10th National Nutrient Databank Conference

July 22-24 3-5, 1985
San Francisco, California

10th NATIONAL NUTRIENT DATA BANK CONFERENCE

JULY 22, 1985

Introductory Remarks. Doris Howes Calloway. University of California, Berkeley.

It is my pleasure to welcome you to this 10th National Nutrient Data Bank Conference, on behalf of all your hosts; the University of California, Berkeley and Davis campuses, Stanford University and the Western Human Nutrition Center of the USDA.

It is appropriate that the sponsors are principal universities in partnership with the USDA. It is that partnership that, for over a century when nutrition has been in and out of vogue more than once - that has sustained the difficult, often monotonous, under-valued and under-rewarded research on what is in foods, what people eat, and what they ought to eat, in scientific and economic terms.

This research underpinned the establishment of nutrient standards before specific nutrient requirements were known and understood, let alone quantified; it has formed and continues to form the basis for much food and drug law, for advertising claims and their regulation, and for public education.

The partnership between academic institutions and agriculture dates from the Morrill Act, signed into law by Abraham Lincoln in 1862. The Morrill Act gave states public land, 30,000 acres for each senator and representative in Congress. Funds derived were to be made available to a designated college, to foster learning about "agriculture and the mechanic arts". California was among the first to act - the legislature authorized the formation of a State University in 1866 and a complete University (including arts and sciences) was created. Agricultural research has been a feature at Berkeley since this beginning and the first permanent building on the campus, which is still standing (South Hall), was the original College of Agriculture. The Agricultural Experiment Station was formally created in 1874, the first in the United States. The first person in the U.S. officially named a Professor of Nutrition was M. Jaffa, appointed to the University of California, Berkeley in 1908. He founded in 1912 and was the Chair of the first Department of Nutrition among the U.S. agricultural universities. Food composition was a principle interest of Jaffa. The University of California, Davis campus began as "The University Farm"; it was later authorized to teach a three-year course in vocational agriculture which it began to do in 1909.

Studies of food and feed composition are of course older than this chauvinistic bit of U.S. history. According to Atwater, (in his 1896 USDA Bulletin, "The Chemical Composition of American Food Materials") The earliest known quantitive analysis of food materials was that made, of potatoes, by Pearson in England in 1795. (Most of the earlier analytical work was by and for apothecaries). Studies of animal feed crops began in Germany in the early 1800's. The first systematic analytical method was developed at the German Agriculture Experiment Station at Weende (est. 1857), under the directorship of Henneberg. The systematic method for proximate analysis was widely adopted in the 1860's, and was the method used by Atwater in its first application in the United States, to analyze maize corn, (this occurred when he was a student at the Sheffield Scientific School of Yale University, in 1869). The Henneberg or Weende method (Z. Biol. 21:613, 1885) has remained virtually unchanged in principle up the the present.

Compositional data were first used to formulate animal feeds. The early enthusiasm for this approach to animal production waned when it was discovered, in cold economic terms, that all nitrogen was not created equal, that feeds of like proximate composition did not support the same growth of or yield from farm animals.

Atwater's classic studies of digestibility were a first step in taking account of physiological availability as a variable. Differences in protein quality were first assessed by animal growth and much of the mystery about balancing and blending of feeds was resolved only when analysis of amino acids and knowledge of essentiality and requirement became known.

Like problems continue to confront us, and I am heartened to see your continued efforts at their solution. Questions that your conference will address are: how to know how much of what is present in a food or feed is physiologically available for meeting nutrient needs - as consumed, in a host of mixtures, and prepared in different wrap and in different containers; and how does what is present in foods and diets relate to nutrient requirements, health and longevity?

This information can be put to fullest use only if we also can know what people do eat, what they formerly ate and what will be available for them to eat in the future. And know all this in an increasingly complex technological food-based system.

I salute your attention to this subject and leave you with a final thought, courtesy of Ogden Nash:

"Our daily diet grows odder and odder. It's a wise child that knows its fodder."

FROM THE EDITORS

Approximately 250 attendees met at the Golden Gateway Holiday Inn to discuss advances in the analysis of food/nutrient data and in dietary assessment methodology. The conference was organized into five sessions, over two and one-half days. The first morning session was devoted to presentations on national and international nutrient data banks. Topics included reports from USDA's Nutrient Data Research Branch, international sources of nutrient composition data, updates on INFOODS and EUROFOODS activities, and the results of a survey on guidelines for nutrient data banks. The afternoon session was oriented to new and prospective users, with discussions of computer issues, nutrient data bank features, and evaluating nutrient data banks.

The second day began with a session on nutrient composition issues. Problems and methods in applying nutrient composition data to current health issues were discussed, with emphasis on vitamin A and dietary fiber. A presentation of nutrient interaction issues followed, and the session was concluded with two reports from USDA's Nutrient Composition Lab. The afternoon session was devoted to small discussion groups, covering nutrient data bank applications (academic education, public awareness, clinical, food service, and industry/marketing) and issues (non-nutrient data and procedures for imputing values).

The theme for the final morning session was dietary assessment methodology. Topics included approaches to the nutritional assessment of population data, the variability of food intakes, dietary assessment of immigrant/refugee children, NHANES III plans, USDA food frequency methodology, the design of the NCI survey diet questionnaire, and the use of a shortened nutrient data bank to analyze national survey data.

The Eleventh Nutrient Databank Conference will be held at the University of Georgia on June 29 - July 2, 1986. For more information, contact: Dr. Wanda Grogan
Georgia Center for Continuing Education
The University of Georgia
Athens, Georgia 30602
Phone: 404/542-5654

Additional copies of this proceeding are available for a minimal charge from :

NTIS - Dept. of Commerce
5285 Port Royal Road
Springfield, VA 22161

TENTH NATIONAL NUTRIENT DATA BANK CONFERENCE

JULY 22 - 24, 1985
SAN FRANCISCO, CALIFORNIA

THE PEOPLE WHO MAKE THIS CONFERENCE POSSIBLE

CONTRIBUTORS:
- Consumer Nutrition Division, United States Department of Agriculture, Robert L. Rizek, director.
- Best Foods
- Borden Foundation
- Campbell Institute for Research and Technology
- Del Monte Corporation
- General Foods Corporation
- Hershey Foods Corporation
- Nabisco Brands
- The Pillsbury Company
- Ross Laboratories
- Stouffer Foods Corporation

STEERING COMMITTEE:
- Al Riley, Chair
- Loretta Hoover
- Suzanne Murphy
- Betty Perloff
- Ken Samonds
- Jack Smith

PROGRAM COMMITTEE:
- Ken Samonds, Chair
- Linda Hicks
- Hugh Joseph
- Suzanne Murphy
- Grace Petot
- Linda Posati
- Rita Warren

ARRANGEMENTS COMMITTEE:
- Ann Coulston, Co-chair
- Suzanne Murphy, Co-chair
- Ann Burroughs
- Pat Crawford
- Joanne Ikeda
- Molly Kretsch
- Arlyn Pittler
- Gerald Russell
- Barbara Schneeman
- Karen Todd

Co-hosts

University of California, Berkeley
Department of Nutritional Sciences
School of Public Health
Cooperative Extension

University of California, Davis
Department of Nutrition
Department of Food Science & Technology

Western Human Nutrition Research Center
USDA/ARS San Francisco

Arrangements Committee Co-chairmen

Suzanne Murphy and Ann Coulston
9 Morgan Hall
University of California
Berkeley, California 94720
(415) 642-5382

DATABASE COMMITTEE:
Loretta Hoover, Chair
Lenore Arab
Joanne Holden
Peter Ingerman
Robert Mathews
Suzanne Murphy
Grace Petot
Nancy Rawson
Ken Samonds
Lillian White

COMMUNICATIONS COMMITTEE:
Betty Perloff, Chair
Michael Montalto
Suzanne Pelican
Christine Shea

GUIDELINES COMMITTEE:
Jack Smith, Chair
Lenore Arab
Connie Dresser
Althea Engle
Frank Hepburn
Connie Hofland
John Klensin
Will Rand
Joyce Wenz
Karen Wilder

CONFERENCE COORDINATOR: Diane Rauchwarter

CONFERENCE SECRETARY: Daisy Wong

VOLUNTEERS:
Jackie Braisted, Coordinator
Charles DiSogra
Kathy Kelly
Esther Kwan
Kathy Mulligan
Wendy Russum
Helen Ullrich

TENTH ANNIVERSARY SUMMARY

Loretta W. Hoover
Dietetic Education
318 Clark Hall
University of Missouri-Columbia
Columbia, Missouri 65211

The first Nutrient Data Bank Conference was convened in Seattle, Washington on April 15-16, 1976. It was jointly sponsored by the American Dietetic Association and the American Academy of Pediatrics. The attendees invited to participate in this initial conference were actively involved in the use of computers in nutrition. The individuals represented health care organizations, universities, and federal agencies. They were nutritionists, dietitians, physicians, economists, and computer programmers. The thirty participants were divided into four task forces to consider the following issues: (1) record formats for data transfer, (2) data needs, (3) software capabilities, and (4) mechanisms for data dissemination.

During the first session of the initial conference, Dr. Robert Rizik discussed the plans for revision of USDA Handbook No. 8 and the creation of data bases to maintain the individual and aggregated food composition data. Much of the time spent in task force sessions focused on these proposed plans. Some recommendations were offered relative to the length of keys on the data records and conventions for loading data on magnetic tapes for distribution.

As conferees we were both excited and frustrated. All were excited by the prospect of more adequate food composition data and by accounts of innovative ways to use the computer in the field of nutrition. We were frustrated because so much work remained before the new data would be available.

Several participants agreed to share samples of their computer outputs by analyzing the same one-day dietary record. Joan Karkeck, editor of the proceedings, appended the sample output to the report of the conference. Much to our surprise, the results differed greatly. This situation caused us to realize that me must be concerned about methodological problems as well as food composition data.

The Nutrient Data Bank Conference became an annual conference when participants were invited to Utah State University in Logan, Utah in 1977 to continue the deliberations initiated in Seattle the previous year. Representatives from the food industry shared information about their role as data generators and data users. Subsequent conferences were held

Annually in Arlington, Virginia, Cleveland, East Lansing, Omaha, Philadelphia, Minneapolis, and Amherst, Massachusetts. Each of these conferences occurred as a result of the dedicated and voluntary effort of individuals living in the vicinity of the meeting site. Contributions to support the conferences have come from government, academic institutions, food industry, and software developers.

The Nutrient Data Bank Conference has remained viable during the past decade as a collaborative effort without a formal, legal structure. Although an organizational structure was proposed in 1979, the majority of the participants favored continuing to convene as a group of mutually concerned users of nutrient data without the requirements of a formal organization. Participants were encouraged to volunteer for a number of committees. A steering committee was formed with representation from each of the committees and the conference hosts. At each conference, individuals are offered the opportunity to volunteer for a committee. Each year more people learn about the conference and become actively involved in the collaboration. A mailing list maintained by Betty Perloff has grown to include over 900 names. After this conference, that number will exceed 1,000. Periodically, newsletters are sent to share information about conference activities and other topics of interest. Instead of dues, participants donate their time and effort and support of the conference activities.

The programs for the conferences have focused on the use of food constituent data in a variety of settings for numerous purposes. The generation and availability of nutrient data has been described by USDA staff. Dietary surveys conducted by USDA and HHS have been described. Data quality and analytic techniques have been reviewed. Methodological issues have been addressed relative to dietary record data collection and computation. Numerous applications for nutrient data bases have been described.

During the past decade, developers have been active in packaging their data bases for use by others. The advent of the microcomputer has hastened expansion in the software marketplace. Through the efforts of Donna Hay and Tony Fisher, the Nutrient Data Bank Directory was initiated as a project of the Conference to provide comparative information about those data bases. The first edition of the directory, prepared in 1980, included 28 systems. In 1982, the second edition contained 39 systems; in 1983, the third edition included 55 systems; in 1984, the fourth edition included 69 systems; (this year's supplement to the fourth edition added 18 new systems for a total of 87); and when the fifth edition is released in 1985, over 90 systems will be included.

The conference also fulfills an educational role for novices who are interested in learning about computerized nutrient analysis functions. Several years ago, a pre-conference program was instituted to meet the needs of inexperienced users. Although this program was elementary in nature, many experienced users continue to attend this portion of the

conference also. This year's conference has devoted a half-day session to the needs of new users.

International visitors have attended the conference to learn about developments in the United States and Canada. As the conference has grown in size and stability, we have been invited to represent North American interests in the International Network of Food Data Systems (INFOODS). Several participants in the Nutrient Data Bank Conference have been invited during the past year to serve on INFOODS working groups.

Responsible use of nutrient data is a concern of individuals active in the Nutrient Data Bank Conference. We have encouraged editors of professional journals to assure that references to nutrient data bases are appropriately documented in manuscripts. In 1983, conference participants began to deliberate the desirability and feasibility of standards or guidelines relative to the use of nutrient data bases. This issue continues to be unresolved.

The conference continues to be a forum relevant to those of us interested in nutrient data and analysis systems. We have seen much transition during the past decade and anticipate much more. We welcome new attendees to the conference. We hope that you will assist others to understand the issues surrounding the use of nutrient data and will help facilitate accomplishments in this area.

Our Bay Area hosts are offering a lovely setting for our Tenth Anniversary celebration. We anticipate an outstanding program supplemented by opportunities to address issues and exchange ideas. We are assured of an enjoyable experience as we conclude our first decade of the National Nutrient Data Bank Conferences here in San Francisco.

REPORT FROM USDA'S NUTRIENT DATA RESEARCH BRANCH

F. N. Hepburn
Human Nutrition Information Service
U.S. Department of Agriculture

The Nutrient Data Research Branch of the Human Nutrition Information Service (HNIS) is responsible for compiling and making available information on the nutrient composition of foods. This work includes not only the publication of reference tables but also the creation of nutrient data bases used for interpreting results of food consumption surveys conducted by another branch of HNIS.

This presentation reports on the status of publications, describes the derivation of the Primary Nutrient Data Set recently created for the Continuing Survey of Food Intakes by Individuals, and briefly outlines activities that will generate new analytical data and improve the reliability of nutrient composition data.

STATUS OF PUBLICATIONS

The Nutritive Value of Foods, Home and Garden Bulletin No. 72 (HG-72), has been completely revised as to food items, nutrients listed, and nutrient values. New foods have been added and foods no longer on the market or no longer popular have been deleted, resulting in a total of 910 items in place of 730 in the previous edition.

Data from published sections of revised Agriculture Handbook No. 8 (AH-8) have been incorporated, and all data have been reviewed, compared to new analytical data, and updated as necessary to reflect current knowledge. For example, new data for iron in beef have been included in accordance with the information on hand for the revision of the AH-8 section on beef.

Responding to popular demand, we have added data on sodium and cholesterol. Vitamin A is given in terms of both International Units (IU), for the convenience of those who continue to work in the traditional units, and Retinol Equivalents (RE), so that reference can be made to the current Recommended Dietary Allowances and foods can be compared in terms of relative vitamin A activity. Fatty acid content is expressed as total saturated, total monounsaturated, and total polyunsaturated fatty acids.

HG-72 is intended primarily for the general public and for educators and health professionals who help American understand the nutrient content of foods they eat. We expect great interest in its use as a data base for microcomputers. The National Technical Information Service will have available both computer tapes and floppy discs of HG-72 as soon as possible after its publication.

The revisions of Agriculture Handbook No. 8 (AH-8) have not progressed as well as predicted at last year's conference. The following table shows current estimated completion dates:

Publication Schedule for Revision of AH-8

Section No.	Food Group	Completion Date
13	Beef products	Winter 1985
14	Beverages	Winter 1985
15	Fish and shellfish	Spring 1986
16	Legumes	Summer 1986
17	Lamb, veal, and game	Winter 1986
18	Bakery products	Summer 1987
19	Sugars and sweets	Winter 1987
20	Cereal grains, flours, and pastas	Summer 1987
21	Fast foods	Fall 1986
22	Mixed dishes	Fall 1987
23	Miscellaneous foods	Fall 1987

The section on beef has first priority. It is entering final computer stages at this time and we are expecting its appearance in printed form sometime during the winter of 1985. Because the section on lamb, veal, and game is being prepared by the same staff now working on beef products, publication of section 17 will depend upon the actual completion date of section 13.

Sections 18 through 23 require later scheduling because they will include results of analyses now being performed under contract. Bakery products and mixed dishes are expected to require longer preparation time because of the complicated nature of the groups. Section 19 is being prepared by the staff working on 14; section 22 by those working on 21.

DATA BASE FOR CONTINUING SURVEY OF FOOD INTAKE BY INDIVIDUALS

A major task this past year was establishing the data base for the new continuing survey. Two factors have made it time consuming. One was the decision to expand the coverage of nutrients from the 14 used in the 1977-78 Nationwide Food Consumption Survey (NFCS) to double that number. The added components are --

Copper	Alpha tocopherol
Potassium	Vitamin A (RE)
Sodium	Carotene (RE)
Zinc	Folacin
Alcohol	Saturated fatty acids
Cholesterol	Monounsaturated fatty acids
Dietary fiber	Polyunsaturated fatty acids

The data base is less reliable for some of these components and interpretation of results will have to consider this limitation. The relative degree of reliability of the various components is discussed below.

The decision to track sodium and the different groups of fatty acids made the task difficult because of the need to provide for foods both with and without salt and for foods prepared with different sources of fat.

The new data base required a new procedure, which is described in the paper by Betty Perloff. Essentially, the procedure involves linking the foods reported by survey respondents to a Primary Nutrient Data Set for Food Consumption Surveys (PDS) through a recipe linking file. The PDS consists of foods that either are consumed as individual items or that are constituents of a food described by a recipe. To the extent possible, data for items are from the USDA Nutrient Data Base for Standard Reference. Additional items are added as required to describe foods being reported by the survey.

It is convenient to regard the PDS as a supplemented Standard Reference Data Base, but this is not strictly true because there are foods in the latter that are not used in the PDS. Values for all nutrients must be provided for new foods and data for those nutrients not included in the Standard Reference Data Base must be supplied as well. The sources of data used for creating the PDS were documented by code according to the following scheme:

Code	Description of Data Source
1	Analytical data, or values calculated directly from analytical data, from revised AH-8 sections.
2	New provisional analytical data from Nutrient Data Bank for food group sections not yet revised.
3	Unrevised analytical data from computer data set 456-3 (based on the 1963 AH-8).
4	Imputed values from USDA Nutrient Data Base for Standard Reference.
5	Nutrient label claims (Breakfast cereals only; AH-8-8).
6	New imputed values for Primary Nutrient Data Set.
7	Assumed zero values.

Various meanings of the word "imputed" are described in Linda Posati's paper on this subject. I am using the word in the same sense as it is used in regard to AH-8 and the Standard Reference Tape. Values in the printed publication are analytical or are calculated in a very direct manner from analytical data; they

are not considered to be imputed. Missing values are left blank. For the Standard Reference Data Base, blanks are filled with imputed values and flagged to differentiate them from other values. These imputed values are our best estimates, usually based on values for a similar food or another form of the same food.

Source codes are attached to each nutrient value in the Primary Data Set. We can examine the data for each nutrient for reliability in terms of data source; that is, the extent to which they are based on analytical as opposed to imputed values. To determine the proportion of analytical to imputed data in the Primary Data Set, data from source codes 5 and 7 were first eliminated. Data from source

codes 1, 2, and 3 were combined as analytical values and those from source codes 4 and 6 were combined as imputed values. About 2,250 total values were found for each nutrient. At least 2,100 remained after dropping code 5 and 7 data except for vitamin B_{12}, carotene, and dietary fiber. These exceptions were due, of course, to assumption of the absence of these components in large numbers of foods. The proportion of analytical data was calculated as the percentage of total data coded as either analytical or imputed. Results are shown in the following table:

Percent of Data from Analytical Sources in Primary Data Set

90 percent or more analytical		Less than 90 percent analytical		
Component	All Foods	Component	All Foods	Major Sources
Calcium	97	Vitamin C	83	92
Protein	97	Vitamin A (IU)	80	89
Fat	96	Magnesium	75	72
Thiamin	91	Zinc	73	79
Riboflavin	91	Copper	67	71
Niacin	91	Vitamin B_6	64	72
Sodium	90	Vitamin B_{12}	64	70
Potassium	90	Vitamin A (RE)	61	73
Phosphorus	90	Folate	56	69
Iron	90	Carotene (RE)	54	88
		Dietary fiber	29	40
		α-tocopherol	28	39

It is evident that the proportion of analytical data is high for the more familiar nutrients that have been tracked over a longer period, equaling or exceeding 90 percent, whereas analytical data of the newest components are below 30 of percent data from all foods. The column on the far right of the table shows results for only major food sources of nutrients. For this analysis, we excluded those food sources that provide less than 5 percent of the U.S. RDA per serving portion for most of the nutrients. For other components, cut off values were vitamin A, 50 IU or 12.5 RE; carotene, 12.5 RE; α-tocopherol, 1.0 mg; copper, 0.1 mg; and dietary fiber, 2.5 g. Portion sizes were based primarily on the default amounts of the NFCS 1977-78, amended for some foods for improved uniformity. Except for magnesium, percent analytical of the better sources is greater than that for all foods, indicating that more of the imputed data are from relatively poor nutrient sources and that more analytical values are available for better sources. The magnesium exception reflects a more general distribution of that element in the food supply. Although the better sources of vitamin E and dietary fiber showed higher percentages of analytical data than for all foods, the values remain relatively low, indicating that our knowledge of these components is much weaker than our knowledge of other food components being studied.

It must be remembered that we have not considered in these calculations the relative quality of analytical data. For example, even though data for the better sources of carotene are 88 percent analytical, these data are based primarily on the AOAC procedures, which generally report total carotenes.

Although amounts may be reported in terms of beta carotene, they may not be specific for beta carotene itself. We expect that results from ongoing contracts supported by the Human Nutrition Information Service and research now going on in the Nutrient Composition Laboratory will provide new data on the separate carotenoid components.

REVISION OF THE NUTRIENT DATA BANK SYSTEM

Under a HNIS sponsored contract, work is nearing completion to revise the Nutrient Data Bank computer system to take advantage of advances in computer technology since the original system was developed ten years ago. The overall effect will be to improve the efficiency of operation. The major change for our staff is that the system will be interactive with our food group specialists. Each specialist will have access to his or her own data sets through a terminal and will be able to examine and compare individual data entries in the process of summarization.

Provision has been made for the future encoding of quality indices of the data so that we will be able to generate confidence codes that apply to the summarized values. This feature is now only in the planning stage, but it will become possible as a result of the revision.

EXTRAMURAL RESEARCH

HNIS funds food composition research in several universities and elsewhere. Most of these research contracts generate data on the nutrient content of foods to provide information needed for the revision of specific food-group sections of the handbook. A few, such as one on carotenoids, are component specific. One project of particular interest to the Nutrient Data Bank Conference, is the development of a Guide for Nutrient Data Users, the product of a cooperative agreement with Grace Petot of Case Western Reserve University. We plan to make this new guide available with the purchase of nutrient data tapes as well as by separate purchase.

New contracts have just been awarded on miscellaneous foods, frozen prepared dishes, and fried foods. These new contracts are the last of those intended for remaining AH-8 sections. Other major research activities will be carried out in cooperation with the Nutrient Composition Laboratory at Beltsville, MD. These will include studies on selenium in foods and a study on the nutrient composition of cookies and sweet-snack bakery items.

FUTURE ACTIVITIES

Efforts will continue on revising AH-8. In addition to following the schedule on remaining sections, work will begin on updating food group sections previously published.

The nutrient data base for the CSFII will be expanded as the survey continues. At the same time we will develop the data base for the household segment of the 1987 NFCS.

We appreciate the opportunity to communicate with data users at the annual Nutrient Data Bank Conference. Your input is especially valuable in providing direction for our efforts.

Recipe Calculations for NFCS Data Base

Betty P. Perloff
U.S. Department of Agriculture
Human Nutrition Information Service
Hyattsville, MD 20782

The U.S. Department of Agriculture (USDA) recently developed an automated system to create nutrient data bases for appraising the nutrient content of food intakes by individuals reported in dietary surveys. The Human Nutrition Information Service used the system for the first time in early 1985 to create the nutrient data base for appraising intakes from the Continuing Survey of Food Intakes by Individuals (CSFII) and will use it again to create the nutrient data base for the 1987 Nationwide Food Consumption Survey (NFCS). The system uses the USDA Nutrient Data Base for Standard Reference (1) as the major source for nutrient values, and it includes procedures for calculating the nutrient content of recipes based on nutrient data for the individual components. This paper describes the recipe calculation method and how it operates within the framework of the new computer system.

THE COMPUTER SYSTEM

A computer program, which forms the nucleus of the system, creates the survey nutrient data base, calculating the nutrient content of recipes as needed. This program and the following data sets make up the new system (Figure 1).

Primary Nutrient Data Set for Food Consumption Surveys. The Primary Nutrient Data Set for Food Consumption Surveys (PDS) contains nutrient values for all food items needed to create a survey nutrient data base, including all items used as ingredients in recipes. The food components for which data are included are listed in Table 1.

Most of the data come from the USDA Nutrient Data Base for Standard Reference. Some changes and additions to data from the Standard Reference Data Base were made. Changes were made to reflect current data soon to be used in the revision to Agriculture Handbook No. 8 (2) if the newer values differed substantially from the older values. Food groups in which these changes were made include beef, beverages, sugars and sweets, bakery products, and fish. Nutrient values were added as needed for nutrients that are not in the Standard Reference Data Base (e.g., dietary fiber), and complete nutrient profiles were added for missing food items. If analytical data were not available, the added values were imputed from other forms of the foods, or estimates were derived from data for similar foods.

All items from the Standard Reference Data Base carry the Standard Reference identification numbers, referred to as NDB numbers. Added food items have been assigned special NDB numbers. The PDS currently contains data for 2,032 foods. Data are expressed as the amount of nutrient in 100 grams of the edible portion of a food.

Table of Nutrient Retention Factors. This data set contains the factors for calculating retention of 18 vitamins and minerals during cooking. It is based primarily on the "Table on Percent Retention of Nutrients in Food Preparation" (3) but contains several additional specific categories of foods and cooking methods. Because analytical data on nutrient retention are not available for all nutrients in each specific category, missing factors were estimated to complete the table. Each category is assigned a code for computer access.

Recipe File. This data set controls the generation of a survey nutrient data base using the PDS and the table of retention factors. The items to be included in a survey data base are designated and survey food codes assigned before this file is constructed. In this file, each survey food code is linked to one or more PDS items through a set of recipe codes. Links to single PDS items are treated as one-component recipes. The information required for each recipe is listed below:

1. Recipe components.
   a. Names.
   b. NDB numbers.
   c. Weights of the components in grams, excluding the weight of any refuse.
   d. Retention codes, where applicable.

2. Changes in moisture and/or fat that occur during cooking, expressed as a percentage (plus or minus) of the total weight of the uncooked recipe.

3. Percent yield of the recipe. This is the final weight of the cooked recipe, expressed as a percentage of the uncooked weight. The percent yield is not used in the recipe calculation but is used for the following edit check: Yield = 100 ± moisture change ± fat change.

4. The NDB number for the type of fat (only for recipes with a fat change).

The recipe file contains approximately 4,450 items: one item for every food listed in the survey code book for the Continuing Survey of Food Intakes by Individuals. Approximately half are one-component recipes--direct links to single items on the PDS. If the food code manual is revised for future surveys, the recipe file will be revised accordingly.

USDA Nutrient Data Base for Individual Intake Survey. This data set is the system's output and is the nutrient data base created for analysis of food consumption survey data. All nutrient values come from the PDS--either directly or through recipe calculations. The program transfers survey food codes to this file from the recipe file as nutrient values are placed here for each item. Nutrient values are expressed on the basis of 100 grams edible portion. This data base may also be used as an input file to the recipe calculation program because values calculated

from recipes may be used for ingredients in other recipes. For example, in the data base created for the CSFII, values for cornbread were calculated from a recipe and subsequently used in calculating values for frozen dinners in which cornbread was an ingredient.

RECIPE CALCULATION METHOD

The recipe calculation method used in the new system is a modification of the procedure described in Bulletin ARS 62-13, "Procedures for Calculating Nutritive Values of Home-Prepared Foods: As Used in Agriculture Handbook No. 8, 'Composition of foods---raw, processed, prepared,' Revised 1963" (4). The method described in that bulletin calls for applying vitamin retention factors to the total recipe nutritive values. For the new system, the method was modified to include retention factors for minerals and to apply retention factors to vitamin and mineral values for each recipe component. This change permits using different retention factors for different components and was made because nutrient retention information is more readily available for individual foods than for mixed dishes.

The calculation procedure involves seven basic steps:

1. Determining the weight in grams of each ingredient and subtracting the weight of any inedible part, such as bone. USDA publications are used as sources of data on weight-volume relationships (5,6,7) and refuse (5,6) of ingredients. This step is not a part of the automated procedure.

2. Determining the nutrients in the specified weight of each ingredient. Nutrient values for 100-gram portions of ingredients are stored in the Primary Nutrient Data Set for Food Consumption Surveys.

3. Applying retention factors to vitamin and mineral values where losses may occur during cooking. Retention factors are contained in the Table of Retention Factors.

4. Determining total uncooked weight of the recipe by summing weights of the ingredients.

5. Determining nutrients in the total recipe by summing nutrient values for the ingredients.

6. Adjusting the total values to account for changes in moisture and fat during cooking. Moisture may be lost through evaporation or drippings, or it may be gained through absorption. The total weight of the recipe and the total moisture value are adjusted at this step. (Vitamin and mineral losses are calculated in step 3.) Fat may be lost through drippings or gained through absorption during frying. Fat changes affect total weight, energy, total fat, and fatty acids and sometimes also affect cholesterol, minerals, and fat-soluble vitamins. These values are adjusted at this step. Information on the amount of moisture and fat changes during cooking are taken from USDA publications (4,8) or are

derived from unpublished materials used in the development of Agriculture Handbook 8 revisions (2,6).

7. Converting nutrient values for the total recipe to the 100-gram basis. Steps 2 through 7 are performed by the computer program.

To illustrate the procedures and how they involve the system data sets, here are the calculations for a sample recipe--flounder fillet, breaded, fried.

Recipe information entered into the Recipe File:

Recipe Components	NDB No.	Retention Code	Weight of Edible Part
1. Flounder, raw	80180	2310	907.2 g
2. Egg, raw	01123	0103	50.0 g
3. Milk	01077		15.2 g
4. Bread crumbs, dry	74750	0305	100.0 g
5. Salt	89630		5.5 g

Moisture change = -20%
Fat change = +8%
Yield = 88%
Fat NDB No. = 04031 (vegetable shortening absorbed during frying)

The recipe program locates each NDB number on the Primary Nutrient Data Set, calculates the nutrients for the specified weight, and applies the appropriate set of retention factors to the resulting nutrient values if a retention code has been designated. Calculations for thiamin are presented below as an example.

	NDB Number	Thiamin in 100 grams (from PDS)		Weight		Thiamin in recipe portion		Retention factor		Thiamin (corrected)
		mg		g		mg				g
1.	80180	0.050	X	907.2 / 100	=	0.454	X	.85	=	0.386
2.	01123	0.087	X	50.0 / 100	=	0.044	X	.85	=	0.037
3.	01077	0.038	X	15.2 / 100	=	0.006			=	0.006
4.	74750	0.350	X	100.0 / 100	=	0.350	X	.75	=	0.262
5.	89630	0.000	x	5.5 / 100	=	0.000			=	0.000

The remaining steps in the recipe calculation procedure are illustrated in Table 2 for energy and five nutrients. The weight and nutrient values for the individual ingredients are summed, and moisture and fat changes are calculated by multiplying the total weight by the input data for "Moisture change" and "Fat change." The nutrient data for the type of fat absorbed during frying is accessed in the PDS by the NDB number entered for "Fat NDB No.," and the individual nutrients in the fat are calculated for the amount of fat absorbed. These values are applied to the subtotals to determine the weight and nutrient content of the cooked recipe, and all nutrient values are converted to the 100 gram basis for storage in the survey nutrient data base.

Recipe Report. In addition to the survey data base created by the program, a recipe report is generated for each recipe. The sample recipe report for the founder fillet is presented in Figure 2. The first part of the report contains the input information. Names of the individual components taken from the PDS are printed next to the name from the input record and can be reviewed to check the NDB numbers. This feature was adapted from a program used in the USDA Lipid Nutrition Laboratory in Beltsville, MD.

AVAILABILITY AND BENEFITS

The nutrient data base created for the CSFII will be available to the public sometime during 1986. The other data sets used by this new automated system will also be made available for public use.

A major benefit of this system to USDA is the ability to automatically create and update nutrient data bases for food consumption surveys. An important part of the automated process is calculation of the nutrient content of recipes. An equally important benefit to USDA and the users of USDA's food consumption survey data is the machine-readable documentation of the recipes used in those calculations.

REFERENCES

1. U. S. Department of Agriculture, Human Nutrition Information Service. 1985. Machine-Readable Data Sets on Composition of Foods and Results from Food Consumption Surveys. Admin. Rep. No. 378, 22 pp.

2. Watt, B.K. and A.L. Merrill. 1963. Composition of Foods...Raw, Processed, Prepared. U. S. Dept. of Agriculture, Agric. Handb. No. 8, rev., 190 pp.

3. U. S. Department of Agriculture, Human Nutrition Information Service. 1984. Provisional Table on Percent Retention of Nutrients in Food Preparation. 2 pp.

4. Merrill, A.L., C.F. Adams, and L.J. Fincher. 1966. Procedures for Calculating Nutritive Values of Home-Prepared Foods: As Used in Agriculture Handbook No. 8, "Composition of Foods--Raw, Processed, Prepared" Revised 1963. U. S. Dept. of Agriculture, ARS 62-13, 35 pp.

5. Adams, C.F. 1975. Nutritive Value of American Foods. U. S. Dept. of Agriculture, Agric. Handb. No. 456, 291 pp.

6. U.S. Department of Agriculture. 1976- . Composition of Foods: Raw, Processed, Prepared. Agric. Handb. No. 8: AH-8-1, Dairy and Egg Products, 1976; AH-8-2, Spices and Herbs, 1977; AH-8-3 Baby Foods, 1978; AH-8-4, Fats and Oils, 1979; AH-8-5, Poultry Products, 1979; AH-8-6, Soups, Sauces, and Gravies, 1980; AH-8-7, Sausages and Luncheon Meats, 1980; AH-8-8, Breakfast Cereals, 1982; AH-8-9, Fruits and Fruit Juices, 1982; AH-8-10, Pork and Pork Products, 1983; AH-8-11, Vegetables and Vegetable Products, 1984; AH-8-12, Nut and Seed Products, 1984.

7. Fulton, L., E. Matthews, and C. Davis. 1977. Average Weight of a Measured Cup of Various Foods. U. S. Dept. of Agriculture, Home Econ. Res. Rep. No. 41, 26 pp.

8. Matthews, R.H. and Y.J. Garrison. 1975. Food Yields Summarized by Different Stages of Preparation. U. S. Dept. of Agriculture, Agric. Handb. No. 102, 136 pp.

Table 1. Food Components in the PDS

Nutrient	Units
Energy	Kilocalories
Moisture	Grams
Protein	Grams
Fat	Grams
Total saturated fatty acids	Grams
Total monunsaturated fatty acids	Grams
Total polyunsaturated fatty acids	Grams
Carbohydrate	Grams
Calcium	Milligrams
Iron	Milligrams
Magnesium	Milligrams
Phosphorus	Milligrams
Potassium	Milligrams
Sodium	Milligrams
Zinc	Milligrams
Copper	Milligrams
Vitamin C	Milligrams
Thiamin	Milligrams
Riboflavin	Milligrams
Niacin	Milligrams
Vitamin B6	Milligrams
Folacin	Micrograms
Vitamin B12	Micrograms
Vitamin A	International Units
Vitamin A	Retinol Equivalents
Carotene	Retinol Equivalents
Vitamin E	Alpha-tocopherol Equivalents
Cholesterol	Milligrams
Alcohol	Grams
Total dietary fiber	Grams

Table 2. Selected Food Components in Flounder Fillet, Breaded, Fried

	Weight	Energy	Moisture	Total Fat	Saturated Fatty Acids	Mono unsaturated Fatty Acids	Poly unsaturated Fatty Acids
	g	Kcal	g	g	g	g	g
1.	907.2	717	737.6	7.2	1.8	1.8	2.7
2.	50.0	79	37.3	5.6	1.7	2.2	.7
3.	15.2	9	13.3	.5	.3	0.1	.0
4.	100.0	392	6.5	4.6	1.0	1.6	1.5
5.	5.5	0	0.0	0.0	0.0	0.0	0.0
Subtotals	1,077.9	1,197	794.7	17.9	4.8	5.7	4.9
Moisture change	-215.6		-215.6				
Fat Change	+86.2	+762		+86.2	+21.6	+38.4	+22.5
Yield	948.6	1,959	579.1	104.1	26.4	44.1	27.4
per 100 Grams	100.0	207	61.0	11.0	2.8	4.6	2.9

Figure 1. System Components

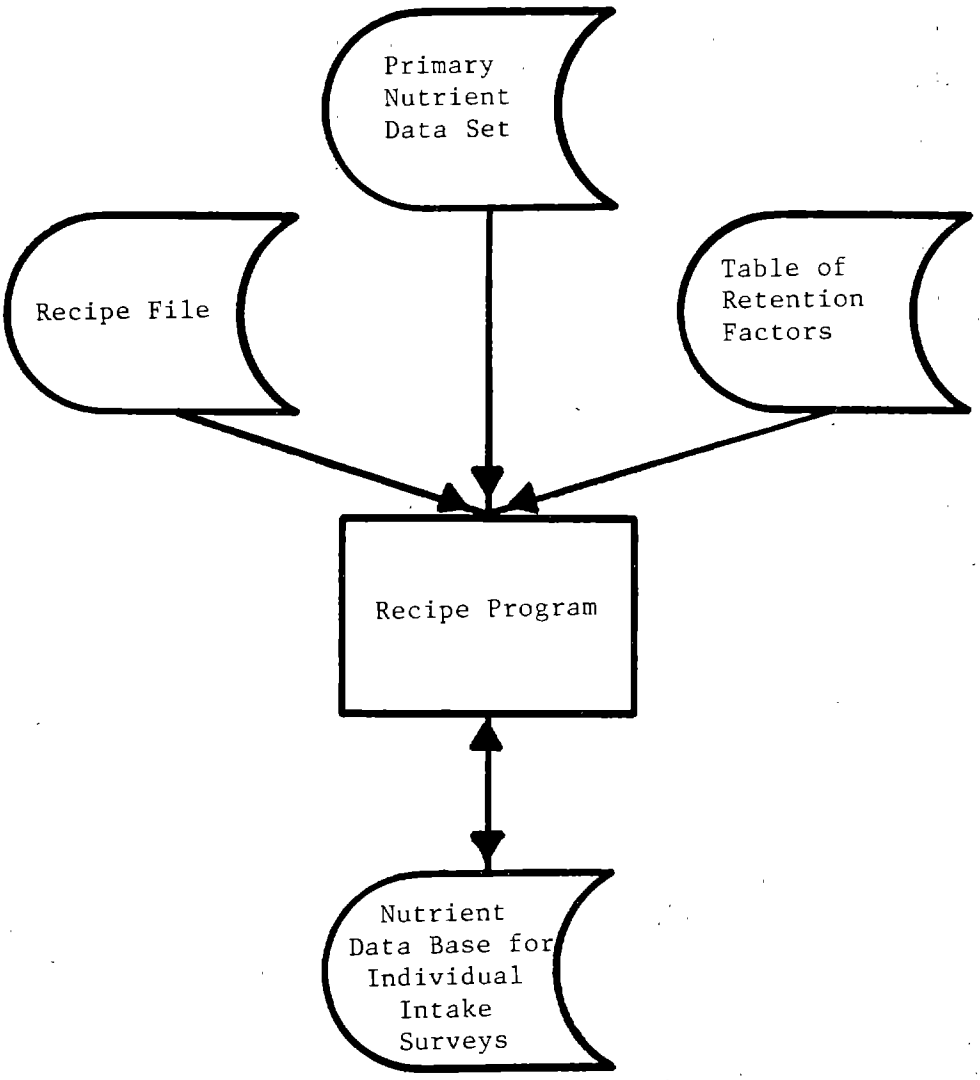


Figure 2. Recipe Report

NOVEMBER 25, 1985 INDIVIDUAL SURVEY NUTRIENT RECORD CREATED FROM STANDARD REFERENCE DATA BASE PAGE 1

FLOUNDER, FILLET, BREADED, FRIED YIELD: 88.0 MOIS CHANGE:-20.0 FAT CHANGE 8.0 FAT ID : 4031 FC: 261-1514

ADD	FLAG	COMP	NDB_NO	INGREDIENT_INPUT_NAME	STD_REF_NAME	RETENTION	MEASURE	GRAMS	PERCENT
		1	80180	FLATFISH,FLOUNDER,RAW	FLATFISHES,RAW	2310	2 LB	907.200	84.164
		2	1123	EGG,RAW	WHOLEEGG FRESHFROZEN	103	1	50.000	4.639
		3	1077	MILK	3.3% FAT WHOLE MILK	0	1 TBSP	15.200	1.410
		4	74750	BREAD CRUMBS,DRY	BREADCRUMB,DRY,GRATE	305	1 C	100.000	9.277
	5	5	89630	SALT	SALT	0	1 TSP	5.500	0.510

COMPONENTS	WEIGHT G	ENERGY CAL	MOISTURE G	PROTEIN G	FAT G	SAT. F.A. G	MONO. F.A. G	POLY. F.A. G
1 FLATFISH,FLOUNDER,RAW	907.200	716.687	737.553	151.502	7.258	1.814	1.814	2.722
2 EGG,RAW	50.000	78.959	37.285	6.070	5.575	1.674	2.228	0.725
3 MILK	15.200	9.339	13.374	0.500	0.508	0.316	0.147	0.019
4 BREAD CRUMBS,DRY	100.000	392.000	6.500	12.600	4.600	1.050	1.580	1.470
5 SALT	5.500	0.0	0.011	0.0	0.0	0.0	0.0	0.0
SUBTOTALS:	1077.900	1196.986	794.723	170.672	17.940	4.854	5.769	4.935
MOIS/FAT CHANGE:	-129.348	762.291	-215.580	0.0	86.232	21.558	38.373	22.507
YIELD:	948.552	1959.276	579.143	170.672	104.172	26.412	44.142	27.441
PER 100 GRAMS:	100.000	206.554	61.055	17.993	10.982	2.784	4.654	2.893

COMPONENTS	CARBO G	CALCIUM MG	IRON MG	MAGNESIUM MG	PHOSPHORUS MG	POTASSIUM MG	SODIUM MG	ZINC MG
1 FLATFISH,FLOUNDER,RAW	0.0	108.864	7.258	272.160	1769.039	3102.624	707.615	4.536
2 EGG,RAW	0.600	28.050	1.045	6.145	90.050	64.950	69.150	0.720
3 MILK	0.708	18.149	0.008	2.043	14.197	23.028	7.448	0.058
4 BREAD CRUMBS,DRY	73.400	122.000	4.100	38.000	141.000	152.000	736.000	0.640
5 SALT	0.0	13.915	0.005	6.545	3.300	0.220	2131.690	0.0
SUBTOTALS:	74.708	290.978	12.416	324.892	2017.586	3342.821	3651.903	5.954
MOIS/FAT CHANGE:	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
YIELD:	74.708	290.978	12.416	324.892	2017.586	3342.821	3651.903	5.554
PER 100 GRAMS:	7.876	30.676	1.309	34.251	212.702	352.413	384.998	0.628

Figure 2--Con.

COMPONENTS	COPPER MG	VIT C MG	THIAMIN MG	RIBOFLAVIN MG	NIACIN MG	VIT B6 MG	FOLACIN MCG
1 FLATFISH,FLOUNDER,RAW	1.814	0.0	0.386	0.431	15.422	1.388	89.813
2 EGG,RAW	0.031	0.0	0.037	0.143	0.029	0.057	24.375
3 MILK	0.002	0.143	0.006	0.025	0.013	0.006	0.760
4 BREAD CRUMBS,DRY	0.204	0.0	0.262	0.315	4.320	0.036	31.850
5 SALT	0.024	0.0	0.0	0.0	0.0	0.0	0.0
SUBTOTALS:	2.075	0.143	0.691	0.914	19.785	1.487	146.798
MOIS/FAT CHANGE:	0.0	0.0	0.0	0.0	0.0	0.0	0.0
YIELD:	2.075	0.143	0.691	0.914	19.785	1.487	146.798
PER 100 GRAMS:	0.219	0.015	0.073	0.096	2.086	0.157	15.476

COMPONENTS	VIT B12 MCG	VIT A IU	VIT A RE	CAROTENE RE	A-TOCO. EQ. MG	CHOLEST MG	ALCOHOL G	TO.D.FIBER G
1 FLATFISH,FLOUNDER,RAW	9.798	385.560	115.668	0.0	5.443	453.600	0.0	0.0
2 EGG,RAW	0.657	260.000	78.000	0.0	0.370	273.800	0.0	0.0
3 MILK	0.054	19.152	4.712	0.456	0.014	2.067	0.0	0.0
4 BREAD CRUMBS,DRY	0.0	0.0	0.0	0.0	0.820	1.000	0.0	1.400
5 SALT	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
SUBTOTALS:	10.509	664.712	198.380	0.456	6.647	730.467	0.0	1.400
MOIS/FAT CHANGE:	0.0	0.0	0.0	0.0	12.504	0.0	0.0	0.0
YIELD:	10.509	664.712	198.380	0.456	19.150	730.467	0.0	1.400
PER 100 GRAMS:	1.108	70.076	20.914	0.048	2.019	77.009	0.0	0.148

FAT USED: SHORTNNG,REG,SOY/COT

INTERNATIONAL SOURCES OF NUTRIENT COMPOSITION DATA

Kenneth W. Samonds, University of Massachusetts, Amherst

Why do investigators and clinicians in the US, where we are supplied with good and ever-improving food composition data, need to know more about data sources from other countries? I think there a four reasons:

1. There is an increasing interest in the nutritional status of ethnic population subgroups, especially those people who have immigrated to the U.S. but continue to consume the foods of their country of origin. The Vietnamese and Hatian refugees are examples of groups who may receive a substantial portion of their nutrients from foods not commonly consumed in the U.S. and therefore not included in the U.S.D.A. data.

