are continuously moving targets with thousands of different commercial products. The levels of nutrients in dietary supplements can easily overwhelm nutrients from foods, and the use of supplements is widespread and variable in terms of the types of supplements that are used and the frequency of their use. Margaret has provided us with the basics about how a database of this nature can be compiled and maintained. When dietary supplement databases become fully developed, researchers will have the opportunity to understand the effects of dietary supplements on nutrient intakes. Perhaps our inability to create and maintain a supplement database and to obtain information on supplement intake from survey

participants is one reason why we see so little direct connection between diet and biochemical measures of health status. It is essential that we be able to separate the populations we evaluate into users and non-users of dietary supplements to begin to make sense of diet, health, and disease relationships.

We are just at the beginning of our understanding of the biochemical and physiological roles of **botanicals and herbal preparations** on human health. This topic, presented by Dr. Bernadette Marriott, is still one of mystique as the extent of use and reasons for use of botanicals and herbal preparations are not well known. Botanicals and herbal preparations include both natural (i.e., raw plants or plant parts) products and brand-named products with mixed and/or processed ingredients. In some respects, botanicals and herbal preparations seem to be at the merging point of foods and drugs. Bernadette has indicated that botanicals and herbal preparations are more like supplements than foods in terms of their form and how they are consumed. Do we need separate database for botanicals and herbs or could they perhaps be merged with dietary supplement databases? What is the composition of these products and do they overwhelm nutrient intake like dietary supplements do? We have similar concerns about how to deal with drugs that contain nutrients (e.g., calcium and magnesium in antacids)? Should all products containing nutrients or other food components be in one database and allow the user to separate foods, supplements, botanicals, herbs, and drugs if they desire to do so? The different products could be coded into these categories, and database users could chose to evaluate the nutrient contributions from these sources together or separately.

We have had presentations on flavonoids and carotenoids at previous Database Conferences (Buffalo, NY in 1995 and Baton Rouge, LA in 1996). At this meeting we have heard about **flavonoids, phytoestrogens, and carotenoids** from Dr. Gary Beecher and David Haytowitz. We seem to be on a voyage of discovery with functional components in foods as their relationships to chronic diseases are uncovered. These compounds are attracting the interest of medical professionals and consumers. The challenge with these compounds is to consider all the different chemical compounds with different biological activities that fall within the definitions of these three types of compounds. Flavonoids alone constitute some 4,000 compounds, some with no apparent biological activity and others with activity related to vascular function, cancer, and/or the immune system. How can we add these compounds to our databases in meaningful ways? Perhaps it is useful to keep these compounds in separate databases until we have more information on their locations and levels in foods and of their physiological effects and actions.

Lessons Learned from National Nutrient Databank Conferences. Databank Conferences serve as forums to share and exchange information about food composition databases. The primary goal of the Conferences is to improve the quality, quantity, and availability of data on the levels of food components in foods. The issues

presented at this conference will take several years of work to resolve. The concerns and problems of our current databases will be solved little by little just as they have for previous databases. We will continue to learn about the complexities of food composition, including naturally-existing components and components that are added to the food supply through agricultural and manufacturing processes and through contamination. Reflecting on the topics and issues of our Conferences over the past 23 years, I think that we have gained some important knowledge:

First, no matter how much we might like to separate food composition data and food consumption data, they are intimately linked. Most database compilers build and design databases to evaluate food and nutrient intake data from food consumption surveys, from diet-related research studies, or from individual patients and clients. The decisions we make about food composition databases affect the nature of the questions that can be asked in food consumption surveys and studies. Similarly, the questions asked about food intake in dietary surveys affect the structure of the database and detail of the food descriptions.

Secondly, as we make decisions and assumptions regarding food composition and food consumption data, we have learned to consider the magnitude of errors that might be introduced. In general, errors in food intake information are far greater than errors in food composition, even when nutrient values for foods are calculated from recipes or imputed. Accurate and reliable food composition data will not yield accurate and reliable dietary intake information if the food composition data are merged with questionable data on food intake.

Third, we now know that the variability of a food component in a food is a reflection of the biological or chemical composition of the materials that we choose to consume as food. (Although some variability between laboratories and analysts may be due to analytical “differences” and not reflect “natural” variability.) Each food component in each food has a variability around the mean value. The variability may be inherent (i.e., due to species characteristics or the age/maturity of the plant or animal tissue), or the variability may be derived from environmental (e.g., due to characteristics related to soil, climate, rainfall, or pesticide application) or processing (e.g., due to cooking methods or food additives) factors. Variability of the levels of food components in foods can and should be measured. Sometimes variability in nutrient values is an important clue telling use that we have a mixed sample. Further investigation may reveal that we need to separate cultivars or separate brand names of the same product. We have yet to determine how to express variability in meaningful ways in food composition databases, and we don’t yet know how to use variability in food composition when we assess daily nutrient intakes.

Fourth, it is difficult for database compilers with backgrounds in nutrition and foods to also be experts in computers and statistics. Therefore, the development of food

composition databases usually requires a collaborative effort among individuals with expertise in these areas (nutrition, foods, computers, and statistics). If we can clearly communicate our needs and expectations for the database to our computer and statistical experts, we are more likely to end up with databases that serve our purposes.

Other Lessons. You can also learn a lot about food composition databases if you attempt to compile one. Many attendees of Nutrient Databank Conferences have database development experience. You learn that the development of food composition data is slow (i.e., requires years) and complex; you must wait for data to appear in the literature. The researchers who want new analytical data must find funding; decide on the foods to be analyzed and the food components to be determined; design a sampling scheme; buy, transport, prepare, and store the foods; find chemists with fully-equipped, modern laboratories; find suitable analytical methods for the food components of interest; determine the statistical strategies for the analytical values that result; document every step of the process; and then write the papers, submit them for clearance and review, and get them published.

If you are also compiling data from food companies and restaurants, you know that the process is also slow and often not productive. You must send letters, call, wait for response, and then try to discern if what you have received reflects analytical data or labeling values. Sometimes you can't get "real" (i.e., non-labeling) data. You must also try to discern what the foods are as industry food names are mostly fanciful and not descriptive. (Consider the names of many ready-to-eat breakfast cereals, candy bars, snack cakes, commercial entrees, fast-food hamburgers, etc.) The response rate for data from food companies (unless you have a friend there) can be low and slow.

By contrast, advances in computer hardware and software move at the speed of light. By the time you have food composition values to put into your database, you may find that the system you had intended to use has become a dinosaur. The lessons learned by database compilers are to be flexible, expect change, and try not to buy into something that can't be updated. Each time you update your database (if you do it at periodic intervals rather than on a routine basis), you will find that the process is different, but not necessarily improved. It is always tedious and requires a special personality type, i.e., someone who is able to deal with minutiae. You will probably also find that you have to work with a new statistician and a new computer expert and that all your contacts at food companies have moved on to other jobs.

There is no doubt that we are making progress on all the issues presented here today. The Databank Conference this year has clearly dealt with very timely and important issues, and I commend Dr. Suzanne Murphy and Dr. Molly Kretsch for the direction of the Conference and for all their efforts in organizing it. I look forward to next year's meeting to see what progress has been made and what new challenges wait for us.