2. There is a greater availability of international foods in the American food supply and a greater dietary diversity among the population in general. Food intake records frequently include foods not in the U.S.D.A. tables.

3. American investigators are becoming increasingly involved in international nutrition studies. In my experience, anthropologists are the most frequent solicitors of food composition information.

4. Most database systems now available make it possible to add data for foods not originally included which may be of particular interest in a database application.

What are the problems of using international nutrient data in order to fill gaps in databases available to us?

1. Sources of international data are difficult to identify. Currently there is no complete bibliography of nutrient data. The bibliography accompanying this presentation depended for a large part on a first draft prepared by Will Rand and INFOODS, for which I am grateful. Presumably this problem, the lack of a comprehensive bibliography, will be eliminated through their efforts in the near future.

2. Once you know that a publication, like those included in the bibliography, exists, it is sometimes very difficult to obtain a copy. Many of these publications are out of print. The INFOODS bibliography will include information concerning the source and cost of many publications.

3. Even with a publication in hand, the interpretation of the data can be a problem. Most tables are published in the country's indigenous language, although many have English versions or English indexes. One can often translate the nutrient names, the column headings, but the food names are usually a problem unless you speak the language.

4. The layout of the tables vary from publication to publication. All information may be in a single table or may be broken up into subtables. Foods may be arranged alphabetically (which is little help if you are unfamiliar with the alphabet) or may be arranged by food groups. These groupings may be different, however, than those to which one is accustomed, eg. olives listed as a fruit. (Olives appear to be neither a fruit or a vegetable in the U.S.D.A. Handbook 8 series!)

INTERNATIONAL SOURCES OF NUTRIENT COMPOSITION DATA

Kenneth W. Samonds, University of Massachusetts, Amherst

Why do investigators and clinicians in the US, where we are supplied with good and ever-improving food composition data, need to know more about data sources from other countries? I think there a four reasons:

1. There is an increasing interest in the nutritional status of ethnic population subgroups, especially those people who have immigrated to the U.S. but continue to consume the foods of their country of origin. The Vietnamese and Hatian refugees are examples of groups who may receive a substantial portion of their nutrients from foods not commonly consumed in the U.S. and therefore not included in the U.S.D.A. data.

2. There is a greater availability of international foods in the American food supply and a greater dietary diversity among the population in general. Food intake records frequently include foods not in the U.S.D.A. tables.

3. American investigators are becoming increasingly involved in international nutrition studies. In my experience, anthropologists are the most frequent solicitors of food composition information.

4. Most database systems now available make it possible to add data for foods not originally included which may be of particular interest in a database application.

What are the problems of using international nutrient data in order to fill gaps in databases available to us?

1. Sources of international data are difficult to identify. Currently there is no complete bibliography of nutrient data. The bibliography accompanying this presentation depended for a large part on a first draft prepared by Will Rand and INFOODS, for which I am grateful. Presumably this problem, the lack of a comprehensive bibliography, will be eliminated through their efforts in the near future.

2. Once you know that a publication, like those included in the bibliography, exists, it is sometimes very difficult to obtain a copy. Many of these publications are out of print. The INFOODS bibliography will include information concerning the source and cost of many publications.

3. Even with a publication in hand, the interpretation of the data can be a problem. Most tables are published in the country's indigenous language, although many have English versions or English indexes. One can often translate the nutrient names, the column headings, but the food names are usually a problem unless you speak the language.

4. The layout of the tables vary from publication to publication. All information may be in a single table or may be broken up into subtables. Foods may be arranged alphabetically (which is little help if you are unfamiliar with the alphabet) or may be arranged by food groups. These groupings may be different, however, than those to which one is accustomed, eg. olives listed as a fruit. (Olives appear to be neither a fruit or a vegetable in the U.S.D.A. Handbook 8 series!)

5. Missing data can be a serious problem. Many of these publications are old and contain data for only the basic nutrients. Some tables include food items for which absolutely no compositional data are supplied. Often, a little bit of data is as bad as no data at all.

6. Conflicting data are as frustrating as missing data. Differing data for the same food item in different sources raises the question "which is the more appropriate for my purposes?" Agreement of data from different publications should not be misconstrued as a sign of validity, however, because there is considerable "borrowing" of values between publications.

7. Many publications do not list sources of data, the "age" of the analyses, the number of samples, the sampling procedures, or the reliability of the analytical methods used. This is particularly a problem with the older publications.

8. Complex food items are a particular problem. Some tables include representative recipes for common dishes, but most do not.

In conclusion, international food composition data, with all their inherent problems, can still play an important role in the expansion of existing databases and the analysis of dietary surveys. We should use the informative, well-documented, and more-complete USDA publications when we can, but we are occasionally forced to supplement these data from other sources. If interpreted correctly and evaluated critically, data from other countries can be a useful supplement to existing databases.

SOURCES OF INTERNATIONAL FOOD COMPOSITION DATA

Prepared by Ken Samonds, with special appreciation to Will Rand and INFOODS for sharing a draft copy of the Directory of Food Composition Data Tables.

INTERNATIONAL

Amino Acid Content of Foods and Biological Data on Proteins.
Food Consumption and Planning Branch
Nutrition Division
FAO, Rome
1969

Chemical and Nutritional Composition of Finfishes, Whales, Crustaceans, Mollusks and Their Products.
NOAA Technical Memorandum NMFS F/SEC-11
Sidwell, Virginia D.
National Marine Fisheries Service
US Department of Commerce

Fatty Acids in Animal and Plant Products
Goddard, V.R. and L. Goodall
Human Nutrition Research Division
Agricultural Research Service
USDA
1959

Food Composition Tables - Minerals and Vitamines - for International Use.
Chatfield, C.
Food and Nutrition Studies No. 11
FAO, Rome
1954

Pantothenic Acid, Vitamin B_6 and Vitamin B_{12}.
Orr, M.L.
Home Economics Research Report No. 36
USDA
1969

Proteins in Foods.
Kuppuswamy, S., M. Srinivasan, and V. Subrahmanyan
Special Report Series No. 33
Indian Council of Medical Research, New Delhi
1958

Table Internationale de Composition Chemique des Aliments.
Brichard, J.P. de, R. Richard, H.Triaca
Francais, English, Espanol, Deutsch, Italiano
Cinquieme ed. revue et augmentee
1957

Tables of Representative Values of Foods Commonly Used in Tropical Countries, revised ed.
Platt, B.S.
Special Report Series No. 253, Medical Research Council
Her Majesty's Stationery Office, London
1962

Tables of Amino Acids in Foods and Feeding Stuffs.
Harvey, D.
Technical Communication No. 19
Commonwealth Bureau of Animal Nutrition, Aberdeen
1970

AFRICA

Africa, General
Food Composition Table for use in Africa
Leung, Woot-Tsuen Wu
FAO and US Dept. Health, Education and Welfare
1968

Cameroon
Les Aliments d'Origine Vegetale au Cameroun
Pele, J. and S. Le Berre
Nutrition Section
Office de la Recherche Scientifique et Technique Outre-Mer
Yaounde
1966

Cameroon
Table Provisoire de Composition des Aliments du Sud-Cameroun
Bergeret, B. and R. Masseyeff
Office de la Recherche Scientifique et Technique Outre-Mer
Vol XI pp 47-69, Ann. Nut. Alim.
1957

Ethiopia
Expanded Food Composition Table for Use in Ethiopia
Ethiopian Nutrition Institute

Ethiopia
Food Composition Table for Use in Ethiopia I
Agren, Gunnar and Rosalind Gibson
Ethiopian Nutrition Institute
Swedish International Development Authority
1968

Ethiopia
Food Composition Table for Use in Ethiopia II: Amino Acid Content and Biological Data on Proteins in Ethiopian Foods.
Agren, Gunnar, Aders Eklund, and Sten-Ake Lieden
Ethiopian Nutrition Institute
1975

Ghana
Food Composition Table
Food Research and Development Unit
Food Research Institute
Accra, Ghana
1969

Kenya
Chemical Composition of Some Kenya Foodstuffs
Records of the Medical Research Laboratory No. 10
Harvey, D.
Medical Department
Government Printer, Nairobi
1951

Senegal
Chemical Composition of 66 Samples of Leaves, Fruits, Seeds and Tubercles Eaten in Senegal
Vol. 8, No. 2, pp. 139-155
Qualitas Plantarum et Materiae
1961

South Africa
NRIND Food Composition Tables 1981
Gouws, E. and M.L. Langenhoven
National Research Institute for Nutritional Diseases
South African Medical Research Council
1982

South Africa
Studies on the Chemical Composition of Foods Commonly Used in South Africa
Fox, F.W.
South African Institute for Medical Research
Johannesburg
1966

Tanzania
The TanzaNIA Food Tables
Marealle
undated

Togo
Table de Composition des Aliments du Togo
Perisse, J. and S. Le Berre
Office de la Recherche Scientifique et Technique Outre-Mer
Vol XI, pp 70-89
Ann. Nutr. Alim.
1957

West Africa
Aliments de l'ouest Africain - Tables de Composition
Toury, J. et al.
Organisation de Coordination et de Cooperation pour le lutte contre les grandes Endemies
ORANA, Dakar
1965

West Africa
Plantes Alimentaires de l'oust Africain - Etude Chimique et Biologique des Vegetaux Alimentaires
Busson, F.F.
Marseilles
1965

NORTH AMERICA (Excluding USA)

Canada
Nutrient Value of Some Common Foods
Department of National Health and Welfare
1980

Caribbean
Food Composition Tables for Use in the English-Speaking Caribbean
Caribbean Food and Nutrition Institute
Kongston 7, Jamaica
1974

Mexico
Composicion de Alimentos Mexicanos
Cravioto, Rene O. et al.
Instituto Nacional de Nutriologia
1955

Mexico
Valor Nutritivo de Los Alimentos Mexicanos: Tablas de Uso Practico
Hernandez, M., A. Chavez, and H. Bourges
La Division de Nutricion
Institute Nacional de la Nutricion
1980

SOUTH AMERICA

Brazil
Tabelas de Composicao dos Alimentos
Sizaret, Francois and Claude Jardin
Estudo Nacional de Despesa Familiar
Secretaria de Planejamento da Presidencia de Republica
Fundacao Instituto Brasileiro de Geografia e Estatistica
1977

Chile
Tabla de Composicion Quimica de Alimentos Chilenos, 6th ed.
Schmidt-Hebbel, H.
Universidad de Chile
Gentileza de Chiprodal
1979

Colombia
Tabla de Composicion de Alimentos Colombianos, 4th ed.
Instituto Colombiano de Bienestar Familiar
Ministerio de Salud Publica
1978

Latin America, General
INCAP-ICNND Food Composition Table for Use in Latin America
Leung, Woot-Tseun Wu
INCAP and NIH
1961

Peru
La Composicion de Los Alimentos Peruanos, Quinta Ed.
Collazos Ch, Carlos, et al.
Ministerio de Salud, Institutos Nacionales de Salud, Instituto de Nutricion
1975

Venezuela
Tabla de Composicion de Alimentos para Uso Practico
Division de Investigaciones en Alimentos
Instituto Nacional de Nutricion
Publicacion No. 42, Serie Cuadernos Azules
1983

ASIA

Bangladesh
Analysis of some Food Stuffs of Bangladesh, Table I
Ahmad, K. and M.A. Malek
undated

Burma
The Nutritive Value of Burmese Foods
The Nutrition Project
Directorate of Health Services
1967

China, People's Republic
Food Composition Table of Peoples Republic of China
1982

East Asia, General
Food Composition Table for Use in East Asia
Leung, Woot-tsuen Wu, et al.
FAO and US Dept. Health, Education, and Welfare
1972

India
Nutritive Value of Indian Foods
Gopalan, C., B.V. Rama Sastri, and S.C. Balasubramian
National Institute of Nutrition
Indian Council of Medical Research, Hyderbad
1984

Indonesia
Food Composition Table
Nutrition Directorate
Health Department, Republic of Indonesia
Penerbit Bhratara Karya Aksara, Jakarta
1981

Indonesia
Observations of Food in Various Districts in Indonesia That Have Not Been Included in Food Composition Tables.
Dewi Sabita Slamet, Merry Komariah, Komari Pusat Penelitian, and Pengembangan Gizi
Komplex Gizi Jalan Semboja, Bogor
1980

Japan
Standard Tables of Food Composition in Japan, 4th ed.
Resources Council, Science and Technology Agency
1982

Korea
Food Composition Table, 2nd ed.
Rural Nutrition Institute
Office of Rural Development
1981

Malaysia
Nutrient Compositioni of Malaysian Foods - A Preliminary Tab
Song, Tee E.
Division of Human Nutrition
Institute for Medical Research
1982

Pakistan
Nutritive Value of Foodstuffs and Planning of Satisfactory Diets in Pakistan: Part I, Composition of Raw Foodstuffs.
Chughtai, M.I.D. and A. Waheed Khan
Institute of Chemistry
Punjab University, Lahore
1960

Philippines
Food Composition Tables: Recommended for Use in the Philippines, Handbook I (5th edition)
Abdon, Isabel C. and Isabel F. del Rosario
Food and Nutrition Research Institute (FNEI)
National Science Development Board (NSDB)
1980

Soviet Union
Marine Fishes: Chemical Composition and Processing Properties, Russian Translation Series #7
ed. V.P. Bykon
A.A. Balkema/Rotterdam

Sri Lanka
Tables of Food Composition for Use in Sri Lanka
Perera, W.D.A., Padma Jayasekera, and Sithy Thaha
World Health Foundation of Sri Lanka
1979

Taiwan
Composition of Foods Used in Taiwan
Ta-Cheng Tung, Po-Chao Huang, and Hong-Chi Li
J. Formosan Medical Association, Nov. 1961
vol 60, no 11, pp 973-1005
1961

Taiwan
Table of Amino Acid Composition of Taiwan Foods, revised
Food Industry Research and Development Institute
Hsinghu, Taiwan
1973

Taiwan
Table of Taiwan Food Composition, revised
Food Industry Research and Development Institute
Hsinghu, Taiwan
1974

Thailand
Amino Acid Content of Thai Foods
Division of Nutrition
Department of Health
Ministry of Public Health
1983

Thailand
Table for Thai Food Composition of 100 gm Edible Portion
Division of Nutrition
Department of Health
1978

Thailand
Table of Thai Food Composition of 100 gm Edible Portion
Division of Red Cross Youth
Department of Physical Education
undated

Thailand
Thai Food Composition Table
Division of Nutrition
Delartment of Health
Ministry of Health
1981

EUROPE

Denmark
Fodevaretabeller
Helms, P.
Akademisk Forlag, Kobenhavn
1973

Denmark
Fodevare-of Ernaeringstabeller
Ege-Nyt, R.
Nordish Forlag, Arnold Busck, Copenhagen
1969

Denmark
Naeringsstoftabeller
Helms, P.
Laegeforeningens Forlag, Kobenhavn
1978

Finland
Kivennaisainetaulukko
Varo, P.
Otava, Helsinki
1981

Finland
Mineral Element Composition of Finnish Foods: N,K,Ca,Mg, P,S,Fe,Cu,Mn,Zn,Mo,Co,Ni,Cr,F,Se,Si,Rb,Al,B,Br,Hg,As, Cd,Pb,and Ash
Pekka Koivistoinen ed.
Supplementum 22, Acta Agriculturae Scandinavica
1980

Finland
Ruoka-ainetaulukko
Turpeinen, O.
Otava, Helsinki
1980

Finland
Ruokalajien Ravintoainesisalto
Ahlstrom, A. and L. Rasanen
Kotitalous 38: 4-9
1974

France
Tables de Composition des Aliments, 3rd ed.
Randoin, L. et al.
Institut Scientifique d'Hygiene Alimentaire
Centre National de Recherche Scientifique
1959

France
Valeur Nutritionelle de Quelques Produits Prets a Etre Consommes
Dosages du Centre de Recherches Foch
1983

Germany
Food Composition and Nutrition Tables 1981/1982
Scherz, H. and G. Kloos
Bundesministerium fur Ernahrung, Landwirtschaft und Forsten
Wissenschaftliche Verlagsgesellschaft mbH
1981

Hungary
Food Composition Tables, 7th ed.
Tarjan, R. and K. Lindner
Medicinakonyvkiado, Budapest
1972

Italy
Tabelle di Composizione Degli Alimenti
Carnovale, E. and F.C. Miuccio
Istituto Nazionale della Nutrizione
Ministero dell'Agricoltura e delle Foreste
1981

Netherlands
Nederlandse Voedingsmiddelen Tabel
de Hartog, C.
Uitgave van het Voorlichtingsbureau vor de Voeding
1973

Netherlands
Nederlandse Voedings Middelen Tabel
Aanbevolen Hoeveelheden Energie en Voedingsstoffen
Uitgave Voorlichtingsbureau vor de Voeding
1981

Netherlands
UCV Tabel: Uitgebreide Voedingsmiddelen Tabel
Voorlichtingsbureau voor de Voeding
1984

Norway
Naerings Middel Tabell, 4th ed.
National Nutrition Council
National Society for Nutrition and Health
1977

Poland
Popularne Tabele Wartosci Odzywczych Zywnosci
Piekarska, J. and A. Szczygiel
PZWL, Warsaw
1979

Poland
Sklad i Wartosc Odzywcza Produktow Spozych
Piekarska, J. and M. Los-Kuczera
PZWL, Warsaw
1983

Portugal
Tabela de Composicao dos Alimentos Portugueses
Goncalves Ferreira
National Institute of Health
1977

Spain
Tablas de Composicion de Alimentos
Andujar Arias, M., O. Moreiras-Varela, and F.G. Extremera
Instituto de Nutricion, CSIC
1983

Sweden
Fodoamnestabeller
Abramson, E. and B-M.Andersson
Esselte
1982

Sweden
Livsmedelstabeller: Energi och Vissa Naeringsaemnen
Statens Livsmedelsverk
1981

United Kingdom
First Supplement to McCance and Widdowson's The Composition of Foods
Paul, A.A., D.A.T. Southgate, and J. Russell
Her Majesty's Stationery Office, London
1980

United Kingdom
McCance and Widdowson's The Composition of Foods, 4th ed.
Paul, A.A., and D.A.T. Southgate
Her Majesty's Stationery Office, London
1978

United Kingdom
Second Supplement to McCance and Widdowson's The Composition of Foods. Immigrant Foods.
Tan, S.P. R.W. Wenlock, and D.H. Buss
Her Majesty's Stationery Office, London
1985

United Kingdom
The Nutrient Composition of Some Cooked Dishes Eaten in Brittain: A Supplementary Food Composition Table.
Wiles, S.J. et al.
Vol 34, No.3, pp 189-224
J. Human Nutrition
1980

Yugoslavia
Tablice: O Sastavu I Prehrambenoj Vrijednosti
Pica, Namirnica I.
Zavod za Zastitu Zdravlja SR Hrtske
Rockefellerova 7, Zagreb
1976

MIDDLE EAST

Middle East, General
Food Composition Tables for the Near East
FAO and USDA
FAO, Rome
1982

Middle East, General
Food Composition Tables for Use in the Middle East
Pellet, P.L. and S. Shadarevian
American University of Beirut
1970

Israel
Characteristics of Israeli Citrus Juice
Cohen, E.
Institute of Quality Control and Extension Services for the Food Industry
1982

Israel
Food Composition Table
1980

Turkey
Gida Komposizyon Cetvelleri
Koksal, O. and A. Baysal
School of Public Health, Ankara
1966

OCEANIA

Australia
Simplified Food Composition Tables
Corden M. and S. Thomas
Australian Government Publishing Service, Canberra
1971

Australia
Tables of Composition of Australian Foods
Thomas, S. and M. Corden
Australian Government Publishing Service, Canberra
1970

Australia
The Nutritional Composition of Australian Aboriginal Bushfoods
Brand, J.C. et al.
Vol 35, No.6, pp 293-298
Food Technology in Australia
1983

Hawaii
Composition of Hawaii Fruits
Wenkam, Nao S. and Carey D. Miller
Bulletin No. 135, Agricultural Experiment Station
University of Hawaii
1965

Hawaii
Nutritive Value of Some Hawaiian Foods in Household Units and Common Measures
Miller, Carey D. and B. Branthoover
Circular 52, Agricultural Experiment Station
University of Hawaii
1957

Hawaii
Foods of Hawaii and the Pacific Basin - Vegetables and Vegetable Products: Raw, Processed and Prepared - Vol. 1
Wenkam, Nao S.
Institute of Tropical Agriculture and Human Resources
University of Hawaii
1983

New Zealand
Composition of New Zealand Foods: Fruits and Vegetables
Visser, F.R. and J.K. Burrows
DSIR Bulletin 235, Dept. Scientific and Industrial Research
1983

Pacific Islands
Food Composition Tables for Use in the Pacific Islands
South Pacific Commission, Fiji School of Medicine
Noumea, New Caledonia
1983

W. Rand
INFOODS
Rm 20A-226
MIT
Cambridge, MA 02139

INFOODS UPDATE -- ONE YEAR OF ACTIVITY

I. Organization of INFOODS

About two and a half years ago, a very diverse, international group gathered to discuss the status and problems of food composition data and to see what should and could be done. This meeting followed a number of preliminary meetings and consultations with people closely involved with food composition data -- with their generation, collection, compilation, and usage. Present at the meeting were representatives of various agencies of the US government, of FAO, WHO, and UNU from the United Nations system, and individuals from IUNS and IUFoST.

It was the consensus of the meeting that the state of food composition data left much to be desired, that the situation was getting steadily worse, and that it was imperative that efforts be started to improve the situation. Moreover, it was felt that these efforts had to be on an international level, and that they needed to address the underlying problems in a unified manner rather than just documenting the status and dealing with the symptoms one by one.

The meeting called for the organization of the International Network of Food Data Systems, INFOODS, as a global collaboration of people and organizations interested both in food composition data themselves, and in working to improve food composition data. The overall mandate for INFOODS was that it work toward improving the amount, quality, and availability of food composition data. The next year and a half was spent developing a plan for how to do this, and the successful search for funding. Our funding currently comes mainly from the US government, with the National Cancer Institute taking the role of lead agency with additional support from USDA, FDA, and NHLBI, and

also funds from private foundations and industry. Additionally, the United Nations University provides support and serves as the lead agency within the United Nations System. Work began in earnest in July of 1984 with the establishment of an INFOODS secretariat at MIT.

II. Problems and Needs

Before detailing just what INFOODS is and hopes to become, I would like to sketch some of the specific problems that called it into being. The major problem most obvious to a potential user of food composition data is not being able to find data: not finding the level of selenium in a potato, the amount of protein in a pot roast, or anything much about chicken curry. If you look in the most obvious place, the USDA Handbook No. 8, you will not find these data.

In situations like this, there are three general options:

First, you can analyze the specific foods of interest for the desired nutrient(s), actually preparing the curry and analyzing it.

Second, you can look in the literature or in other tables or databases for the data, e.g., the German tables, Souce, Fachman, and Kraut, does have the selenium content of potatoes, and the second supplement to McCance and Widdowson has chicken curry.

Your third option is that you can use what data you have to make an intelligent estimation of the value (you can impute the data), e.g., Handbook No. 8 has lots of different cuts of beef, you need only choose one, and figure out what cooking does to the protein (on a per 100g basis) and you have what you want for a pot roast entry.

Each of these options illustrates different aspects of some of the general difficulties inherent in the current status of food composition data.

(1) If you choose the straightforward option of direct analysis you must decide

-- how to choose and prepare the food - what, precisely, is chicken curry?

-- how to conduct the analytic procedure - how do you analyze for selenium?

-- how to summarize the results - is the average of all the samples analyzed sufficient?, should it be weighed?

-- how to make the results available to others - do you leave them buried in your lab books?

(2) If you choose to look for the data elsewhere you face the problems of

-- where to look for the tables,

-- how to identify the food - is their curry yours? or their potato?

-- how to interpret the values - some tables give only a range,

-- how to judge the values - did they do the work carefully?

(3) If you choose the third, to 'impute' the data, you face the decisions of

-- how to actually do it,

-- how to assess how well it was done,

-- how to tell someone what you've done.

While each user knows how to do some of the above, very few can do all, and many would do things differently. What is obvious is that the lack of readily available, high-quality data is really symptomatic of a number of underlying problems:

-- too many foods,

-- foods are inherently quite variable,

-- too many nutrients and biologically active components of foods,

-- too few easy, accurate, inexpensive analytic methods,

-- too little documentation of what has been done,

-- too few widely accepted standard and guidelines for how food composition data should be gathered, saved, or even used, and

-- too little communication between workers in the field.

III. The tasks of INFOODS

It was to find ways to deal with these problems that INFOODS was organized. It was felt that the best approach was to focus on linking existing food composition data and in so doing set up an environment for their general improvement. Thus the fundamental intention of INFOODS is to create a loose, international collaboration of the generators, compilers and users of food composition data, and to develop channels of communication and guidelines for operation. Structurally, INFOODS is coordinated by a small secretariat which works with regional groups which, in turn, interact with individuals and groups within their own regions and with other regional groups (Figure 2). To accomplish this, INFOODS is in the process of setting up two networks:

-- First is a network of people interested in food composition data, linking them together and drawing upon their expertise. The goal here is to develop a sense of unity in the field and to increase awareness of the importance and limitations of the data, and to keep everyone informed of what is being done.

-- Second is a data network, a linkage of food composition data around the world, set up so that anyone can find just what data exist, where they are, and get them, and know just what it is they have gotten.

Considerable machinery is needed to make and keep these networks operational; thus INFOODS has initiated a number of activities (see Figure 2).

-- Development of a standardized terminology and nomenclature. Stewart Truswell, a nutritionist at the University of Sydney, Australia, is leading a group which is examining the problems of international nomenclature and classification. Several meetings have been held, leading to a preliminary design of a terminology which will permit international exchange of food composition data. This system is conceptually similar to the Factored Food Vocabulary of the FDA and builds to a certain extent on the work and experiences of the International Network of Feed Information Centers (INFIC). We expect to have detailed prospectus of this ready for review in September; we are planning an international coding meeting early next year to "try it out," and we hope to have a solid, preliminary version of the whole thing to present to you by this time next year.

-- Development of standards and guidelines for how food composition data should be collected. Dr. David Southgate, a British chemist who produced a brief manual of food data collection in 1974, has undertaken the complete revision and extensive expansion of this document. A first draft has been formally reviewed. The IUNS meeting in Brighton next month will feature a workshop on a second draft which will also be widely distributed for review. A final draft is expected by the end of this year, and I hope to present it to this group in detail at next year's meeting.

-- Exploration of how modern information systems ideas and technology can be involved with the whole field of food composition data. The working group in this area is being set up and chaired by Dr. John Klensin, a computer scientist at MIT. Initial tasks of this group include development of standards for data interchange and design of prototypical regional centers. Additionally, this group will develop computational routines that embody the recommendations about terminology, data quality, and interchange, and facilitate electronic communication among workers in the field. This activity

is central to the whole concept of a network of food composition data. As part of this, this group expects to be able to recommend and provide technical assistance for end-user systems development. By next year, we hope to present you with firm versions of interchange protocols and specifications for regional centers and programs.

-- The production of an international directory of existing data bases. FAO produced, in 1975, an international inventory of food composition tables. Since they currently have no plans for updating this valuable document, INFOODS is preparing an expanded and updated version; a preliminary version is available and being circulated. This will complement the work of Loretta Hoover, who, in the context of this conference, produces an annual inventory of the data that exist within the United States.

-- Regional INFOODS liaison groups. Within the industrialized world there already exists some, although not good, communication between people involved with food composition data. Outside those parts of the world communication between such individuals is very limited. INFOODS is working to link, foster, and even organize where necessary, regional groups which are involved with food composition data. These groups aid INFOODS in determining regional needs and resources; they coordinate regional activities such as workshops and seminars, and in some cases with with INFOODS to establish and maintain regional computer centers for food composition data. We currently have strong links to groups in Scandinavia (NORFOODS, centered in Uppsala, Sweden), Europe (EUROFOODS, centered in Wageningen in the Netherlands), and in the countries bordering the Mediterranean (MEDIFOODS, based in Italy). We are organizing groups in Asia (ASIAFOODS, based in Bangkok), Latin America (LATINFOODS at INCAP in Guatemala), and the US and Canada (NOAFOODS). Other regional groups are in the planning stage; by next year we hope to have a good start on an AFRICAFOODS and an OCEANICFOODS for the south Pacific, and to have

increased the involvement of the countries of eastern Europe in the INFOODS network. Our ultimate goal is that everyone should have easy access to at least one regional group.

-- Development of a detailed description of users' needs. Last March, a users and needs meeting was held in Logan, Utah, gathering together individuals involved with the generation, compilation, and especially usage of food composition data. This meeting, organized for INFOODS by Carol Windham and Guarth Hansen of Utah State University, reviewed the initial plans of INFOODS and went on to formulate operational specifications of what was needed in terms of food composition data, and how INFOODS could best accommodate these needs and desires. We expect publication, within the next year, of the proceedings of this meeting.

Two important tasks for INFOODS were identified by this group. These were production of (1) guidelines on how best, temporarily, to fill the gaps in food composition tables and (2) guidelines on how to statistically process and present the data. These tasks are now being organized by the secretariat, and we expect to report preliminary results within the next year.

-- Establishment of a Secretariat. In order to keep all this moving along in a consistent, compatible, and timely fashion, a small secretariat was established at MIT. This consists of myself, John Klensin, and Vernon Young, a nutritional biochemist. In addition to dealing with the specific tasks already mentioned, the secretariat is involved with organizing and coordinating the taskforces and meetings and serving as a general, international clearinghouse and resource. To let people know what we are doing, a newsletter is issued quarterly, being sent to those who have shown interest. We intend to continue these activities during the next year; perhaps the most exciting plan is the starting of an INFOODS journal of food composition. This will be published by Academic Press and initially subsidized by the United Nations University.

IV. Summary

In summary food composition data are of fundamental importance, especially from a global perspective, the food composition data now available leave much to be desired. INFOODS was organized as an international collaboration to improve the situation by linking those working in the field, and identifying, organizing, initiating, and encouraging work in a number of specific areas. After the first year, we feel that we are well begun; we have started to formulate the problems and outline potential solutions. Our next step is the review and revision of what we are coming up with, and then we must start filling in the details. At this stage it is especially critical that we get input from people involved with all aspects of food composition data. INFOODS is a collaborative, and I would like to close this talk by issuing a general call for volunteers, for people who are seriously interested in critically reviewing our effort in all areas.

EUROFOODS UPDATE

Lenore Arab
Klinisches Institut für Herzinfarktforschung
Heidelberg, FRG

Significant advances towards the goals of EUROFOODS have been made in its first year of existence. This presentation is a brief glimpse of the state of affairs including recent progress and current plans, and a sneak preview of an upcoming EUROFOODS meeting in Norwich, United Kingdom in August 1985 hosted by Dr. D.A.T. Southgate and the Food Research Institute.

WHAT IS EUROFOODS?

EUROFOODS is an organization of individuals and institutes throughout Europe involved in dietary assessment, analysis, distribution, or use of food composition information. A common goal of all involved is to improve the compatibility of European food composition tables and of studies relating diet to disease. Seventeen countries are actively represented currently, including the Scandinavian countries, the Mediterranean, Central and Eastern Europe.

The executive committee consists of five individuals, Dr. Ostrowski (France), Dr. West (Netherlands), Ms. Bergström (Sweden), Dr. Southgate (United Kingdom) and the author (FRG). Currently five working groups exist, addressing the following topics: Nutrient losses and gains, and missing values, analytic laboratory exchanges, inventorization of data base producers and users, tourist food tables and computer issues. The activities of each will be further detailed below.

THE WORKING GROUPS

The committee on nutrient losses and gains has formulated a proposal to collect and computerize information on nutrient changes during preparation in a data base which can be referenced widely to obtain information on changes by cooking method, food and food group or nutrient. Additionally, collection of formulas for calculation and comparisons of calulated changes as compared with measurements in different countries are being undertaken.

Missing nutrient information from various European tables is being collected and categorized for 100 commonly eaten foods to determine the availability of and type of values in the published tables for various nutrients. The classification scheme of Greenfield and Southgate (1) is being adopted with separate definitions of analysed, calculated, imputed, or borrowed values or combination of values, represented as missing or simply absent in the tables. This exercise, which reveals the extent to which certain tables borrow the majority of their values (and often from other countries) is reported on in the proceedings (2).

The committee on analytic laboratory comparisons has organized an interlaboratory trial in which 6 foods were distributed to 20 recognized national food laboratories. Protein, fat, available carbohydrates, dietary fiber and ash were measured by each lab's own routine method; moisture by the prescribed standard method. These results, indicating large interlaboratory differences, particularly for fat measurements, will be presented in Norwich, and also reported on in the proceedings (2).

Inventorization of the data bases existing in 15 countries was begun in 1982 and is presented in the report "EUROFOODS: Towards compatibility of Nutrient Data Banks in Europe" (3). At that time 32 groups reported having systems with 21-3000 food entries and 4-120 food components. This

effort is being updated, and a guide to the available food tables, their publishers and prices; as well as the existing (prominent) nutrient data bases in Europe is being composed.

The tables for tourists is forseen to include useful information on nutrients of known health importance in general nutritional guidelines as well as those of critical importance to specific risk groups (hypertonics , hyperlipidemics, lactose intolerant individuals and diabetics). Progress on this front depends on further development in the area of translation, recipe exchange and harmonization of the nutrient data bases.

The computer committee has received financial support from the European Community to develop a common coding system, identifying the dissimilarities between tables and test the feasibility of developing a common European nutrient data base, including the maintenance of such a system, and its distribution. Two workshops have been held on these topics; in Luxembourg and in Heidelberg, with representatives of 17 countries participating. The German Federal Health Office (Bundesgesundheitsamt) has volunteered to assist in the development of the common coding system, through the committee on dietary assessment.

To date, translations of the native language entries of 15 European tables have been computerized in Heidelberg. They are presented in Table 1. This has been difficult and does not solve the problem of foreign table use. Seldom as they may be consulted, introductions are also important keys to responsible application of food tables. Translations of these for 15 major European tables have been compiled (see Table 1) and will be distributed through EUROFOODS. Common recipes are also being collected and stored from many countries. Comparisons of the format of 16 different tables have been undertaken, and the units of measure, conversion factors, and modes of expression have been collated. A sample of the printout for vitamins and their units/100 g as found in a booklet constructed for comparative use can be found in Table 2. The complete comparison of table contents can be acquired from EUROFOODS.

The merger of the complete tables including names, translations and nutrient information from Holland, the United Kingdom, the Federal Republic of Germany, Finland and Sweden is underway.

EUROCODE

First steps in the development of a widely acceptable EUROCODE have been made this year through development, circulation and discarding of successive drafts. Agreement on the main food groups was achieved at a EUROFOOD coding workshop in Heidelberg in February, and the result can be seen in Table 3. It has already been proven useful in comparisons of the distribution of foods in 10 tables.

The current design for a coding and descriptor system as proposed by the commission on dietary assessment of the German Federal Health Office involves 2 independent codes and standard variables for supplemental description of foods as seen in Diagram 1. Code one will be an unstructured sequential numbering of all foods available in Europe. Code two is a semi-hierarchical, semi-informative code involving 3 to 4 fields of information for use in cross referencing food groups between countries and bypassing the translational problems. It is hoped that this will be standardly used in future national food tables. Code three will be an agreed upon set of information necessary to completely describe food or dishes for nutritional use - similar in concept to the factored food vocabularies of the Food and Drug Administration.

The greatest amount of effort has gone until now into development of code 2 under the following premises listed in Table 4. It is meant primarily for coding by basic food group and cross referencing tables. A sample printout of food table entries for the group apple as identified by this code can be found in Table 5.

Early on it became apparent that although there was tremendous enthusiasm and willingness to cooperate within Europe on some standardized coding system, many different expectations were held by different groups - some wanted a system for collection and evaluation for surveys, others for computing, still others for identifying of foods. One simple code and data base could not serve all purposes well and led to the current conception of an "entire" nutrient system. The computer committee of EUROFOODS envisions various components allowing for multiple uses, in answering such questions as when one can consider two foods from different sources identical. The answer depends on the purpose of the user, but the necessary information should be available through the coding and descriptor system. The EUROFOOD concept of information and software program needed to conduct nutrient analyses and use foreign tables responsibly are presented in Diagram 2.

Not only information but programs for various purposes are also needed, as well as guidelines for use. Information files, including nutrient analyses related information (on sample, method, laboratory, recipe information related to mixed foods, descriptive information on handling, processing, part of plant or animal, and the like) are required for identification and comparison and language files with scientific and native names as well as local synonymes.

Conclusion

The author's interest in these efforts is in terms of the potential for better understanding the relationships between diet and disease through international epidemiology. There is still much to be learned about the risk differentials for diseases between countries which can better help us understand what we are doing right and what wrong in food offerings and food selection. And there are still many problems impeding studies of diet and nutrient intakes between countries. Many EUROFOOD projects are directed at improving this situation. I would like to conclude by reporting that in this effort a usually high level of willingness to collaborate within Europe has been demonstrated which has made the rapid and exciting results already achieved possible.

Literature

1. Greenfield H, Southgate DAT. Guidelines for the preparation of food composition tables (in press)

2. Proceedings of the Second EUROFOODS general meeting, August 25–28, 1985, Norwich, England (in press)

3. EUROFOODS: Towards Compatibility of Nutrient Data Banks in Europe. Ann Nutr Metab 29/S1/85, Karger, Basel 1985

TABLE 1.
MAJOR EUROPEAN FOOD TABLES :
TRANSLATION TO ENGLISH

Foods		Introduction	
*		DENMARK	Moller
*	*	FED.REP. GERMANY	Souci/Fachmann/Kraut
	*	FINLAND	Koivistoinen
	*		Turpeinen
*	*		Varo
*	*	FRANCE	Ostrowski/Josse
	*		Randoin et al.
*	*		Renaud et al.
*	*	GERMAN DEM. REP.	Haenel
*	*	GREECE	Trichopoulou
*	*	ITALY	Carnovale/Miuccio Fidanza/Versiglione
*	*	NORWAY	The National Nutrition Council
*	*	POLAND	Piekarska/Los-Kuczera
*	*	PORTUGAL	Gonsalves Ferreira/da Silva Graca
*		SPAIN	Arias/Moreiras-Varela/Extremera
*	*	SWEDEN	Statens Livsmedelsverk
*		THE NETHERLANDS	UCV-Kommissie
*	*	UNITED KINGDOM	Paul/Southgate

Table 2 PRESENTATION OF VITAMINS IN VARIOUS EUROPEAN FOOD TABLES

	Vitamin A	Retinol	Retinol-equiv.	Carotene	total Carotinoids	active carotene Provit. A	ß-carotene	Vitamin D	Vitamin E	α-tocopherol	α-tocopherol equiv.	Vitamin K
Austria	Foreign tables in use											
Belgium	ER*											
Denmark		ug	ug				ug	ug		mg	mg	KI ug
FRG	mg			mg	mg			ug	mg			
Finl. Varo												
Turp.			ug									
Koiv.												
France: Ostr.		mg						mg	mg			
Randoin	mg					mg		mg	mg			
Renaud	mg					mg		mg	mg			
GDR		mg					mg	mg	mg			
Greece												
Ireland	Foreign tables in use											
Italy: Carn.			ug									
Fidanza		ug	ug	ug								
Norway	IU											
Poland		ug	ug	ug								
Portugal	IU			ug								
Spain			ug					ug				
Sweden:												
1. Skolupplage		mg	mg			mg		ug	mg			
2. Large tab.		mg/kg	mg/kg			mg/kg		ug/kg	mg/kg	mg/kg		
The Netherl.		mg										
UK		ug		ug				ug	mg			

*1 ER = 1 ug Retinol = 6 ug ß-carotene

Table 3: EUROCODE MAIN FOOD GROUPS

1. Milk + products, mixed foods incl. dishes
2. Eggs + products, mixed foods incl. dishes
3. Meat + products, mixed foods incl. dishes
4. Poultry, game birds, game + products, mixed foods incl. dishes
5. Fish, molluscs, reptiles, crustaceans + products, mixed food incl. dishes incl. dishes
6. Oils and fats + products, mixed foods and dishes
7. Grain + products, mixed foods incl. dishes
8. Pulses, seeds, nuts, kernels + products, mixed food incl. dishes
9. Vegetables + products, mixed foods incl. dishes
10. Fruits, -products, -mixed foods incl. -dishes
11. Sugar and sugar products
12. Beverages (except milk)
13. Miscellaneous, soups, sauces
14. Food for special nutritional use

Table 4: PREMISES FOR EUROCODE 2

a practical code for intake assessment and table cross references

1. Incorporating mixed foods integrally with natural foods or unprocessed foods (because the intake form is regularly as a mixture and many hypothesis are directed at the food as consumed).
2. A system in which the foods are easy located. A system whose usage does not require practical food knowledge (so that no extensive training in use is needed).
3. A code in which no detail in food identifying information reported by the subject is disgarded or aggregated (even if nutrient compositoral data is not yet available).
4. A brief numeric code for each food (since alphanumeric systems show preference for a single language).
5. A system with only one code per food or dish.
   (The advantage for the coder is that there is no need of decision which one is which. An alphabetical index list will help with the problem of locating foods.)
6. A food group oriented system to enable international food comparisons.
7. A code open for new developments in the food field as well for new industrial products as for new botanical and zoological hybrides.

Table 5: NUTRIENT VALUES FOR THE SAME FOODS FROM 4 DIFFERENT TABLES

EUROCODE	Foodname	Land	Sodium Na mg/100 g	Potassium K mg/100 g	Magnesium Mg mg/100 g	Calcium Ca mg/100 g
10.303.0	Apples	Netherlands	2.0	150	.	10.0
10.303.0	Apple	Finland	0.8	140	6.0	7.0
10.303.0	Apple	Sweden	1.0	110	8.0	7.0
10.303.0	Apple	FRG	3.0	144	6.4	7.1
10.303.1	Apple unpeeled	Netherl.	3.0	279	.	10.0
10.303.1	Apple dehydrate water : 2.5 %	Sweden	7.0	730	22.0	40.0
10.303.2	Apples dried water : 24 %	Sweden	5.0	569	22.0	31.0
10.303.1	Apple sauce - can or glass -	Netherl.	9.0	130	.	5.0
10.303.1	Apple purree canned	FRG	2.7	114	9.8	4.4

Diagram 1: EUROCODE SYSTEM DESIGN

CODE ONE: 1, n + 1 — non-informative, multilingual list of European foods with a unique code number

CODE TWO: ___ . ___ . ___ . ___

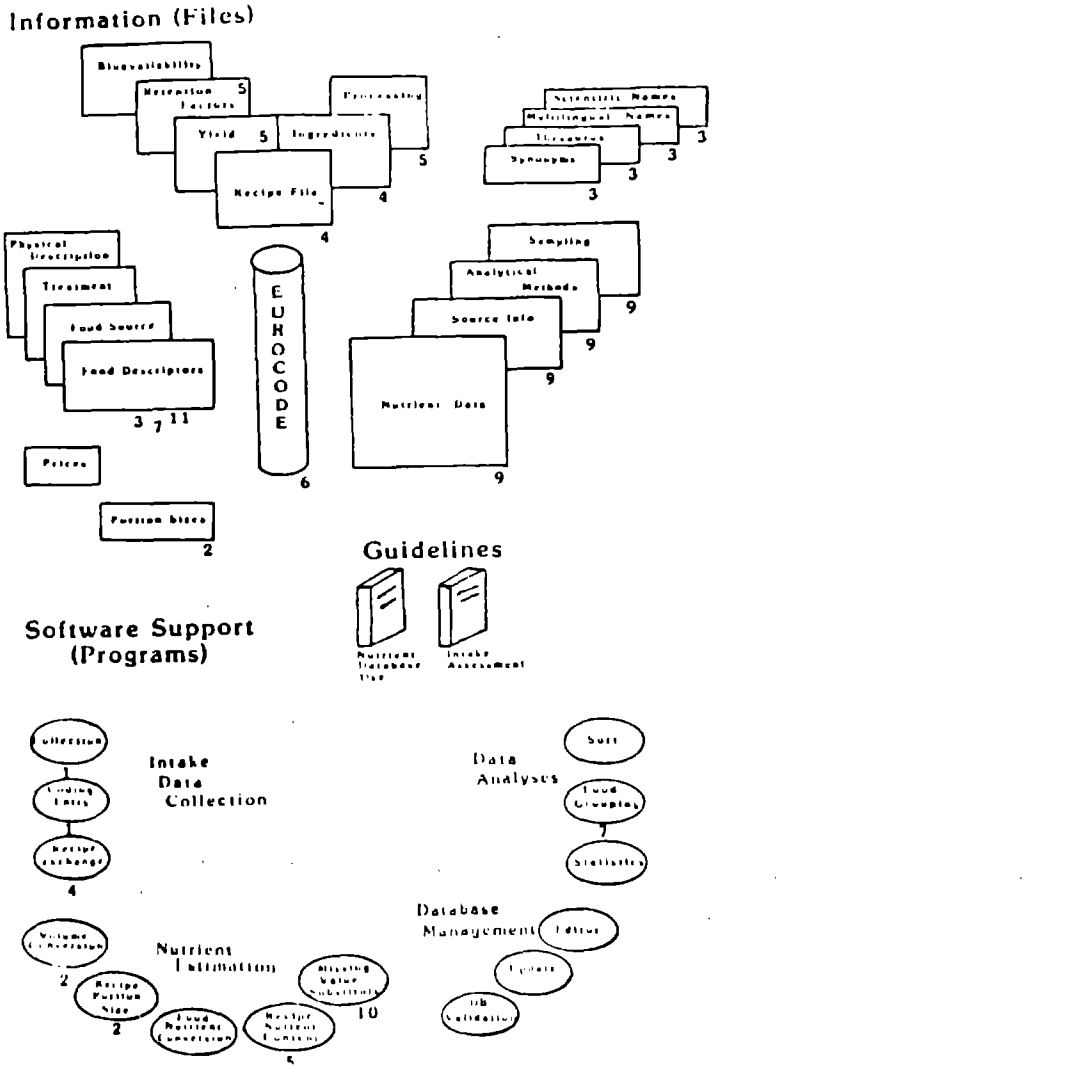
field Indiv.	1	2	3	4
	main group	sub group	national items	variant

CODE THREE: A123, B24, C1, D47 — Descriptor system , a set of variables considered necessary to describe a food for nutritional use

A - preparation
B - packaging
C - form

Diagram 2

Components of an extensive Dietary Assessment System



GUIDELINES FOR NUTRIENT DATA BANKS

Jack L. Smith, Ph.D.
STRATEGIC MARKETING ASSOCIATES, INC.
307 Fellowship Road
Suite 110
Mt. Laurel, New Jersey 08054

The Guidelines Committee developed a questionnaire to examine the needs and priorities of nutrient data base users. As reflected in the responses, guidelines for standardization of data bases would advance reliability and comparability between different data bases or studies. Of the 925 questionnaires mailed to nutrient data base users, 39 were undeliverable. Of the remaining 886, fifty-three (6.0%) responded.

For each item, number of respondents who marked that line were tabulated and additional suggestions, comments and answers to questions were summarized. Information on a variety of packaging (fresh, canned, frozen, dry) and preparation methods (fried, steamed, roasted, baked) appeared important to many respondents. For food items, measurement by household measure or by 100 gram edible portion were preferred by more respondents. Values per 100 grams were most preferred for nutrient measure. In addition to standard foods, those foods indicative of ethnic groups would be helpful to many data base users. Documentation of a data base was important, especially sources of information. Respondent comments emphasized that documentation aids in interpretation, validation, and updating of a data base.

The Guidelines Committee can now consider several options: 1) feasibility of a workshop to explore what standardization is possible and impacts of such, 2) changing the focus of the Committee to help new users, or 3) identifying projects which will increase the number of credible nutrient programs and the credible uses of the data.

QUANTIFIED RESPONSES

Total responding questionnaires = 53.
= number of questionnaires having that item marked.
% = percent of 53 responding questionnaires having that item marked.

#	%	Item
		I. FOODS - Check the items you want information for in the nutrient data base.
		A. DESCRIPTION
#	%	
22	41.5	1. Packaging
49	92.5	a. fresh
47	88.7	b. canned
48	90.6	c. frozen
48	90.6	d. dry
30	56.6	e. microwave pack
38	71.9	f. enrichment noted
15	28.3	2. Preparation
49	92.5	a. fried
47	88.7	b. steamed
47	88.7	c. roasted, baked
45	84.9	d. smoked, cured
43	81.1	e. pickled, packed in brine
9	17.0	3. Source of Consumption
34	64.2	a. home prepared
23	43.4	b. school lunch
31	58.5	c. restaurant
34	64.2	d. fast food restaurants
21	39.6	e. vending machines
7	13.2	4. Data from:
39	73.6	a. USDA, (#8, 456,72, etc.)
34	64.2	b. manufacturers
16	30.2	c. other data bases
28	52.8	d. composite of the above

#	%	
8	15.1	5. Measurement
37	69.8	a. 100 gram edible portion
48	90.6	b. household measures and gram weights
8	15.1	c. pounds
14	26.4	d. nutrient density/cubic inch
10	18.9	e. grams per ounce
21	39.6	f. actual gram weights for baked goods
11	20.8	6. Recipes - should be based on:
24	45.3	a. USDA samples values
37	69.8	b. edible portion
16	30.2	c. assay values
15	28.3	d. composite values
5	9.4	e. ranges of values
9	17.0	f. imputed values

Check Items in this section only if they would enhance your data base.

#	%	
3	5.7	B. STANDARD FOOD REFERENCE
13	24.5	1. Core foods section
3	5.7	1. Core foods should include:
14	26.4	a. all known food items or,
20	37.7	b. only foods most often consumed by most people
12	22.6	c. foods contributing the majority of macro and micronutrients
4	7.5	2. Modular Foods section - this section should include foods indicative of:
23	43.4	a. religious groups
36	67.9	b. ethnic groups
15	28.3	c. groups who report fad diets
30	56.6	d. regional variations

II. CODING OF FOODS -

I prefer the following system:

#	%	
18	34.0	A. library classification system
4	7.5	B. bar code system
19	35.8	C. universal codes
5	9.4	D. FDA's Factored Food Vocabulary

B. COMPONENTS OF A STANDARDIZED NUTRIENT DATA BASE

I would like the following items included:

#	%	
36	67.9	1. water
44	83.0	2. calories
16	30.2	3. ash
37	69.8	4. caffeine
41	77.4	5. total protein
33	62.3	6. essential amino acids
28	52.8	7. non-essential amino acids
44	83.0	8. total fat
43	81.1	9. total SFA
37	69.8	10. total USFA
42	79.2	11. total poly USFA
21	39.6	12. individual fatty acids
45	84.9	13. total carbohydrates
23	43.4	14. crude fiber
44	83.0	15. dietary fiber
34	64.2	16. alcohol
33	62.3	17. total sugars
29	54.7	18. starch
29	54.7	19. simple sugars
24	45.3	20. complex sugars
42	79.2	21. water soluble vitamins
42	79.2	22. fat soluble vitamins
41	77.4	23. B_6
40	75.5	24. B_{12}
26	49.1	25. biotin
18	34.0	26. choline
22	41.5	27. vitamin K
14	26.4	28. inositol
44	83.0	29. calcium
43	81.1	30. phosphorus
44	83.0	31. sodium
44	83.0	32. potassium
44	83.0	33. iron
37	69.8	34. magnesium
38	71.7	35. zinc
35	66.0	36. copper
21	39.6	37. chlorine
25	47.2	38. iodine
34	64.2	39. chromium
17	32.1	40. cobalt
19	35.8	41. manganese
16	30.2	42. molybdenum
17	32.1	43. sulfur
21	39.6	44. flourine

I would like guidelines for the following:

#	%	
25	47.2	1. number of digits in a food code
24	45.3	2. minimal number of codes necessary in a data base
10	18.9	3. code and description for:
31	58.5	a. edible plain food item, e.g. steamed fish
27	50.9	b. edible food item with fat, salt, etc.
29	54.7	c. major ingredients in mixed dishes, e.g., chili
17	32.1	d. generic foods
18	34.0	e. brand named foods
25	47.2	f. a combination of generic and brand name foods

III. NUTRIENTS - I would like information for the following:

#	%	
8	15.1	1. Sources
32	60.4	a. manufacturer
23	43.4	b. independent laboratory
29	54.7	c. USDA, FDA, or other agency
23	43.4	d. composite of data bases
20	37.7	e. imputed data
14	26.4	f. representation of multiple samples of food items
26	49.1	g. procedures for assignment of values
5	9.4	2. Units of Measure - values
42	79.2	a. per 100 grams
37	69.8	b. per household measure
8	15.1	c. per pound
31	58.5	d. per edible parts only
13	24.5	e. on products labels
7	13.2	3. Methods of Measure
35	66.0	a. laboratory analysis
22	41.5	b. imputed values
19	35.8	c. composite values
15	28.3	d. range of values
19	35.8	e. confidence intervals for each nutrient
16	30.2	f. confidence codes
4	7.5	4. Content - include:
21	39.6	a. values for all known nutrients and dietary components
34	64.2	b. only reliable values for nutrients and dietary components

3. The following means of documentation are important. Check all items applicable to your own data base.

#	%	
43	81.1	a. sources of information
26	49.1	b. methods of analysis
23	43.4	c. footnotes
35	66.0	d. meaning of symbols
30	56.6	e. % of missing values
29	54.7	f. % of blank values
27	50.9	g. % of imputed values
26	49.1	h. % of codes with zero values
18	34.0	i. % of values analyzed in the laboratory
33	62.3	j. definition of zero
33	62.3	k. definition of trace
18	34.0	l. quality control procedures
22	41.5	m. confidence levels or standard deviation of nutrient values
33	62.3	n. sufficient space to add additional codes or values
14	26.4	o. disclaimers
18	34.0	p. confidence codes

V. QUALITY CONTROL - I think guidelines are needed for evaluating the limitations of a data base in the categories of :

#	%	
33	62.3	a. foods
40	75.5	b. nutrients
40	75.5	c. updating the nutrient data base
22	41.5	d. 24 hour recall data resulting from use of a data base

DEFINITION OF A NUTRIENT DATA BASE

AN ORGANIZED COLLECTION OF NUTRITIVE AND NON-NUTRITIVE CONTENTS IN SPECIFIED AMOUNTS OF FOODS USED TO EVALUATE THE FOOD SUPPLY AND INTAKES OF THE GENERAL POPULATION AND SPECIFIC ETHNIC GROUPS. THE COMPILATION OF THESE FOODS AND ASSOCIATED NUTRIENT VALUES SHOULD BE UPDATED TO REFLECT CURRENT KNOWLEDGE OF THE NUTRIENT VALUES OF THE FOOD PRACTICES OF INDIVIDUALS AND GROUPS.

GENERAL DEFINITION OF STANDARDIZATION

-- WEBSTER, 1968

APPLIES TO SOMETHING ESTABLISHED FOR USE AS A RULE OR BASIS OF COMPARISON IN MEASURING OR JUDGING CAPACITY, QUANTITY, CONTENT, EXTENT, VALUE, QUALITY, ETC.

DEFINITION OF STANDARDIZATION AS IT APPLIES TO NUTRIENT DATA BASES

A SET OF GUIDELINES (TERMINOLOGY, RULES AND PROCEDURES) FOR COMPILING AND REPORTING DATA THAT PROVIDE UNIFORM AND CONSISTENT RESULTS.

COMPONENTS FOR STANDARDIZATION OF A NUTRIENT DATA BASE

1. Foods
   - Description
   - Measure
   - Recipes
   - Coding
   - Data Sources
2. Nutrients
   - Macro- And Micronutrients
   - Data Sources
   - Methods of Measure
   - Units Of Measure
   - Reliability Of Values
3. Documentation/Quality Control
   - Analytical Methods Of Food Analysis
   - Procedures For Imputation
   - Methods For Recipe Calculation
   - Procedures For Assigning Missing Values
   - Use Of Generic And Brand-Named Foods

REASONS FOR STANDARDIZATION OF NUTRIENT DATA BASES

- Produce A Common Interpretable Language For Communication
- Enable Users To Compare And Interface Data Base Systems
- Establish Reliability
- Enable Systematic Updating And Documentation
- Minimize And/Or Understand Variation In The Nutrient Sources Used In Food Intake Calculation
- Improve Accuracy And Reliability Of Information During Periods Of Nutrition Monitoring
- Improve Reproducibility
- Enable Comparisons From Different Studies
- Be Cost And Time Effective
- Ensure Uniform Procedures For Imputing Data
- Be Useful In Assessing The Validity Of Data Collection Methods

PERFORMANCE MEASURES OF STANDARDIZATION

- Compare Results Of Nutrient Analyses Among Laboratories Or With "Standard" Values
- Code And Evaluate Diet Recalls Using Different Data Bases
- Calculate Recipes

NUTRIENT DATA BASE COMPONENTS PRIORITIZED BY USER RESPONSES

Food Description

1. Packaging

	Response Rate
• Fresh	92.5%
• Frozen	90.6%
• Dry	90.6%
• Canned	88.7%
• Enrichment noted	71.7%
• Microwave pack	56.6%

Additional Suggestions

- Commercial or home pack
- Chilled
- Concentrate
- Irradiated
- Fermented

2. PREPARATION

	RESPONSE RATE
• FRIED	92.5%
• STEAMED	88.7%
• ROASTED, BAKED	88.7%
• SMOKED, CURED	84.9%
• PICKLED, PACKED IN BRINE	81.1%

ADDITIONAL SUGGESTIONS

- RAW
- BROILED
- MICROWAVE
- BOILED
- BRAISED
- STIR-FRIED

- THE MOST POPULAR METHODS USED FOR ALL FOODS,

  FOR EXAMPLE:

  EGGS - FRIED
  HARD-COOKED
  SOFT-COOKED
  POACHED
  ETC.

3. Source of Consumption

	Response Rate
• Home-Prepared	64.2%
• Fast-Food Restaurants	64.2%
• Restaurant	58.5%
• School Lunch	43.4%
• Vending Machines	39.6%

Additional Suggestions

- Institutional Food Service
- Cafeteria

4. Data From:

	Response Rate
• USDA Handbooks	73.6%
• Manufacturers	64.2%
• Composite of all Others	52.8%
• Other Data Bases	30.2%

Additional Suggestions

- Non-U.S. Data Bases, e.g. Asian, Swedish

5. Measurement

	Response Rate
• Household Measures and Gram Weights	90.6%
• 100 Gram Edible Portion	69.8%
• Actual Gram Weights for Baked Goods	39.6%
• Nutrient Density/Cubic Inch	26.4%
• Grams Per Ounce	18.9%
• Pounds	15.1%

6. Recipes - Should Be Based On:

	Response Rate
• Edible Portion	69.8%
• USDA Sample Values	45.3%
• Assay Values	30.2%
• Composite Values	28.3%
• Imputed Values	17.0%
• Ranges of Values	9.4%

Coding of Foods - System Preferred

	Response Rate
Universal Codes	35.8%
Library Classification System	34.0%
FDA's Factored Food Vocabulary	9.4%
Bar Code System	7.5%

NUTRIENT DATA BASE COMPONENTS PRIORITIZED BY USER RESPONSES

Standard Food Reference

1. Core Foods Section

	Response Rate
• Only Foods Most Often Consumed By Most People	37.7%
• All Known Food Items	26.4%
• Foods Contributing The Majority Of Nutrients	22.6%

2. Modular Foods Section

	Response Rate
• Ethnic Groups	67.9%
• Regional Variations	56.6%
• Religious Groups	43.4%
• Groups Who Report Fad Diets	28.3%

Additional Suggestions

- Formulas
- Nutrient Supplements
- Vitamin And Mineral Supplements
- Commercial Dietetic Foods

COMPUTER CONSIDERATIONS

PREPARED FOR THE

10th NUTRIENT DATA BANK CONFERENCE

by
Peter Zilahy Ingerman
Systems Consultant
40 Needlepoint Lane
Willingboro NJ 08046-1997
(609) 871-7474

This paper should be viewed as a check-list of concerns when considering the acquiring of a micro-computer, rather than a list of recommendations.

Computer Selection Considerations

There are three possible choices with respect to a new on-site computer system that is based on microcomputers:

- Stand-alone microcomputers
- Networked microcomputers with a file-server as the central unit
- Terminals or microcomputers connected to a multiprocessing central unit

Stand-alone microcomputers

A stand-alone microcomputer is, as its name implies, an entirely self-contained unit. It consists of a display, a keyboard, the computing unit itself, some internal memory (RAM and ROM), and some secondary storage (floppy disk and, possibly, a hard disk).

A single stand-alone microcomputer may be sufficient for your purposes. However, it is important to keep in mind that if there is only one computer available, and it goes down for any reason, you are out of business. It is therefore reasonable to consider either having more than one microcomputer, or having arranged for access to another microcomputer on an emergency basis, if it is essential that you be operational.

The major advantage to this approach is that, if care is taken in the purchase of these computers and their accessories, so that every capability is available on at least two computers, then each computer has a "secondary" computer to serve as back-up in case the "primary" computer fails for some reason.

The major disadvantage of this approach is that the only way of transporting data from one computer to another without reentering it is by physically transporting a floppy disk from one machine to another. Unless great care is taken to control this environment, it will be difficult to determine which copy of a particular file is the current one, and to be sure that all current files are backed up. (See below, however, under Telecommunications.)

An additional disadvantage is that the size of a reasonable stand-alone microcomputer is (comparatively speaking) limited, and hence it becomes difficult to handle large quantities of data: either the data is on a number of floppy disks, which must be continually interchanged, or the data is on a hard disk, which can become inaccessible if the hard disk or the microcomputer breaks down.

Networked microcomputers

This configuration requires a number of independent microcomputers, as described in the preceeding section, connected ("networked") to a "file-server" through a Local Area Network (LAN). The file-server is also a computer, but is not used as a terminal. Rather, it acts as the custodian for copies of programs and data that will be required for use on the individual computers; when something is required, a message is sent via the network, the required file is moved from the file-server to the requesting microcomputer, and work then proceeds. When the work is completed, the updated copy of the data (as appropriate) can be sent back to the file-server for storage.

The major advantage to this approach is that the problem of manual control of the current copy of data is eliminated, since there is only one current copy at any point in time. Similarly, there need be only one current copy of any program that is to be used in the network.

In addition, the terminals, since they are complete microcomputers, can continue to function even if the file-server breaks down; if careful backup has been performed from the file-server, the temporary loss of the file-server means a reduction in efficiency, not a total inability to work.

There are several disadvantages to this approach.

There is no single neworking standard, so that your organization may be at risk if the vendor selected is no longer available for support.

Certain proprietary software packages, because of the protection schemes used internally to them, may not operate in a networked environment.

There is a potential problem with multiple users updating the same data base. This can be solved by careful programming, together with a suitable operating system, but it is frequently not addressed, and can lead to badly garbled data.

The individual terminals, since they are microcomputers in their own right, are more expensive than the "dumb" terminals that can be used in the multiprocessing environment described next.

Multiprocessing

This approach presumes that the individual terminals have little or no computational ability and internal storage (that is, they are

explicitly terminals, rather than microcomputers), and are all connected to a single computer that performs all of the computations required for all of the tasks that are currently in process, as well as providing the capabilities of a file-server with respect to data and program files. This is, of course, the typical environment with a "large main frame"

The major advantage of this approach is that the incremental cost to add a work station is very much smaller ($700 to $1500 depending on the bells and whistles), compared with the cost of a full microcomputer ($1500 to $8000, depending on the bells and whistles).

The major disadvantage of this approach is that if the central computer goes down for any reason, the entire system goes down. Even if an entire back-up system is available, data that is stored internally to the central computer (e.g. on a hard disk) may not be available without considerable effort and time.

This approach also has the disadvantage that the central processing unit must be rather larger than that required for the file server (and hence more expensive). At the same time, it is uneconomical to get a central processer that is capable of providing all of the computational power of a stand-alone microcomputer for each work station. Hence, when the work stations are being heavily used, there is an increasing likelyhood that performance will be degraded. This will be particularly true if compute-intensive tasks (e.g. statistical work, reformatting while word-processing) are all being done at once.

Hybrid Systems

If your environment is one that includes a "large main frame" computer, most of the discussion above does not apply directly to you. However, both microcomputers and terminals can be attached to a main frame as terminals. If microcomputers are attached to your main frame, the situation is similar to that discussed under "Networked microcomputers", and all of the considerations discussed elsewhere in this paper with respect to microcomputer ancillary devices and peripherals apply.

There is an additional disadvantage that arises in this environment. If microcomputers are used as terminals, there will be a temptation to have "personal" copies of data and programs, that have been downloaded from the main frame, or developed individually on microcomputers. If this situation is not controlled with great care, it will be difficult to determine which copy of a particular file is the current one, and to be sure that all current files are backed up.

Telecommunications

Most microcomputers presently on the market can be adapted for communications on ordinary telephone lines, by using either a built-in or external modem (q.v.). It is probably not reasonable to purchase a computer that does not have this capability.

However, it is reasonable to consider the implications of having telecommunications capabilities, particularly if the microcomputer is left unattended, operating, and connected to the telephone lines. In this situation it is possible that unauthorized users may access the microcomputer. In the best case, nothing will happen. In the worst case, data and programs can be erased

Hence, in a telecommunicating environment, there are additional questions with respect to security, privacy, and integrity that must be addressed in a manner appropriate for your particular installation.

Software and Data Considerations

The concerns with respect to software (that is, the programs that will process data to produce the desired information) and data (the base from which the programs produce the desired information) are essentially the same:

- is the software, or data, purchased "as is", with no representation as to its usefulness or correctness, or is there some kind of guarantee?
- if there are problems (bugs in software, errors in data), will they be able to be fixed, and, if so, by whom?
- will some kind of service contract have to be purchased, to assure that updates will be received as available, or will each update have to be purchased separately, as a new product?
- will the author of the program or compiler of the data, whether a company or an individual, be available for consultation?
- if the author of the program or compiler of the data, whether a company or an individual, chooses to abandon support (or not to provide support) will the original source coding for the software, or the maintenance tools for the data, be available to present owners of the software or data?
- is the documentation for the software or data sufficiently complete with respect to the intended user?
- what is the cost of the software, or data? If that cost does not include maintenance and updates, what is the cost of maintenance and updates?

These same considerations apply whether the author of the software or the compiler of the data is an outside vendor or a part of the same organization as the user.

Peripheral Selection Considerations

Floating Point Hardware

Most of the computers that are presently available are inherently capable only of arithmetic on integers When real numbers (e.g. 1.25)

are used, it is necessary to provide a set of programs that perform the arithmetic on these numbers.

If your environment is one where significant numerical processing will be done, an additional device that should be considered is a "floating point processor". This is a piece of equipment that can plug into the computer, and that increases the speed of arithmetic on real numbers by approximately a factor of 10.

Modems

Modems are devices used when two computers communicate with each other over telephone lines. It is necessary to have a modem at each end of the communication link, and the two modems must be compatible with each other.

In addition to the modem itself (which is a piece of hardware), it will be necessary to have communciations software at each end of the communication link, and the software (as well as the hardware) must also be mutually compatible.

Printers

There are four fundamentally different types of printers on the market: dot matrix and fully-formed are the two types of contact printers, and ink-jet and laser are the two types of non-contact printers.

The dot-matrix printers are the most flexible of the contact printers. In the price range of $1000 to $2000 it is possible to get a dot matrix printer that can print in a draft mode (at upwards of 400 characters per second) and in a near-letter-quality ("NLQ") mode (at about one-fourth the draft speed) as well as having extensive graphics capabilities. In addition, it is possible to have multiple fonts, and to change fonts on a letter-by-letter basis. Dot-matrix printers that print in colors are available.

Fully-formed printers (true letter quality) print at about the same speed as dot-matrix printers in NLQ mode. However, their graphics capabilities are usually more limited, and, since font-changes involve changing a print element, it is not practical to make numerous font changes. Also, there is some software available that can print a wide report "sideways" on a dot matrix printer; this capability is simply not available on a fully-formed printer.

Non-contact printers are significantly faster and quieter than contact printers, since there is no requirement that there be the mechanical impact of an element (or a wire) against the paper

Ink-jet printers are reasonable when they are in use more often than not (so that the ink does not dry and clog); they are also particularly useful when printing on irregularly shaped objects (many brands of soda have a bottling date "sprayed" on the side of the cap). One major advantage of an ink-jet printer is that it can print in colors.

Laser printers are quite fast (8-12 pages per minute), have most of the flexiblity of the dot-matrix printers, and are expensive ($2500 and up).

Stand-alone Printer Buffers

A stand-alone printer buffer is a device that appears to be a printer to a computer, but appears to be a computer to a printer. It has its own internal memory (typically, from 32K characters to 256K characters), and can receive data from a computer as fast as the computer can generate it. The data is then stored in the internal memory, and is retransmitted to the printer as fast as the printer can accept it.

This device has the effect of freeing the computer from having to slow down to printer speed, and, therefore, permits the use of the computer while a long document is printing, rather than requiring the operator to wait

Floppy Disks

Floppy disks are the storage media used in floppy disk drives. They are considered to be "expendable" items, in the sense that they are thrown out when they are no longer functional.

It is a poor economy to use cheap floppy disks in an environment where reliability is important. If the only purpose for a particular floppy disk is to be written on once and then, shortly thereafter, to be read once (as when a floppy disk is used to send data or a program through the mails), then bulk-packed, off-brand disks (at a cost of about $1.00 each) may be sufficient. High-quality disks, on the other hand, can be read and written many times without deterioration, and are stable (when stored in reasonable environments) over long periods of time. (High-quality disks can cost $2.00-$4.00, depending on size of diskette and the number purchased at any one time.)

Floppy disks must be stored in an environment that does not exceed 125 degrees F, nor 80% relative humidity (see Media Safes).

Floppy Disk Drives

A floppy disk drive uses a single floppy disk as the recording medium. Floppy disks (and their associated drives) come in 8-inch, 5.25-inch, and a variety of 3-inch sizes. There are a number of different ways of organizing the data when it is recorded on a floppy disk, and the fact that the physical diskette is interchangable with a machine other than the one on which it was recorded is no assurance that the data will be readable on that machine. Many programs are available, however, that can be used to transcribe the data from one format to another.

The organization of the data on a floppy disk is determined by the operating system of the computer on which the disk was written, as well as the computer itself. Hence, if more than one operating system

is available on a given computer, it is possible that disks written by using one operating system will not be readable by the other operating system. (An "operating system" is a program that provides the working environment for other programs. CP/M, MS-DOS, and UNIX are examples of operating systems for microcomputers, and CMS and TOS are examples of operating systems for IBM main frames.)

The advantage of a floppy disk is that it is easily removed from the computer and stored in a safe place. (It can also be used to transport files from one computer to another, similar, computer.) Also, it is easy to make a copy of a floppy disk (assuming that a computer with two disk drives is available), and a disk can be completely duplicated in less than 120 seconds.

The disadvantage of a floppy disk is that it is very sensitive to the way in which it is handled, and can be easily damaged by being scratched or having its surface contaminated (e.g. with smoke particles). Also, if floppy disks are the only "bulk" storage in use on a system, there is a tendency for copies of programs and data to proliferate, with the consequent difficulties of maintaining control.

Hard Disk Drives

A hard disk drive is the logical equivalent of a number of floppy disks, packaged with a drive mechanism. The smaller drives use the "Winchester" technology, in which the drive and the media are sealed in a box, so that the only interfaces are electrical. Drives of this type are available in the 10 megabyte to 50 megabyte range, at a cost of about $100/megabyte.

A hard disk has the advantage of compact storage, without requiring the handling of individual volumes (e.g. floppy disks).

A major disadvantage, however, is that it is necessary to back up the data on a hard disk regularly, so that in case of disaster it will be possible to reconstruct the hard disk.

Streaming Tape Drives

A streaming tape drive is a special-purpose tape drive that can be used to back up a hard disk more rapidly than simply transcribing the contents of the hard disk to floppy disks. (Note that it takes on the order of 25 5.25" floppy disks to hold the complete contents of a 10 megabyte hard disk.)

Certain computers that have hard disks have built-in streaming tape drives. This is a consideration if the purchase of a computer with a hard drive is contemplated.

Reel-to-Reel Tape Drives

A reel-to-reel tape drive can also be used for backup of a hard disk.

This device has, however, the additional capability of being able to

read tapes that are written by other computers, and being able to write tapes that other computers can read. Such a capability might allow your organization to utilize data bases that are presently available only on tape. (It should be noted that certain data bases are available from the Government on both magnetic tape and floppy diskettes.)

(Data can, of course, be sent over telephone lines --- but this is a slow process at best, and can take hours for large data bases.)

There is clearly no necessity for the $10,000 investment that is required for a reel-to-reel drive at the initial stages of computerization. Its long-term desirablily, however, should be considered carefully at the initial stages, and, if there is any expectation that reel-to-reel tape capability will be required, only those computers that have the proven ability to have such a drive attached should be considered.

Color and Graphics Displays

Graphics displays permit the display of pictures (e.g. bar graphs, pie charts) as well as letters and numbers. Graphics displays are now common, and it would be unreasonable to select a computer that did not provide that capability.

Color displays, while more expensive than monochrome displays, can be significantly more effective in displaying data. There are camera attachments that permit 35mm slides to be made directly from the computer display

Plotters

Certain dot-matrix printers can also be used for graphic displays; if the dot-matrix printer is a color printer, the graphics can be printed in color. (There are programs that will "dump" the screen of an Apple Computer to a suitable color dot-matrix printer with a single key-stroke.)

Plotters, that are designed to put graphic displays on paper, usually in multiple colors, may be an alternative to color dot-matrix printers. They are usually best suited to the displaying of colored lines, rather than masses of color, and the inexpensive models do not have the flexibility of a color dot-matrix printer. On the other hand, they can be easier to use, since a line in any direction can be specified by giving its end-points and its color, and the plotter will do the arithmetic. (Note that some color printers also have this capability.)

Ancillary Considerations
--------- -------------

Furniture

It may be desirable to purchase specially designed tables for the

work stations. If ergonomic stands are not included in the terminals or microcomputers that are acquired for the work stations, then they can be purchased independently, as part of the tables.

The chairs that are to be used at the work stations should be designed for such use; this implies a somewhat greater degree of adjustability than may otherwise be required.

Stands for printers will also be convenient, as well as acoustic hoods if the printers are not isolated.

Static mats, to protect the computer and media from static electricity, are advised for each work station.

Diskette files, for the orderly storage of floppy disks at each work station, will be desirable.

Suitable lighting should be assured for each work station.

Cables

Cables will be required to interconnect pieces of equipment such as printers and microcomputers. Some of the interconnection cables may be provided as part of a "package" price, while others may have to be purchased separately.

Media Safes

An ordinary "fire safe" is not satisfactory for the storing of magnetic media. Magnetic media must not exceed a temperature of 125 degrees F, nor a relative humidity of 80%; exceeding these limits will do permanent damage to the disks, and will make them unreadable. A media safe is designed to maintain the internal environment below these limits in case of a fire.

Enough media safes will be required to store all of the required on-site back-up copies of magnetic media. It may also be reasonable to provide capacity for the storage of all on-site working disks for nights and weekends, although if backing up is done carefully and regularly, it may not be strictly necessary.

It may, however, be reasonable to have a media safe off-site so that a secondary back-up copy of all data will exist.

Media safes are available for floppy disks (list price approximately $1300) and for magnetic tapes (list price approximately $2000).

Supplies Storage

Adequate space must be provided for the storage of supplies (e.g. disks, paper) prior to their use. Paper, in particular, is heavy, and should be stored near the printers (or, alternatively, paper dollies are available, and should be acquired).

Power Supplies

The electrical power supply that is conventionally available is usually not adequate, for a computer system, in terms of reliability and "noise" on the line.

The first type of protection for computer components to be considered is a surge/spike protector. This device tends to eliminate sudden changes in voltage caused by lightning and by the starting and stopping of other equipment (typically, those containing motors). A top-of-the-line surge/spike protector costs on the order of $300.

Next is a line conditioner. A line conditioner tends to remove high-frequency noise (such as that caused by fluorescent lighting) from the line, and also tends to protect against "brown-outs" by producing a near-constant output voltage even when there are fairly wide variations in the input voltage. A reasonable rule of thumb for price for line conditioners is $200 plus $400 per KVA.

Finally comes an uninterruptible power supply (UPS). This device contains batteries that provide power to the computer for some number of minutes (8 to 15, typically) when the supply power fails entirely. This allows for the orderly shut-down of the computer system in the event of a power outage, without losing data. A reasonable rule of thumb for price for a UPS is $100 plus $1000 per KVA.

(A KVA, which stands for kilovolt-ampere, is approximately a kilowatt.)

(There are larger UPS available --- including some with diesel generators --- that can maintain power to the computers for 24 hours or more. They are probably not reasonable in the kinds of environments contemplated here.)

A power line monitor should be rented, and the power monitored for at least a week (longer, if possible) to determine the exact nature of any problems with the power as supplied. This, together with the kinds of usage that will be made of the computer system, will determine the exact kinds of power conditioning equipment that will be required

Personnel Considerations

Day-to-Day Usage

All personnel who will be using the computer system on a day-to-day basis should be trained to perform such routine maintenance as the cleaning of display screens on terminals and the changing of paper and ribbons on printers.

If this is an environment using stand-alone microcomputers, then, in addition, at least two persons should be trained as "key operators", in the sense of being familiar, e.g., with option switch settings, clearing paper jams, replacing fuses, and the like.

Programming

It is probably not cost-effective to have a staff programmer, at least for the short term. Rather, the standard report programs should be contracted out, and the users of the system should become sufficiently familiar with the software tools available to be able to prepare one-time reports; if it becomes desirable to make these reports "pretty", it will be reasonable to contract to have that done.

Ultimately, if the dollar cost of contract programming becomes high enough, it will become reasonable to hire a staff programmer.

Maintenance

One of the considerations in the selection of both hardware and software is the availability of maintenance.

While it is possible to get deep discounts on hardware, they are usually only available from firms that provide little or no maintenance support. Hence, for an organization that is interested in using computers, rather than repairing them, it is reasonable to shop for price --- but only in the context of the price including hardware maintenance.

Most dealers, and some independent organizations, will provide after-warranty service contracts. Unless your organization contemplates having an in-house person with the technical competence to perform hardware repair, a service contract will be essential. The type of service contracts vary from service on a bring-in basis to round-the-clock service on location, with a guaranteed response time. (Obviously, the latter is significantly more expensive than the former.)

If your organization has sufficient equipment so that every piece of equipment is duplicated, then there is, by definition, a degree of in-house back-up for all of the hardware, and it may not be strictly necessary to purchase the most extensive possible service contract.

With software, on the other hand, the maintenance typically involves the acquiring of upgraded versions of software that has already been purchased and installed. A single person at your organization should be designated as "software coordinator", and should be responsible for upgrading the in-house software as appropriate. Note that this implies careful control of all copies of each software product, so that all copies can be upgraded at once.

(On the other hand, whenever an update is obtained, it is important that it be checked out before being put into the general environment. Unfortunately, much of the available software is not carefully debugged before it is distributed, and there is no sensible reason to upgrade a piece of software that is working to one that has not been shaken down.)

FOOD COMPOSITION DATA BASE FEATURES

Jean A. T. Pennington, Ph.D., R.D.
Food and Drug Administration, Washington, D.C.

Introduction

As new or prospective users of food composition data bases, you may be (or will soon be) developing your own data base or adopting or modifying a previously developed data base. The large number of food composition data bases within government, academic, food industry, and private institutions emphasizes the necessity of designing data bases with unique features to serve specific needs. Because the development of a data base is a costly and time-consuming venture, it is important to identify at an early stage those features that will serve your specific needs so that you may incorporate them into your data base as you develop it, or you may select a data base that contains these needed features.

The features of a data base are usually a direct reflection of the functions served by the base. The most common use of food composition data bases is to assess the nutrient levels of diets. These diets may have been collected in food consumption surveys or they may be the diets of patients or clients who are consulting with dietitians or nutritionists. However, even the data bases developed for the purpose of evaluating diets vary considerably depending upon the population groups evaluated, the nutrients evaluated, and the purpose of the analysis.

In addition to assessing the nutrient content of diets, food composition data bases have numerous other uses. The Food and Drug Administration (FDA) uses food composition data bases to develop definitions for imitation foods and food substitutes; to develop standards of identity for specific foods; and to develop regulatory policies concerning nutrient fortification and use of food additives. The United States Department of Agriculture (USDA) uses food composition data bases to determine the adequacy of the United States (U.S.) food supply to meet nutritional needs. Food composition data bases are used to plan menus for hospitals, military feeding facilities, school lunch programs, and other group feeding programs. The food industry uses food composition data bases to develop nutritional labeling and labeling claims for their products.

To serve these various purposes, food composition data bases have characteristics that are specific for user needs. These characteristics include the types of foods; the number of foods; the food descriptions; classification or coding schemes; the types and number of nutrients; and the expression of the nutrient values.

Development of a Data Base

The characteristics of a data base generally reflect the effort and decisions made by the data compiler(s). Food composition data for data bases are gathered primarily from USDA publications, the scientific

literature, the food industry, and unpublished laboratory reports. At the first level of data collection, the compiler gathers the original reports or publications and transfers specific information to a file which may be written or computerized. This information includes the name of the food, its descriptors, and nutrient data in the original units. The compiler should also record the bibliographic reference, analytical methodology, quality control information (e.g., duplicates, blanks, standard reference materials), number of samples, method of sampling, and any other information (e.g., maturity, season, soil type, geographical location) that may be pertinent to the ultimate use of the data base. The best method for a hand-written version is to use one page per food per source. The nutrient composition data should then be converted to standard units (usually values per 100 grams wet weight of food and/or values per typical serving or other practical portion).

An evaluation of the compiled data may reveal that there are many or several sources for the nutrient content of some foods, but only partial data for other foods. There will probably also be conflicting data from various sources. The data compiler (a person knowledgeable about food composition, analytical methodology, and statistics) must carefully scrutinize the data from the various sources and make decisions regarding: combining partial data from various sources; aggregating data from various sources; and not using data that appear erroneous. If data are to be compiled from several sources, it is essential that the food descriptions and levels of some of the more basic nutrients (e.g., water, protein, calories) agree. The aggregation process may involve calculating averages or weighted averages for nutrient values. For practical purposes, the end result should be a separate single entry for each nutrient in each unique food product. The manner in which the values are derived must be documented, and the compositional data must continuously be reevaluated as newer data become available.

Foods to Include

The foods included in a food composition data base will reflect the functions of the base. A data base developed for a food company may only need compositional data for company and perhaps competitor products. A hospital data base should include the institutional foods, special recipes, and medical foods used by that institution. Similarly, data bases used by other institutions (military feeding facilities, school lunch programs, dormitories, prisons, boarding schools) should reflect the foods commonly served by these institutions. Data bases used for assessment of patient/client diets, epidemiological studies concerning diet and health, or other nutrition studies that include the general population should reflect the current food supply to include traditional, commercial, homemade, and restaurant foods.

To select the foods for a data base to be used in food consumption studies, it must be clear what dietary information will be elicited during the interview or by the food record instructions. If the subject is not asked the appropriate questions regarding the identification of each food, the subject will probably not provide the information. The code selected

for a product such as a cheeseburger could reflect a generic product as easily as a homemade or fast food brand name product. Likewise, low-fat milk may be a single code, or it may have several codes depending upon the level of fat (1/2%, 1%, or 2%) and the presence of nonfat dry milk solids.

The concern is the level of detail needed for your work. Do you need to know the brand name of the cheeseburger or the percent of fat in low-fat milk, or will the more generic terms suffice? It is possible that for some foods you will need more detail than others, depending upon the nutrients you wish to emphasize. You must also be realistic about the level of description you can expect from the participants in a food consumption survey. If your sample is large and the amount of time with each person is short, the amount of information to be elicited will be much less than if you have a small sample, repeated subject contact, and enough time to ask detailed questions.

If your data base only has specifically described food items, and the subject is unable to remember food details, the coder must guess at appropriate codes. Many such guesses might affect the validity of your results. The greater the number of food codes, the greater is the burden upon the coders and the greater is the chance of error.

If the purpose of your data base is to evaluate diets, you need only include foods as consumed. In this case you will not want to waste space with food items like raw meat, dry oatmeal, unpopped popcorn, etc. Your system should allow for easy addition and deletion of foods, as appropriate, to keep pace with the ever changing food supply.

Some thought should be given to including water as well as vitamin/mineral supplements, other supplements (protein powder, bone meal, bran, wheat germ oil, cod liver oil), and medications that have energy or nutrient values (e.g., cough syrup or antacids containing calcium carbonate) or that may interfere with nutrient absorption. Regional or local analyses for the nutrient content of water will probably be needed to assure the accuracy of the water data. The inclusion of water, supplements, and medications in a data base is optional and depends upon the purposes for which of the data base will be used.

Number of Foods

The number of foods in a food composition data base becomes a function of the detail of your food descriptors. The greater the level of detail per food, the more foods you will need to include. For example, it may be sufficient to have raw spinach and cooked spinach in your data base, or you might need more detail about the cooked spinach such as whether it was fresh cooked; canned; frozen, leaf; frozen, chopped; or frozen, creamed. Virtually every simple food item in a data base can be expanded in this way. The expansion is even greater if you choose to include special dietary products (e.g., foods with low or reduced sodium, fat, or cholesterol) or brand names of products.

Brand names for some products are important for product identification. Subjects undergoing a dietary interview might be more likely to report brand names for some candy bars or ready-to-eat cereals than to identify them as "chocolate covered nougat and caramel" or "O-shaped oat rings." Brand names may also be important for frozen entrees and fast foods because of differences and proportions of ingredients and hence nutrient values from one brand to another. However, a data base for a food consumption study will probably not need to include different brands of foods with similar nutrient values (e.g., corn flakes, chocolate cake from mix, mashed potatoes from instant, macaroni and cheese from box mix, frozen orange juice). A data base for a regulatory agency or for a food industry may, however, need such brand name data, especially if there is concern with label compliance, nutrient fortification, and/or label claims.

It may be that nutrient data for a certain brand named product (e.g., Brand Y macaroni and cheese) is used for a group of similar products (e.g. all macaroni and cheese mixes). This fact should be documented in your data base files at the first level of data collection. The brand name should not, however, be associated with the food description record in the coder's manual as this will cause confusion for the coders.

The number of foods in currently available data bases vary considerably. The data base used by the Neonatal Intensive Care Unit at the Milwaukee County Medical Complex contains 57 foods, all of which are infant formulas; the FDA's Total Diet Study has 234 foods; and the Minilist of the University of California, Berkeley has 235 foods. In contrast, the data base of the Ohio State University contains over 8,500 foods. The data bases used in the two most recent national food consumption surveys, the USDA 1977-78 Nationwide Food Consumption Survey (NFCS) and the Second National Health and Nutrition Examination Survey (NHANES II) contained approximately 3,700 and 2,600 foods, respectively. These data bases were sufficient to evaluate 24-hour recalls and two-day diaries for 30,000 persons (NFCS) and 24-hour recalls for 20,000 persons (NHANES II). The diets were those of subjects selected to be representative of the non-institutionalized U.S. population with regard to age, sex, income, race, region, and urbanization.

The number of foods in a data base should reflect the unique nutritional differences of individual foods and the importance of these nutritional differences to the uses of the data base. For example, a study concerned with fat intake and its relationship to cardiovascular disease may have replicate entries for the same item (e.g., fried eggs, homemade chocolate cake, pancakes, french fries) depending upon the type of fat used (e.g., butter, margarine, corn oil, vegetable shortening, peanut oil, palm oil, lard, etc.). This information could be very important in evaluating the relationship between intake of fatty acids and cholesterol with the incidence of heart attacks or strokes. This same data base may, however, have broad food aggregations for foods that are low in fat (e.g., fruits and vegetables). A data base developed for an epidemiological study of diet and cancer would probably focus on foods that are major sources of carotenes, dietary fibers, vitamin C, and fat. Again, there are no right or wrong choices for data bases, only appropriate choices to suit specific needs.

FDA's Total Diet Study monitors the levels of contaminants and minerals in the diets of selected age-sex groups though yearly analysis of 234 "core" foods. Each food represents a group of similar foods. For example, the frozen commercial apple pie represents all commercial, homemade, and restaurant fruit pies, turnovers, pastries, and strudels. These "core" foods are collected, prepared for consumption, and analyzed four times per year for 11 essential elements and over 200 pesticides residues, industrial chemicals, and toxic elements. The heavy analytical burden of this program requires that the number of food samples be kept relatively low. Though the numbers of foods are low, the Total Diet Study has successfully monitored the levels of contaminants and nutrients in the U.S. food supply since 1961. Included among the Total Diet Study nutrient findings were increases in iodine from dairy and grain products and a decrease in iron intake of infants due to the decrease in the fortification level of this mineral in infant cereal.

Food Descriptions

Depending upon the uses of the data base, food descriptions may range from the general (e.g., fried chicken) to the very specific (e.g., chicken, roaster, thigh, batter dipped, fried in hydrogenated cottonseed oil). Some data bases such as those for national food consumption surveys must include the full range of descriptions from the general to the specific. The data base for the NFCS contained many "not further specified" (NFS) entries (e.g., milk, NFS; meat, NFS; sandwich, NFS) because many of the participants could not adequately describe the foods they consumed. Nutrient values for these NFS foods were estimated so that nutrient intakes could be estimated. This feature is far superior to having only specifically defined foods and forcing the coder to select what they consider to be the "best" food code.

The level of detail you want in your food descriptions is dependent upon the types of questions you wish to address. If your data base is used to estimate nutrient intake, the level of food description should parallel the information collected from study participants. This is dependent upon the food consumption methodology and the level of intelligence and patience of the subject and interviewer. Different food consumption methodologies (24-hour recalls, food diaries, food frequencies, etc.) do elicit somewhat different food descriptions and detail.

For the food descriptions of your data base, consider whether you need to know the source of the food (e.g., homemade, commercial, fast food, deli, restaurant); their preservation method (fresh, frozen, salted, dehydrated); the packaging material (metal, paperboard, glass, cellophane); the preparation method (boiled, roasted, fried, microwaved, pressure cooked, steamed, barbequed, stir fried). For mixed dishes you may want a recipe file which lists the ingredients of commercial and restaurant foods and quantities of ingredients for home prepared items.

It is often difficult to determine what is a single food (requiring no recipe) as opposed to a mixed dish (which requires a recipe). Foods to which only water, salt, herbs, and spices (including garlic and onion) have been added are usually considered single foods. Foods with added fat

or prepared with a fat, sauce, or gravy are often considered single foods as well. For some foods you need to decide if you will have entries for the mixed food or will code added ingredients separately. For example, for a baked potato with added butter, margarine, sour cream, or cheese sauce, the baked potato and topping may be coded separately or there may be separate codes for baked potato with a specific topping. The same is true for coffee with milk or cream and/or sugar.

You may wish to establish an order to the descriptive terms of foods in your data base. It is often convenient to list the raw/fresh item first, followed (if applicable) by the cooked item, followed by various processed forms of the item listed alphabetically:

peas, green, immature
    raw
    fresh, cooked
    canned
    frozen
    frozen, boil-in-bag
    frozen, in butter sauce
peas and carrots
    canned
    frozen
peas and mushrooms, frozen
peas and onions
    frozen, in butter sauce
    frozen, in cream sauce

It is extremely useful to develop a dictionary of descriptive terms. This is important if several persons are maintaining and updating the data base. The dictionary should also include "use" and "used for" terms (e.g., salted: use salt added; cowpeas: used for black-eyed peas). The dictionary is most easily handled as a computerized file.

Classification or Coding Schemes

A classification system is of particular importance to the person who must code food consumption data. To assess the nutrient content of a diet, the coder must find the appropriate food code for each food using a code manual. The food codes and quantities of foods consumed may then be entered into a computer, and a software program may perform the necessary calculations to determine nutrient intakes. The most common classification scheme used for food composition data bases is based on food groups (meat, dairy, fruit, vegetable, etc.). There may be subsequent categories within these major food groups. For example, meat may be subdivided among beef, pork, lamb, veal, and game, and beef may be further categorized by round, rump, loin, hamburger, porterhouse steak, T-bone steak, etc. The foods within a subcategory are usually listed alphabetically, and each food is assigned a code number.

The USDA NFCS classified foods into 10 major groups and various major and minor subgroups. These three types of classifications are denoted by the first three numbers of a seven digit code number. This classification and coding scheme allows data retrieval for the major food groups, major food subgroups, and minor food subgroups in addition to retrieval of data for individual foods. For example, one could use appropriate codes to estimate consumption of whole fresh, fluid milk (a specific food); all fluid milks (a minor food subgroup); all milk and milk drinks (a major food subgroup); or all milk and milk products (a major food group).

Classification of foods by food groups produces a strict hierarchical scheme. Although useful for some food consumption studies, it is not efficient for other data base uses such as identifying foods by other characteristics such as food source (soy, beef, apple) preservation method (freezing, canning); or packaging materials (plastic, paperboard, metal). Also a food group classification leads to problems in placing many foods. For example, Irish coffee could be grouped with coffee or with alcoholic beverages; cafe au lait could be grouped with coffee or milk beverages; mixed dishes could be grouped together as a "mixed dish group" or each mixed dish could be grouped according to its major ingredient. The NFCS uses the latter method. It is often difficult to determine the major ingredient in mixed dishes such as beef-vegetable stew or lasagne. Fast foods may be grouped together as a "fast food group" or placed according to individual food type (e.g., sandwiches, vegetables, beverages). Commercial baby foods may also be grouped together as a "baby food group" or placed in appropriate food categories (e.g., meats, fruits, vegetables, cereals). Soups could be placed in a "soup group" or placed according to major ingredient (vegetable, meat, pasta, bean, etc.). These are decisions to be made by the data base compiler.

The FDA has developed an internal coding sytem to allow retrieval of data on the basis of food descriptors. This system can be applied to any food composition or food consumption data files. The FDA system describes foods on the basis of 11 factors: product type; food source; part of plant or animal; physical state, shape or form; degree of preparation; treatment applied; preservation method; packing medium; container or wrapping; food contact surface; and user group. Product type refers to food group. Food source and part of plant or animal describe the origin of the food or the major ingredient if it is a mixed dish. Degree of preparation; treatment applied; and preservation method refer to industry processing of the food. Packing medium, container or wrapping, and food contact surface refer to industry packaging materials. User group refers to the major consumers of the food. There are almost 2,000 factor terms which are fully defined by the FDA thesaurus which can be used in various combinations to retrieve foods from data files. For example, one could retrieve all canned foods that contain mushrooms or all soy-based foods in plastic containers.

If you are developing a data base for international use, you will probably need to include the scientific name (genus and species) in the food

description. If you are coding foods from other countries, you will need to retain the original name (in the native language), and determine the English translation, the scientific name, the language used, and the geographical location.

Types and Numbers of Nutrients

The nutrients included in a data base (like the foods included) remain a function of the purposes for which the base will be used. A data base used for institutional meal planning may include only the proximate nutrients and the major vitamins and minerals. A data base used by a food industry may include those nutrients used for food labeling. Data bases for epidemiological studies will focus on those specific nutrients believed to be associated with the diseases/disorders of concern to the study. A multipurpose data base tends to be all-inclusive (i.e., to include all available nutrients for all available foods). This often leaves many missing values in the data base as nutrient values for some foods are quite extensive, while values for many other foods are only available for the major nutrients.

The number of nutrients available in currently developed data bases ranges from 4 to over 100. Many data bases retain only the basic 17 nutrients of the original USDA Handbook 8. Others may also include vitamin B-6, vitamin B-12, folacin, carotenes, dietary fiber, and trace minerals such as zinc, copper, manganese, magnesium, iodine, and selenium. Other data bases include individual sugars, fatty acids, and amino acids. The necessity of including crude fiber and ash in the data base may be questioned, although you may wish to retain them in the first level of data collection. Some data bases have separate entries for plant and animal sources of protein and/or iron.

Expression of Nutrient Values

It is important that the data in the data base be on a fresh (wet) weight basis. The compiler may convert dry weight values to fresh weight values, but unfortunately, most references do not give the residual moisture of a dried food. The best that can be done (if communication with the authors of the references is not feasible) is to assume zero percent moisture in the dry product and footnote the fact that your nutrient values are estimated from dry weight values. One must be careful to discern between nutrient values listed as "percent dry weight" versus "percent dry ash."

For most purposes, nutrient values per 100 grams are most useful in a food composition data base. Information on the weight of standard servings or single service portions should also be collected. This information is essential for converting food consumption data into nutrient consumption data. If you are evaluating food frequency data which is not quantitated by serving portions, you will need to retain the nutrient data in your base by standard serving portions.

Data that have been imputed or calculated from recipe ingredients should be footnoted as such. Missing values should be imputed, especially for

foods that are recognized sources of particular nutrients. Otherwise tabulations of total daily intakes should be denoted as lacking values from one or more foods. This is of particular concern when intakes appear to be marginal compared to Recommended Dietary Allowances or Estimated Safe and Adequate Daily Dietary Intakes. The imputed and calculated values should remain flagged in the data base and replaced when possible by analytical values.

The ideal expressions for nutrient levels of foods are means with standard deviations and medians. Unfortunately, most data bases can only include means, and certainly those who use data bases to assess intakes of nutrients are primarily concerned with mean values for each nutrient in each food. However, those who use data bases for other purposes (food labeling, regulatory purposes, data comparisons on single products) may need to know the extent of nutrient variability and the reliability of the data. USDA has begun a practice of using confidence codes to indicate the reliability of nutrient data.

Data compilers also need to be concerned about the units used for nutrient measures. For several nutrients, there are several acceptable units (e.g., international units and retinol equivalents for vitamin A, international units and milligrams for vitamin E). Vitamin E may be expressed as total or as alpha-tocopherol. Energy may be expressed as kilocalories or kilojoules. Consideration must be given to the units desired and the units most commonly used and/or available. In some cases, it might be desirable to include values for some nutrients in several units.

Summary

There are many currently used food composition data bases. Even though most of them incorporate food composition data from similar sources, there are unique characteristics and features of these data bases that reflect their individual functions. If you are at the point of developing, adopting, or modifying a data base, think carefully about the functions that your base will serve. If you are in academia, your base may need to serve the needs of research in several departments plus analysis of student diets; if you are in private practice your base may serve mainly to analyze the diets of individual clients and patients. Data bases for industry may need only company products and only those nutrients listed on food labels. Data bases for large diet-health epidemiological studies or national food consumption studies must consider variables associated with eating habits such as age, sex, race, region, urbanization, season, income, religion. Food consumption studies directed at specific population groups may contain a smaller number of foods.

With regard to nutrients, consider the questions you wish to address and consider all the nutrients that might be of concern. It is difficult and inaccurate to go back and add nutrients to an already developed data base. One might attempt to include all available nutrient values in a data base, but only use those that are essentially complete for evaluating food consumption data. A review of the data bases that are currently in use may help to determine those characteristics regarding foods, descriptors, nutrients, and data expressions that would be appropriate for your data base. An awareness of these features at an early stage will aid in the development or selection of a data base to serve your needs.

EVALUATING NUTRIENT DATA BANKS

Loretta W. Hoover
Dietetic Education
318 Clark Hall
University of Missouri-Columbia
Columbia, Missouri 65211

As computers become more affordable, professionals are realizing the benefits of computerization for tasks that formerly consumed much valuable time. With computational tasks relegated to a computer, expertise can be focused on professional activities that require judgement and interpersonal skills. Selecting a suitable nutrient data base system and associated software can seem like an overwhelming task to an uninitiated user. However, with an orientation to some of the issues and an understanding of how the computer can serve as a professional tool, one can more easily proceed to evaluate the numerous options.

Two trends in computer technology should be kept in mind when considering hardware and software. These trends are (1) increased use of communications technology and (2) the increase in networking of computing devices. Transferring data from one system to another and sharing peripheral devices is becoming common in many settings. Microcomputers are being structured into local area networks to facilitate data sharing, to provide greater computing capability at an economical price, and to overcome the limitations of some single-user software and hardware devices.

Assessing the professional need for a computer is the first stage of the evaluation process. As a starting point, the potential user should enumerate tasks that could be accomplished with computer assistance. Current and future needs should be considered. In some situations, a nutrient data base may be the principle data collection, while in another case, a nutrient data base may be part of an integrated network of data bases. The broad range of activities that might be undertaken should be considered. The user who can forecast future applications for the nutrient analysis system will be better able to make a suitable selection. An inventory of potential uses for an analysis system is also helpful in determining the type of hardware required to support the range of activities. In addition, one should consider the resources that may already exist in the work setting. The existence of hardware capability and support personnel may influence one's strategy and software choices. The background work done at this stage can assist in making sound decisions.

Nutrient data bases vary widely with respect to size, contents, and source of data. The major portion of data should come from a reliable source that compiles data generated by sound analytical methods and routinely updates

the data base to reflect changes in the food supply. The option to update a data base or to receive frequent updates including user-specified foods permits tailoring the data base to specific needs. Users frequently add data to nutrient data bases to incorporate more recent data, to add other foods, to add other nutrients, to reflect regulatory changes, to supply missing values for nutrients, to add data for brand name foods, and to add nutritional profiles for mixed dishes. Maintenance of a data base is time consuming and expensive. The availability and cost of updated data bases should be determined before a final decision is made.

Estimating nutrients for recipes involves several considerations. The calculation methodology should be examined to assure that the weights of ingredients are adjusted to reflect the loses or gains in yield occurring in preparation and that the nutrient values are adjusted to reflect retention after preparation. Recalculation of the nutrient values by computer when new data become available facilitates keeping the mixed dish portion of a nutrient data base up-to-date.

The size of the nutrient data base chosen is influenced by its intended use. While computational efficiency may be gained by using a smaller data base, the absence of foods appearing frequently on some dietary records may result in numerous substitutions which distort the results of the analysis. A large data base may include many foods occurring on most dietary records but require excessive time and effort for coding and entering data.

When evaluating a software package, the users should consider how well it will meet their needs, the quality of the documentation, and it's compatibility with existing hardware and software. The features of the software should be appraised to identify computational options, comparisons of nutritional profiles with standard values, the technique for handling missing values, and the process for data entry.

The operational requirements of the software should be ascertained. Incompatibilities may arise out of differences in hardware, operating systems, programming languages, and teleprocessing systems. Sometimes different versions of the same product may require different computing environments.

The design of a system affects the reliability and flexibility of that system. The theoretical basis and assumptions in a software package should be reviewed. Also, the options for interfacing to other systems or adding other modules at a later time should be explored. The capability for real-time processing is determined by the system design. The competence level required of the end user is another important consideration.

Establishing a set of specifications facilitates the software selection process. Some additional aspects to consider are response time, functions, data files, interfacing requirements, and the contents of reports and screen displays. Comparable information about various products is useful

in identifying those likely to fulfill one's needs. The Nutrient Data Bank Directory (1) is one source of comparative information.

The performance of a nutrient analysis system may be difficult to determine without some actual experience using the system. The accuracy of several functions can be appraised in a systematic fashion. For example, the accuracy of updating functions and recipe calculations can be examined. The vintage of the data can be identified by retrieval of data for several different types of foods. The accuracy of portion adjustment and dietary record computations can be determined with desk checks. A sample methodology for this type of assessment is presented in the Model for Review of Nutrient Data Base System Capabilities (2).

Ongoing maintenance and management is required to assure the integrity of any nutrient analysis system. Updating policies and procedures should be stipulated. The data base represents a considerable investment in effort and dollars. Back-up procedures are needed to protect against its inadvertent loss. Quality control procedures are needed to verify accurate data entry. Effective communication with programming staff is necessary to overcome software limitations and to enhance system capabilities.

After careful evaluation, a suitable nutrient analysis system can be acquired to function as a professional tool. The more experience one has using a system, the more ideas one can generate for the use of a system. Thus, flexibility and the availability of new features should be considered.

References

1. Hoover, L.W., ed.: Nutrient Data Bank Directory. 4th Edition and Supplement. Columbia, MO: University of Missouri-Columbia Printing Services, 1984, 1985.

2. Hoover, L.W. and Perloff, B.P.: Model for Review of Nutrient Data Base System Capabilities. 2nd Edition. Columbia, MO: University of Missouri-Columbia Printing Services, 1984.

PROBLEMS AND METHODS IN APPLYING NUTRIENT COMPOSITION DATA TO CURRENT HEALTH ISSUES: VITAMIN A, BETA-CAROTENE AND CANCER

JEAN H. HANKIN

School of Public Health and Cancer Research Center
University of Hawaii

The role of vitamin A and its precursors in carcinogenesis has been investigated in several experimental and epidemiologic studies during the last ten years. The experimental studies have identified a role of vitamin A in the differentiation of epithelial cells and have shown that a deficiency of vitamin A may increase susceptibility to carcinogenesis in animals (1). From epidemiologic research, there is a growing body of evidence that the risk of particular cancers is inversely related to the consumption of the preformed vitamin (retinol) or the provitamin (carotene). This has been demonstrated in several case-control studies of various sites, such as cancers of the lung (2-4), bladder (5), larynx (6), and others (1), as well as in a few cohort studies (7-9). Of recent interest is the hypothesis that dietary beta-carotene may function as an anti-cancer agent. This was postulated by Peto et al. in 1981 (10) and has led to further testing of the associations of dietary and serum beta-carotene with cancer incidence and to implementation of a few intervention trials.

In planning epidemiologic studies with dietary components, it is essential that the methods of data collection and analysis be considered jointly. The data analysis, in turn, requires an up-to-date relevant set of food composition data that is adequate for testing dietary hypotheses. This should include a comprehensive set of values on the various retinoids and carotenoids on a weight basis, similar to the current procedures used by the U.S. Department of Agriculture for fatty acids and amino acids.

Because of the interrelationship of data collection and analysis, I will review the procedures reported in epidemiologic studies on vitamin A and cancer, discuss some of the problems in these methods, and propose a procedure for estimating the consumption of retinol, beta-carotene and other carotenes prior to the availability of improved analytical food composition data.

Dietary Methods

Most of the investigators collected data on the frequencies of consuming food items that were good sources of retinol, beta-carotene, and probably other carotenoids. As you will note, the procedures for selecting the particular food items and analyzing the dietary intake data were generally not specified.

Cohort Studies. One of the first studies was reported by Bjelke, who sent a mailed questionnaire to a cohort of approximately 8,200 Norwegian males to examine the relationship of vitamin A intake to lung cancer risk (7). Respondents were asked to indicate their current frequencies of consuming 25 selected food items. There were five frequency intervals, ranging from less than once a month to more than 14 times a month. To obtain quantitative data on vitamin A intakes, each food was assigned the vitamin A value of 100 grams of the item. Scores or indices were then estimated by summing the products of frequencies and vitamin A values of the food items. After 5 years, the data analysis revealed an inverse association of vitamin A with the incidence of lung cancer. Subsequently, the cohort was expanded to cover almost 17,000 men and women with more than 11 years of follow-up (11). Mailed questionnaires were again used to obtain frequencies of eating the 25 items, and the negative association of vitamin A with lung cancer risk was again demonstrated.

Hirayama has been following a cohort of 265,000 adults in 25 health districts of Japan since 1965 (8). At that time, interviewers collected information on several demographic and lifestyle characteristics, including the frequencies of consuming various green and yellow vegetables that contained more than 1,000 I.U. of vitamin A per 100 grams. Frequencies were recorded as daily, occasionally, rarely, and never. After 13 years of follow-up, Hirayama reported that daily consumption of these vegetables, as compared to less frequent intakes, was associated with a lower risk of cancer of several sites.

A third prospective study conducted among 2,100 men in Chicago was reported by Shekelle et al. (9). Dietary interviews among these men were conducted in 1958, and food profile scores were developed to record the frequencies of consuming food items in 26 food groups. These scores ranged from 0 to 3 for all groups. Zero always indicated no intake. The ranges in the "1 to 3" scores varied for each group of items. For example, as shown in Table 1, a score of "1" for vegetables was used for a consumption of 1 to 27 units per 28-day period. The ranges were subsequently replaced by a single value, such as "9" for the 1 to 27 units of vegetables. The amount of vitamin A per food unit was estimated by averaging the vitamin A values of all selected items in the food group. For vegetables, this was the mean number of International Units of vitamin A in 100 grams of a variety of items. After computing the individual intakes of vitamin A from each food group, these were combined to form retinol and carotene indices (see Table 1). Nineteen years later analysis of the cancer deaths in this cohort showed no association with the retinol index, but there was a strong inverse association of the carotene index with lung cancer risk. It should be noted that the items in the retinol index also included proportional amounts of carotenoids. Consequently, the

finding of a protective association between the "carotene index" and lung cancer risk could be due to factors other than carotene in fruits and vegetables.

A recent paper by Colditz et al. focused on the frequencies of consuming 41 fruits and vegetables among a cohort of 1,271 elderly persons free of cancer (12). From these data, a "green and yellow vegetable score" was derived for each person. (Parenthetically, this term is misleading as the foods also included dried fruits, strawberries and melon.) After five years, the investigators found that persons with the highest "score" had a significantly lower risk of death from cancer than those with lower "scores".

Case-Control Studies. One of the first case-control studies of vitamin A and cancer was conducted among Chinese men and women in Singapore by MacLennan et al. (2). They obtained dietary data on the frequencies of eating eight dark-green leafy vegetables. An index of four of these vegetables was shown to discriminate between cases and controls as well as an index of all eight of these items. Subjects were dichotomized and classified as eating at least two of the four items more than once a week (high) or eating less than two of the items in a week (low). The results showed that persons consuming vegetables less frequently were at higher risk of lung cancer than those consuming them more frequently.

Investigators at Roswell Park Memorial Institute in New York collected dietary data from all patients admitted during the period of 1957 to 1965 (3). Vitamin A intake was estimated from questions concerning the usual frequencies of eating 21 selected foods that were good sources of vitamin A during the months one year before the onset of symptoms. Each item was assigned the vitamin A value of a standard serving as listed in the U.S. Department of Agriculture food composition tables (13), and the estimated individual consumption was the sum of the products of the vitamin A values and the frequencies of the 21 items. Using this method, Mettlin et al. reported inverse associations of vitamin A with lung cancer (3) and with bladder cancer (5). This general procedure has been followed in the analysis of various nutrients and risk of site-specific cancers at Roswell Park (14).

Another case-control study of diet and lung cancer was conducted by Gregor et al. in London (15). Data on the current and past intakes of eggs, butter, margarine, milk, liver, carrots, and green vegetables, as well as vitamin supplements, were obtained by interview. It appears that amounts consumed were estimated and that weekly frequencies were ascertained, although these procedures were not described in the paper. The investigators reported that male cases consumed less vitamin A than controls, due to a lower intake of liver and vitamin pills.

Data from females were inconsistent.

Stehr et al. estimated vitamin A intakes in a case-control study of gastric cancer among proxy informants (16). (Next-of-kin were selected for deceased cases of cancer, and the age and sex-matched control group comprised relatives of deceased heart disease patients.) Similar to the preceding paper, the methodology was described vaguely. Evidently, a vitamin A index based on frequencies of foods that were good sources of vitamin A was derived. Nutrient values were based on standard size servings of the food items. Differences between cases and controls were in the expected direction.

In our first case-control study of vitamin A and lung cancer risk, we selected 84 food items which would account for 85 to 90% of the total vitamin A intakes of the subjects in a usual week (4,17). Amounts consumed were estimated by the subjects from photographs showing three serving sizes of each item. The particular items and the serving sizes were derived from measured three-day food records of 330 persons representative of the study population. Using the frequencies and nutrient values of the portions consumed, we obtained estimates of the usual dietary intakes before symptoms for the cases and the same time period for the age- and sex-matched controls. At that time, vitamin A was listed as "International Units" in our data base, without specific values of the vitamin A components. The published results included data on total vitamin A from foods and supplements, vitamin A from foods only, and vitamin A from retinol precursors or carotenes in foods. The latter comprised the vitamin A in fruits and vegetables and the estimated proportion of provitamin A in eggs, dairy products and a few additional items. Our findings showed that men consuming the highest intakes of vitamin A and particularly provitamin A had a lower risk of lung cancer than men consuming lower intakes of these nutrients. These results were not apparent in women.

Problems in dietary methodology

It is obvious that the reported methods of estimating dietary intakes of vitamin A and beta-carotene in several epidemiologic studies have many limitations. First, the rationale for selecting food items to be included in the questionnaires was generally not stated. Yet, it is important that the particular items are representative of the eating patterns of the study population and account for an estimated 85 to 90% of the vitamin A intakes. This can be derived from accurately measured or recalled food intake data on a similar population.

Second, free-living persons are unlikely to consume 100-gram portions or standard size servings of foods on a regular basis. Such assumptions will lead to inaccurate conclusions.

For example, we recently completed a study to determine if frequency data could be substituted for quantitative intakes in the assessment of dietary histories among a group of 340 men participating in a case-control study of prostate cancer (18). This was tested by converting the frequencies to 100-gram portions or to standard household servings. We found that the frequency methods in both instances failed to yield the same diet-disease associations as did the quantitative method at the individual level.

Third, I question the justification of grouping frequencies of intake into a single value. Considerable information is lost by this process. Such individuality can and should be retained in any research study of diet and disease. Furthermore, the use of average nutrient values for analysis of a diverse group of vegetables or fruits serves no purpose. It also seems unnecessary with the availability of nutrient data banks and computer facilities.

There are other issues to be resolved in planning the data collection and analysis, such as:

1. Wide variation of beta-carotene in leafy green vegetables
2. Difference in fortification levels of breakfast cereals
3. Variation in retinol and carotene contents of soups
4. Effects of seasonal foods on carotene intakes
5. Variation in retinol content of different kinds of liver
6. Difference in serving portions of various vegetables

These factors also point out the need for developing a food composition data base that enables the investigator to analyze the intakes of vitamin A as well as the retinoids and the carotenoids. A number of investigators, including Beecher and his colleagues at the Nutrient Composition Laboratory, U.S. Department of Agriculture are currently refining high pressure liquid chromatography (HPLC) methods to measure these components (19). This research will be extremely useful for future epidemiologic research on vitamin A components and cancer.

Although the data collection issues I raise need to be answered, it should be noted that the reported studies distinguished in a relative manner low and high consumers of sources of vitamin A, and most likely carotene. The negative associations of these nutrients with lung cancer risk certainly suggest a protective role for one or more of these dietary factors.

Recommended procedures for estimating retinol and carotene

After publishing our initial results on diet and lung cancer, we decided to add specific data on retinol, beta-carotene and other carotenes to our data base. We began by reviewing

published food composition data such as those of Paul and Southgate (20). However, there were limitations in their data. For example, their values do not distinguish between beta-carotene and other carotenes; the importance of these variations is unknown. Also, because this reference does not include a major portion of the foods consumed in Hawaii, there would need to be considerable imputing of retinol and carotene values. To avoid the use of a diverse array of published data, we followed the procedures recommended by the FAO/WHO Expert Group who published a percentage distribution of vitamin A into retinol, beta-carotene and other carotenes (21). Because all food items could not be classified neatly into the published groupings, we added a few additional categories for mixed dishes by calculating the proportions of the major ingredients in these items. Table 2 shows the food groups of FAO/WHO, and Table 3 includes the food groups with imputed distributions of the components. After computing the I.U. for each component, we utilized the formulas, published by the National Research Council (22), for deriving the micrograms (mcg) of retinol, beta-carotene and other carotenes: 1 I.U. vitamin A = 0.3 mcg retinol, 0.6 mcg beta-carotene, or 1.2 mcg other carotenes.

These derived values are certainly not precise, but they do provide a systematic procedure for estimating the intakes of vitamin A and its components. Until more comprehensive food composition data become available and the roles of these dietary factors in carcinogenesis are elucidated, this method, along with greater attention to data collection, will increase the reproducibility of dietary data and provide further evidence on the roles of vitamin A and beta-carotene in carcinogenesis.

References

1. Committee on Diet, Nutrition, and Cancer, Assembly of Life Sciences, National Research Council. Diet, nutrition, and cancer. Washington, D.C.: National Academy Press, 1982.

2. MacLennan R, Da Costa J, Day NE, et al. Risk factors for lung cancer in Singapore Chinese, a population with high female incidence rates. Int J Cancer 1977;20:854.

3. Mettlin C, Graham S, Swanson M. Vitamin A and lung cancer. J Natl Cancer Inst 1979;62:1435.

4. Hinds MW, Kolonel LN, Hankin JH, et al. Dietary vitamin A, carotene, vitamin C and risk of lung cancer in Hawaii. Am J Epidemiol 1984;119:227.

5. Mettlin C and Graham S. Dietary risk factors in human bladder cancer. Am J Epidemiol 1979;110:255.

6. Graham S, Mettlin C, Marshall J, et al. Dietary factors in the epidemiology of cancer of the larynx. Am J Epidemiol 1981;113:675.

7., Bjelke E. Dietary vitamin A and human lung cancer. Int J Cancer 1975;15:561.

8. Hirayama T. Does daily intake of green-yellow vegetables reduce the risk of cancer in man? An example of the application of epidemiological methods to the identification of individuals at low risk. In Armstrong B, Bartsch, H, eds. Host factors in human carcinogenesis. Lyon, France: IARC, 1982:531.

9. Shekelle RB, Lepper M, Liu S et al. Dietary vitamin A and risk of cancer in the Western Electric Study. Lancet Nov. 28, 1981; 1185.

10. Peto R, Doll R, Buckley JD et al. Can dietary beta-carotene materially reduce human cancer rates? Nature 1981;290:201.

11. Kvale G, Bjelke E, Gart JJ. Dietary habits and lung cancer risk. Int J Cancer 1983;31:397.

12. Colditz GA, Branch LG, Lipnick RJ et al. Increased green and yellow vegetable intake and lowered cancer deaths in an elderly population. Am J Clin Nutr 1985;41:32.

13. United States Dept. Agriculture. Nutritive value of American foods in common units. Agriculture handbook 456: Washington D.C.: U.S. Govt. Printing Office, 1975.

14. Graham S. Epidemiology of retinoids and cancer. J Natl

Cancer Inst 1984;73:1423.

15. Gregor A, Lee PN, Roe FJC et al. Comparison of dietary histories in lung cancer cases and controls with special reference to vitamin A. Nutr and Cancer 1980;2:93.

16. Stehr PA, Gloninger MF, Kuller LH. Dietary vitamin A deficiencies and stomach cancer. Am J Epidemiol 1985;121:65.

17. Hankin JH, Kolonel LN, Hinds MW. Dietary history methods for epidemiologic studies: application in a case-control study of vitamin A and lung cancer. J Natl Cancer Inst 1984;73:1417.

18. Chu S, Kolonel LN, Hankin JH et al. A comparison of frequency and quantitative dietary methods for epidemiologic studies of diet and disease. Am J Epidemiol 1984;119:323.

19. Beecher GR. Personal communication.

20. Paul AA, Southgate DAT. McCance and Widdowson's the composition of foods, 4th ed., New York: Elsevier/North Holland Biomedical Press, 1978.

21. WuLeung WT, Butrum RR, Chang FH et al. Food composition table for use in East Asia. Rome: FAO, 1972.

22. Food and Nutrition Board, Assembly of Life Sciences, National Research Council. Recommended dietary allowances, 9th ed., Washington DC: National Academy of Sciences, 1980:56.

Table 1
Unit of measurement, vitamin A per unit, and number of units in each food-profile score for food groups forming retinol and carotene indices[a]

Food groups	Unit of measurement	Vitamin A (IU/food unit)	Number of units in food-profile scores 1	2	3
Forming the retinol index:					
Whole milk	480 ml	780	1-27 (14)	28 (28)	≥29 (56)
Cream	30 ml	249	1-13 (7)	14-84 (49)	≥85 (168)
Butter	14 g	460	1-27 (14)	28-84 (56)	≥85 (140)
Margarine	14 g	460	1-27 (14)	28-84 (56)	≥85 (140)
Cheese	28 g	400	1-7 (4)	8-16 (12)	≥17 (32)
Ice cream, custard, pudding	120 ml	330*	1-3 (2)	4-12 (8)	≥13 (24)
Eggs	54 g	550	1-11 (6)	12-28 (20)	≥29 (56)
Liver	120 g	52,680	<1 (0.5)	1-2 (1.5)	≥ 3 (4)
Forming the carotene index:					
Vegetables	100 g	2560**	1-27 (9)	28-84 (42)	≥85 (98)
Soup	240 ml	1113***	1-11 (3)	12-28 (16)	≥29 (42)
Fruit	100 g	940****	1-27 (9)	28-84 (42)	≥85 (98)

*Mean value of the 3 items

**Average value of asparagus, green beans, beets, broccoli, cabbage, carrots, cauliflower, corn, eggplant, leafy green vegetables, other green and yellow vegetables, onions, peas and tomatoes

***Value for composite soup

****Average value of avocado, apple, banana, cantaloupe, citrus fruit, other fresh or canned fruit and dried fruit

[a]Data from Shekelle et al. (Ref. 9)

Table 2

Food groups and estimated distribution of vitamin A components*

Food Group	(% from) Retinol	(% from) β-Carotene	(% from other) Carotenoids
Meat and meat organs	90	10	
Poultry	70	30	
Fish and shellfish	90	10	
Eggs and egg products	70	30	
Milk and milk products	70	30	
Animal fats and margarine	90	10	
Grains: corn products		40	60
Other grains and breads		50	50
Legumes, seeds, nuts		50	50
Green vegetables		75	25
Deep-yellow vegetables and tomatoes		85	15
Other vegetables		50	50
Deep-yellow fruits and products		85	15
Other fruits and products		75	25
Vegetable oils		50	50

*Data from FAO/WHO (Ref. 19)

Table 3

Food groups and imputed distributions
of vitamin A components

Food Group	(% from) Retinol	(% from) β-Carotene	(% from other) Carotenoids
Mixed entrees with meat, milk, veg. (1/3 each)	53	30	17
Mixed entrees with grains, cheese or eggs, meat or fish (1/3 each)	53	30	17
Soups with meat or fish and veg. (1/3 meat, 1/3 green veg., 1/3 other veg.)	30	45	25
Soups with poultry or milk and veg. (1/3 poultry/milk, 1/3 green veg., 1/3 other veg.)	23	52	25
Other soups		50	50
Soups and sauces with milk or cheese base	70	30	
Fortified breakfast cereals	90	10	
Cakes, cookies, pancakes, waffles	70	30	

Problems and Methods in Applying Nutrient Composition Data to Current Health Issues: Dietary Fiber

by

Barbara O. Schneeman
Nutrition Department
University of California
Davis, CA 95616

Dietary fiber consists of the plant material which is resistant to hydrolysis by the enzymes of the mammalian digestive tract. The publication of Burkitt and Trowell's epidemiological associations between lack of fiber in the diet and chronic diseases prevalent in Western and developed countries stimulated the interest of clinicians, nutritionists, and gastroenterologists in fiber as an important component of foods. Because of the complexity of dietary fiber, the etiology by which it may affect disease remains uncertain; however, evidence has been accumulating that fiber supplements or fiber-rich foods can modulate the digestive and absorptive process and may have benefit as therapeutic agents. Given the increasing emphasis placed on the importance of fiber in the diet, we have been faced with the problem of trying to accurately determine the fiber content of foods and particularly to determine the components of fiber which may by important in the gastrointestinal tract. In my presentation I plan to cover some of the methods that have been developed to determine the fiber content of foods. To illustrate some of the problems and complexities in determining dietary fiber, it is necessary first of all to consider some of the diversity in the chemical and physical properties of dietary fiber and to discuss the variations in physiological effects which have been associated with sources of fiber.

Fiber as a group are found in plant materials. The most abundant compounds associated with fiber are associated with the plant cell wall. Some of the compounds identified as fiber are part of the intracellular cement in plants, others are secreted by the plant in response to injury, or are used to prevent seeds from dessicating. Because of these functions the general group dietary fiber can be divided into 3 sections: Structural Polysaccharides, which include those polysaccharide that are associated with the cell wall and include cellulose and the noncellulose polysaccharides (hemicelluloses and some pectins); Structural Nonpolysaccharide of which lignins are the main category; and the Nonstructural Polysaccharides, which include the gums and mucilages which are secreted by cells and the polysaccharides such as carrageenan and agar from algae and seaweed. Table 1 gives the average distribution of the cellulose, noncellulosic polysaccharide, and lignin fractions found in cereal, vegetables, and fruits. In all of these the noncellulose polysaccharides are likely to be present in a greater proportion than cellulose or lignin, and fruits and vegetables tend to be higher in cellulose than cereals. Some of the fruits have a very high proportion of lignin, which is likely to occur if seeds are a part of the edible portion of the fruit, as with strawberries. When examining the distribution of fiber in plants, it is important to keep in mind that the fibrous composition will vary by the species of plant, part of the plant (e.g. root vs. leave), and the maturity of the plant. For example, the

lignin content generally increases significantly as the plant matures.

Just as their functions in the plants vary, the chemical composition of each fraction varies widely. Table 2 describes the chemical components found in various fiber. The simplest is cellulose which is a glucose polymer without branching. The noncellulose polysaccharides are based on a variety of carbohydrates and can contain a high degree of branching. Lignin is a highly complex nonpolysaccharide polymer which contains phenylpropane units. The structures of the major polysaccharide components of fiber are shown in figure 1. Cellulose is a linear polymer of glucose with beta 1-4 links, it is the main structural component of plant cell walls, and it is considered relatively insoluble. D-galacturonic acids are the major component of pectins. The carboxyl groups on the galacturonic acids are partially methylated which are important to the properties of pectic substances. Pectins can have various carbohydrates linked to it. They are generally considered as highly soluble and are found as part of the cell wall and as intercellular cement. Hemicelluloses are a heterogenous group; the component sugars which make up its backbone and side chains are shown in figure 1. Hemicelluloses as a group are soluble in dilute alkali; however, there is a wide range in solubility with a greater solubility being associated with a high degree of branching. The structure of a purified lignin is highly complex and has a three dimensional structure. Lignins are considered very inert, insoluble, and resistant to digestion.

The description above provides a brief overview of the chemical diversity of dietary fiber. Sources of fiber also have unique physical properties which are clearly important in determining the physiological response to fibers. Table 3 summarizes some of those physical properties and indicates the type of compound for which that physical property may be important in eliciting a physiological response. Although dietary fibers cannot be degraded by the enzymes in the mammalian small intestine, they can be fermented by the bacteria in the large bowel. The polysaccharides and not lignin are the compounds degradable by bacteria. Among the polysaccharides, there is considerable variations in the degree of degradation. For example, pectins, mucilages, and hemicelluloses appear to be completed degraded by bacteria whereas cellulose is only partially broken down. The physical structure within the plant will also determine the extent of breakdown, hence fiber from fruits and vegetables appear to be more fermentable than that from cereals and grains. The extent of bacterial breakdown of fibers is important for several reasons. During the breakdown short chain fatty acids are produced which may be important metabolically, the fermentation process can lower the pH which could be important for microbial metabolism, and bacterial cells can account for a significant portion of the fecal weight hence their growth can contribute to fecal bulk.

The second physical property of importance associated with dietary fibers is their water-holding capacity. The presence of sugar residues with free polar groups in polysaccharides is associated with a significant hydration capacity. Hydration of the fibers results in formation of a gel matrix, which can also result in a higher viscosity of the small intestinal contents. The pectins and mucilages and, to a limited extent hemicelluloses, have the greatest water-holding capacity. Within the small intestine water holding capacity may have an important effect on nutrient absorption. Presumably diffusion of nutrients for absorption will be slowed by partitioning of water-soluble nutrients into the gel matrix and because of increases in the viscosity of the intestinal contents. Water-holding capacity has also been associated with increased fecal bulk;

however, the relationship is not straightforward because of the degradation of fiber by bacteria in the colon.

The third physical property of interest with dietary fiber is the ability to bind organic molecules. Molecules of interest include bile acids, cholesterol, and toxic compounds. In vitro studies have demonstrated that lignin is a potent bile acid adsorbent. Pectin and other acidic polysaccharides also seem to sequester bile acids. In contrast cellulose has little bile acid binding ability. In vivo bile acid adsorption is primarily measured as the ability to increase fecal bile acid steroid excretion. The ability to increase fecal excretion has been associated with the plasma cholesterol lowering effect of certain soluble, noncellulose polysaccharides, such as pectin and guar gum. The potential ability of some fibers to bind toxic compounds has been proposed as a mechanism for fibers to protect against gastrointestinal cancers; however,this property has not been studied extensively.

The fourth physical property concerns the cation exchange properties of certain fibers. The lower mineral availability and electrolyte absorption associated with diets high in fiber is undoubtedly due to the ability of some sources of fiber to bind minerals and electrolytes and increase their fecal excretion. The number of free carboxyl groups on the sugar residues and the uronic acid content of polysaccharides appears to be related to the cation exchange properties of fibers. These are 4 of the physical properties of fibers which are important in explaining the physiological response to fiber in the diet and it is evident that there is considerable diversity in these properties which is related to the diversity in chemical composition of dietary fibers.

Some of the physiological effects of dietary fibers include increased fecal bulk, decreased nutrient availability, lowering of plasma cholesterol, and reduction in the glycemic response to a meal. For each of these effects there are examples of fibers which are effective and those which are not. Two examples, fecal bulk and plasma cholesterol response, are useful to illustrate this point. The data in Table 4 were obtained from several studies which reported the percentage increase in fecal bulk with various fiber supplements. Coarsely ground wheat bran is the most effective in for increasing fecal bulk (80-120% increase). Disrupting the physical structure of wheat bran eliminates its effectiveness as a bulking agent, probably because it no longer holds water or because it is more readily degraded by bacteria. This is an example where it is not sufficient to know the absolute amount of fiber consumed but also know the physical form of the fiber. It is also possible to contrast different sources of fiber. The data indicate that the soluble, noncellulose polysaccharides tend not to be effective as fecal bulking agents. In the case of pectin and gums such as guar, the lack of bulking effect is because they are degraded by bacteria in the large bowel. The other example is shown in table 5 in which the ability of several fiber sources to lower plasma cholesterol is shown. The sources of noncellulose polysaccharide which are viscous appear to be most effective in lowering plasma cholesterol. These are also the fiber sources which are most effective in blunting the plasma glucose response. This suggests that the ability to form a gel matrix may be important in mediating the physiological response to these fiber sources. The importance of viscosity for lowering the postprandial glycemic response has been demonstrate in several studies. Hence that physical property must be maintained for the fibers to be effective.

It is clear that the term dietary fiber is used to encompass a very diverse group which of course becomes a tremendous problem for the analysis of the fiber content of foods. The methods for determining fiber include three basic approaches. One has been to use extraction methods to isolate various fractions and quantify gravimetrically. Extraction with acid and alkali is the basis of the crude fiber analysis, which has been the current AOAC method. The crude fiber procedure does not measure any specific carbohydrate or group of carbohydrates and consequently it does not accurately estimate the dietary fiber content nor does it have a quantitative relationship to fiber content of foods. The other extraction method is based on neutral and acid detergent extraction to isolate fractions. This method was originally developed by Van Soest for analysis of animal forages, and has had to be modified for analysis of human foods which are high in fat, protein, or starch. It has the advantage of being relatively rapid and useful for estimating the insoluble structural polysaccharides and lignin. A modified NDF is an official method of the American Association of Cereal Chemists. The second approach to determining the fiber content of food has been to analyze the individual components in the fiber residue. The best known of these procedures has been the Southgate procedure which involves a series of extraction steps to remove individual fractions, followed by hydrolysis and determination of the component sugars by gas-liquid or liquid chromatography. Other investigators have proposed extraction and chromatographic schemes to analyze the fiber content of foods in this way. This approach is difficult and time-consuming; however, in view of the variability in physiological response to different sources of fiber there is clearly a need for the detailed analytical data provided by these methods. In addition, analysis by this more rigorous and accurate method is needed to provide reference values to evaluate the results of simpler procedures. The third approach has been the development of rapid enzymatic procedures to provide a single value for the soluble and insoluble fiber content of the food. Enzymes are used to remove protein and starch from fat-extracted food. The residue is corrected for ash and protein content and the fiber determined gravimetrically. AOAC currently has this method under review. The enzymatic procedure supplies the need for a rapid gravimetric procedure but does not replace the need for a more comprehensive method to determine individual fiber components. Table 6 presents some of the results from the interlaboratory study to evaluate the enzymatic fiber method. For most of the samples the coefficient of variation is between 5 and 25 % and the lower CV tended to be associated with the higher level of fiber. This range of variability is in agreement with the variability reported by other fiber methods. For 2 of the samples the CV were unacceptably high (64 and 100 %). In the case of rice, incomplete digestion of the starch may have caused problems and in the case of soy protein isolate, the protein content may not have been determined and reported accurately by some laboratories. The method was also compared to the method of Englyst (nonstarch polysaccharides) and of Theander and Aman (component analysis). The methods appeared to give good agreement for the total dietary fiber content of the foods.

Table 7 contains a chart to provide a relative comparison among methods for the determination for the fiber content of foods. The Southgate or component analysis procedure was not included; however, the method was designed to determine each component which it does. The main problem encountered has been that incomplete removal of starch can interfere with the analyses. If we compare the 4 methods most commonly used we can see the relative value of each for fiber estimation. The crude fiber method detects cellulose and lignin;

however, they are not estimated accurately nor are the noncellulose polysaccharides estimated. This method seriously underestimates the fiber content of food and is not related in any systematic way to the dietary fiber content of foods. The neutral detergent method primarily estimates the structural components of the cell wall hence it does not provide values on the soluble fiber sources which must be estimated by other methods. The procedure by itself underestimates the total fiber content of a food, but in conjunction with an analysis of soluble polysaccharides such as pectins or gums is useful. The acid detergent fraction mainly contains the lignin and cellulose components with perhaps some small amounts of hemicellulose and pectin. Originally it was thought that the difference between NDF and ADF would provide the hemicellulose content but that has not proved to be a accurate method. Both of these methods then will underestimate the fiber content primarily because of the loss of soluble fibers. The final method is the enzymatic procedure in which all of these fractions will be part of the total fiber estimation. This procedure is most likely a slight overestimatation of fiber content because of potential contributions from starch or heated proteins.

At the present time we are hampered in providing nutrient information by not having accurate, complete information on the fiber content of food. It is clear that the values for crude fiber are of little use in trying to evaluate dietary fiber intake and the association between fiber intake and disease patterns. If the enzymatic method reaches the final stages of approval, we will at least have a more accurate means to report total fiber content. There are many applications in which this information will be useful such as for labelling and for evaluating trends in dietary consumption. At present, individuals who are interested in some of the therapeutic uses for fiber will need more accurate information on the components of various fiber sources.

REFERENCES

J. W. Anderson and W. L. Chen, 1983. In: Unconventional Sources of Dietary Fiber (edited by I. Furda), ACS Symposium Series, pp. 49-60.

D.J.A. Jenkins, D. Reynolds, A.R. Leeds, A.L. Waller, J.H. Cummings, 1979. Am. J. Clin. Nutr. 32: 2430-2435.

R.M. Kay, 1982. J. Lipid Res. 23: 221-242.

R.M. Kay and A.S. Truswell, 1977. Brit. J. Nutr. 37: 227-235.

R.M. Kay and A.S. Truswell, 1977. Am J. Clin. Nutr. 30: 171-175.

R.W. Kirby, J.A. Anderson, B. Sieling, E.D. Rees, W.L. Chen, R.E. Miller, R.M Kay, 1981. Am. J. Clin. Nutr. 34: 824-829.

T.A. Miettinen and S. Tarpila, 1977. Clin. Chim. Acta 79: 471-477.

L. Prosky, N.G. Asp, I. Furda, J.W. Devries, T. F. Schweizer, B.F Harland, 1984. J. Assoc. Off. Anal. Chem. 67: 1044-1052.

H.P. Roth and M.A. Mehlman, editors, 1978. Symposium on Role of Dietary Fiber in Health. Am. J. Clin. Nutr., Suppl. 31.

D.A.T. Southgate, 1976. Determination of Food Carbohydrates, Applied Science Publishers.

G.V. Vahouny and D. Kritchevsky, editors, 1982. Dietary Fiber in Health and Disease, Plenum Publishing.

K.L. Wrick, J.B. Robertson, P.J. Van Soest, B.A. Lewis, J.M. Rivers, D.A. Roe, L.R. Hackler, 1983. J. Nutr. 113: 1464-1479.

Composition of Dietary Fiber Sources

	Noncellulose Polysaccharides %	Cellulose %	Lignin %
Cereals, 8 samples			
Average	75.7	17.4	6.7
Range	71-82	12-22	Tr-15
Raw Vegetables, 11 samples			
Average	65.6	31.5	2.98
Range	52-76	23-42	Tr-13
Fruits, 10 samples			
Average	62.9	19.7	17.4
Range	46-78	9-33	1-38

Southgate, 1978

Dietary Fiber Estimation by Enzymatic Analysis

	Dietary Fiber, % Average	Std. Dev.	Reproducibility (cv), %
Corn bran	89.02	2.63	2.95
Wheat bran	82.25	2.23	5.29
Lettuce, freeze-dried	23.31	2.75	11.79
Whole wheat flour	12.92	1.43	11.08
Oats, quick-cooking	12.47	3.20	25.64
Lacto-ovo vegetarian mix	8.59	1.90	22.07
Soya isolate	7.51	7.58	100.93
Potatoes, instant	7.22	0.96	13.24
Nonvegetarian mix	7.19	1.90	26.39
Rye bread, dried	5.90	1.45	24.41
Raisins, seedless	4.43	1.03	23.12
Rice, powdered	3.67	2.35	64.15
White wheat flour	3.07	1.01	39.95

Prosky et al., 1984; Analysis by 30-32 laboratories.

Chemical Classification of Dietary Fiber

Fiber	Chemical Components: Main Chain	Side Chain
Polysaccharides		
Cellulose	Glucose	None
Noncellulose Hemicellulose	Xylose Mannose Galactose Glucose	Arabinose Galactose Glucuronic acid
Pectic substances	Galacturonic acid	Rhamnose Arabinose Xylose Fucose
Mucilages	Galactose-mannose Glucose-mannose Arabinose-xylose Galacturonic acid rhamnose	Galactose
Gums	Galactose Glucuronic acid-mannose Galacturonic acid-rhamnose	Xylose Fucose Galactose
Algal polysaccharides	Mannose Xylose Guluronic acid Glucose	Galactose
Lignin	Sinapyl alcohol Coniferyl alcohol p-Coumaryl alcohol	3-dimensional structure

Fecal Bulk Associated with Fiber Supplements

Fiber Source	% Increase in fecal wet weight
Oat bran	15
Pectin	16-35
Guar gum	20
Apple	40
Carrot	59
Cabbage	67
Cellulose	75
Wheat bran, coarse	80-127
Wheat bran, fine	28

Ref: Jenkins et al., 1979; Kirby et al., 1981, Kay and Truswell, 1977; Miettinen and Tarpila, 1977; Wrick et al., 1983.

Change in Serum Cholesterol with Plant Fiber Supplements

Source of Fiber	Plant fiber g/day	Cholesterol % change
Cellulose	16	0
Wheat bran	17	+1
Whole oats	15	+11
Oat bran	27	+17
Pectin	25	+13
Guar gum	24	+16
Bengal gram	unknown	+22
Beans	30	+19

Anderson and Chen, 1983

Physicochemical Properties of Fibers in the Gut

Property	Type of Fiber	Result
1. Bacterial Degradation	Polysaccharides	Short chain fatty acids Flatulence, Chemical environment.
2. Water-holding Capacity	Polysaccharides with polar groups	Nutrient absorption, Fecal weight, transit in stomach and small intestine
3. Adsorption of organic materials	Lignin Pectin	Bile acid binding and excretion
4. Cation exchange	Acidic polysaccharides	Increase mineral excretion

ANALYSIS OF FIBER IN FOODS

	Cellulose	Hemicelluloses	Lignin	Pectin	Total Fiber Content
Crude Fiber	+++	+	++	0	underestimates
Neutral Detergent Fiber (NDF)	++++	+++	++++	0	underestimates soluble fibers
Acid Detergent Fiber (ADF)	++++	+	++++	+	underestimates
Enzymatic Procedure	✓	✓	✓	✓	slight over-estimation

Key + indicate degree to which compound is determined
0 not estimated
✓ part of total fiber estimation

NUTRIENT INTERACTION ISSUES

Judith R. Turnlund, Ph.D., R.D.
Research Nutritionist
Western Human Nutrition Research Center

This presentation covers nutrient interactions, and the impact of these interactions, on nutrient bioavailability. Nutrient bioavailability includes two important components, absorption and utilization (1). Absorption is the process by which a nutrient moves from the intestinal lumen into the body. Until nutrients are absorbed, they are still outside of the body and are not available to perform their functions. Utilization of the absorbed nutrients includes transport to various parts of the body, assimilation by cells, and conversion to biologically active forms.

The major effects of nutrient interactions on bioavailablity and thus nutrient requirements, can be illustrated with zinc and iron. Daily endogenous losses of these minerals has been estimated at 2.2 mg of zinc, 1 mg of iron for men and postmenopausal women, and 1.5 mg of iron for women during their reproductive years. If zinc absorption were 40%, only 5.5 mg of dietry zinc per day would replace body losses, but if absorption were only 10%, 22 mg per day would be needed in the diet. One mg of zinc in the first diet would be equivalent to 4 mg in the second. Women would need about 6.5 mg of iron in a highly available (23% absorbed) form, but they would need 50 mg a day if the iron were in a poorly available form (3% absorbed). One mg of highly available iron would be equivalent to 6 mg of a poorly available form. The higher levels of both minerals are considered nearly impossible to obtain routinely in a diet without supplements. Thus, nutrient interactions which affect bioavailability adversely can impair nutritional status. Some of the examples of nutrient interactions I will discuss have been known for some time and are relatively well understood. However, most are newer research observations. The mechanisms of these interactions and their implications are not yet well understood.

Interactions can affect all of the major categories of nutrients; protein, carbohydrates, fats, vitamins, and minerals. Nutrients in each of these categories can also affect bioavailability of other nutrients. Interactions of nutrients with non-nutrient components of foods can alter availability. The effects of dietary fiber were discussed in the previous presentation. Other examples of non-nutrients which interact with nutrients include oxalate, avidin, and phytate. Oxalate combines with calcium to form an insoluble salt. The avidin in raw egg white renders biotin unabsorbable.

High levels of phytate have been known for some time to impair zinc availability in rats. However, the effect of phytate on zinc absorption in humans had not been confirmed until recently. In a study conducted in a human metabolic unit, young men consumed diets with no phytate or diets with sodium phytate added (2). Zinc absorption was determined with ^{67}Zn, a stable isotope of zinc. The results are shown in the table below.

NUTRIENT INTERACTION ISSUES

Judith R. Turnlund, Ph.D., R.D.
Research Nutritionist
Western Human Nutrition Research Center

This presentation covers nutrient interactions, and the impact of these interactions, on nutrient bioavailability. Nutrient bioavailability includes two important components, absorption and utilization (1). Absorption is the process by which a nutrient moves from the intestinal lumen into the body. Until nutrients are absorbed, they are still outside of the body and are not available to perform their functions. Utilization of the absorbed nutrients includes transport to various parts of the body, assimilation by cells, and conversion to biologically active forms.

The major effects of nutrient interactions on bioavailablity and thus nutrient requirements, can be illustrated with zinc and iron. Daily endogenous losses of these minerals has been estimated at 2.2 mg of zinc, 1 mg of iron for men and postmenopausal women, and 1.5 mg of iron for women during their reproductive years. If zinc absorption were 40%, only 5.5 mg of dietry zinc per day would replace body losses, but if absorption were only 10%, 22 mg per day would be needed in the diet. One mg of zinc in the first diet would be equivalent to 4 mg in the second. Women would need about 6.5 mg of iron in a highly available (23% absorbed) form, but they would need 50 mg a day if the iron were in a poorly available form (3% absorbed). One mg of highly available iron would be equivalent to 6 mg of a poorly available form. The higher levels of both minerals are considered nearly impossible to obtain routinely in a diet without supplements. Thus, nutrient interactions which affect bioavailability adversely can impair nutritional status. Some of the examples of nutrient interactions I will discuss have been known for some time and are relatively well understood. However, most are newer research observations. The mechanisms of these interactions and their implications are not yet well understood.

Interactions can affect all of the major categories of nutrients; protein, carbohydrates, fats, vitamins, and minerals. Nutrients in each of these categories can also affect bioavailability of other nutrients. Interactions of nutrients with non-nutrient components of foods can alter availability. The effects of dietary fiber were discussed in the previous presentation. Other examples of non-nutrients which interact with nutrients include oxalate, avidin, and phytate. Oxalate combines with calcium to form an insoluble salt. The avidin in raw egg white renders biotin unabsorbable.

High levels of phytate have been known for some time to impair zinc availability in rats. However, the effect of phytate on zinc absorption in humans had not been confirmed until recently. In a study conducted in a human metabolic unit, young men consumed diets with no phytate or diets with sodium phytate added (2). Zinc absorption was determined with ^{67}Zn, a stable isotope of zinc. The results are shown in the table below.

Table 1. Zinc-phytate interaction

	Basal diet	Basal + Phytate
Dietary zinc (mg)	15	15
Absorption (%)	34.0	17.5
Total fecal zinc (mg)	12.6	15.6
Unabsorbed dietary zinc (mg)	9.9	12.4
Endogenous fecal zinc (mg)	2.7	3.2
Urinary zinc (mg)	0.5	0.4
Zinc balance (mg)	+1.9	-1.0

Average zinc absorption in the young men was over 30% with no phytate in the diet. When phytate was added, at a level resulting in a phytate:zinc molar ratio of 16, zinc absorption was cut in half. When the diet contained phytate, the men were in negative zinc balance, or were losing more zinc than they look in. Not only was zinc absorption impaired, but it appeared that losses of endogenous zinc also increased. The phytate added to the diet not only interfered with absorption of dietary zinc, but impaired reabsorption of endogenous zinc which had been excreted into the gastrointestinal tract. The impairment in zinc absorption is probably due to formation of an isoluble phytate-zinc complex, which can not be absorbed. In contrast, the addition of phytate neither impaired copper absorption nor increased copper losses (3). The effect of foods containing high levels of phytate on zinc absorption in humans is now being investigated. Animal studies suggest an additional variable may play a role in the zinc-phytate interaction. It appears that high levels of dietary calcium enhance the negative effect of phytate on zinc availability (1).

Nutrient-nutrient interactions may affect bioavailability in either a positive or negative way. These interactions may either enhance or inhibit nutrient absorption or utilization. High or low levels of one or more nutrients may affect bioavailability of other nutrients. Examples of enhancement and inhibition of bioavailability and the effects of very high and low levels of nutrients are described later.

The relative proportion of nutrients present in the diet may be the most important factor in determining the impact of nutrient interactions on nutritional status. Using the earlier example of the zinc-calcium-phytate interaction, a diet high in all three may not affect zinc status. Even though the phytate and calcium present form isoluble complexes with zinc, enough zinc is present in uncomplexed, absorbable form to supply the needs of the body. A diet with a marginally adequate level of zinc, but low in calcium and phytate would not produce insoluble zinc complexes, would not impair zinc absorption, and would not result in zinc deficiency. But if a diet with the same marginal level of zinc were high in phytate and calcium, nearly all the zinc might be complexed in an insoluble form and zinc deficiency could result.

There are several mechanisms in addition to complex formation which may be responsible for nutrient interactions. One is competition (4). Competitive interactions cause problems most often when one nutrient is present in high amounts. Competitive interactions are common between minerals. In the

intestinal lumen, two nutrients could compete for a common binding ligand. A large excess of one would leave little binding ligand accessible to the other. The two nutrients could also compete for a common receptor site or carrier protein for entry into the intestinal mucosa, or compete for transport through the mucosal cell and into the blood stream. An interaction between iron and zinc has been studied by Solomon. The presence of high levels of iron appears to inhibit zinc absorption. Which type of competitive interaction is responsible for this altered zinc absorption is not yet known. Excess iron at the mucosal cell could prevent much of the zinc present from entering the cell. The interaction could occur within the mucosal cell. If iron status were adequate, so less iron entered the blood stream, the effect on zinc could be reduced. But if the excess iron produced a block to zinc, zinc status could still be affected. The interaction could present a problem when iron supplements are used, particularly in large amounts. Iron supplements are prescribed routinely for pregnant women, who may also have an increased need for zinc. Conclusive results on the effect of iron supplement on the zinc status of pregnant women have not yet been obtained.

An interaction between zinc and copper has been demonstrated. This interaction is also likely to be competitive. Copper deficiency, rarely observed in humans, was seen when very high levels of zinc were used therapeutically and copper intake was low (1).

Another mechanism for nutrient interactions is substitution. It is often a favorable type of interaction. In the presence of inadequate vitamin E, selenium is known to substitute for the vitamin in the inactivation of potentially harmful oxygen radicals. Folic acid supplements will prevent the anemia associated with vitamin B-12 deficiency. However, folate will not prevent the irreversible neurological damage resulting from vitamin B-12 deficiency. Observation of anemia leads to diagnosis of vitamin B-12 deficiency before irreversible neurological damage occurs, so the effect of folate is not considered advantageous in this situation. Another example of substitution is the substitution of tryptophan for niacin (5). Niacin deficiency will not result from a diet low in niacin, but with sufficient tryptophan, since tryptophan can be partially converted to niacin.

Function changes can result in nutrient interactions. For example, a deficiency of folic acid over a sufficient period of time results in changes in intestinal mucosa morphology. The intestinal villa become flattened, with less area. The function of the intestinal mucosa is impaired, resulting in inefficient absorption of most nutrients. Copper deficiency results in anemia in the presence of adequate iron. One hypothesis for the mechanism of this interaction is that copper deficiency produces a functional defect in the bone marrow. This defect impairs formation of red cells by the marrow. Some scientists think the mechanism responsible for anemia is a lack of copper oxidase activity which is necessary for iron transport.

Biochemical changes can result in interactions. Protein energy malnutrition (PEM) and other nutrient deficiencies can result in decreased levels of enzymes. These enzymes catalyze chemical reactions necessary for proper utilization of nutrients. Some of the deficiency symptoms may not be readily apparent in the presence of general malnutrition. During recovery

from PEM, symptoms of deficiency of other nutrients may become apparent. Requirements are exaggerated by rapid growth during repletion which results in a greatly increased need for most vitamins and minerals. The xerophthalmia of viamin A deficiency, and even blindness, have been observed when PEM is treated with protein and calories, but no vitamin A supplements.

The chemical or biochemical interactions produced by a nutrient may be advantageous for one nutrient, but deleterious for another. Ascorbic acid supplements will enhance the absorption of iron, probably by assuring it is in the reduced, best absorbed form. However, recent work suggests that these same supplements may have an adverse effect on copper status. They may have reduced the oxidase activity of ceruloplasmin, a copper-containing enzyme (6). The effect of iron on copper absorption is being evaluated with a stable copper isotope.

Often the mechanism of an interaction is not understood. Some recently discovered interactions whose mechanisms are not yet understood are described below.

Recent research with laboratory rats in Beltsville suggests that diets high in a carbohydrate, fructose, exaggerate the effect of copper deficiency (7). The impact of this interaction on copper status in humans is under investigation.

A recent experiment at the Western Human Nutrition Research Center was conducted to study the effects of vitamin B-6 deficient diets on vitamin B-6 status and also on the metabolism of several minerals. The study revealed interactions between vitamin B-6 and zinc and between vitamin B-6 and calcium (8, 9). When the diet was deficient in vitamin B-6, zinc absorption and retention were higher than usual. With higher levels of B-6, zinc absorption was lower and zinc retention declined. While zinc absorption and retention were higher with low vitamin B-6, the additional amount of zinc did not appear to be available for the body to use. The plasma level of zinc declined significantly with low B-6, then increased as vitamin B-6 was added to the diet. These results suggest that the enhance absorption was accompanied by impaired utilization.

Another type of effect was observed with calcium. Though calcium balance did not change markedly, urinary calcium fell to very low levels during B-6 depletion, suggesting a change in calcium metabolism.

In the above study, as in other studies conducted earlier, we used stable isotope of minerals as tracers to obtain definitive information on mineral absorption. Stable isotopes are valuable new tools to nutritionists, which allow the metabolic fate of minerals to be followed without exposure to radioactivity. They can therefore be used safely in pregnant women and in children. Interactions between several minerals can be determined simultaneously.

The importance of nutrient interactions on bioavailability suggest that

information on the nutrient content alone of the diet is not sufficent. Requirements vary depending on bioavailability. The addition of bioavailability factors to recommended intakes of nutrients or to nutrient data bases would be advantageous. Enough data are available on some nutrient interactions to introduce such factors. Dietary iron absorption can be calculated based on five factors: total iron, heme iron, non-heme iron, ascorbic acid, and amount of meat (5). Tryptophan can be expressed in niacin equivalents (5). Sixty mg of tryptophan is equivalent to 1 mg of niacin. Evidence in rats and now in humans suggests that a phytate:zinc ratio in excess of 20 adversely affects zinc status in rats (10). A ratio of 16 impaired zinc absorption in humans. A zinc:copper ratio which exceeds 10 may adversely affect copper status (4). These examples, as well as most other interactions become a problem when a nutrient or other dietary component is present in unusually high or low amounts.

Insufficient data are currently available to establish bioavailability factors for most nutrient interactions. In addition, as more data become available, an approach to introducing and applying these factors must be agreed upon and consistent to avoid confusion.

New techniques are available and others are being developed to facilitate evaluation of nutrient interactions. These include use of stable isotopes or radioactive tracers to determine mineral absorption and utilization; the slope-ratio assay, often used in laboratory animals; and identification of sensitive indicators of nutritional status, so minor changes in status can be detected.

As the data become available on interactions and bioavailability, I urge those developing nutrient data bases to work toward taking bioavailability factors into account in nutrient composition data.

REFERENCES

1. O'Dell BL. Nutr Rev 42:301-8, 1984.

2. Turnlund JR, King JC, Keyes WR, Gong B, and Michel MC. Am J Clin Nutr 40:1071-7, 1984.

3. Turnlund JR, King JC, Gong B, Keyes WR, and Michel MC. Am J Clin Nutr 42:18-23, 1985.

4. Solomon NW. In: Nutritional Bioavailability of Zinc. Inglett GE, ed. ACS Symposium Series 210, Wash DC, 1983.

5. National Research Council. Recommended Dietary Allowance, 9th ed. Wash. D.C., 1980.

6. Jacob RA, Skala JH, Omaye ST, and Turnlund JR. Fed Proc 44:937, 1985.

7. Fields M, Ferretti RJ, Smith JC, and Reiser S. Am J Clin Nutr 39:289-295, 1984.

8. Turnlund JR. Accord LL, Betschart AA, Kretsch MJ, and Sauberlich. Fed Proc 44:1278, 1985.

9. Turnlund JR, Khouri CM, Betschart AA, Kretsch MJ, and Sauberlich HE. Am J Clin Nutr 41:83, 1985.

10. Oberleas, D, Harland BF. J Am Diet Assn 79:433-6, 1981.

CRITICAL EVALUATION OF NUTRIENT DATA:

A PREREQUISITE FOR GENERATING NEW DATA

Joanne M. Holden
Nutrient Composition Laboratory
Beltsville Human Nutrition Research Center
Agricultural Research Service
U.S. Department of Agriculture
Beltsville, Maryland 20705

The research goal of the Nutrient Composition Laboratory (NCL), ARS, USDA is to provide essential data on the nutrient content of foods as consumed in the United States (U.S.). In order to make efficient use of limited fiscal and human resources, we must evaluate the quality of available data for a given nutrient. If high quality composition data exist for levels of a given nutrient in a wide variety of foods there is no need to generate more. If the quality of composition data has been determined to be inadequate, foods to be analyzed must be ranked in order of their priority for analysis. The NCL is particularly interested in the list of foods known as 'core foods,' that is, those foods most frequently consumed by the United States population. In particular, we have attempted to construct a list of 'core foods' for selenium (Se), defined as those frequently consumed foods which contribute the bulk of Se to the American diet. In addition we are interested in foods which may not be frequently consumed but are significant sources of Se.

Nutrient composition data are obtained from many sources - the food industry, published scientific literature, and from studies, published or unpublished, which are specifically designed to yield nutrient composition data. The job of evaluating these data from various sources for inclusion in the National Nutrient Data Bank is accomplished by the Nutrient Data Research Branch, Human Nutrition Information Service, USDA, under the direction of Frank Hepburn. In the past we have used their evaluation of such data as a guide to establishing analytical priorities within NCL for determining the nutrient composition of a specific food group or product type. Past projects have included fresh beef and pork, salty snacks, and fruit juices.

Recently, we have initiated a research project to establish guidelines for the evaluation of data for a specific nutrient - selenium (Se). The objective of this paper will be to describe the development of a system of general criteria to be specifically defined or qualified for this nutrient. In addition, I will briefly describe the method for selecting the list of 'core foods' which were the focus of the evaluation. Finally, I will show the application of the specific criteria requirements to published Se data to yield a list of 'core foods' for Se, together with a 'confidence code,' an indicator of quality for each mean value for each food.

As mentioned previously, the evaluation of available data is critical to the development of strategies for generating new data. The criteria which were developed and will be discussed were applied to published Se data. Selenium data were selected as a test case due to the current interest in the possible relationship between cancer incidence and Se intake (1,2). As a result of this interest, many human studies of Se intake and utilization are in progress and in need of Se composition data. In the 1980 edition of the Recommended Dietary Allowances a range of 50 to 200 ug/day has been given as the Estimated Safe and Adequate Daily Intake of Se for age 7 and over (3). An additional reason for focusing on Se is that good analytical methods for its determination are available. Finally, the data set is relatively small.

In order to evaluate published Se composition data for foods, two major tasks needed to be accomplished. First, we needed to develop a list of foods ranked in order of their importance to Se nutriture. The importance of a food to Se nutrition is based on its frequency of consumption, the amount consumed, and the Se concentration in that food. Frequency is defined as the rate or number of times a food is reported per day or per 3-day period. For example, if margarine was consumed at each of three meals then the frequency per day is 3. The list of foods was needed in order to establish priorities for consideration. A second but parallel activity concerned the development of a set of criteria to be used to evaluate the data from the scientific literature for the frequently consumed foods.

The frequency distribution for all foods consumed over a 3-day period by the 28,032 participants in the 1977-78 Nationwide Food Consumption Survey was obtained from the Food Monitoring Division, HNIS, USDA. This was considered to be the first step in an attempt to determine the contribution of any food to Se nutriture. We multiplied the frequency/day for each food by the average portion size to get the total intake/day of each food by the population. Frequency alone does not reflect the importance of a food's contribution since a food may be frequently consumed but may be consumed only in small quantities. Other foods may be consumed infrequently but their portion sizes may be relatively large. After evaluating the published data for various foods and determining a mean Se value for each, we incorporated that value into the frequency data for each of the top ranking foods to determine the Se contribution of each food to the U.S. diet.

Gram weights of some foods were summed to yield a total quantity of food aggregate consumed based on similarity of product type, the availability of Se data for each food, and the proximity of their Se levels.

Nationwide Food Consumption Survey food item	Total food intake/day
	g
Bread, whole wheat Bread, whole wheat, toasted Bread, whole wheat, high fiber Bread, whole wheat, high fiber, toasted	141,977

A food aggregate was regarded as a single item when the list of items was ranked by the total food intake per day. The amounts of the individual foods within the aggregate were retained in the file to permit re-grouping, if necessary.

An equally important and parallel activity was the development of the criteria for evaluating data. Five criteria were identified as significant in the evaluation of a data source: Analytical Method, Sample Handling, Sampling Plan, Analytical Quality Control, and the Number of Samples analyzed. These general criteria were similar to the criteria developed for the provisional Iron Table (5). In that publication three criteria were used. For each of the general criteria described here, a rating scale of '0' to '3' was developed. A range of 4 intervals was selected because it provided a level of resolution necessary for discriminating among various data sets and yet was not too detailed for available data. Requirements for each of the ratings were written to be specifically applicable to Se data. Writing the requirements required a knowledge of analytical methods for Se as well as a knowledge of sampling and statistics. In general a rating of '3' was considered to be optimum for a given criterion; a rating of '0' was assigned when there was no documentation or when the documentation was not acceptable. Table 1 lists the criteria and their specific requirements.

It is important to note some of the issues which are addressed by each of the criteria. For Analytical Method the emphasis is on method validation. The use of a validated official method or a well documented and validated new method is desirable. In order to obtain a rating of '3' the scientist must document the use of a validated method, including the use of tested quality control materials such as the National Bureau of Standards-Standard Reference Materials. Nutrient concentrations should be determined at quantifiable levels in the prepared sample extract or digest.

Sample Handling includes complete documentation of the validation of the homogenization method as well as details of sample preparation (cooked or uncooked, peeling, etc.) and storage. Evidence must be provided to indicate monitoring of the moisture content of samples during sample preparation and storage. Such precautions can assure that the sample which was analyzed is representative of the product which was purchased.

The criterion Sampling Plan addresses the representativeness of data for a specific nutrient and food aggregate. Our evaluation was conducted to determine the suitability of data for use in assessing the Se content of the American diet. With that purpose in mind, a study of the Se composition of samples obtained on a nation-wide basis with evidence of the representativeness of brands and varieties selected will be rated '3.' Minimal credit, '1', will be given to the references where samples were obtained only at the local level and where no evidence of the representativeness of brands or varieties is given.

Number of Samples Analyzed addresses the number of analyses for a given food in a particular study. Assuming that all other factors were equal, a study which reported multiple analyses of a food would be a stronger study

Table 1. Data Quality Criteria Requirements

Criteria\Ratings	3	2	1	0
Number of Samples	>10; SD, SE, or raw data reported	3-10	1-2; explicitly stated or not specified	-
Analytical Method	Official fluorometric (ref. given) or other method documented by a complete published write-up with validation studies for foods analyzed, including use of appropriate SRM where available, 95-105% recoveries on food similar to sample analyzed in same or other paper; Se concentration above quantitation limit of the method	Modified fluorometric or other method, some documentation, incomplete validation studies for foods analyzed; must include 90-110% recoveries on food similar to sample analyzed (or good recovery but no statistics given), and/or use of other method (official fluorometric, isotope dilution, or NAA) on same sample with good agreement (within 10%)	Non-fluorometric method, partially described; 80-90% or >110% recoveries on food similar to sample; or use of comparison method or recoveries on food only somewhat related to sample (animal/plant)	No documentation of method, no ref. or inaccesible ref. given, no validation studies, or poor agreement (>10%) of test method with comparison method on same sample
Sample Handling	Complete documentation of procedures incl. validation of homogenization method, details of food preparation and storage and moisture changes monitored	Pertinent procedures documented, seem reasonable, but some details not reported	Only edible portion analyzed	Totally inappropriate procedures or no documentation of criteria pertinent to food analyzed
Sampling Plan	Multiple geographical sampling with complete description; sample is representative of brands/varieties commonly consumed or commercially used	1 or 2 geographic areas sampled; sample is representative	Sample representative of small % of U.S. and/or origin not clear	Not described or sample not representative
Analytical Quality Control	Optimum accuracy and precision of method monitored and indicated explicitly by data	Documentation of assessment of both accuracy and precision of method; acceptable accuracy and precision	Some description of minimally acceptable accuracy and precision of method	No documentation of accuracy and/or precision

than one which reported single measurements. With multiple measurements (greater than 4-5) it is possible to generate some statistical indication of variability as well as a mean value.

Analytical Quality Control evaluates the level of accuracy and precision achieved in the day-to-day execution of the method. Both accuracy and precision are explicity defined in a paper to be published concerning the development of the criteria (6).

Greater detail on each of the criteria can be found in the two papers to be published within the next year (6,7).

The worksheet for whole wheat bread illustrates the assignment of ratings to each criterion and study. The reader should note that some references include data from more than one study. After all the data for a specific food have been rated, the numeric scores for each study are averaged over the five criteria to obtain a Quality Index (QI). For the Se data evaluation we decided that the presence of a '0' in Analytical Method for any study would automatically override the usual procedure for calculating the QI and yield a QI of zero for that study. The rationale was that if the analytical method was undocumented or not validated, the quality of data from that reference would be questionable. Similarly, ratings of zero in three criteria for a given study would yield a QI of zero.

WORKSHEET FOR WHOLE WHEAT BREAD

	Data Quality Criteria						Se	
Ref. No.	No. of Samples	Anal. Method	Sample Handling	Sampling Plan	Anal. Quality Control	Quality Index	Mean	SD
							µg/100g	
1	1	2	1	2	0	1.2 ↓	66.5	
2	1	2	0	2	0	1.0 ↓	53.5	
3	2	2	2	3	1	2.0 ↓	28.3	10.1
3	2	2	2	3	0	1.8 ↓	29.3	7.9
4	1	2	0	0	0	0	350	
5	1	2	1	2	0	1.2	67.0	
5	1	0	2	2	0	0	75 - 150	
						Σ = 6.0		
				Confidence Code = b		Grand Mean =	44.4	

↓ An index >1.0 is required for inclusion of an individual Se value in the calculation of the grand mean

After QIs were determined for each study, the sum of the QIs equal to or greater than 1.0 was calculated for each food. The worksheet for whole wheat bread indicates a QI sum of 6.0. The QI sum is used to determine a Confidence Code (CC) for the mean Se value for each food. The CC for a food indicates the degree of confidence a user can have in the mean value. The table below lists three ranges of QI's and their corresponding CC.

Assignment and Meaning of Confidence Codes

Sum of Quality Indices	Confidence Code	Meaning of Confidence Code
>6.0	a	Considerable confidence
3.4 to 6.0	b	Some confidence/ some problems
1.0 to <3.4	c	Less confidence/ limited data

Whole wheat bread received a CC of 'b' based on its QI sum of 6.0.

The mean Se concentration was calculated from those studies with QIs equal to or greater than 1.0. Those studies receiving a QI of less than 1.0 were excluded. None of the acceptable references was weighted more than another. The mean Se value for whole wheat bread was 44.4 ug/100 g.

The mean Se value for each food as determined by the evaluation procedure was combined with the total intake (g/day) of each food (as described above) to determine the Se intake/day for the population contributed by each food or food aggregate.

Determination of Se Intake/Day

Nationwide Food Consumption Survey food item	Total food intake/day	Se concentration	Se Intake/ day*
	g	ug/g	ug
Bread, whole wheat Bread, whole wheat, toasted Bread, whole wheat, high fiber Bread, whole wheat, high fiber, toasted	141,977	0.444	63,038

*Se intake/day by 28,030 survey respondents

This list of foods was ranked by the Se/day contribution to obtain the 'core foods' list specific for Se.

The evaluation of 65 references published since 1960 yielded mean minimum, and maximum Se values with Confidence Codes for more than 120 food items. The five highest ranked foods (beef, white bread, eggs, chicken, and pork) contribute one-half of the Se in American diets. Table 2 lists the first 16 foods which have been evaluated.

Based on this evaluation of available published data, NCL has begun to plan for further analyses of foods to obtain additional Se data. There were no Se data available for some foods in the NFCS list. Similarly, there were no data or frequency information for foods which have been introduced since the 1977-78 survey and may be important sources of Se. Select foods from these two groups were combined with high ranking foods which had been assigned a confidence code of 'c.' This group of foods was sampled and analyzed by NCL to supplement published data. In the future these data will be published and submitted to the Nutrient Data Bank, USDA.

Table 2. FIRST SIXTEEN SELENIUM CORE FOODS RANKED BY CONTRIBUTION TO U.S. DIETS

Rank	Food Item	Mean	Min. - Max.	Confidence Code	n/N+	Cumul. %¥
		µg Se/100g*				
1	Beef, ckd.	26	15 - 52	a	11/17	16.66%
2	White bread	32	23 - 54	a	7/9	31.05%
3	Eggs, ckd. (incl. scrambled, fried, soft-boiled)	25	16 - 38	a	7/11	39.02%
4	Pork/ham, fresh/cured, ckd./cnd. (incl. roasted, pan-ckd.)	35	19 - 92	a	8/13	46.46%
5	Chicken, ckd. (incl. fried, roasted)	21	17 - 26	a	5/10	52.90%
6	White rolls	34	21 - 61	a	4/4	57.00%
7	Whole milk	1.6	1.1- 2.5	b	4/15	60.20%
8	Whole wheat bread	44	28 - 67	a	4/7	63.25%
9	Tuna, cnd. (incl. light, white, packed in oil or water)	72	37 - 115	a	14/23	65.17%
10	Egg noodles, macaroni, spaghetti, ckd.++	20	14 - 42	a	20/23	67.03%
11	2% fat milk	2.6	---	c	1/3	68.88%
12	White rice, enriched, ckd.	9	---	c	1/5	70.59%
13	Macaroni & cheese, prep. from box mix	14	13 - 15	b	2/2	71.97%
14	Mayonnaise	60#	---	c	1/3	73.12%
15	Spaghetti with meat sauce, homemade	9	---	b	2/3	74.27%
16	Luncheon meat (incl. Spam¶)	28	27 - 28	c	2/5	75.42%

* Edible portion
\+ n = number of Se values from acceptable studies which were used to derive the mean value; N = total number of studies evaluated
¥ Cumulative % of total Se intake from core foods.
† Total Se intake/person/day from 89 core foods based on Nationwide Food Consumption Survey data and the mean Se values presented here = 74 µg
++ Se values calculated from dry and cooked values
\# Questionable based on the data from the other two studies which had acceptable QIs but reported semiquantitative Se levels ± 8 µg
¶ Geo. A. Hormel & Co., Austin, MN 55912

REFERENCES

1. Shamberger, R.J., Tytko, SA., and Willis, C.E.: Antioxidants and cancer. Part VI. Selenium and age-adjusted human cancer mortality. Arch Environ Health 31:231-5, 1976.

2. Schrauzer, G.N., White, D.A., and Schneider, C.J.: Cancer mortality correlation studies--III. Statistical associations with dietary selenium intakes. Bioinorg Chem 7:23-34, 1977.

3. Food and Nutrition Board, National Research Council: Recommended Dietary Allowances. 9th Rev. Ed. Washington, DC: National Academy of Sciences, 1980.

4. United States Department of Agriculture, Human Nutrition Information Service: Food Intakes; Individuals in 48 States Year 1977-78. Nationwide Food Consumption Survey 1977-78 Report No. I-1, 1983.

5. Exler, J. Iron Content of Food. Home Economics Research Report Number 45, Consumer Nutrition Division, Human Nutrition Information Service, USDA, Washington, DC, 1985.

6. Holden, J.M., Schubert, A., Wolf, W.R. and Beecher, G.R. A system for evaluating the quality of published nutrient data: Selenium, a test case. Food and Nutrition Bulletin, UN University 1986 (In press).

7. Schubert, A., Holden, J.M. and Wolf, W.R. Selenium content of a core group of foods based on a critical evaluation of published analytical data. Submitted to J. Amer. Diet. Assn.

PROGRESS IN METHODS DEVELOPMENT AND NEW NUTRIENT DATA

Gary R. Beecher
USDA, Nutrient Composition Lab

The earlier papers in this session give you some appreciation for the very confusing and very difficult research areas we are facing in terms of nutrition, nutrients and foods. Today, I would like to discuss several items pertinent to food composition. Firstly, I'd like to give an overview of the history and organization of the Nutrient Composition Lab (NCL). Secondly, I want to reveal what's on the horizon for food composition data. Lastly, I will talk about what we're doing in the lab in terms of analyzing foods.

The Nutrient Composition Lab was developed in 1975 as the result of the lack of good data on fats and fatty acids in foods. NCL is a part of the Beltsville Human Nutrition Research Center which is one of the five human nutrition research centers in the U.S. that's associated with the Agricultural Research Service.

We have two goals : One, to develop analytical procedures for the analysis of nutrients in foods and two, to provide nutrient composition data for foods that are commonly eaten. Part of this last goal is to assess the variability of the nutrient content of foods.

We have 8 scientists who have been trained in such diverse areas as Food Science, Physical Chemistry, Organic Chemistry, Analytical Chemistry, and Biochemistry. In addition, the laboratory has two groups: Joanne Holden heads the Food Sampling Group which is responsible for developing and executing food sampling plans. The computer group, headed by Robert Doherty, is responsible for the computer system and the integration of computers with the scientific equipment.

Funding for NCL comes from a variety of sources including the Agricultural Research Service, NHLBI, and HNIS/USDA. Recently, funds from the National Cancer Institute (NCI) have been provided to work on selenium and carotenoids.

Next, I would like to discuss ongoing research activity to improve data on the carotenoid and fiber content of foods. There are really two problems with the vitamin A data as they are presented. Dr. Hankin discussed cancer epidemiological studies and what confidence you might have in the vitamin A data that are available. One problem is that vitamin A activity comes from many components including retinol, alpha-carotene, beta-carotene, and perhaps several minor components. The second problem is that there are many carotenoids in fruits and vegetables that do not have any or very little vitamin A activity, yet these components may be important in maintaining optimum health in human beings.

Let me point out where we are and what you can expect in terms of carotenoid data. Dr. Buzzard and her group at the Nutrition Coding Center in Minneapolis have evaluated existing vitamin A data and have converted vitamin A activity into carotenoid data using the FAO/WHO 1967 tables that Dr. Hankin referenced.

These data are available for your use. Dr. Hankin has recently extended this concept to other foods and mixed dishes.

In terms of carotenoid data, HNIS has acquired a contractor who is generating data on the carotenoid and retinol levels in several foods. He's analyzing a variety of foods for retinol, alpha-beta carotene and cryptoxanthin. These data will be available as soon as they can be organized into a table.

The National Cancer Institute has awarded a contract to Arthur D. Little, Inc. in Cambridge, Massachusetts to develop methods and to analyze a number of foods for carotenoids and retinoids. Dr. Richard Taylor is the Chief Scientist responsible for analyzing not only foods for vitamin A carotenoids, but for other abundant carotenoids in foods also.

NCL is developing methods for the analysis of the abundant carotenoids in fruits, vegetables and mixed diets and is analyzing fruits and vegetables for these components.

In terms of new fiber data, Dr. Elaine Tanga at NCI has compiled a table on fiber data which will be published in the Journal of the American Dietetic Assocation in the next few months. Ms. Ruth Matthews and her group at HNIS are compiling fiber data and will provide it in a provisional table which should be available in early 1986.

NCI has awarded contracts to 2 different laboratories with the specific purpose of developing methods and analyzing the fiber content of foods. The laboratories are at Cornell University and the University of Wisconsin in Madison. A large amount of data will be available from these contractors as soon as the analysis can be conducted, data verified and presented in a useable format.

Also, Dr. Betty Li at NCL is developing an analytical method for pectin in food.

Finally, I would like to discuss food sampling activity currently ongoing at NCL. Fast-food chicken was first sampled in the Baltimore-Washington area to assess the variability between such parameters as brand, type of cooking and cut of meat. After application of appropriate statistical procedures to those data, a nationwide sampling plan was developed and executed. Approximately one hundred samples of the most popular brands of fast-food chicken were acquired. Starch, moisture, fat and fatty acids were analyzed at NCL. In collaboration with Mr. Hepburn at HNIS, a contractor was selected to provide analytical data for the majority of the rest of the nutrients.

NCL is in the process of conducting the preliminary sampling for cookies and snack cakes. Analysis for selected nutrients, including fat, fatty acids, starch, sugars, niacin, thiamin, and riboflavin will be conducted.

Finally, NCL in collaboration with HNIS has begun to develop plans to improve selenium data on foods. We anticipate it will take 2-3 years to complete our efforts with the available resources.

Dr. Turnlund indicated a potential interaction between normal levels of dietary copper and high dietary fructose. Scientists at NCL have begun to carefully evaluate published

copper data on foods as the first step in a major effort to improve these data. If cardiovascular health is affected by a dietary copper-fructose interaction, we would like to be able to present a table of the best available copper data to the public health community as soon as it is needed.

PROCEDURES FOR IMPUTING VALUES

Linda P. Posati
U.S. Department of Agriculture

Procedures for calculating and imputing nutrient values are used as circumstances prescribe by USDA's Human Nutrition Information Service in preparing portions of nutrient data bases. Major nutrient data bases available from USDA include the following: The revised sections of Agriculture Handbook No. 8; releases of USDA's Nutrient Data Base for Standard Reference; and the newly prepared USDA Nutrient Data Base for Individual Food Intake Surveys (1985). These data bases and others available from USDA may contain three types of nutrient values--analytical values which are usually arithmetic or weighted means, calculated values, and imputed values.

The distinction between analytical and calculated values and between calculated and imputed values is not clear cut. A goal in preparing data bases is to derive values which reflect the year-round, nationwide contribution of nutrients from foods. To achieve this goal, analytical data are sometimes weighted with factors obtained from production or marketing statistics. Calculated values are not themselves averages of analytical data but are derived from analytical data using specific procedures. They are no less valid than the analytical data upon which they are based and the soundness of the calculation procedures. Imputed values are quite often calculated values but they carry a degree of uncertainty. Analytical data upon which imputed values are based may be inadequate or assumptions that are without a sound research basis may have been made during the calculation steps. Below are data for cholesterol for three foods as presented in revised AH-8.

Food	Cholesterol (mg/100g)		
	Mean	Standard error	Number of samples
Whole milk	14	0.17	113
Dried egg yolk	548	--	133
Fresh ham, rump, lean and fat, raw	66	1.54	--

The mean values shown for whole milk and dried egg yolk are the arithmetic means of 113 and 133 samples, respectively. The standard error is given for the milk value to provide an estimate of variability. The standard error for the egg value is not calculated because the means were based on data that included averages of more than one analysis and for at least one of those averages no estimate of variance could be obtained. Analytical data contained in AH-8 are usually presented in the manner shown here for the milk and egg items. Calculated values are usually presented as means only without inclusion of a standard error or number of samples. The presentation of data for the fresh ham, rump cut, represents a deviation from this approach. The cholesterol content of this item is calculated using analytical data for cholesterol in the rump separable lean, the rump separable fat, and physical composition data for proportions of lean and fat in the particular ham cut. The mean is calculated using four variables which are estimated independently of each other using different numbers of replicates. Because of this there is no intrinsic number associated with the standard error.

Six specific procedures for calculating nutrient values from analytical data are based on or employ the following: Analytical and physical composition data; the content of a nutrient in a nutrient fraction of similar forms of a food; weighting factors; regression analysis; retention and yield data; and recipes. A commonly used procedure involves the use of the second item listed. Examples of nutrient fractions of foods are the lipid or fat fraction; the total solids; moisture-free, fat-free solids; the milk solids-not-fat fraction of dairy products; and moisture-free, fat-free, ash-free solids. Similar forms of a food would include the separable lean tissue from pork blade, rib, and top loin cuts. Similar forms of turkey are the flesh from the thigh, drumstick, and back, all of which are dark meat cuts. The pork cuts, on a total edible basis, would not be considered to be similar because their proportions of lean and fat differ, and the turkey cuts, if with skin, would not be similar because proportions of skin on the cuts differ.

A typical calculation based on analytical data for nutrients and physical composition data for cholesterol in raw chicken thigh with skin is as follows:

Tissue	Weight % of cut (EP)	Cholesterol (mg/100g)
Thigh meat	73.0	83
Skin	16.1	109
Separable fat	10.9	58

Cholesterol in cut =

(83 x .730) + (109 x .161) + (58 x .109) = 84 mg/100g

The cut consists of three tissues-thigh meat, skin, and separable fat. The mean percentages of the tissues are listed as weight percent of cut, edible portion. In the right hand column are listed means of the analytical data for cholesterol for each tissue. A calculated value for the chicken thigh with skin is obtained by multiplying the cholesterol content of each component tissue in the cut by the weight percent of each tissue and summing the results.

Below is a variation of calculations based on physical composition in which the analytical data are those for a nutrient in a nutrient fraction of the tissue. The value to be calculated is the grams of oleic acid per 100 grams of flesh and skin of a raw, whole chicken without neck and giblets.

Tissue	C18:1 (g/100g fat)	Fat %	Weight % of cut (EP)
Light meat	20.50 (27)*	1.65 (26)	31.95 (256)
Dark meat	25.84 (24)	4.31 (25)	39.49 (256)
Skin	34.22 (33)	32.34 (36)	17.09 (256)
Separable fat	37.22 (70)	67.95 (40)	11.45 (256)

% fat in flesh and skin:
By analysis = 15.06 (82)
Calculated = 15.54

C18:1 = [(20.50 x .0165 x .3195) + (25.84 x .0431 x .3949)
+ (34.22 x .3234 x .1709) + (37.22 x .6795 x .1145)]
x (15.06/15.54) = 5.17 g/100g Food

*Numbers in ( ) = number of samples

Most of the fatty acid data compiled for poultry were obtained from analyses conducted on separated tissues which consisted of the breast and thigh muscle, skin, and separable fat. The four tissue components of the whole chicken are the light meat (breast and wing meat), dark meat (drumstick, thigh, and back meat), and the skin and fat. The second column contains normalized means for oleic acid on a gram-per-100 grams fat basis. Data for breast meat are used to represent light meat and data for thigh meat to represent dark meat. The oleic acid data were obtained as weight percent of methyl esters and converted to a gram-per-100 grams of fat basis by applying factors to account for the non fatty acid constituents of the lipid. The third column contains the means of analytical data for fat content of the four tissues. Percent fat for light meat is that for breast and wing meat in natural proportions and percent fat for dark meat is that for drumstick, thigh, and back meat in natural proportions. Samples analyzed for oleic acid and fat are not the same. The weight percent of each tissue in the cut represents a separate data set. The percent fat in raw chicken flesh and skin is 15.06%. A calculated fat value for flesh and skin obtained using figures in the two right hand columns would be 15.54%. The calculated oleic acid value for this particular cut is the sum of the products of oleic acid, fat, and weight percent of the four tissues adjusted by the ratio of analyzed to calculated fat content. Thirteen independent sets of analytical values are used to derive the value of 5.17 grams of oleic acid in the food.

A simple example of a calculation based on similar forms of food is for calcium content of 1% fat lowfat milks.

Known nutrient content: 13.88 mg calcium per gram milk-solids-not-fat (MSNF) of whole milk

Lowfat milk	MSNF (%)	Calcium (mg/100g)
1% fat	8.86	123 (13.88 x 8.86)
1% fat with added MSNF	9.20	128 (13.88 x 9.20)
1% fat, protein fortified	10.23	142 (13.88 x 10.23)

The mean calcium content per gram of milk solids-not-fat of whole milk is 13.88 mg based on a large body of analytical data for pooled milk. The nutrient contribution of different forms of milk depends primarily upon the level of milkfat and milk solids-not-fat in the products. Lowfat milks may be of three types-plain, with added MSNF, and protein fortified. The latter contains over 10% MSNF in the final product. A large amount of information on the percent of milk solids-not-fat in commercial lots of the three types of 1% fat lowfat milks was available. Very little analytical data were available for calcium in these milks as described. Rather than use the means of few

data to represent the calcium levels in these foods, levels were calculated as shown using means of data for calcium in whole milk and for the milk solids-not-fat content of the three products.

HNIS sponsored research to obtain data on turkey for use in revising food composition tables for this commodity. Data were collected on proximate components, minerals, and four vitamins. Analyses were conducted on three turkey classes-fryer-roasters, young hens, and young toms. Vitamin B_6 was not among the nutrients examined. Levels in raw dark meat for each turkey class were calculated as shown in the following example:

Known nutrient content: B_6 in raw thigh meat of 18 young hens
0.351 mg B_6/100g wet tissue
20.22 g moisture-free, fat-free solids (MFFS)
0.0173 mg B_6/g MFFS

Turkey class	Number	MFFS (%)	B_6 (mg/100g)
Fryer-roaster	6	21.34	0.37 (0.0173 x 21.34)
Young hen	16	21.02	0.36 (0.0173 x 21.02)
Young tom	20	20.97	0.36 (0.0173 x 20.97)

Information from published literature was located for B_6 content in raw thigh meat of 18 young hens. The mean of these analytical values was 0.351 mg B_6/100g wet tissue. These particular samples contained an average of 20.22 grams of moisture-free, fat-free solids. The mean B_6 level in the analyzed samples was 0.0173 mg/g MFFS. The percentages of MFFS in the dark meat (flesh from back, drumstick, and thigh) are listed. The number of samples analyzed represents composites from 60 fryer-roasters, 60 hens, and 80 toms. The B_6 content for the dark meat of these three turkey classes was calculated by multiplying the known B_6 content of 0.0173 mg/g MFFS by the percentage of MFFS in the respective tissues. Differences in vitamin content between light and dark meat of poultry have been noted. However, effect of age or sex on vitamin levels is not well defined. If B_6 levels should be influenced by these factors, calculated values shown for fryer-roasters and young toms would be estimates.

The basic nutrient and physical composition data that HNIS had available for turkey were for the three classes previously mentioned. Nutrient values for a fourth classification of turkey, designated as "all classes" turkey, were calculated for use when the specific class was unknown and to provide a single set of nutrient values that would most appropriately reflect the nutrient contribution of turkey to our food supply. Below is an example of a procedure for deriving weighting factors for use in calculating nutrient values.

Turkey class	% of lb r-t-c (AP)		% EP in bird/100		Lb EP	Lb EP as %
Fryer-roaster	5	x	.7536	=	3.77	4.76
Young hen	45	x	.7836	=	35.26	44.53
Young tom	50	x	.8030	=	40.15	50.71
Sum					79.18	

Turkey class	Flesh and skin % of EP of bird		% edible contribution per class/100		Edible contribution by class to flesh and skin	Factors as %
Fryer-roaster	90.73	x	.0476	=	4.32	4.68
Young hen	92.08	x	.4453	=	41.00	44.40
Young tom	92.75	x	.5071	=	47.03	50.92
Sum					92.35	

The weighting factors used to derive the values for "all classes" turkey are based on the total pounds of ready-to-cook turkey of each of these three classes held in frozen storage over a certain time period. The proportions by weight contributed by the three classes are 5% fryer-roasters, 45% young hens, and 50% young toms. These values are for the as-purchased form of turkey including bone. The edible portion of each turkey class differs. Fryer-roasters contain more bone than do hens and toms. The pounds of edible portion contributed by each class expressed as a percent are shown in the right column with the top set of lines.

In order to use these factors to calculate specific cuts, the percentage of edible portion of each cut is taken into account. The first column on the lower set of lines contains the percent of edible portion contributed by the flesh and skin of each class. The difference between these values and 100 represents the edible portion of the neck and the giblets. To obtain weighting factors to calculate flesh and skin for an "all classes" turkey, these values are multiplied by the percent edible contribution per class as a decimal. The products expressed as a percent are shown in the lower right hand column. The factors are applied to nutrient data for flesh and skin of each class to calculate the "all classes" cut.

A procedure recently used to calculate nutrient values using analytical data for beef involved use of regression analysis. Below is an example of a calculation of fat content for a weighted market average value for beef top round steak, separable lean, raw.

OBS	Marbling Score	Observed Fat (%)	Predicted Fat (%)
1	5	2.78	2.47
3	8	3.10	3.13
6	9	3.31	3.35
7	10	3.55	3.57
9	12	4.14	4.02
10	13	4.77	4.24
11	14	4.35	4.46
12	15	4.94	4.68
14	23	5.91	6.45
16	24	7.73	6.69
18	25	7.17	6.89
19	26	7.27	7.11

Weighted market average: Marbling score = 12
Fat content = 4.02 g/100

Available for use were data for top round steaks from 19 beef carcasses. Only selected values are shown. The marbling score was available for each animal. This score is the primary quality grade determinant. The quality grades assigned to animals with these marbling scores were as follows: Standard grade-scores 5-7; good-scores 8-11; choice-scores 11-19; and prime-scores 20-26. Regression analysis showed a high correlation of both fat and water with marbling score. The observed fat values for the top round steaks are shown in the third column in order of increasing marbling scores of the animals. The right hand column contains the predicted fat values for the marbling scores shown. Data collected by USDA over the period 1980-1985 on graded beef for mean marbling scores within quality grades reflected a weighted market average marbling score of 12, equivalent to a grade of low-choice. For an overall value representative of the market, a predicted fat value of 4.02 g/100g associated with the score of 12 is appropriate.

It has been our experience to find more analytical data available on raw forms of food than on cooked forms. Therefore, cooked values are frequently calculated from raw data using information on nutrient retentions and cooking yields. A sample calculation for thiamin in light meat without skin for a mature stewed chicken follows:

Analytical data:

Thiamin in raw meat = 0.132 mg/100g
Thiamin retention = 51%
Yield of cooked meat = 72%

Thiamin in cooked meat:

$$\frac{0.132 \times 0.51}{0.72} = 0.094 \text{ mg/100g}$$

Analytical data available are the thiamin content of the raw light meat of mature chicken; the percent retention of thiamin in light meat upon stewing based on cooking studies on broiler-fryer chickens; and the yield of cooked light meat from mature chicken upon stewing. The thiamin content of the cooked meat is calculated by multiplying the thiamin content of the raw meat, 0.132 mg/100g, by the retention fraction, and dividing the product by the cooking yield to express the value on a 100-gram basis.

HNIS compiles a considerable amount of data on cooking yields and the retention of nutrients for various cooking methods for use in these types of calculations. An example of the determination of cooking yield factors for chicken breast using one particular sample from a study on chicken that HNIS had sponsored follows.

	Left Breast	Right Breast		
	Raw Wt	Raw Wt	Roasted Wt	
	g	g	g	g x 194/192
Part	194	192	134	135
Meat	132		97	98
SKin	17		14	14
Fat	6		—	—
Bone	37		21	21

Part yield = (135 x 100)/194 = 70%
Meat yield = (98 x 100)/132 = 74%
meat + skin yield = [(98 x 14) x 100]/(132 + 17 + 6) = 72%

Anatomically matched cuts representing opposite sides of the same chicken were utilized. The breast from one chicken was split into left and right halves. The left raw breast was weighed at 194 grams and then separated into its component parts-meat, skin, separable fat, and bone and each were weighed. The right raw breast weighed 192 grams. It was then roasted, weighed after cooking at 134 grams, and finally separated into meat, skin, and bone. The right hand column contains the cooked weights adjusted by the ratio of the raw weights of the left and right breast to eliminate discrepancies due to cutting. The cooking yields are calculated using the data in the left and right hand columns. The cooked yield of the entire breast, that is the part with bone, equals 70%. The yield of cooked meat from raw meat is the weight of cooked meat, 98 grams, divided by the weight of raw meat, 132 grams, and equals 74%. The yield of the total edible portion of the breast is the sum of the weights of cooked meat and skin divided by the sum of the weights of raw meat, skin, and separable fat and equals 72%. The different tissues, including bone, lose different amounts of weight on cooking.

Cooking yield data are used in developing retention factors. The preferred method for calculating nutrient retentions is based on measuring the proportion of nutrient remaining in a cooked food in relation to the amount of that nutrient originally present in a given weight of the food before cooking. Shown is an example of changes in proximate components upon roasting of the chicken breast meat and skin sample used for the previous cooking yield calculations.

	Raw	Roasted	% Retention
Weight g	155	112	
Water %	70.01	62.89	64.9
Protein %	22.06	28.12	92.1
Fat %	7.28	7.45	73.9
Ash %	.99	.99	72.3

	Raw (g/100g)		Retention	=	Cooked (g/72g)	Loss or gain (g)
Water	70.01	x	.649	=	45.44	-24.57
Protein	22.06	x	.921	=	20.32	- 1.74
Fat	7.28	x	.739	=	5.38	- 1.90
Ash	.99	x	.723	=	.72	- .27
Sum	100.34				71.86	28.48

The raw weight of 155 grams is the sum of the weights of raw meat, skin, and separable fat. The roasted weight of 112 grams is the sum of the weights of cooked meat and skin. The weight of 112 grams divided by the weight of 155 grams represents the cooking yield and is equivalent to 72%. Below the weight values are listed the analytical values for the raw and roasted tissue. The percent retention for water, for example, is the amount of water in the cooked breast with skin, 62.89%, multiplied by the 72% yield factor and divided by the percent water in the raw breast with skin, 70.01%. Nutrient retentions shown are about 65% for water, 92% for protein, 74% for fat, and 72% for ash. The proximate data for the raw sample are repeated in the first column on the lower set of lines. These values, when multiplied by the retention factors, yield the amounts of the four components in 72 grams of cooked product. The differences between these numbers and the values in the left hand column represent the changes in nutrients upon cooking. A 100-gram sample of raw chicken breast with skin would lose about 25 grams of water, 1.7 grams of protein into the drippings, 1.9 grams of fat, and 0.27 grams of ash to drippings. The solids lost on cooking, that is protein, fat, and ash, represent 14% of the total weight loss of 28 grams. Nutrient retentions calculated using data on a dry-weight basis are generally higher than those calculated using the actual yield or weights of raw and cooked foods because weight loss is presumed to be due to water loss alone. Such is not the case for meat products which lose fat to drippings.

The following examples of procedures for calculating nutrient values are based on recipes. An important aspect of any recipe calculation is the determination of ingredient proportions. Below is an example of development of a formulation based on a partial nutrient profile. The food item is a chocolate coated ice cream bar.

Analytical Data		Ingredient	NDB No.
Pro	3.4%	1 Vanilla ice cream	01061
Fat	22.3%	2 Coconut oil	04047
Carb	25.7%	3 Sugar	92300
Iron	0.2 mg	4 Cocoa, high-fat Dutch	77820
Vit A	308.0 IU		

Ingredient (%)	
1	79.0
2	13.5
3	6.4
4	1.1

Calculated Values	
Pro	3.0%
Fat	22.3%
Carb	25.7%
Iron	0.2 mg
Vit A	322.5 IU

Analytical data are shown for protein, fat, carbohydrate, iron, and vitamin A. Ingredients listed on the label for the product are those for ice cream and the chocolate coating. Coating ingredients in decreasing order are coconut oil, sugar, and high-fat Dutch cocoa. Each ingredient corresponds to an item contained in USDA's Nutrient Data Base for Standard Reference. The Nutrient Data Bank (NDB) numbers are shown for the item contained in the Standard

Reference data file. Using as input the analytical data shown for the five nutrients listed and the NDB numbers for the ingredients in decreasing order, the percentages of the ingredients are determined using a computer program which adjusts proportions of ingredients until calculated nutrient values converge as closely as possible with the analytical values. Use of the percentages of ingredients shown in the lower left hand column yields the calculated values shown on the right.

If the proportions of cooked ingredients in a mixed dish are known, nutrient values may be calculated for a mixture using data for cooked foods, thereby eliminating the need to apply retention and yield factors. Alternatively, recipe calculations may be based on raw ingredients. An example of a recipe calculation based on raw ingredients follows:

Egg dessert, baked — Food code:321-2010
Yield: 90.0 — Fat change: 0.0
Moisture change: - 10.0 — Fat ID:

NDB No.	Ingredient	Retention	Measure	Grams
92300	Sugar	0	3 c	600
01125	Egg yolk	101	16	272
01123	Whole egg	101	2	100
97010	Water	0	3/4 c	177

The format for the input information is that used with the HNIS recipe linking program system for automatic calculation which was described at this conference. The item for which nutrient values are to be calculated is a baked egg dessert. The food code for this item as contained in the Nutrient Data Base for Food Intake Surveys is recorded. The cooked yield of the product is 90%. Fat change is indicated as zero, fat being neither absorbed nor lost during preparation. Cooking losses for this item are due to evaporation. Moisture change is indicated as minus 10%. The NDB numbers for the four ingredients of the item are listed in the left hand column. A retention code is entered for two of the ingredients in this example. The code 101 is used to access retention factors for vitamins and minerals in egg cooked by a dry-heat method. These factors are contained in a separate data file. This particular recipe is the standard cook book type — 3 cups of sugar, 16 egg yolks, 2 whole eggs, and 3/4 cup of water. Both the measures and their corresponding gram weights are part of the entry information. The calculation steps are outlined below.

Ingredient	Weight (g)	Energy (Cal)	Water (g)	Fat (g)	B_{12} (mcg)
Sugar	600	2310	3.0	0	0
Egg yolk	272	1004	132.6	89.5	10.34 x .8
Whole egg	100	158	74.6	11.2	1.55 x .8
Water	177	0	177.0	0	0
Subtotals:	1149	3472	387.2	100.7	9.51
Mois/fat ch:	-115	0	-114.9	0	0
Yield:	1034	3472	272.3	100.7	9.51
Per 100 grams:	100	336	26.3	9.7	0.92

Nutrient values for energy, water, fat, and B_{12} on a 100-gram basis are accessed from the Standard Reference Data file. The values are converted to the amounts in the specific weight of each ingredient. Retention factors of 0.8 (80%) are applied to the B_{12} data. The values are summed as shown on the subtotal line. The next line shows the changes in weight and water content during preparation. Ten percent of the weight is deducted from both the sum of ingredient weights and the grams of water. The yield line contains amounts of nutrients in 1,034 grams of food. These values are converted to the per-100-gram basis. If these calculations were done with the recipe linking program system, the final values on the 100-gram basis would automatically be written to a data file.

A final example of a recipe calculation shows how changes in fat, in this case a loss of fat, may be taken into account. The item to be calculated is stewed turkey.

Turkey, stewed — Food Code: 242-0140
Yield: 75.6 — Fat change: -3.3
Moisture change: -21.1 — Fat ID: 04575

Ingredient	Weight (g)	Energy (Cal)	Water (g)	Fat (g)	Cholest (mg)
Turkey, flesh & skin, raw	100.0	160	70.4	8.0	68
Mois/fat change:	-24.4	-30	-21.1	-3.3	-3
Yield:	75.6	130	49.3	4.7	65
Per 100 grams:	100.0	172	65.2	6.2	86

The starting ingredient is the raw item. Fat loss is shown as 3.3%; moisture loss is 21.1%. The yield of product is 100 minus the sum of the fat and moisture loss and equals 75.6%. Fat identification, on the upper right, is the NDB number 04575 which is the code for turkey fat. Nutrient values for energy, water, fat, and cholesterol for 100 grams of raw turkey flesh and skin are listed on the first lower line. The second line shows adjustments made to the weight and nutrients to account for water and fat losses. The cooking loss of 24.4 grams is subtracted from the raw weight. Thirty calories from fat are deducted. Grams of water and fat lost are subtracted. The 3.3 grams of turkey fat contained 3 mg of cholesterol, which is also deducted. The amounts of the nutrients contained in 75.6 grams of cooked food are shown on the line for yield. These figures are converted to the 100-gram basis by division by 0.756.

The typical calculation procedures shown can also be used to impute nutrient values. Imputed values bear some degree of uncertainty. They may be calculated from label declarations for fortified foods and by using closely related but different foods or recipes with uncertain ingredient proportions. In some cases it may be necessary to use estimated retention factors, cooking yields, or physical composition. In order to impute values for certain nutrients it may be necessary to extrapolate data from a parent food to processed forms of the food when little information on the processing effects is available. Such factors affect the validity of the final results.

CLINICAL APPLICATION

COMPUTERIZED NUTRITIONAL ANALYSIS SYSTEM FOR BURN CARE

Jan Wyatt M.A. R.D.
Bothin Burn Center
St. Francis Memorial Hospital
San Francisco, California

To aid the dietitian and the staff of a burn center in monitoring and evaluating the patient's nutritional care, the Bothin Burn Center at St. Francis Memorial Hospital (San Francisco, California) has developed a nutrition analysis program for burn care which utilizes a microcomputer system. This microcomputer system is designed to help improve patient care by increasing the efficiency and accuracy of assessing the daily nutritional status of each burn patient. The data generated by this system allows the burn staff to evaluate the success of the care plan in providing the necessary nutritional support for the thermally injured patient.

The Bothin Burn Center's goal for the computerized nutrition analysis system was to update and maintain a high level of accuracy. It is important to note that the program is directly dependent on the accuracy of data entry as well as the accuracy of the data itself. Although the Burn Center attempts to maintain these standards, there will always be varying factors that will influence the outcome of the data.

So what will the Bothin Burn Center's Nutritional Analysis System do? The system has been programmed:

1. To determine and update the patient's nutritional requirements using various published formulae.

2. To analyze daily nutrient content of all oral, enteral and parenteral feedings used at St. Francis Memorial Hospital. This is accomplished by weighing each food item consumed by the patient. The computer substracts this weight from the pre-established weight, and analyzes it for all 73 nutrients or for a selective group of nutrients depending on the need. The expendable data base allows the dietitian the ability to add new recipes or food items, to make changes in the menu, and to analyze food from home or fast food restaurants.

3. To identify and report a lack or excess of essential nutrients in the patient's diet. Trends in this data stimulate alterations in the diet and/or method of delivery.

4. To provide long term patient histories, which should increase the burn staff's knowledge of metabdism and nutritional support of the burn patient.

The Bothin Nutrition Analysis System is easy to use and requires no programming experience. All burn team members share access to the system, enabling them to have a more active role in the patient's nutritional care.

Application of the nutrition analysis system is not solely limited burn care. Although St. Francis Memorial Hospital has not tested this system in other health care areas, this system should prove to be an asset in areas such as renal dialysis and oncology, where a more detailed nutritional analysis may be required in aiding patient recovery.

FOODSERVICE APPLICATIONS

Marianne Wilcox, M.S., R.D.
Menu and Food Systems Manager, Saga Corporation

Many of the clinical dietetian's functions can effectively use the personal computer. Particularly patient nutrient intake studies, menu planning for the individual patient and menu planning for the general diet category.

Two critical elements of these functions, menus and recipes, link the clinical and foodservice aspects of the Dietetics Department. Menu planning must be a team effort so that concerns of the clinical dietitian as well as the production manager are met. The menu plan, i.e., what the patient will consume, is dramatically influenced by the recipe used to prepare the food.

Therefore, it is essential that the computer software supports both the clinical and foodservice aspects of the Dietetics Department. The software must include four files: 1) food item or ingredient file, 2) recipe file, 3) menu file, all reflecting what happens at your specific account -- not some theoretical account. The fourth is a nutrient data base. All this is necessary if the nutrient analysis is to accurately reflect what the patient consumes.

The recipe data base and software for nutrient analysis must be able to handle purchasing an item in one state but using it in a different form and food items consumed in both raw or cooked states, depending on the recipe. Also, the software must be able to exclude the inedible portion of the ingredient. These features are necessary to account for varying stages of preparation and consumption for the accuracy of the nutrient analysis.

Once the software and data are in place, actual food production is the next step. Recipes must be scaled or adjusted and produced correctly. An ingredient room ensures that ingredients are measured accurately. Proper portioning based on the standard or specified portion must be followed because the nutrient analysis is based on the standard recipe and the specified portion size.

The USDA Nutrient Data Base for Standard Reference is the basis for Handbook 8 revisions and for many of the nutrient data bases available today. A goal for this discussion group will be to develop recommendations for improvements to this data base. To list a few problem areas:

1. Lack of nutrient data for different stages of consumption, i.e., drained vs undrained fruits.
2. Cooked data only given for all components of the recipe including oil and breading or batter, i.e., fish and shellfish.
3. Lack of data for manufactured products, i.e., croissants, crepes, wonton wrappers, etc.

NEW APPROACHES TO THE NUTRITIONAL ASSESSMENT OF POPULATION DIETARY DATA

George H. Beaton
Department of Nutritional Sciences, Faculty of Medicine
University of Toronto, Toronto, Canada

Background: A New National Academy of Sciences Report

For the last two years, a National Academy of Sciences Committee and its Subcommittee have been addressing the issues surrounding the design and interpretation of national surveys. The first report of the Committee was released last year (1). It addressed the broad issues of the need for both the NFCS and NHANES surveys and issues of improvement of compatability of the data bases arising from these two surveys. The second report, arising from the Subcommittee, is due to be released in July, 1985 (2). This report, "Food Consumption Surveys: Criteria For Assessing Dietary Adequacy", deals with matters that are very germane to this conference and should be of considerable interest to the audience of the present Conference. The present paper can do no more than provide a sampler of some of the material to be found in that report. It is hoped that this sampler will be enough to encourage the reader to acquire a copy of the full report and to consider its content in detail.

Four years ago, on the occasion of the opening of the USDA/ARS, Western Human Nutrition Research Center, I had the opportunity to present a paper describing the application of a probability approach to the assessment of observed dietary intake. This approach, which has since been presented elsewhere (3-6), accepts that human nutrient requirements vary between individuals. It emphasises, therefore, that there can be no single numerical criterion of adequacy applicable to all individuals in a class (e.g. adult men or women). Rather, the approach recognizes that as observed intake moves across the distribution of requirements, the likelihood or probability of inadequacy changes. That is, if observed intake lies in the upper tail of the requirement distribution, then there are very few individuals with requirements above that level of intake. It follows that the risk or probability that the intake is inadequate for a randomly selected individual is very low (7). By knowing the areas under the requirement distribution to the right and left of the observed intake, it is possible to make a probability statement with regard to adequacy of the particular intake. By making such assessments for each intake observed in a population study and summing the individual probabilities, one can generate an estimate of the prevalence of inadequate intakes in the population - one can estimate <u>how many</u> individuals are likely to have inadequate intakes but one cannot identify <u>which</u> individuals have inadequate intakes. The recent FAO/WHO/UNU report on Energy and Protein Requirements (7), released this summer, elaborates a number of statistical issues relating to the probability assessment of observed intake. This is the basic approach that has been developed in the NAS report. That report provides a greatly extended examination of issues and questions surrounding the application of the probability approach to population data.

During the past five or so years, a second major issue rose to considerable prominence in the nutrition community. That was the matter of day-to-day variation in intake and the implications for "reliability" of dietary data (8-20). It was recognized that for many applications, it was an estimate of the individual's usual intake, not intake on a particular day or in a particular week, that was needed. Commonly used quantitative dietary methods give a poor estimate of the usual intake of an individual. In data analyses, the error term in the estimate of usual intake can have profound effect on regression coefficients or correlation coefficients, potentially giving rise to false negative conclusions (8,14,16,17,19). There are issues also in relation to the interpretation of nutritional adequacy of population dietary data. The inclusion of day-to-day variation in population data leads to an overestimation of the variance of intakes in the population - a spuriously high proportion of very high and very low intakes (21). If this effect is not eliminated either by choice of method or by adjustment during data analysis, there will be an error in the estimation of the prevalence of inadequate (or excessive) intakes in the population.

The Subcommittee considered this issue very carefully - indeed it was one of the first issues on the agenda. It was recognized that if the data set included a sufficient number of replicates, an analysis of variance (ANOVA) could be run and an estimate of the inter- and intra-individual variances could be obtained. With this information in hand, it would be possible to derive an estimate of the <u>distribution</u> of usual intakes in the population group - to adjust the observed distribution. The National Food Consumption Survey (NFCS) contains more than enough replications of intakes. From that data base it is quite feasible to generate adjusted distributions describing usual intakes. Unfortunately, the NHANES data bases of the past include no replications; no internal adjustments to those data can be made and they cannot be used in the manner described in this paper. It may be that future HANES studies will make provision for obtaining replicate observations for an appropriate sample of the population. The Subcommittee was able to work with data sets from the NFCS thanks to the work of staff of USDA who ran ANOVA's on logarithmically converted data required for the adjustment of the distributions and who also formatted the data sets into sequential files with 200 intervals each containing an equal number of individuals. These data were made available in log form, in untransformed form for all single days (as if they were independent) and for the 3 day means for young adult men and for young adult women. With these reduced data sets representing some 2400 women and 1700 men, it was quite feasible to mount the data in an Apple computer and to then write programs to conduct the analyses presented in the report and sampled in this paper. Thus, it was possible to demonstrate that the approaches both to adjustment of distribution and application of a probability approach were entirely feasible of implementation with population data.

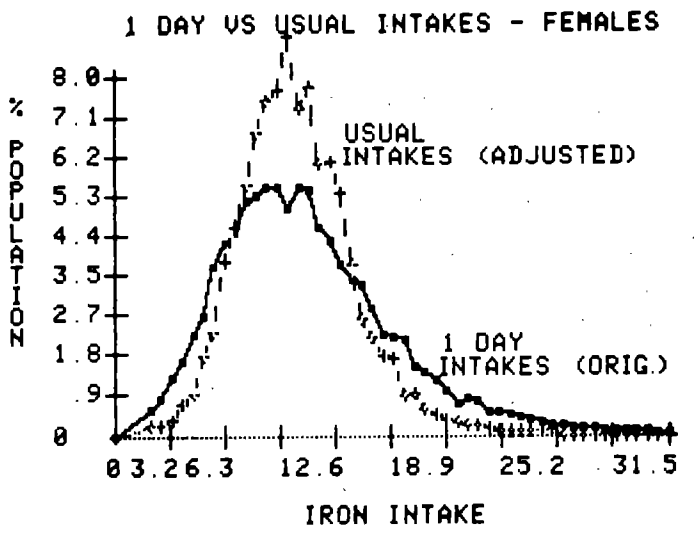
The report, however, does not stop at presenting the approach. Rather, the Subcommittee attempted to address all of the possible sources of error in the estimation of prevalence of inadequate intakes

by this method and then, through a process of sensitivity testing and statistical modelling, it assessed the potential impact of these individual sources of error. Availability of the data set on a microcomputer readily permitted the introduction of error terms in either requirement or intake distributions, shifting the locations of the distributions, and conducting other manipulations to see their impact on the prevalence estimate - to assess the sensitivity of the estimate to these variations and hence to ascertain which assumptions and inputs were really critical to the assessment.

The new NAS report, then, also makes major contribution to an understanding of "error terms" in population data sets and provides an approach to the assessment of these error terms - which have major import and which are of little significance for a defined use of the data set. The presentation of this approach to analysis of error terms - a "sensitivity analysis approach" - may be a more important contribution to our field than the specific proposals concerning the nutritional interpretaton of NFCS intake data. It is an approach that is applicable to the examination of other uses of dietary data.

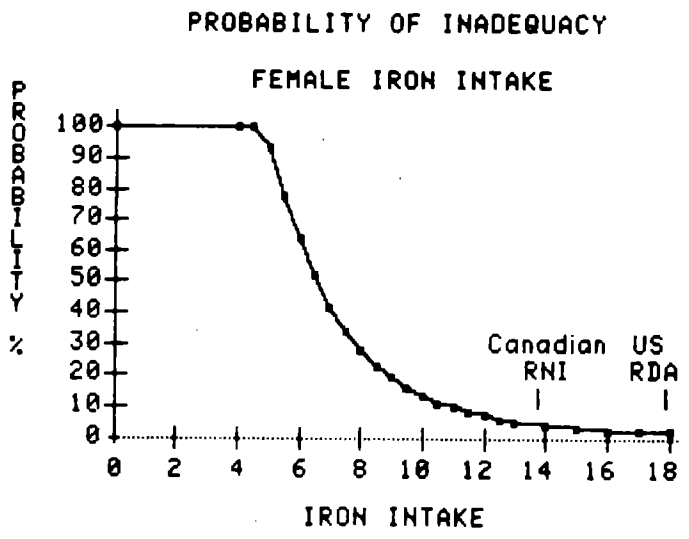
The Probability Approach to Assessment of Population Intake Data

The composite approach can be seen by examining one example - assessing the apparent adequacy of reported iron intakes among adult women. (Note that the NFCS data do not include intake from supplements. Thus total intake has been underestimated and prevalence of inadequate intake has been overestimated in the following examples.) Fig. 1 illustrates the distribution of single day intakes among the adult women and the distribution of "usual" intakes among these same women. To make this adjustment, the distributions were first normalized by logarithmic transformation (a more correct approach would be to use Box-Cox transformation appproaches (22)), the ANOVA was performed, the standard deviation of interindividual (between subject) variation was obtained and then mean intakes for each of the 200 intervals used in the reduced data sets were shifted toward the mean of the whole distribution by applying the following equation:



$$\text{Adjusted intake} = (\text{Mean Intake} - \text{Observed Intake}) \times SD_{inter}/SD_{observed}$$

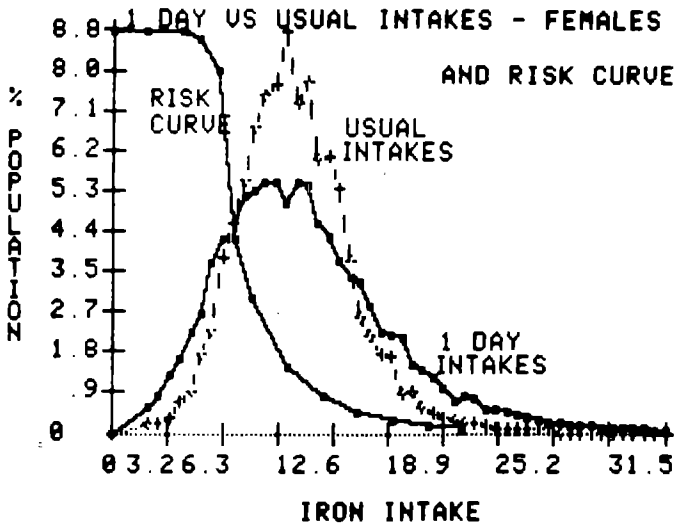
The exponentials of the adjusted data were then computed to yield the new distribution shown in the figure. This procedure effectively removes the day-to-day variability of recorded intake. (As will be seen later, this variability includes true variation in intake by the same individual and random error terms in the data collection methodology.) By using a nonparametric approach it was possible to avoid an assumption of perfect fit of the converted distribution to the normal curve.



The next step requires a distribution of requirements. Fig. 2 portrays the cumulative distribution of iron requirements among menstruating women. This curve is based on measured menstrual losses plus an allowance for basal losses and assumes a constant upper limit for dietary iron bioavailability (23). When plotted in the manner shown, this curve portrays the probability that any given level of intake is inadequate to meet actual needs.

Fortunately the distribution of menstrual iron losses approximates quite closely the log normal distribution. Following logarithmic conversion of data, it is quite possible to read from a table, or compute with available algorithms, the area under the normal distribution to the right of a given intake - the probability of inadequacy. This is what is portrayed in Fig. 2.

The final step involves application of this probability estimate to the adjusted distribution. This is portrayed graphically in Fig. 3. It was done with the computer for each of the 200 intervals of intake. Probabilities were then summed and an estimate of the population prevalence of inadequate intakes was derived.



The impact of this approach is vividly seen in Table 1 which compares assessments obtained by a common present approach with those obtained for the same population by the probablility approach.

Table 1. Comparison of Estimates of the Prevalence of Inadequate Iron Intakes - Menstruating Women

Approach to Assessment	Estimated Prevalence
Proportion with intake below the US RDA (18 mg/d)	98 %
Proportion with intake below the Canadian RNI (14 mg/d)	88 %
Predicted prevalence of inadequate intakes (probability approach)	23 %

Examination of Potential Sources of Error - Sensitivity Testing

The Subcommittee, in accepting this approach first satisified itself that the approach was statistically valid. It then satisfied itself that if indeed requirement were perfectly described and intake measurements were error free, the prevalence estimates derived in this manner would have acceptable confidence limits. Indeed, the limits are extremely tight - often the standard error of the prevalence estimate is in the order of ± 1-2%. This merely asserts that with large data sets, the ANOVA approach yields a statistically sound estimate of the interindividual distribution and that the error term in the final estimate of prevalence reflects this. After satisfying itself on these points the Subcommittee systematically identified and examined possible sources of error with regard to the potential magnitude of their effect. The sources of error considered are summarized graphically in Fig. 4 below.

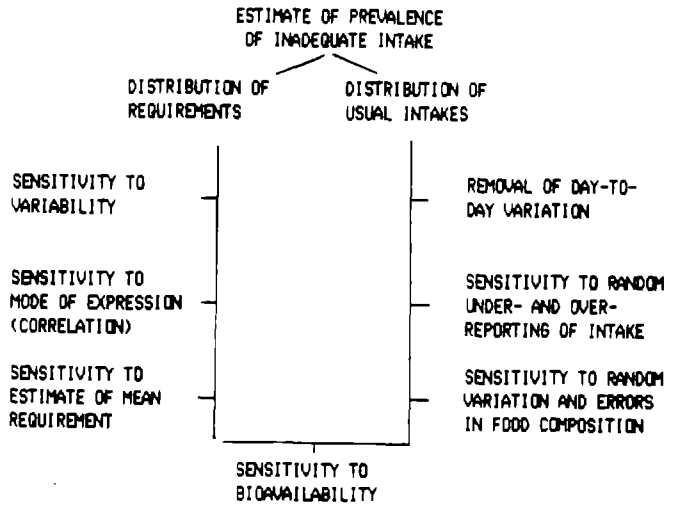


Fig. 4. A schematic portrayal of some potential sources of error

THE REQUIREMENT DISTRIBUTION

Sensitivity to Variability of Requirement

As portrayed earlier, the probability approach applied to individuals requires a knowledge of the distribution of requirements and a method of estimating the fractional areas under the distribution. It has been suggested by many authors that our present knowledge of the variability of human nutrient requirements among individuals within a relatively homogeneous class (e.g. young adult men or young adult women) is very limited. This criticism implies that a probability approach based on judgements about the distribution of requirements could be subject to major error. The NAS subcommittee examined this my making various assumptions about the magnitude of the variation of requirement and about the shape of the requirement distribution. For each of these assumptions, the prevalence of inadequate intakes was again estimated. Table 2 presents the results of two such sets of sensitivity trials.

Table 2. Sensitivity of the Prevalence Estimate to Variability of Nutrient Requirement

A. Protein in Young Adult Men

Mean Requirement g/d	Assumed CV of Requirement SD as % of Mean	Implied RDA* g/d	Prevalence Estimate** % of men
43	10	52	1.9
43	12.5	54	2.0
43	15	56	2.2
43	17.5	58	2.3
43	20	60	2.4

B. Ascorbic Acid in Young Adult Men

Mean Requirement mg/d	Assumed SD of Requirement mg/d	Implied RDA* mg/d	Prevalence Estimate** % of men
45	2	49	40.2
45	4	53	39.9
45	6	57	39.7
45	8	61	39.4
45	10	65	39.2

* Assuming the RDA had been set at mean + 2 SD to meet needs of all but 2.5% of individuals

** This is an estimate of the proportion of individuals expected to have intakes below their own requirements

Other analyses presented in the report examine the import of the shape of the requirement distribution. The main conclusion is that as long as there is reason to believe that requirements are distributed reasonably symmetrically about the mean the estimate of prevalence is relatively insensitive to the magnitude of the variability of requirement or to the shape of the distribution curve (normality is not a necessary assumption). Clearly these conditions are not met for iron requirements of menstruating women where the requirement distribution is known to be highly skewed. In the two examples presented above, the positions of the requirement distributions are quite different in relation to the intake distributions (see difference in prevalence estimates), nevertheless, the impact of variability of requirement is quite small in both. This was a surprising, but reassuring observation.

The analyses presented in Table 2 suggest also that it would be inappropriate to rely on the published RDA as the underpinning of a nutritional assessment of population data. What is needed is an estimate of the mean requirement and a knowledge of, or reasonable judgement about, the distribution of requirements (or a reasonable assurance that the distribution is not markedly skewed).

Sensitivity to Mode of Expression of Requirement - Correlation

A necessary assumption of the probability approach is that requirement and intake are not correlated (or that the correlation is known). It is necessary therefore to eliminate common variables that would lead to spurious correlation during analysis. Certain of these are obvious. The approach must be applied to reasonably homogeneous groups such as those conventionally described in dietary standards - specified age, sex and physiological state groupings. If young children and adults were included in the same analysis, it would be found that the children would have lower requirements and lower intakes than the adults simply because of major difference in body size; a spurious correlation between requirement and intake would be present in the data set snd the prevalence estimate would be wrong.

Other situations of correlation may be less obvious yet they too must be eliminated or analysed differently. A clear example is found in assessing thiamin intake in young adult men. Experimental evidence demonstrates that thiamin requirement per day is affected by the energy flux of the body and this in turn associates with energy intake (total food intake). Observational evidence clearly indicates that total thiamin intake associates with the level of total food intake. Thus, in population data a correlation between thiamin requirement per day and thiamin intake per day is to be expected. This can be eliminated if both requirement and intake are expressed in relation to the common variable, energy intake.

When these relationships are taken into account, the results are quite dramatic. Again using adjusted NFCS data, the prevalence of inadequate thiamin intakes is estimated to be about 37% when requirements and intakes are both expressed on a per day basis. However, when both are expressed on a per 1000 kcal basis, the estimate of prevalence of inadequate intakes falls to about 3.5%! (This latter estimate does not include provision for a "floor" of thiamin requirement below which need is said to be unrelated to energy flux. Such a floor could be built into the computations.) A moment's consideration will lead to the recognition that protein requirements and intakes should be expressed per kg per day, and that vitamin B_6 requirements and intakes should be expressed per gram protein intake per day, since these are recognized variables of requirement and associates of intake. The lesson that emerges is that for a correct assessment of intake, as many of the variables of requirement as possible should be incorporated.

For exactly the reasons described above, the probability approach cannot be applied to energy unless there is a direct estimate of the correlation between intake and requirement (7). There is <u>much</u> evidence to suggest that in free-living subjects consuming food without imposed restraint, energy intake and energy utilization (requirement) are strongly correlated. Application of the probability approach without taking this into account would be very misleading.

Sensitivity to Mean Requirement - Requirement for What?

Analyses of the type presented above served to demonstrate that the shape of the requirement distribution (unless skewed) has little impact on the prevalence estimate and that with care in mode of expression, the issue of correlation between intake and requirement can be avoided (except for energy). What then about the position of the requirement distribution. What is the impact of change in the estimate of mean or median requirement?

This can be portrayed in a particularly meaningful manner by considering the question "requirement for what?" The NAS Report places great emphasis on the need to address this question in any assessment of dietary data.

Consider ascorbic acid. Currently published requirement estimates in both Canada and United States base requirements on the intakes needed to maintain metabolic pools or metabolic turnover rates deemed to be desirable. For such a criterion of adequacy, mean requirements have been estimated at about 45 mg/d. In depletion studies, eye lessions were suggested at intakes of about 25 mg/d (not a clear estimate of a requirement level but used here as an example). The literature abounds with information to suggest that the average requirement for the prevention of clasical signs of scurvy is 10 mg/d or perhaps even lower. We then have three answers to the "requirement for what?" question. There is no reason why we cannot, and should not, provide three estimates of the prevalence of inadequate intakes:

Table 3. Prevalence of Inadequate Intakes of Ascorbic Acid in Men Defining Requirement on Differing Criteria of Nutriture

Criterion	Prevalence
A. inadequate to maintain metabolic pools	40%
B. inadequate to prevent eye lesions *	12%
C. inadequate to prevent scorbutic signs	2%

* a very uncertain criterion, used only for example

Similarly, for thiamin, expressed per 1000 kcal/d, the prevalence of intakes that would be inadequate to meet the criterion of maintenance of metabolic pools (and associated enzyme activities) was estimated to be about 3.5%. This falls to 0 when requirement is defined in terms of prevention of lesions suggestive of beriberi.

Clearly the prevalence estimate is sensitive to the estimate of the average requirement. It would be sensitive to any error (bias) in that estimate. Equally, or perhaps more, important, the prevalence estimate is very sensitive to the answer to the question "requirement for what?" In practical terms, the NAS Report opens the door to a multiple criteria assessment of observed dietary intake and points out that such an approach would add a great deal of information pertinent to

understanding the situation of the population and to the formulation of policy. It would at least begin to make sense of the expected relationship between dietary assessments and biochemical and clinical assessments: one should not expect agreement unless all of the assessments accept the same underlying definition of adequacy - the same criterion of adequate nutriture!

THE INTAKE DISTRIBUTION

Sensitivity to Day-to-day Variation of Apparent Intake

Figure 2, portraying the adjustment of iron intake in adult women, illustrated the impact of removing day to day variation from the data set. The procedure of adjusting the distribution was described previously. Only two comments need be made here. First, as will be developed below, this step has other important advantages since it eliminates not only true variability in intake but also greatly reduces the impact of a number of sources or random error in the intake estimate. Second, the impact of day-to-day variation depends upon the nutrient (and probably also the population group and dietary methodology choice). The NAS report summarizes available information on the magnitude of this variation. Table 4 presents only a limited example drawn from one study (8,9). The point to note is that the ratio of variances changes with the nutrient. This is a measure of the relative error term in the estimate of usual intakes.

Table 4. An Example of the Ratio of Variances in Dietary Data

Nutrient (units)	Ratio of Intraindividual:Interindividual Variances	
	Adult Males	Adult Females
Energy (kcal/d)	1.1	1.4
Protein (g/d)	1.5	1.5
Carbohydrate (g/d)	1.6	1.4
Starch (g/d)	2.9	1.4
Fat (g/d)	1.2	1.6
SFA (g/d)	1.1	1.4
PUFA (g/d)	2.8	4.0
Cholesterol (mg/d)	3.4	4.3
Vitamin A (IU/d)	- *	24.3 *
Vitamin C (mg/d)	3.5	2.0
Thiamin (mg/d)	2.5	4.4
Riboflavin (mg/d)	2.4	2.2
Niacin (NE/d)	1.6	4.0
Calcium (mg/d)	2.2	0.9
Iron (mg/d)	1.7	2.5

* The data of Beaton suggest an extremely high day-to-day variation for vitamin A. Other studies cited in the NAS report suggest much lower ratios, in the order of 3 to 11.

The available data suggest that, for adjustement of observed intake distributions, the partitioning of variance should be estimated from within the study data set. That is, it would be inappropriate to use the estimates presented in Table 4 as a basis for adjusting another data set. The implication is that the survey design must provide for the collection of a statistically appropriate sample of replicate intake estimates. This does not mean collection of a reliable estimate of usual intake for each individual - an objective that would require many replicates for each individual. The objective is the development of a reliable estimate of the distribution of usual intakes. This is much less demanding. (For example, the 1971-74 NFCS contained many more replicates that would be needed solely for this purpose. They may have been needed for other applications of the data.)

Sensitivity to Random Under- and Over-Reporting of Intake

If an individual randomly under- and over-reports intake across days, the error will appear as a part of the day-to-day variation discussed above and will be removed by the procedure of adjustment of the distribution. Therefore, it will have little or no impact on the prevalence estimate.

In contrast, if some individuals consistently under-report and other individuals consistently over-report intake, this will appear as a part of the interindividual variation and will not be removed by the adjustment procedures. The NAS Report presents an examination of the impact of this type of reporting error. It examined it on the assumption that such individuals were randomly distributed across the population (see later comments on bias associated with misreporting by a defined group).

To gain some perspective, consider the type of data that we have all seen presented in dietary validation studies - the distribution of the magnitude of misreporting (differences in reported intake between a reference and test method). In Table 5, this is presented from simulation modelling in relation to assumed random (between individual) reporting error.

Table 5. Impact of Random Reporting Error on the Distribution of Deviations Between Test and Reference Dietary Methods

Assumed Magnitude of Random Error CV (% of Mean)	Expected Distribution of Deviations Between a Test and Reference Method (Observed Intake vs. True Intake) Proportion of Population Showing Deviation of At Least					
	30%	25%	20%	15%	10%	5%
5	2.7	3.4	4.2	5.2	6.4	8.2
10	5.3	6.7	8.4	10.4	12.8	16.5
15	7.8	10.1	12.6	15.5	19.2	24.7
20	10.5	13.5	16.8	20.7	25.6	32.9
25	13.1	16.8	21.0	25.9	32.0	41.1
30	15.8	20.2	25.2	31.1	38.4	49.4

The above table is useful in deciding the outer limits of a reasonable assumption about potential error terms incorporated in a population data base. From the literature comparing methods, it appears that a realistic worst case situation would be that portrayed by the incorporation of a 20% random reporting error term. Table 6 then presents the impact of various magnitudes of this error term on the estimated prevalence of inadequate intakes. It is done for two nutrients (one with high prevalence estimates and one with low prevalence estimates) and the impact of both adding and removing a random error term is examined through simulation analyses. As would be expected the effect is greater when the prevalence estimate is low; here the tail of the intake distribution is more critical and adding or removing variance has substantial impact on the tails of the distribution. When the prevalence estimate approaches 50%, no matter what happens to the tails of the intake distribution, a similar proportion of the population will be estimated to have inadequate intakes. Considering the suggested worst case situation, removal of the error term would change the prevalence estimate for protein from 2.1% to 0.8% and for ascorbic acid from 41.0% to 39.8%.

Table 6. Impact of Random Between Subject Reporting Error on the Estimate of the Prevalence of Inadequate Intakes

Assumed Random Error CV (% of Mean)	Impact of Adding Error		Impact of Removing Error	
	Protein	Vitamin C	Protein	Vitamin C
0	2.1% *	41.0%	2.1%	41.0%
5	2.3%	41.1%	2.0%	41.0%
10	2.8%	41.3%	1.5%	40.8%
15	3.7%	41.6%	0.8%	40.4%
20	4.9%	42.1%	0.2%	39.8%
25	6.5%	42.6%	0%	39.1%
30	8.3%	43.1%	- **	38.0%

* Values are apparent prevalences of inadequate intakes computed by probability approach.

** Cannot be computed. The "error term" being removed is greater than the interindividual variation in the data set, an impossible situation.

These analyses suggest that a reasonable level of random under- and over- estimation of intake should not consitute a major concern for this type of application. While it will affect prevalence estimates, it will not affect the policy-relevant assessment that apparent prevalence of inadequate intakes is quite high or is very low. The issue for the designer and analyst to bear in mind is the needed precision of the estimate of prevalence. In turn this will establish the acceptable magnitude of random error in the reporting of intake.

The situation is very different if there is reason to believe that the error is systematic. The obvious example is a bias toward

under- or over- reporting by the whole population. This can have major impact on the prevalence estimate just as did change in mean requirement estimates. It is clear, for example, that all prevalence estimates presented in this paper are too high for the simple reason that the NFCS data base does not include nutrient intake from pharmaceutical supplements. All that is being examined is nutrient intake estimated from food intake - a systematic underestimation of total nutrient intake. If it were established that certain segments of the population systematically misreported their intakes, and if these groups were a focus of interest, then obviously there would be errors in the estimates of prevalence of inadequate intakes among these groups!

The concern in design of surveys to be used in this type of analyses should focus upon the elimination of bias - consistent misreporting of intake - more than the elimination of random errors although these too should be minimized.

Sensitivity to the Variability of Food Composition

When one calculates nutrient intake using a food composition data base, at best one is assigning an average composition to the particular item of food consumed. We know that composition varies between foods falling into the same classification. Therefore, it follows that there is an unavoidable error in the estimation of nutrient intake even if the estimation of food intake is perfect! This error term does not appear within the data set. That is, we underestimate the true variability of nutrient intake because we assume the average composition rather than true composition. How large is this effect?

In Table 7 an estimate of the variability of composition of foods is presented. This has been generated from an examination of the new USDA food composition data base. For many foods, that data base provides an estimate of the standard error and the number of samples examined. From this the standard deviation of composition can be estimated. In turn, from this the coefficients of variation presented in Table 7 can be estimated. Empirically, it appears that the relative variability (CV) is higher at low concentrations of nutrient than at high concentrations. This may be a real biological phenomenon or it may be simply the result of methodologic errors (a constant error would be relatively greater at low concentration than at high concentration). For this reason, the table presents estimates of variability above and below the break point at which the CV appears to change for the particular nutrient. The variability terms appear to be very substantial. One might wonder whether, with this magnitude of "error", we can make any reliable interpretation of intake data.

Table 7. Apparent Variability of Food Composition
(Based on USDA Food Composition Data Base)

Nutrient Examined	Cut-off Point /100 g	Apparent Range of CV's Below Cut-off	Above Cut-off
Protein	2 g	5 - 50	5 - 15
Calcium	20 mg	5 - 50	5 - 15
Iron	1 mg	5 - 65	10 - 30
Magnesium	10 mg	5 - 50	10 - 30
Zinc	1 mg	5 - 65	10 - 30
Thiamin	.05 mg	5 - 50	10 - 30
Niacin	.5 mg	5 - 65	5 - 15
Vitamin C	7.5 mg	5 - 50	10 - 30
Folacin	20 mcg	5 - 65	10 - 30
Vitamin A	300 IU	5 - 65	10 - 30
Vitamin B_6	0.1 mg	5 - 65	10 - 30

To test the impact of this variation on a one day food intake estimate an actual record of intake was used. For each item of food a random member of the population of possible compositions for each nutrient was selected. To do this, either the data presented in the USDA tables were used as a descriptor of the population of compositions having mean and SD as described, or the population mean was taken as shown in the food composition table and the variability was inferred by selecting a random CV from the ranges shown in Table 7. In either case, a random member of a normal distribution having these characteristics was selected. The nutrient contents of the individual foods were then added to estimate the one day intake. The exercise was then repeated using new random members of the populations of possible compositions. This was repeated 1000 times and then the standard deviations and CV's of the computed one day nutrient intakes were computed. The results are presented in Table 8.

The table suggests that the error term in the one day intake estimate is appreciably less than might have been feared from an examination of the variability of composition of individual foods (Table 7). Statistical theory leads to the further conclusion that the more items of food included in the diet, the smaller will be the final relative error. This assumes that the individual error terms are indeed random and hence that while composition may be underestimated for one item, it may be overestimated for the next. If there is systematic error across all items, then of course the calculation would be misleading.

Table 8. Expected Variability of Composition of a One Day Diet Based on an Actual One Day Diet With 1000 Computations

Nutrient	Mean Content	Standard Deviation	Coefficient of Variation
Protein	105 g	6.2	5.9%
Calcium	1540 mg	80.8	5.2%
Iron	8.0 mg	1.2	14.9%
Magnesium	250 mg	15.7	6.3%
Zinc	11.6 mg	0.9	7.9%
Thiamin	2.1 mg	0.4	17.9%
Riboflavin	2.6 mg	0.2	7.9%
Niacin	15.9 mg	0.9	5.7%
Vitamin B_6	1.5 mg	0.13	9.4%
Vitamin C	153 mg	11.9	7.8%
Folacin	184 mcg	19.8	10.4%
Vitamin A	3800 IU	280	7.4%

Given the approach to adjustment of intake distributions to remove the impact of day-to-day variation, it should be recognized that part of the error remaining in Table 8 would be removed as a part of day-to-day variation. In the final distribution estimate, the contribution of variability of food composition would be quite small as long as there is reason to believe that it is random.

Conversely, it is also correct to recognize that the average compositions presented in the food composition data bases are not true averages - they are estimates of the true average composition. Each of these has an associated error represented in the USDA tables by the standard errors. In a manner analogous to that presented above, the standard error of the estimate of a one day intake may be calculated. Through statistical modelling the NAS Report examined the potential impact of this error on the prevalence estimate. For the nutrients examined, the standard error of the prevalence estimate ranged from $\pm$ 0.2 to $\pm$ 5%. This SE estimate included not only the error attributable to food composition but also the error terms assopciated with population sampling and intake reporting. These represent acceptably tight confidence intervals for the presumed purpose of the analyses.

In summary, then, the situation is not unlike that of random under- and over- reporting of intake. The random variation of food composition should not be a major concern <u>for this application</u>. Designers and users of data bases should be much more concerned about sources of systematic bias in the composition data than about random variation in food composition. An obvious example would be a methodologic error or an imputation error that consistently under- or over-estimates nutrient content for foods contributing a substantial part of the intake. Such an error is known to be present with regard to the iron content of meats used in the composition data base of the NFCS survey; it has been corrected in current USDA data bases and will appear in future food composition tables. A similar problem would arise if

soil composition affects food composition and if all or most foods consumed were grown in the same region (e.g. selenium in wheat). It has been suggested that brand allegiance and differences in manufacturing practices might have similar impact although this should be subjected to sensitivity testing before any firm conclusion is drawn.

When there are very few analyses for a particular food item, or when the tabulated value in a food composition table has been imputed, there is obvious advantage in improving the reliability of the estimate of average composition. In either case, a priority among foods might be based on their relative contribution to total intake of the nutrient in question. The larger the contribution, the more important is the reliability of the food composition estimate.

Considerations of this type are very germane to this meeting. Sensitivity testing of the type portrayed offers an approach to considering the import, or non import, of reducing food composition data bases by lumping more foods into single descriptors. Such actions would alter both the reliability of the mean composition and the variability of possible compositions. The impact on prevalence estimates can be examined through simulation exercises and then a strong base for judgement about the appropriateness of the modification of the data base can be reached. A similar approach would hold for consideration of possible benefit of increased specificity in the description of ingested foods (in effect, an increase in the scope of the composition data base). As will be mentioned below, such considerations must take into account the intended purpose of data collection. Although the analyses presented herein suggest that the variability of food composition is a limited issue for the estimation of population prevalence of inadequate intakes, it remains a major source of error in the estimation of the composition of a particular individual's actual intake.

Mention should be made also of the impact of this variation in composition in validation studies involving the comparison of estimated food intake and direct chemical analysis of duplicate meals. One should not expect full agreement. One should not expect a regression slope of 1 in such comparisons. The so-called "flat slope syndrome" in comparison of dietary methodology may be explained, in part by the simple fact that the random error terms differ between methods of estimating intake (see Tables 10 and 11 for examples of effects that would be expected). The moral is that it is necessary to consider the nature and sources of variation in data sets before one can interpret them or before one can interpret validation studies.

BIOAVAILABILITY

Sensitivity to Nutrient Bioavailability in Ingested Diets

Dietary nutrient bioavailability can be considered as a variable affecting nutrient requirement or as a variable affecting effective intake. Either way the import is similar. It is illustrated in Table 9 for iron in young adult women. In this model, the upper limit of iron absorption is varied and the prevalence of inadequate intakes is estimated. This, in effect, generates a family of requirement curves each of which can be applied to the adjusted distribution of total iron intake. The same results would have been obtained were requirement described as the absorbed iron need and intake were discounted by taking into account the bioavailability limit.

Table 9. Impact of Change in the Estimate of Bioavailability on Apparent Prevalence of Inadequate Iron Intakes in Young Adult Women

Assumed Upper Limit of Iron Absorption	Estimated Prevalence of Inadequate Intakes
14%	50%
16%	39%
18%	30%
20%	23%
22%	18%
24%	14%
25%	13%

The effect shown in Table 9 is fully expected and has been the reason for major attention to the estimation of a bioavailability figure applicable to the population under consideration, i.e. to the nature of the usual diet.

Of considerable interest is the question of variability of bioavailability across the diets consumed by individuals. It is known that for iron at least, bioavailability is heavily influenced by the mixture of foods consumed at the same time. As the typical diet of individuals varies, so also would vary the appropriate bioavailability figure to apply to that diet. Thus, as a minimum, it is to be expected that there is a component of variation associated with bioavailability that is not taken into account in Table 9. If this is random across individuals, then its impact should be no more than that described for random under- and over- reporting or random variation in food composition. In this circumstance, a single bioavailability figure might be applied to the whole population without serious detriment to the prevalence estimate.

This would not be the case if relative bioavailability changed systematically with the level of total iron intake. That is, it is at least possible that diets that characteristically contribute a high level of total iron intake also characteristically have a high (or low)

% bioavilability and diets characteristically contributing a low level of total iron intake exhibit a relative bioavailability in the opposite direction. At present no studies examining this type of effect have been identified. One is underway in Toronto at present, hopefully again with the cooperation of USDA which has a data set that can address this issue.

While this may seem to be an esoteric question, it is really an issue that has considerable practical importance. There has been a major urging for expansion of food composition data bases and increase in precision of identification of foods ingested so that bioavailability algorithms can be applied on a meal by meal basis. To implement these recommendations is likely to mean a major expenditure of funds for food composition analyses, for data collection, and for computer analysis of collected data. Would such expenditures be justified? Would the acquisition of such increased precision be likely to change the final answers in material ways? This is the type of issue that can be addressed through sensitivity analyses. Undoubtedly the answer will depend upon the research question being asked. Nevertheless, the conduct of preliminary sensitivity analyses is going to be much cheaper than simply plunging into a commitment to the creation of new food composition data bases.

The Probability Approach to Estimation of the Prevalence of Inadequate Intakes: Data Requirements and Limitations

The foregoing discussion should have provided the reader with some assurance that the approach presented in the NAS Report is indeed viable. Many of the traditional concerns about nutritional interpretation of survey data have been set aside as having minimal import for this application. Some of these were surprising. For example, it was very surprising to this author that variability of requirement was not a major issue in actual practice. It is not surprising in hindsight!

A critical need that remains is the availability of estimates of mean requirements of nutrients for the various age-sex groups and reasonable assurance that requirements are distributed in a generally symmetrical manner about these means (or an estimation of the actual distribution of requirements as in the case of iron in women). It is the position of the present author that for many if not most nutrients, we can generate reasonable estimates of these mean requirements and that the major issue we must address and declare is "requirement for what?" A recent FAO/WHO committee convened to address requirements for iron, folate, vitamin B_{12}, and vitamin A addressed this issue by defining two levels of requirement:

> Basal Requirement The intake of a nutrient required to prevent any clinically demonstrable impairment of function. Individuals meeting this requirement will be well and will maintain normal growth and reproduction. They would not be expected to have any reserve of the nutrient.
>
> Normative Storage Requirement This refers to the requirement of nutrient to maintain a reserve in body tissues. The reserve is seen as a supply of nutrient that can be mobilized without detectable impairment of function. The amount of such reserve deemed to be appropriate and desirable is a normative judgement.

I sincerely hope that this initiative of the FAO/WHO group is continued. It is consistent with the recommendation found in the NAS Report and will permit a multi-level assessment of observed nutrient intake in population data. Such assessments should be much more effective in facilitating consideration of the need or otherwise for interventions.

At the same time, it must be recognized that for some nutrients we really lack sufficient data to generate what we can all accept as reasonable estimates of average requirement. Calcium may be an example of this situation. It is extremely difficult, with existing methods, to estimate calcium requirement. This may or may not improve in the near future. Clearly if we cannot estimate requirements, or offer reasonable judgements about average requirement, we cannot apply the probability approach with confidence - nor can we apply any other assessment approach except one that gives assurance that all intakes above a stipulated level are adequate (i.e. a cut-point set above the highest suspected requirement). Thus, for example, one might suggest that if all intakes in a population fell above the RDA for a nutrient, there is no problem of inadequacy, one cannot say that if some intakes fall below the RDA there is a definable prevalence of inadequacy. On this rationale, it might be suggested that for some nutrients, and for public health purposes, the absence of requirement estimates is unimportant - intakes are known to be adequate thus no assessment is needed.

It is emphasized that this discussion focuses upon information about requirements. The published RDA's are not estimates of requirements although the text of the RDA reports may provide information about the underlying distribution of requirements. This is not a criticism of theh RDA report. Up until now it has not been a perceived mandate to the RDA committees to define average requirements or even to answer the question "requirement for what?" This may change in the future.

The NAS Report does emphasize, as should be apparent from the present paper, that there is no scientific justification for use of the RDA's, or some fixed proportion of the RDA"s, as a criterion of adequacy in assessing a population. There is no way in which RDA's or even a full knowledge of the requirement distribution can be used to classify the intake of a particular individual as either adequate or inadequate

unless that intake falls outside of the range of requirements. At most a probability statement can be made.

The second major need for application of the probability approach, or any other approach to assessment of intake, is an unbiased estimate of the distribution of intakes. At present this may represent more a judged feeling of confidence, or non-confidence, about dietary data being collected than actual proof. If and when we detect bias in food intake data or in food composition data, the persons involved take steps to eliminate the source of bias. The judgement question arises from the fact that one cannot detect the absence of bias - one can only respond to the demonstration of bias.

Methodologic efforts must continue if we are to improve the estimation of nutrient intake; a major focus must be the detection and control of potential sources of bias. This is probably more important (for the present purpose) than eliminating all sources of variation although the two efforts will often move together.

A Warning: Always Remember the Purpose of Data Collection

The foregoing discussion and sensitivity testing presentations have addressed a defined purpose of data collection and analysis - the estimation of the prevalence of inadequate intakes in population data. For this purpose, a number of sources of variance, a number of "error" terms, have been demonstrated to have very limited impact.

This is not necessarily the case for other applications, other purposes of data collection.

It is critical that in the design of surveys, the ultimate purpose be borne in mind and the proposed analytical strategies be considered. It is only in this way that sound decisions can be made about which error terms must be reduced and which are acceptable.

To illustrate the importance of considering purpose and analytical strategy, let us consider a hypothetical example in which one wished to examine the relationship between usual ascorbic acid intake (independent variable) and white cell ascorbic acid (dependent variable). An analytic strategy might be the application of either regression analysis or correlation analysis. Several authors have pointed out that an error term in the independent variable will bias the regression coefficient toward 0 and error terms in either variable will attenuate the correlation coefficient. Tables 10 and 11 show the magnitude of these effects. (See reference 8 for equations to derive this table.)

Table 10. Bias of the Regression Coefficient Associated With Random Error in the Variables

Variance Ratio in Dependent Variable (Y)	Variance Ratio - Independent Variable (X) *							
	0	0.4	0.8	1.2	1.6	2.0	2.4	2.8
0	1.0 **	0.71	0.56	0.46	0.39	0.33	0.29	0.26
0.4	1.0	0.71	0.56	0.46	0.39	0.33	0.29	0.26
0.8	1.0	0.71	0.56	0.46	0.39	0.33	0.29	0.26
1.2	1.0	0.71	0.56	0.46	0.39	0.33	0.29	0.26
1.6	1.0	0.71	0.56	0.46	0.39	0.33	0.29	0.26
2.0	1.0	0.71	0.56	0.46	0.39	0.33	0.29	0.26
2.4	1.0	0.71	0.56	0.46	0.39	0.33	0.29	0.26
2.8	1.0	0.71	0.56	0.46	0.39	0.33	0.29	0.26

* Ratio of intraindividual variance/interindividual variance (See Table 4 for examples)

** The observed regression slope would be the true slope multiplied by the factor shown

Table 11. Attenuation of Correlation Coefficient by Random Error Terms in the Independent and Dependent Variables

Variance Ratio in Dependent Variable (Y)	Variance Ratio - Independent Variable (X) *							
	0	0.4	0.8	1.2	1.6	2.0	2.4	2.8
0	1.0 *	0.85	0.75	0.67	0.62	0.58	0.54	0.51
0.4	0.85	0.71	0.63	0.57	0.52	0.49	0.46	0.43
0.8	0.75	0.63	0.56	0.50	0.46	0.43	0.40	0.38
1.2	0.67	0.57	0.50	0.46	0.42	0.39	0.37	0.35
1.6	0.62	0.52	0.46	0.42	0.39	0.36	0.34	0.32
2.0	0.58	0.49	0.43	0.39	0.36	0.33	0.31	0.30
2.4	0.54	0.46	0.40	0.37	0.34	0.31	0.29	0.28
2.8	0.51	0.43	0.38	0.35	0.32	0.30	0.28	0.26

* Observed correlation coefficient would be the true correlation coefficient multiplied by factor shown

In the example chosen, the variance ratio for dietary ascorbic acid with one day data would be about 2.4 (a generalized estimate from Table 4). If one examined the relationship between dietary ascorbic acid and white cell ascorbate using one day data, the observed slope would be only 29% of the true slope; unless the relationship were very strong, this might not achieve statistical significance. As one increases the number of days of observation, the intraindividual variance would be decreased while the interindividual component would not change. With 3 days of data for each individual, the variance ratio would fall to 0.8 and the observed slope would rise to 56% of the true slope; after 6 days the variance ratio would be 0.4 and the observed slope would be about 70% of the true slope. Liu et al (16) calculated that if one wished to demonstrate statistical significance for the relationship between dietary cholesterol and serum cholesterol about three weeks of dietary data would be needed. Analogous calculations can be made for the correlation coefficient but here the error term in the dependent variable also contributes to the attenuation and would have to be taken into account. There are other statistical approaches, such as reversing the dependent and independent variables that can minimize some of these effects. Nevertheless, the point should be clear. In statistical analyses that use the individual's intake as a variable, the error terms assume a much greater importance than they do in the population assessments presented in this paper.

Actually the impact estimates presented above are conservative. Only measured day-to-day variability is taken into account. To the error term of the independent variable would have to be added the variability attributed to food composition and perhaps also the variability attributed to over- and under- reporting. As discussed previously, some of these error components would be factored out with increasing numbers of days while some (those attributed to interindividual variance in previous discussions) would remain as an error term in the independent variable.

In this situation, a viable alternative might be to select a dietary methodology that estimates usual intake rather than intake on a particular day - a food frequency or dietary history approach. This should minimize the error terms discussed above but these qualitative or semi-quantitative methods will tend to introduce imprecision in the estimate of intake (24).

The moral to be learned from this discussion is that there is no perfect dietary methodology. However, there are preferred methodologies for specified purposes. The secret then is to consider the purpose and analytical strategy before selecting the dietary methodology and conducting the survey or other dietary study.

Acknowledgement

The material presented in this paper, prior to the final section, is based on simulation analyses performed during the deliberations of a National Academy of Sciences Subcommittee. The author is indebted to the US Department of Agriculture for providing NFCS data in a format that could be used on a microcomputer and for conducting analyses of variance on the original data. He acknowledges very strongly the input of all members of the Subcommittee in the evolution of concepts and approaches. A more complete discussion of these analyses will be found in the report of that Subcommittee: "Food Consumption Surveys: Criteria for Assessing Dietary Adequacy", National Academy of Sciences, 1985.

The members of the Subcommittee were:

L.J. Filer, Jr. (Chairman)	R. Havlik
G.H. Beaton	D.M. Hegsted
J.J. Feldman	K.K. Stewart
H.A. Guthrie	H. Smicklas-Wright
J.P. Habicht	A.A. Tsiatis

Staff Officer: Dr. V.H. Laukaran.

The members of the parent Cooordinating Committee on Evaluation of Food Consumption Surveys were:

J.P. Habicht (Chairman)	H.A. Guthrie
G.H. Beaton (Vice-Chairman)	R. Havlik
P. Pinstrup-Anderson	D.M. Hegsted
J.J. Feldman	H.D. Hurt
O.R. Fennema	K.K. Stewart
L.J. Filer	H. Smiciklas-Wright
P. Greenwald	

Staff Officer: Dr. S. Crocco, followed by Dr. V.H. Laukaran

While acknowledging the contribution of all of the above persons, the present author accepts full responsibility for interpretations presented in the present paper. Readers are urged to examine the NAS Report for a full discussion of Subcommittee interpretations and recommendations.

References Cited

1. National Academy of Sciences. National Survey Data on Food Consumption: Uses and Recommendations. Coordinating Committee on Evaluation of Food Consumption Surveys. Food and Nutrition Board. National Academy Press, Washington, D.C. 1984.

2. National Academy of Sciences. Food Consumption Surveys: Criteria for Assessing Dietary Adequacy. Subcommittee on Criteria for Dietary Evaluation. Coordinating Committee on Evaluation of Food Consumption Surveys. Food and Nutrition Board. National Academy Press, Washington, D.C., in press. 1985.

3. Anderson, G.H., R.D. Petersen, and G.H. Beaton. Estimating nutrient deficiencies in a population from dietary records: The use of probability analyses. Nutr. Res. 2:409-415. 1982

4. Beaton, G.H. The use of nutritional requirements and allowances. Pp 356-363 in P.L. White and N. Selvey, eds. Proceeding of the Western Hemisphere Nutrition Congress III, Futura Publishing, Mount Kisko, New York.

5. Beaton, G.H. Nutritional assessment of observed dietary intake: An interpretation of recent requirement reports. In H.H. Draper ed. Recent Advances in Nutritional Sciences, in press, 1985.

6. Beaton, G.H. Uses and limits of the use of the Recommended Dietary Allowances for evaluating dietary intake data. Am. J. Clin. Nutrit. 41: 155-164. 1985.

7. FAO/WHO/UNU. Energy and Protein Requirements. Report of a Joint FAO/WHO/UNU Meeting. WHO Tech. Rept. Ser. in press, Geneva, 1985.

8. Beaton, G.H., J. Milner, P. Corey, V. McGuire, M. Cousins, E. Stewart, M. de Ramos, D. Hewitt, P.V. Grambsch, N. Kassim, and J. A. Little. Sources of variance in 24 hour dietary recall data: implications for nutrition study design and interpretation. Am. J. Clin. Nutrit. 32: 2546-2559. 1979.

9. Beaton, G.H., J. Milner, V. McGuire, T. Feather, and J.A. Little. Sources of variance in 24 hour dietary recall data: Implications for nutrition study design and interpretation. Carbohydrate sources, vitamins and minerals. Am. J. Clin. Nutrit. 37: 986-995. 1983.

10. Garn, S.M., F.A. Larkin, and P. Cole. The real problem with 1 day food records. Am. J. Clin. Nutrit. 31: 1114-1116. 1978.

11. Hackett, A.F., A.J.. Rugg-Gunn, and D.R. Appleton. Use of a dietary diary and interview to estimate the food intake of children. Hum. Nutr. Appl. Nutr. 37A: 293-300. 1983.

12. Hauser, H.B., and H.T. Bebb. Individual variation in intake of nutrients by day, month and season and relation to meal patterns: Implications for dietary survey methodology. Pp 155-179 in Assessing Changing Food Consumption Patterns. Committee on Food Consumption Patterns, Food and Nutrition Board, National Academy Press, Washington, D.C., 1981.

13. Hunt, W.C., A.G. Leonard, P.J. Garry, and J.S. Goodwin. Components of variance in dietary data for an elderly population. Nutr. Res. 3: 433-444. 1983.

14. Jacobs, D.R., Jr., J.T. Anderson, and H. Blackburn. Diet and serum cholesterol. Do zero correlations negate the relationship? Am. J. Epidemiol. 110: 77-87. 1979.

15. Kavetti, R.-L., and L.-R. Knuts. Agreement between dietary interviews. J. Am. Dietet. Assoc. 79: 654-660. 1981.

16. Liu, K., J. Stamler, A. Dyer, J. McKeever, and P. McKeever. Statistical methods to assess and minimize the role of intra-individual variability in obscuring the relationships between dietary lipids and serum cholesterol. J. Chronic Dis. 31: 399-418. 1978.

17. McGee, D., G. Rhoads, J. Hankin, K. Yano, and J. Tillotson. Within-person variability of nutrient intake in a group of Hawaiian men of Japanese ancestry. Am. J. Clin. Nutrit. 36: 657-663. 1982.

18. Rush, D., and A.R. Kristel. Methodologic studies during pregnancy: The reliability of the 24 hour dietary recall. Am. J. Clin. Nutrit. 35: 1259-1268. 1982.

19. Sempos, G.T., N.E. Johnson, E.L. Smith, and C. Gilligan. Effects of intraindividual and interindividual variation in repeated dietary records. Am. J. Epidemiol. 121: 120-130. 1985.

20. Todd, K.S., M. Hudes, and D.H. Calloway. Food intake measurement: Problems and approaches. Am. J. Clin. Nutrit. 37: 139-146. 1983.

21. Hegsted, D.M. Problems in the use and interpretation of the recommended dietary allowances. Ecol. Food Nutr. 1: 255-265. 1972.

22. Box, G.E.P., and D.R. Cox. An analysis of transformation. J. Roy. Stat. Soc. Britain. 26: 211-252. 1964.

23. FAO/WHO. Requirements of Ascorbic Acid, Vitamin D, Vitamin B_{12}, Folate, and Iron. Report of a Joint FAO/WHO Expert Group. WHO Tech. Rept. Ser. No. 452. Geneva, Switzerland. 1970.

24. Beaton, G.H. What do we think we are estimating? Pp 36-48 in V.A. Beal and M.J. Laus, eds. Proceedings of a Symposium on Dietary Data Collection Analysis, and Significance. Research Bulletin No. 675. Massachusetts Agricultural Experiment Station, College of Food and Natural Resources, University of Massachusetts at Amherst, Massachusetts. 1982.

VARIABILITY OF FOOD INTAKES: ANALYSIS OF DATA FOR 12 DAYS

Karen J. Morgan, Ph.D.
Department of Human Nutrition, Foods
and Food Systems Management
University of Missouri-Columbia

S.R. Johnson, Ph.D.
Center for Agricultural and
Rural Development
Iowa State University

Basile Goungetas
Department of Agricultural Economics
University of Missouri-Columbia

INTRODUCTION

There are important questions about the reliability of estimated mean daily intake levels of food energy and nutrients (1-15). One issue of concern is the number of days of intake data (sample size) required to estimate with a given reliability level mean daily intake per individual. Results from analyses of one and three day survey data have shown substantial day-to-day variation in intakes for a number of the nutrients (7-11, 13-15). Beaton et al. (11) and other researchers (8, 9) have estimated inter and intraindividual components of intake variance. One study (10) suggests the intraindividual standard deviation of food energy intake is about 25 percent of the mean.

The purpose of this study was to evaluate individual intake data for day to day patterns and relate these patterns to the reliability with which mean daily energy and nutrient intakes can be estimated. The study extends the results of Beaton and others utilizing a special survey data base that includes 12 daily records per individual.

DATA

The data for this investigation were obtained from the Exploratory Study of Longitudinal Measures of Individual Food Intake conducted in 1982 under the auspices of the Nutrition Monitoring Division of Human Nutrition Information Service, U.S. Department of Agriculture. This "methodology study" was designed to evaluate effects of different numbers of daily intake records and different methods of recording intake on estimated intake levels. To date, analyses of these data have evaluated only method effects for estimated mean individual daily intakes (16). The present analysis is restricted to data from the methodology study for the NFCS standard three day

intake procedure replicated in four quarters. The data include 12 observations on daily intake per subject.

The subjects in the survey were female homemakers between the ages of 19 and 70 years. These female homemakers were thought to be those in the household who could provide the most accurate intake data. Thus, the study was not designed to evaluate intake levels of different household members. For the NFCS standard method, 100 of the potential 150 females completed the questionnaires in each of the quarters; thus, the sample size for the present analysis was 100.

Although the subjects for the study were all female household heads and meal planners, some information on differentiating characteristics was recorded. These characteristics were age, education level, employment status, and household size. Variables reflecting these characteristics are utilized in the subsequent analysis along with the recorded daily intakes to provide an evaluation of the influence of numbers of days on the reliability of the estimated mean intake.

MODELS FOR VARIABILITY ANALYSIS

Clearly if observations from the different days were independent, standard statistical methods could be used to relate numbers of days to the reliability of mean intake estimates. The standard deviation of the mean in this case would be calculated from the estimated variance of the underlying distribution and the sample size. There is, however, a question about the applicability of this simple statistic for estimating reliability of mean daily intakes. Specifically, a cursory examination of the available results from the methodology study (16) shows that the variances of estimated mean daily intakes did not decrease with increased numbers of observation days as rapidly as if the observations had been independent. Either the distributions of intakes were not constant and/or there were patterns in individual intakes across days.

The estimators for this investigation were modified to reflect the fact that individuals' daily intake records exhibited day-to-day correlations. Specifically, estimators were modified to reflect persistence in consumption behavior. Generally, this persistence means, other things equal, that a greater number of days are required to achieve a particular reliability level for the estimate of mean daily intake.

A regression analysis framework is convenient for generating expressions for the estimators incorporating persistence. Assume that $v_i = \mu + \varepsilon_i$ (i = 1, 2, ..., 12), i.e., that the reported intake for the i-th day (y_i) is equal to the mean intake (μ), plus an error ($\varepsilon_i = y_i - \mu$). The 12 observations per individual can be written as

$$\underline{y} = \underline{x}\mu + \underline{\varepsilon} \qquad (1)$$

where $\underline{y} = (y_1, y_2, \ldots, y_{12})'$, $\underline{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{12})'$, and $\underline{x} = (1, 1, \ldots, 1)$. The day to day pattern in intakes or the relationship among the elements of $\underline{\varepsilon}$ can take different forms. For the present analysis, it is assumed that the elements of $\underline{\varepsilon}$ are related to each other by a first-order autoregressive process, i.e.,

$$\varepsilon_i = \rho\varepsilon_{i-1} + u_i \; ; \; i = 1, 2, \ldots, 12 \qquad (2)$$

and that the u_i are independently and identically distributed with mean zero and variance σ^2. Under these assumptions the generalized least squares (GLS) estimators for the mean, standard deviation and standard error of the mean are,

$$\tilde{\mu} = (\underline{x}'\Omega^{-1}\underline{x})^{-1}\,\underline{x}'\Omega^{-1}\underline{y}$$

$$\tilde{\sigma}^2 = \frac{\tilde{\underline{\varepsilon}}'\Omega^{-1}\tilde{\underline{\varepsilon}}}{N-1}$$

$$\tilde{\sigma}_\mu = \left|\tilde{\sigma}^2(\underline{x}'\Omega^{-1}\underline{x})^{-1}\right|^{1/2},$$

where $\tilde{\varepsilon}_i = y_i - \tilde{u}$ and Ω is the Prais-Winsten matrix (17) depending on the value of ρ. If ρ is unknown, it can be estimated by

$$\hat{\rho} = \left(\sum_{i=2}^{N} \hat{\varepsilon}_i\hat{\varepsilon}_{i-1}\right) \Big/ \sum_{i=2}^{N} \hat{\varepsilon}^2_{i-1} \; .$$

Then the "feasible" GLS (with $\hat{\Omega}^{-1}$ instead of Ω^{-1}) can be applied to estimate the parameters of equation (1). Notice that if $\rho = 0$, $\hat{\mu} = \tilde{\mu} = \bar{y}$, $\hat{\sigma}^2 = \sigma^2$, and $\hat{\sigma}_\mu = \sigma_\mu = \sigma_{\bar{y}}$.

An important implication of assumed structure for the "persistence" in consumption patterns is for forecasting. The forecasting question is: given the sample to day i, what is the "best" estimate of the individual intake for the day i+1? Using models (1) and (2), the expected value of the individual intake for the day i+1 is

$$\tilde{y}_{i+1} = (1 - \hat{\rho})\tilde{\mu} + \hat{\rho}y_i \; . \qquad (3)$$

Observe that the forecast estimator (3) incorporates "persistence" in consumption through both $\tilde{\mu}$ and $\hat{\rho}$. In the absence of intake persistence, $\rho = 0$, and $\hat{\mu} = \tilde{\mu}$, and $\hat{y}_{i+1} = \tilde{y}_{i+1}$, with the hats denoting the simple ordinary least squares estimators.

Estimated mean daily intakes for the diet components selected for the analysis can be evaluated individually or plotted as cumulative distributions for indicating impacts of added sample size for the set of individuals sampled. These cumulative distribution functions, when constructed using estimators of mean daily intakes for the subjects based on different numbers of days, provide for generalizing the results for individuals to the population. Specifically, comparisons between these cumulative distributions for different numbers of days utilized in estimating the individual daily mean intakes show how the set of individual estimates shifts as numbers of days used in estimation increases. These shifts in the cumulative distributions can be quantitifed easily. The shifts in the cumulative show the contribution to the accuracy of the set of estimated mean daily intakes from the increased numbers of days.

METHODS

Data for the 100 sampled subjects were used to estimate autocorrelation coefficients based on 12 intake records per person to quantify the persistence in daily intakes (ρ). These estimates were made for food energy, fat, iron and vitamin A. These four dietary components were selected for evaluation because it was believed that food energy intake levels would be relatively consistent across days, fat intake would be representative of macronutrient intake consistency and iron and vitamin A intakes would represent micronutrient intake for a widely distributed and a more food specific micronutrient, respectively.

First the autocorrelation coefficients for each of the 100 females were estimated. Then estimates of the expected next day intakes for each subject were calculated using three days' intake to predict the fourth day intake, six days' intake to predict the seventh day intake, eight days' intake to predict the ninth day intake and 11 days' intake to predict the twelfth day intake. All of these forecasts were calculated with and without ρ, the persistence factor. The difference between the actual (known) intake for each individual on days three, six, nine, and 12 and the forecasted intake based on the days in the sample up to these "test" days were used to estimate the absolute value of the "error". The forecasts were made assuming no autocorrelation and with autocorrelation. Finally, for each individual, the absolute error was determined for the difference between the GLS forecast and the GLS estimated mean intake including the added sample day.

After these calculations were made for each of the 100 sample subjects, absolute errors for estimated daily intakes were averaged across the sample to obtain a mean absolute error estimate for the total sample for each diet component. These estimates were made using both GLS and OLS forecasts. To further illustrate the importance of the persistence factor in forecasting intake levels and evaluating added days of intake data, the sample was

partitioned into two subgroups based on values of the estimated autocorrelation coefficients for the individuals. That is, one subgroup contained individuals with estimated autocorrelation coefficients greater than 0.3 and the other subgroup was composed of individuals whose autocorrelation coefficients were less than 0.3.

Previous research (16) has shown that individuals who usually consume large amounts of food have larger standard deviations of mean daily intake than those who usually consume small quantities of food. It follows then physiologically that if the lower consumers of food energy are near maintenance levels, predicted intake levels should be more accurate than those for "large" eaters. To test this hypothesis, the 100 subjects were arrayed from highest to lowest in estimated average food energy intake. Then the top quartile and bottom quartile in this distribution were selected as subsamples. GLS estimates for mean daily intakes and mean absolute error values for GLS and OLS forecasts were calculated for these two subsamples.

Several studies (e.g., 18, 19, 20) have suggested that characteristics of individuals such as age, sex, education, employment, household income, location of residence and other factors, influence estimated intake levels. An analysis for effects of selected socioeconomic characteristics in forecasting intake levels were conducted. Recall that the sample was rather homogeneous. That is, all subjects were adult females with moderate household income levels and residing in one region of the U.S. Therefore, the analysis for socioeconomic characteristic impacts was somewhat limited. This analysis was restricted to food energy intake.

The socioeconomic characteristics examined were age, household size, employment status and education level. The sample partitions were specified as follows:

Age
- 19 to 28 years
- 29 to 38 years
- 39 to 48 years
- 49 to 58 years
- 59 to 70 years

Household Size
- 1 or 2 members
- 3 members
- 4 members
- 5 or more members

Employment Status
- unemployed
- employed part-time (< 39 hours/week)
- employed full-time (> 40 hours/week)

Education Level
- < high school
- high school graduate
- > high school

RESULTS FOR INDIVIDUALS

Estimates of autocorrelation showed that 27 of the 100 subjects had coefficients for food energy intake greater than 0.3, indicating these individuals exhibited considerable persistence in caloric intake (Table 1). Interestingly, similar numbers of subjects had relatively high estimated autocorrelation coefficients for fat (26), iron (28) and vitamin A (29). Forty-five of the subjects had negative autocorrelation coefficients for food

energy. These negative values indicated that eating patterns alternate, from high to low consumption levels, while the 55 positive values showed consistency in levels of food energy intake. The numbers of positive values for fat, iron and vitamin A were somewhat less indicating intake levels of these three nutrients were somewhat less consistent than food energy intake levels.

Mean absolute error values for estimated daily individual intakes for the sample are summarized in Table 2. Observe that these results indicate that the major gains in accuracy of mean daily intake estimates occurred prior to day seven. That is, subsequent to day six, added observations generally contributed relatively less to the accuracy of the estimates. Accuracy in this case is measured in an operational way--the forecast based on previous days contrasted with the actual intake value in the comparable day. From Table 2 note that from the root mean square errors the differences in the estimates when no autocorrelation was included (which assumes days are independent) were larger. Thus, the patterns in consumption are valuable in estimating next day intakes and, in general, detract from estimates of the additional accuracy that can be obtained by adding another day to the sampling design.

Since the larger the autocorrelation coefficient, the greater is the persistence in consumption, it follows that forecasts for individuals with larger autocorrelation coefficients should require fewer days for accurate estimates of next day intake than estimates for subjects with small autocorrelation coefficients. Results for the empirical testing of this proposition are in Tables 3 and 4. The values in Table 3 for the absolute errors between GLS forecast and GLS mean intake show a limited decrease in accuracy between day three and day six. Similar conclusions can not be drawn from results provided in Table 4 until days nine and 12.

Tables 5 and 6 provide the forecast evaluation results for "small eaters" and "large eaters," respectively. These results show clearly that small eaters have much more predictable consumption patterns than "large eaters". In fact, very little additional information is gained after day six for "small eaters," i.e., the average root mean square error estimates are nearly the same for days six, nine and 12. The "large eaters" also showed greatest gains in accuracy of estimated mean daily intake prior to day seven. However, added observations after day six did improve, albeit in a limited way, the accuracy of the forecasts relative to observed intake levels.

Results for the sample partitioned into five age groups are provided in Table 7. Comparisons of these mean square error values with the ones in Table 2 for the total sample showed that the accuracy of predicted intake is not respondent age dependent. That is, the mean absolute errors for the age partitioned sample are not increased or decreased with increasing/decreasing age. The possible exception is for females ages 59 to 70 years. Results in Table 7 show improved accuracy of predicted intake compared to actual intake for the oldest age group; however, this result may be confounded with the lesser quantities of food consumed by the older individuals.

As indicated in Table 8, household size was not found to be an important characteristic for influencing the accuracy of predicted food intake levels. That is, comparisons of results provided in Table 2 with results in Table 8 show that accuracy of predictions, whether for days three, six, nine or 12 were not improved significantly for households of different sizes.

Table 9 provides evidence that food intake patterns of subjects employed full time are more variable, and thus more difficult to predict accurately, than the food intake patterns of subjects unemployed or employed part-time. This conclusion is based on the mean absolute errors for individuals employed full time being consistently greater than similar estimates for the total sample (Table 1). Eighty-three percent of the mean absolute errors for subjects employed part-time or unemployed were smaller than comparable estimates for the total sample.

Education level of the respondent was also an important factor for predicting accurately food energy intake levels. For both the GLS and OLS forecasts, the mean absolute errors for subjects having less than a high school education were significantly less than similar results for the total sample. Less educated individuals likely had less variety in their diets than more educated respondents. There were no differences observed in the accuracy of predictability of food energy intakes between respondents with a high school education and those with an education beyond high school.

RESULTS FOR TOTAL SAMPLE

Cumulative distributions for the four selected dietary components plotted for three, six, nine and 12 day GLS estimated means are exhibited in Figures 1 through 4. From the figures, it is clear that the big contribution to the accuracy of the set of estimated mean daily intakes is achieved by the increase in numbers of days from three to six days. Alternatively, the increase in numbers of days from nine to 12 days produced a relatively smaller improvement. The big gains occurred at lesser numbers of days for food energy than for the other three dietary components but, in general, do not differ appreciably. This analysis could be extended to partitions of the sample for age, household size, employment status and education level. Comparisons between the cumulative distributions for these partitions would demonstrate how, for specific subgroups, the estimated mean daily intakes for the sample improve with added numbers of days or sample.

CONCLUSIONS

Results of this analysis demonstrate: 1) the importance of reflecting appropriately patterns in day-to-day food consumption in the estimation of mean daily intake levels, 2) the importance of these consumption patterns in evaluating incremental contributions of added days of intake information to the accuracy of estimated mean daily intakes, 3) the potential for conditioning sample sizes for observable characteristics of survey subjects for reliability of mean intake estimates and 4) the estimated improvement of

results for the sample subjects as a group that can be obtained by comparisons of cumulative distributions.

The general results on contributions of numbers of days or sample size to the accuracy of mean intake estimates suggest that for the four diet components examined, the benefits fall off importantly after six days. This conclusion holds across the nutrients and for different patterns of food consumption indicated by the estimates of the coefficient of autocorrelation. In addition, the plots of the cumulative functions for estimates of mean daily intakes for the total sample show that the larger changes in accuracy occur between the three and six days of sample sizes. These results are altered appreciably if eating patterns which make the assumption of independence between days inappropriate are not incorporated in the estimation and evaluation processes.

The exact nature of the patterns in intakes between days of individuals warrants more careful investigation. In the present analysis, this pattern assumed a first-order autoregressive form. It is clear from the results, however, that this assumption on the pattern in day to day intakes was only a gross approximation, more appropriate for some individuals than others. With larger samples, permitting analyses of day to day effects in more detail, alternative models of persistence should be investigated as well as perhaps physiological and institutional reasons for this persistence or patterning of daily individual intakes.

The evidence of patterns in day to day individual intakes for all diet components raises many questions about previous estimates of reliability of estimated mean daily intake and numbers of recorded days intakes. Most of the previous analyses have been conducted using estimators for means and variances of daily intake that presume independence of daily intakes. They may have substantially underestimated the true variances of estimators of mean daily intakes. In statistical terms, the estimated means are unbiased but not efficient and the variance estimators biased. Clearly, the reduced efficiency for the estimator of mean daily intake and bias in the variance estimator depend upon as yet not well known day to day eating patterns.

REFERENCES

1. Madden JP, Goodman SJ, Guthrie HA. Validity of the 24-hr recall. J Am Dietet Assn 1976; 68:143-7.

2. Gersovitz M, Madden JP, Smiciklas - Wright H. Validity of the 24-hr dietary recall and seven-day record for group comparisons. J Am Dietet Assn 1978; 73:48-55.

3. Campbell VA, Dodds MI. Collecting dietary information from groups of older people. J Am Dietet Assn 1967; 51:29-33.

4. Linusson EE, Sanjur D, Erickson EC. Validating the 24-hour recall method as a dietary survey tool. Arch Latinoam Nutr 1974; 24:277-94.

5. Carter RL, Sharbaugh CO, Stapell CA. Reliability and validity of the 24-hour recall. J Am Dietet Assn 1981; 79:542-547.

6. Yound CM, Hagan GC, Tucker RE, Foster WD. A comparison of dietary study methods. II. Dietary history vs. seven-day vs. 24-hr recall. J Am Dietet Assn 1952; 28:218-221.

7. Stuff JE, Garza C, O'Brian Smith E, Nichols BL, Montandon CM. A comparison of dietary methods in nutritional studies. Am J Clin Nutr 1983; 37:300-306.

8. McGee D, Rhoads G, Hankin J, Yano K, Telletson J. Within-person variability of nutrient intake in a group of Hawaiian men of Japanese ancestry. Am J Clin Nutr 1982; 36:657-63.

9. Todd KS, Hudes M, Calloway DH. Food intake measurements: problems and approaches. Am J Clin Nutr 1983; 37:139-46.

10. Beaton GH, Milner J, Corey P, McGuire V, Cousins M, Stewart E, deRamos M, Hewitt D, Grambsch PV, Kassim N, Little JA. Sources of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. Am J Clin Nutr 1979; 32:2546-59.

11. Beaton GH, Milner J, McGuire V, Feather TE, Little JA. Sources of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. Carbohydrate sources, vitamins and minerals. Am J Clin Nutr 1983; 37:986-995.

12. Lechtig A, Yarbrough C, Martorell R, Delgado H, Klein RE. The one-day recall dietary survey: A review of its usefulness to estimate protein and calorie intake. Arch Latinoam Nutr 1976; 26:243-71.

13. Liu K, Stamler J, Dyer A, McKeever J, McKeever P. Statistical methods to assess and minimize the role of intra-individual variability in obscuring the relationship between dietary lipids and serum cholesterol. J Chronic Dis 1978; 31:399-418.

14. Acheson KJ, Campbell IT, Edholm OG, Miller DS, Stock MJ. The measurement of food and energy intake in man - an evaluation of some techniques. Am J Clin Nutr 1980; 33:1147-54.

15. Rush D, Kristal AR. Methodologic studies during pregnancy: the reliability of the 24-hour dietary recall. Am J Clin Nutr 1982; 35:1259-68.

16. Morgan KJ, Johnson SR, Rizek RL, Reese R, Stampley GL. Collection of food energy intake data: An evaluation of methods. (In press.)

17. Fomby TB, Hill RC, Johnson SR. Advanced Econometric Methods. New York: Springer-Verlag, 1984.

18. Basiotis P, Brown M, Johnson SR, Morgan KJ. Nutrient availability, food cost, and food stamps. Am J Agr Econ 1983; 65:685-693.

19. Adrian J, Daniel R. Impact of socioeconomic factors on consumption of selected food nutrients in the United States. Am J Agr Econ 1976; 58:31-38.

20. Brown M, Johnson SR, Morgan KJ. Socioeconomic factors and the nutritional intake of children. J Econ 1980; 6:216-218.

Table 1. Summary of estimated coefficients of autocorrelation for food energy, fat, iron and vitamin A.

Dietary Component	Absolute Value		Sign	
	<0.3	>0.3	-	+
Food energy	27*	73	45	55
Fat	26	74	58	42
Iron	28	72	59	41
Vitamin A	29	71	56	41

*Number of subjects with autocorrelation coefficient estimators for food energy greater than 0.3.

Table 2. Mean absolute errors for estimated daily individual intakes of food energy, fat, iron and vitamin A.

Dietary Component	Day and Mean Absolute Error Estimator			
	Day 3	Day 6	Day 9	Day 12
	GLS Forecast and Actual Intake			
Food Energy (kcal)	458	456	361	315
Fat (gm)	30	26	21	18
Iron (mg)	3.8	3.0	2.8	3.3
Vitamin A (IU)	3155	3163	3080	3021
	OLS Forecast and Actual Intake			
Food Energy (kcal)	450	493	386	333
Fat (gm)	30	27	23	20
Iron (mg)	3.8	3.3	3.2	3.4
Vitamin A (IU)	3185	3278	3219	3277
	GLS Forecast and GLS Mean Intake			
Food Energy (kcal)	173	117	74	86
Fat (gm)	11	7	5	5
Iron (mg)	1.4	1.0	1.0	1.0
Vitamin A (IU)	1093	823	703	784

Table 3. Mean absolute errors for estimated daily individual intakes of individuals with autocorrelation coefficients of greater than 0.3.

Dietary Component	Day and Mean Absolute Error Estimator			
	Day 3	Day 6	Day 9	Day 12
	GLS Forecast and Actual Intake			
Food Energy (kcal)	394	326	370	289
Fat (gm)	26	18	21	16
Iron (mg)	3.3	2.1	2.3	2.3
Vitamin A (IU)	3108	2755	2146	4001
	OLS Forecast and Actual Intake			
Food Energy (kcal)	387	456	395	323
Fat (gm)	26	24	28	20
Iron (mg)	3.4	3.2	3.4	2.9
Vitamin A (IU)	3068	3247	2899	4426
	GLS Forecast and GLS Mean Intake			
Food Energy (kcal)	157	164	110	152
Fat (gm)	12	10	9	9
Iron (mg)	1.4	1.1	1.2	1.1
Vitamin A (IU)	1231	1047	1042	937

Table 4. Mean absolute errors for estimated daily individual intakes of individuals with autocorrelation coefficients of less than 0.3.

Dietary Component	Day and Mean Absolute Error Estimator			
	Day 3	Day 6	Day 9	Day 12
	GLS Forecast and Actual Intake			
Food Energy (kcal)	482	504	357	324
Fat (gm)	31	28	21	19
Iron (mg)	3.9	3.3	2.9	3.6
Vitamin A (IU)	3175	3330	3461	2621
	OLS Forecast and Actual Intake			
Food Energy (kcal)	473	506	383	336
Fat (gm)	31	28	22	20
Iron (mg)	3.9	3.4	3.1	3.7
Vitamin A (IU)	3233	3290	3350	2808
	GLS Forecast and GLS Mean Intake			
Food Energy (kcal)	179	100	60	51
Fat (gm)	11	6	4	4
Iron (mg)	1.4	0.7	0.5	0.5
Vitamin A (IU)	1037	731	565	721

Table 5. Mean absolute errors for estimated daily individual intakes of small eaters.

Dietary Component	Day and Mean Absolute Error Estimator			
	Day 3	Day 6	Day 9	Day 12
	GLS Forecast and Actual Intake			
Food Energy (kcal)	350	213	311	258
Fat (gm)	26	14	17	16
Iron (mg)	3.0	2.5	2.3	3.1
Vitamin A (IU)	1800	1781	2535	2272
	OLS Forecast and Actual Intake			
Food Energy (kcal)	355	241	327	245
Fat (gm)	26	14	19	15
Iron (mg)	3.1	2.9	2.4	3.2
Vitamin A (IU)	1862	1925	2533	2185
	GLS Forecast and GLS Mean Intake			
Food Energy (kcal)	113	60	56	69
Fat (gm)	8	4	4	4
Iron (mg)	1.0	0.7	0.5	0.7
Vitamin A (IU)	702	610	586	508

Table 6. Mean absolute errors for estimated daily individual intakes of large eaters.

Dietary Component	Day and Mean Absolute Error Estimator			
	Day 3	Day 6	Day 9	Day 12
	GLS Forecast and Actual Intake			
Food Energy (kcal)	488	558	390	358
Fat (gm)	36	33	23	23
Iron (mg)	3.1	3.4	3.0	3.5
Vitamin A (IU)	4663	3013	3057	2471
	OLS Forecast and Actual Intake			
Food Energy (kcal)	472	645	401	416
Fat (gm)	35	35	25	27
Iron (mg)	3.1	3.8	3.8	3.8
Vitamin A (IU)	4401	3057	3046	3274
	GLS Forecast and GLS Mean Intake			
Food Energy (kcal)	198	163	93	111
Fat (gm)	14	9	5	6
Iron (mg)	1.3	0.9	0.8	0.6
Vitamin A (IU)	1843	649	578	1461

Table 7. Mean absolute errors for estimated daily individual food energy intakes for individuals partitioned by age classifications.

Age (Years) Group	Day and Mean Absolute Error Estimator			
	Day 3	Day 6	Day 9	Day 12
	GLS Forecast and Actual Intake			
19 to 28	481	407	327	316
29 to 38	390	461	331	318
39 to 48	481	554	368	343
49 to 58	577	409	448	334
59 to 70	407	420	336	249
	OLS Forecast and Actual Intake			
19 to 28	506	505	335	381
29 to 38	385	498	367	321
39 to 48	451	579	415	377
49 to 58	553	419	442	312
59 to 70	408	450	361	286
	GLS Forecast and GLS Mean Intake			
19 to 28	161	183	90	111
29 to 38	140	101	67	76
39 to 48	212	124	65	95
49 to 58	225	89	84	94
59 to 70	138	121	72	64

Table 8. Mean absolute errors for estimated daily individual food energy intakes for individuals partitioned by household size classifications.

Household Size Groups	Day and Mean Absolute Error Estimator			
	Day 3	Day 6	Day 9	Day 12
	GLS Forecast and Actual Intake			
1 and 2	488	351	461	236
3	501	468	347	283
4	370	594	344	362
5 or more	449	447	262	414
	OLS Forecast and Actual Intake			
1 and 2	486	352	465	247
3	494	524	368	332
4	351	646	406	357
5 or more	442	493	285	423
	GLS Forecast and GLS Mean Intake			
1 and 2	175	114	78	80
3	188	135	68	88
4	146	99	71	108
5 or more	177	116	77	72

Table 9. Mean absolute errors for estimated daily individual food energy intakes for individuals partitioned by employment status classifications.

Employment Status Groups	Day and Mean Absolute Error Estimator			
	Day 3	Day 6	Day 9	Day 12
	GLS Forecast and Actual Intake			
Unemployed	456	453	354	290
Employed Part-Time	399	462	320	320
Employed Full-Time	498	454	393	332
	OLS Forecast and Actual Intake			
Unemployed	448	468	365	300
Employed Part-Time	377	486	349	295
Employed Full-Time	499	518	427	384
	GLS Forecast and GLS Mean Intake			
Unemployed	174	104	85	82
Employed Part-Time	161	94	71	76
Employed Full-Time	181	143	65	96

Table 10. Mean absolute errors for estimated daily individual food energy intakes for individuals partitioned by educational status classifications.

Educational Status Groups	Day and Mean Absolute Error Estimator			
	Day 3	Day 6	Day 9	Day 12
	GLS Forecast and Actual Intake			
< High School	408	246	343	210
High School Graduate	473	502	350	352
> High School	446	454	404	270
	OLS Forecast and Actual Intake			
< High School	411	339	380	218
High School Graduate	464	529	369	371
> High School	434	481	442	291
	GLS Forecast and GLS Mean Intake			
< High School	140	85	68	79
High School Graduate	177	121	71	95
> High School	183	126	85	64

FIGURE 1: CUMULATIVE FREQUENCY DISTRIBUTIONS OF THE 3-DAY, 6-DAY, 9-DAY, AND 12-DAY "ADJUSTED" MEANS OF FOOD ENERGY (KCAL) INTAKE FOR ALL INDIVIDUALS.

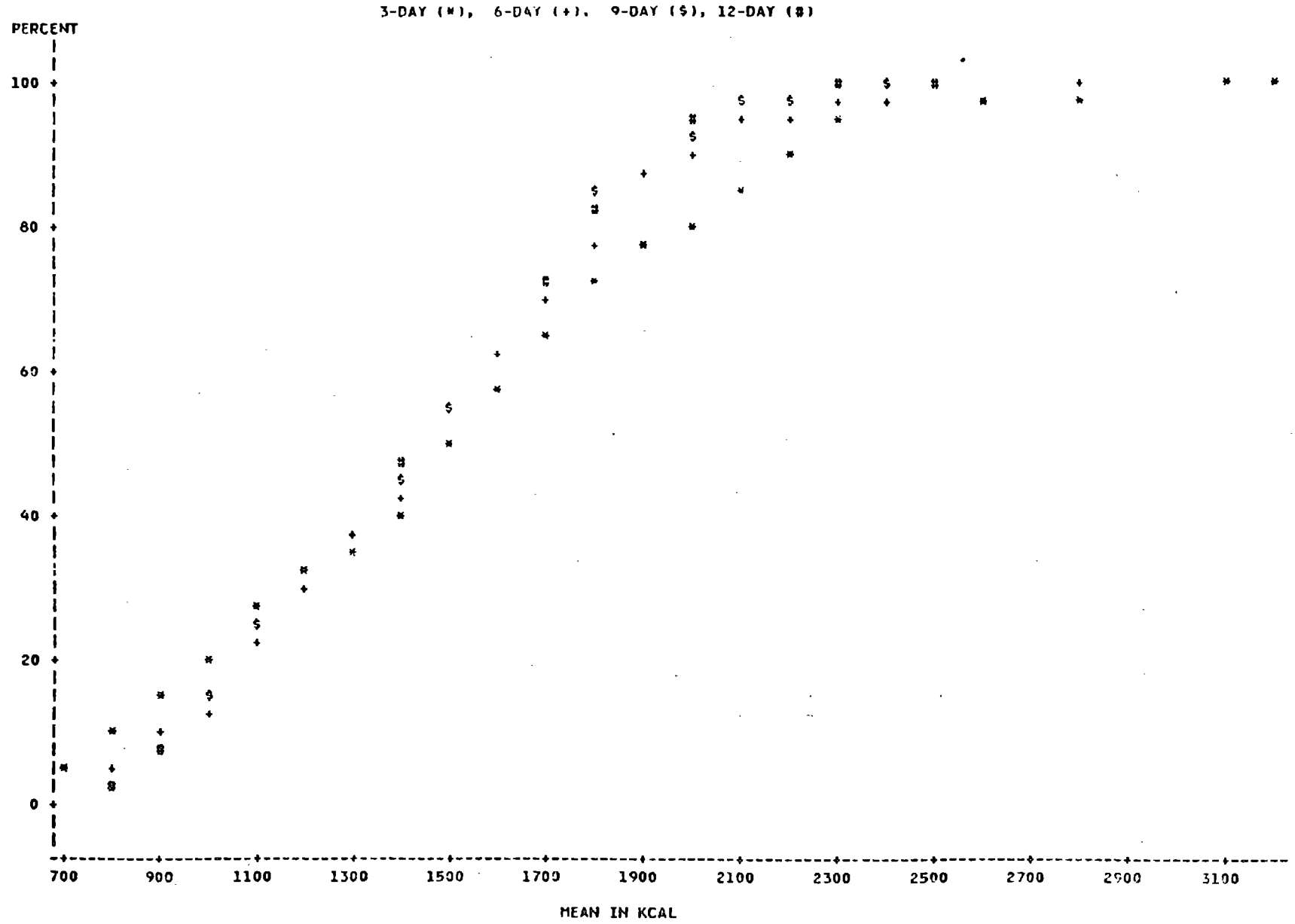


FIGURE 2: CUMULATIVE FREQUENCY DISTRIBUTIONS OF THE 3-DAY, 6-DAY, 9-DAY, AND 12-DAY "ADJUSTED" MEANS OF FAT (GM) INTAKE FOR ALL INDIVIDUALS.

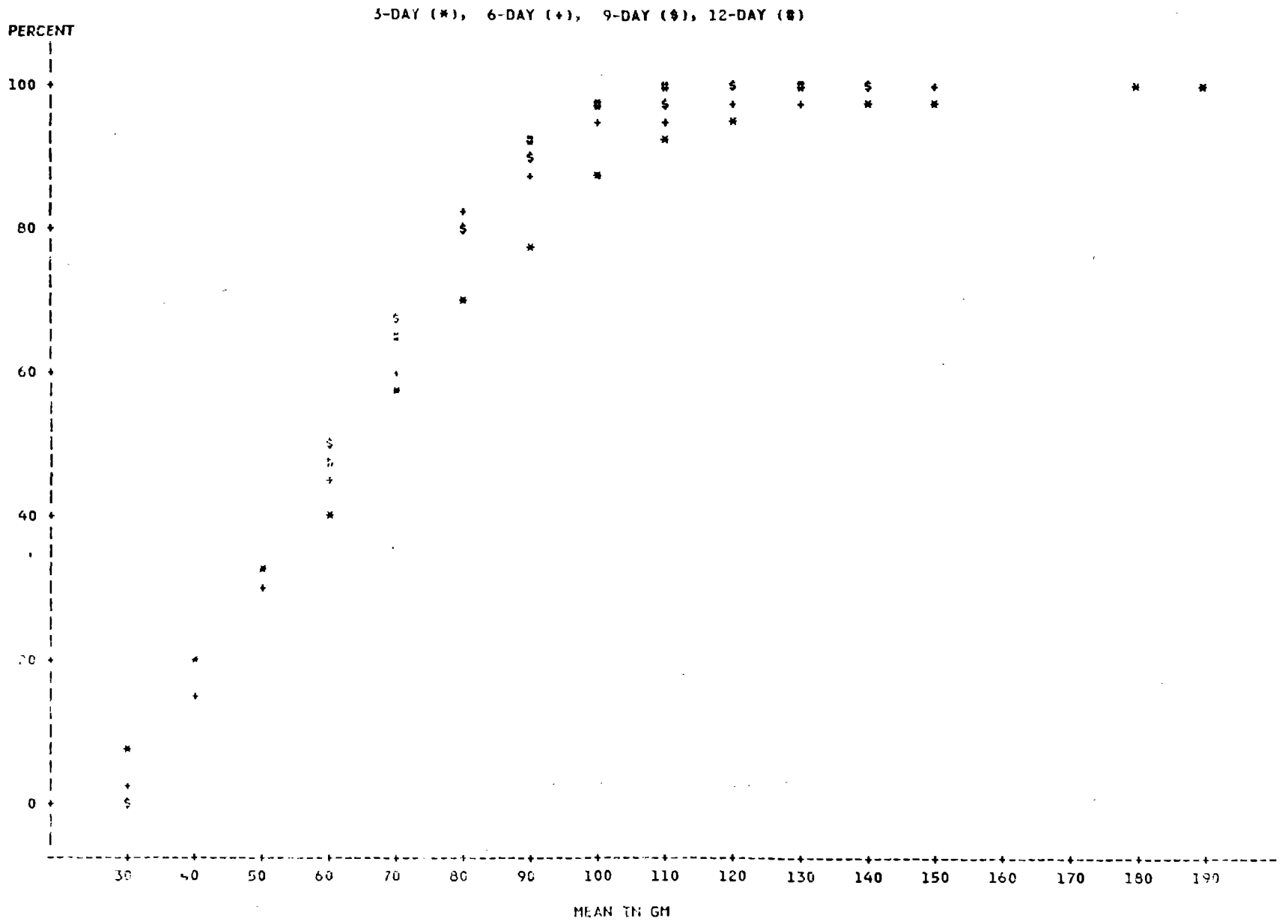


FIGURE 3: CUMULATIVE FREQUENCY DISTRIBUTIONS OF THE 3-DAY, 6-DAY, 9-DAY, AND 12-DAY "ADJUSTED" MEANS OF IRON (MG) INTAKE FOR ALL INDIVIDUALS.

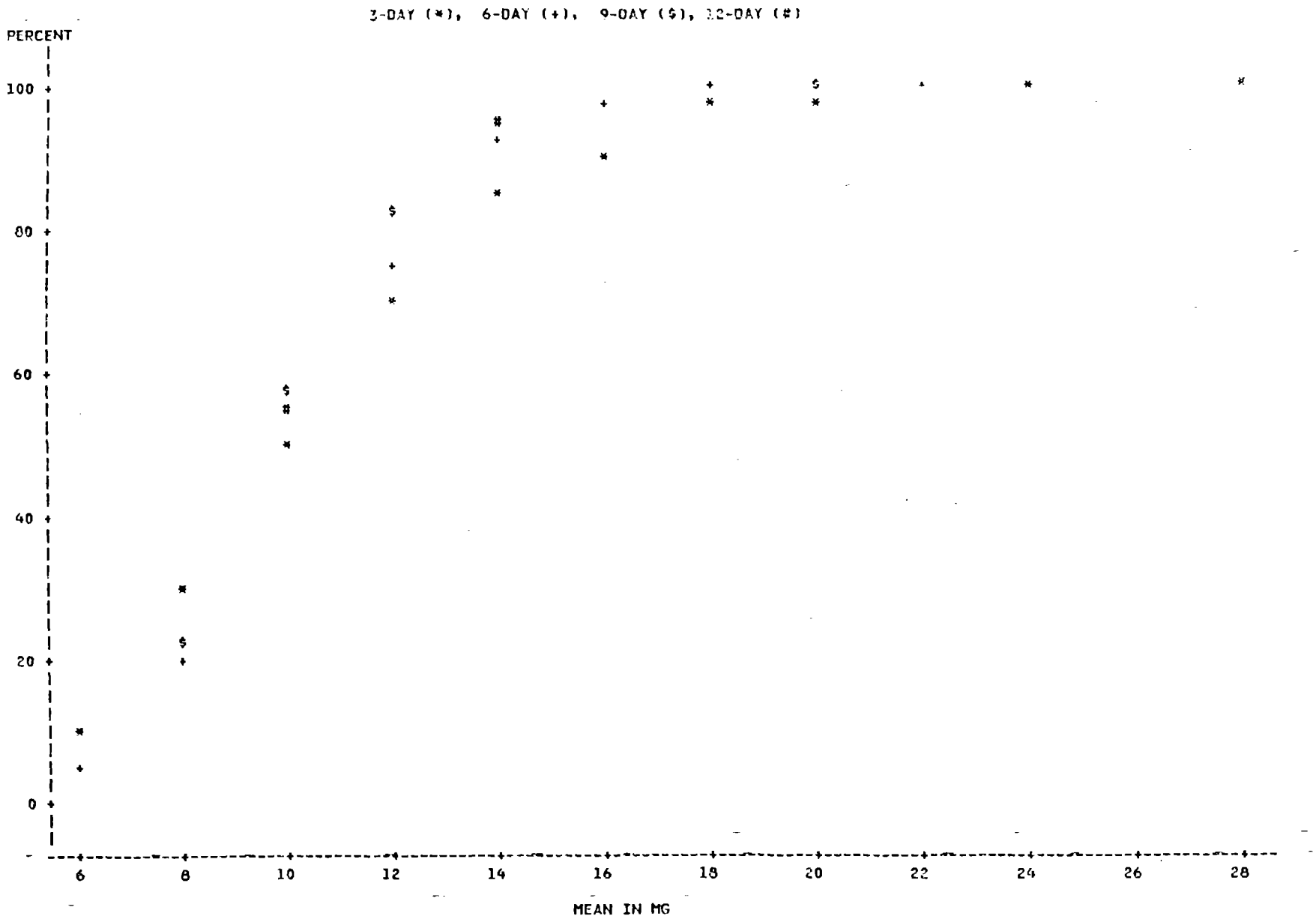
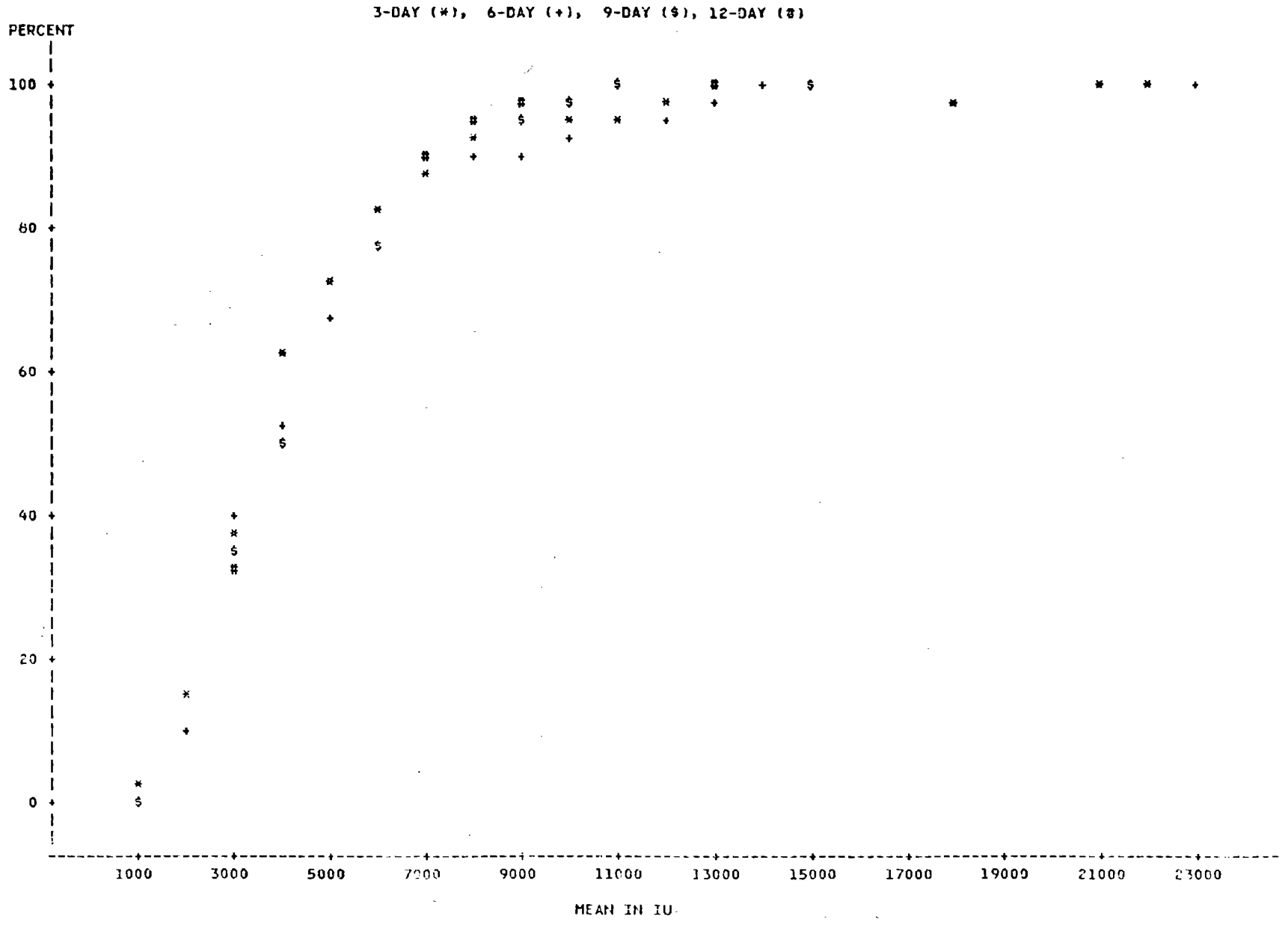


FIGURE 4: CUMULATIVE FREQUENCY DISTRIBUTIONS OF THE 3-DAY, 6-DAY, 9-DAY, AND 12-DAY "ADJUSTED" MEANS OF VITAMIN A (IU) INTAKE FOR ALL INDIVIDUALS.



DIETARY ASSESSMENT OF IMMIGRANT AND REFUGEE CHILDREN

Mary J. Kretsch, Ph.D. and Karen M. Todd, M.S., R.D.
USDA/ARS, Western Human Nutrition Research Center
Presidio of San Francisco, California 94129

The study results which I am presenting today are part of a larger cooperative study being conducted by the Western Human Nutrition Research Center, United States Department of Agriculture and the Department of Nutritional Sciences, University of California, Berkeley, to determine the nutritional status of newly arrived immigrant and refugee school children in San Francisco, California. Due to difficult environmental conditions in their countries of origin and the disruptive effects of cross-cultural migration, these children were believed to be at a heightened risk of malnutrition. Dietary, anthropometric, biochemical, dental and socio-demographic information were collected from these children to evaluate their nutritional status. We at USDA have been responsible for the collection and evaluation of the nutrient and food intake data.

The objectives of the dietary intake evaluation were to determine: (1) the nutritional adequacy of the children's diets; (2) the accuracy of the recall technique for collection of dietary information from these children; (3) the effects of acculturation on the food and nutrient intake; and (4) the effect of ethnicity on the acceptability of foods served in the school food program. However, since the focus of this morning's presentations is dietary methodology, I will limit my presentation to the evaluation of the recall technique as a method to collect dietary information from immigrant and refugee children.

San Francisco has a Newcomer School system to aid immigrant and refugee children in adapting to life in the United States. Upon arrival in San Francisco, the children enter one of the Newcomer schools for approximately one year. Bilingual education and a curriculum adapted to the needs of the children are provided. Children from three Newcomer schools were studied: the school for Chinese speaking children, the school for Spanish speaking children, and the school comprised mainly of Southeast Asians. Table 1 shows the country of origin by ethnic group for the children studied. Since this study was conducted during 1984, the children's country of origin breakdown reflects the immigration and refugee patterns of that time. Eighty-five percent of the Chinese children were from China; the remainder being from Hong Kong and Vietnam. Sixty percent of the Hispanic children were from El Salvador; the next largest segments were from Mexico and Nicaragua. All of the Filipino children were from the Philippines and the Cambodian children from Cambodia. The children were from varying social and educational backgrounds.

School breakfast and lunch were provided to all the children attending the Newcomer Schools. Therefore, potentially 10 meals per week or about half of a child's weekly meals were provided by the school system. However, not all of the children ate the school breakfast and about 20% ate breakfast at home as well as at school. Foods served were an

American-type menu rather than being culturally adapted. All school meals were prepared at a central kitchen and delivered to the schools approximately 30 minutes before meal time. At the Hispanic school, food was delivered in bulk and portioned out to the children as they passed through the tray line in the school cafeteria. At the Chinese and Southeast Asian schools, foods were provided pre-portioned and pre-packaged.

The accuracy of the recall technique to collect dietary information from these children was determined by comparing the actual food consumed in the cafeteria (weighed intake) with the cafeteria food intake as recalled the next day by the child. Interviews were conducted in the child's native language by bilingual and bicultural interviewers who had been trained in the recall technique adapted for children. All methodology was pilot-tested prior to data collection to minimize differences between interviewers.

In the interview, the children were asked to recall all foods and beverages consumed in the last 24-hour period. For the study being presented today, only data on cafeteria foods consumed at the breakfast and lunch meals will be presented. As a point of reference, the children were asked about their intake in relation to activities during the day such as playing, watching T.V. etc. A food picture book had been prepared by each interviewer to prompt the children if they had difficulty identifying foods. Pictures of American and culturally specific foods, appliances, and cooking untensils, as well as color samples were available to help identify foods and preparation techniques. To help the child describe the portion of food consumed, each child was asked to draw the size of the food item on a paper plate or use dried beans to fill one of the containers to the level they thought accurate. No food models were used. The interviewers had been instructed not to correct the child even if they knew the child had inaccurately identified a food or beverage or misjudged the quantity.

Each interviewer coded for computer processing the dietary data of his/her assigned subjects and verified the correctness of the coded data. Nutrient intakes for the recall and observed methods were calculated using a computerized nutrient data bank based on Handbook 8. School recipes were obtained from the foodservice director and their nutrients calculated.

The age, sex and ethnic breakdown of the children studied are shown in Table 2. Children between the ages of eight and eleven were studied which encompassed the 4th through 6th grades. Half of the children were female and half were male. More Chinese and Hispanic children were available for study and therefore, the group sample sizes are larger than for the Filipino and Cambodian children. Since some of the children skipped school breakfast, less children were studied at breakfast than at lunch.

Analysis of variance was used to test for significant effects of age, sex, and ethnicity on the nutrient intake differences between the two methods. No significant nutrient differences were found for age or sex, but significant differences were found between ethnic groups. Therefore, the data in this paper are presented by ethnic group. As a point of

interest, however, there was a trend, although not significant, for the younger children (8 year olds) to be more accurate than the older children (10 and 11 year olds).

Table 3 shows the mean calculated energy intake for each method by ethnic group and meal. At breakfast, there were no significant differences between the methods. The Chinese children tended to underreport whereas the other groups overreported. At lunch, however, there was a significant difference ($p<.01$) between the methods for the Chinese and Filipino groups. The Chinese children underreported energy intake and the Filipino children overreported. Although not all differences were significant, the Chinese children underreported at both meals, the Hispanic and Filipino children overreported at both meals, and the Cambodian children overreported at one meal and underreported at the other.

Table 4 shows the mean correlations between the weighed and recalled methods for select nutrients consumed at the breakfast meal. For energy, the Filipino chldren were the most accurate, and the Cambodian children the least accurate. Breakfast usually consisted of milk, fruit or juice, and a bread type item (i.e. roll, waffle, coffeecake etc.). Failure to recall beverages at the breakfast meal was the principal source of error for all ethnic groups. About half of the Cambodian children failed to recall milk which is reflected in their low correlation coefficients for the nutrients shown.

When a larger number and variety of food items were served, as at the lunch meal, the overall accuracy for all ethnic groups, except the Cambodian children, declined. Table 5 shows the mean correlations between the weighed and recalled methods for nutrients consumed at the lunch meal. The correlations for energy intake were: Chinese, .36; Hispanic, .43; Filipino, .54; and Cambodian, .56. These are low to moderate correlation coefficients. However, to put this in perspective, in evaluations of the dietary recall technique, correlations between the recalled and weighed methods rarely exceed .70. The magnitude of the correlation coefficients for the nutrients other than energy varied by nutrient and ethnic group. Particularly low correlations were found for calcium for the Chinese children and calcium and vitamin A for the Filipino children. However, it should be remembered that in addition to being children, these children are recalling food items that for the most part were "new" foods to them. Therefore, it would be expected that the children would be more accurate when recalling familiar foods such as those eaten at home. Unfortunately, we were not able to assess the validity of the recall method for at home meals.

The "flat-slope syndrome" has been reported in adult and pediatric populations when recalled energy intakes are regressed against actual energy intakes. When this occurs, those individuals consuming lower energy intakes are overreporting their intake and those consuming higher intakes are underestimating their intake. By regression analysis, we investigated this in the four ethnic groups for energy intake at lunch. We found the "flat-slope syndrome" for the Chinese and Hispanic children but not for the Filipino or Cambodian children. Instead, the Filipino children consistently overreported their intake and the Cambodian children

consistently underreported theirs. In addition, the Filipino and Cambodian data was scattered close to the regression line; this was not found for the Hispanic and Chinese groups.

Limited data is available in the literature on the accuracy of dietary recall interviews conducted with children. In addition, the studies conducted have shown conflicting results. With increasing numbers of women entering the workforce, the ability of the mother to provide an accurate accounting of the child's food consumption is diminished. Therefore, it may become necessary to rely on dietary interviews with school age children. This study has shown that ethnicity of the child significantly affects the accuracy of the self-reported recall. This is reasonable since a child's culture strongly influences food habits and may also influence the child's ability to function in an interview setting. In addition, accuracy of recall tends to decrease as the number of food items served at a meal increases. It should be remembered, however, that these children were recalling unfamiliar food items and therefore, their accuracy may be higher when recalling foods served at home. Additional studies need to be conducted with school-age children in the United States to determine their ability and accuracy to self-report dietary intake.

TABLE 1. COUNTRY OF BIRTH

CHINESE	
China	85%
Hong Kong	10%
Vietnam	5%
HISPANIC	
El Salvador	60%
Mexico	18%
Nicaragua	16%
Guatemala	2%
Peru	2%
Columbia	2%
FILIPINO	
Philippines	100%
CAMBODIAN	
Cambodia	100%

TABLE 2. SEX, AGE, AND ETHNICITY OF STUDY SUBJECTS

	BREAKFAST SUBSET (N)	LUNCH SUBSET (N)
SEX		
Male	51	59
Female	53	58
AGE		
8 yrs.	32	36
9 yrs.	30	32
10 & 11 yrs.	42	49
ETHNICITY		
Chinese	30	33
Hispanic	31	34
Filipino	22	28
Cambodian	21	22

TABLE 3. MEAN (±SD) CALCULATED ENERGY INTAKE FOR EACH METHOD BY ETHNIC GROUP AND MEAL

	ETHNIC GROUP			
	CHINESE	HISPANIC	FILIPINO	CAMBODIAN
	(KCAL)	(KCAL)	(KCAL)	(KCAL)
BREAKFAST				
RECALL METHOD	225 ± 120	252 ± 164	171 ± 90	235 ± 212
WEIGHED METHOD	243 ± 141	232 ± 128	142 ± 120	185 ± 103
DIFFERENCE	-18	+20	+29	+50
LUNCH				
RECALL METHOD	475 ± 136	427 ± 186	400 ± 244	261 ± 152
WEIGHED METHOD	555 ± 123	403 ± 180	260 ± 148	333 ± 141
DIFFERENCE	-80*	+24	+140*	-72

*SIGNIFICANT DIFFERENCE AT $P<.01$

TABLE 4. CORRELATIONS BETWEEN THE WEIGHED AND RECALL METHODS FOR SELECT NUTRIENTS CONSUMED AT THE BREAKFAST MEAL

	ETHNIC GROUP			
	CHINESE (N=30)	HISPANIC (N=31)	FILIPINO (N=22)	CAMBODIAN (N=21)
ENERGY	.51	.44	.70	.27
PROTEIN	.57	.60	.71	.28
FAT	.19	.54	.66	.35
CALCIUM	.66	.61	.69	.58
VITAMIN A	.69	.67	.77	.45

TABLE 5. CORRELATIONS BETWEEN THE WEIGHED AND RECALL METHODS FOR SELECT NUTRIENTS CONSUMED AT THE LUNCH MEAL

	ETHNIC GROUP			
	CHINESE (N=33)	HISPANIC (N=34)	FILIPINO (N=28)	CAMBODIAN (N=22)
ENERGY	.36	.43	.54	.56
PROTEIN	.43	.32	.24	.53
FAT	.36	.27	.36	.46
CALCIUM	.16	.65	.03	.62
VITAMIN A	.68	.53	.02	.66

NHANES III Plans: Issues and Approaches

Clifford Johnson
Chief, Nutrition Statistics Branch
Division of Health Examination Statistics
National Center for Health Statistics

The National Health and Nutrition Examination Survey conducted by the National Center for Health Statistics (NCHS) is an expansion of the National Health Examination Survey, which was authorized under the National Health Survey Act of 1956 and fielded in 1960. The survey was designed to collect data by direct standardized examination of a sample of the U.S. civilian noninstitutionalized population.

Beginning in 1960, data from household interviews and extensive physical examinations were collected through the National Health Examination Survey. In 1971, responsibility for monitoring the nutritional status of the population was added and the National Health Examination Survey became the National Health and Nutrition Examination Survey (NHANES). The first National Health and Nutrition Examination Survey (NHANES I) was conducted during 1971-1975. The second NHANES (NHANES II) was conducted during 1976-1980. In 1982-1984, a special study was conducted to collect similar information on the U.S. Hispanic population. The Hispanic Health and Nutrition Examination Survey (HHANES) was conducted because the national surveys included too few Hispanics to enable adequate estimation of the health and nutritional status of this subpopulation.

The results of these surveys have been published in numerous NCHS publications and journals. In addition, computer data tapes have been released to the public through the National Technical Information Service.

The next national survey (NHANES III) is scheduled to begin in 1988. But before I tell you about the proposed plans for NHANES III, I feel it is important to make you aware of the kinds of issues that must be confronted before a new survey can be fielded. My comments will be limited to the nutrition component of the survey, and more specifically, the dietary component, since that is the focus of this conference.

NHANES dietary data have been put to four major uses: relating diet and demographic characteristics, relating diet and health characteristics, determining interactions of diet and nutritional status indicators, and monitoring trends in diet and nutrient intakes over time.

Since dietary intakes were calculated for population subgroups in NHANES I and NHANES II by means of the results from a single 24-hour recall, we had the opportunity to compare results over time; the fourth major use of the data as described previously. This appeared to be a straightforward simple operation until it came time to interpret the findings.

When we compared sodium intakes between the two survey periods, for example, we observed a large increase in all age-sex groups in NHANES II. In order to try to explain this difference we first listed possible sources for the change.

Even though the data collection procedures had remained the same, we identified four other possible sources that might explain the observed results:

1. the total intake, or quantity consumed, by all population subgroups could have increased,
2. the choice of foods eaten by the population could have changed,
3. the enrichment, fortification, or preparation of the foods by manufactures could have changed, or
4. the nutrient data base used to process the information could have changed.

In a worst possible case, it could be some or all of the above possible sources.

In this example, our analyses have led us to believe the most of the difference is likely due to changes in the nutrient data base between NHANES I and NHANES II. This may mean there has been no significant change in sodium intake between the two surveys. At the least, interpretation of any changes by the population will be difficult to determine. What this example shows us is that we must be very careful when comparing dietary results from two similar studies and even more cautious when comparing results between two different surveys. The implications of incorrect or inappropriate interpretation of the data may be far-reaching.

Any survey conducted periodically over time must contend with this kind of an issue. Is it more important to use the same methodology and data base over time to allow for comparability, or to adopt new, and probably improved, methodologies and information as it becomes available? This is not an easy or obvious decision to make. However, this type of decision must be confronted and dealth with when planning a new survey.

This brings me to the discussion of where we are now in our planning for NHANES III and more specifically the dietary component of NHANES III. We are still very early in the planning stages for the survey. The major health components won't be selected until early 1986. What we plan to do is to select dietary methodologies that will be most useful when related to the health components selected and at the same time maintain some degree of comparability with past NHANES.

We hope that NHANES III will be conducted over six years with a total of about 60,000 examined persons. In addition, each two year cycle of the survey would be a nationally representative sample of the population. We would like to automate the data collection to the extent possible and to build into the sample design a longitudinal component. These will take a lot of planning and appropriate budgets to execute and complete. Only the future will tell how successful we will be.

The USDA Food Frequency Study

Frances A. Larkin, Ph.D.
School of Public Health, University of Michigan

The Food Frequency Study was designed to develop a questionnaire that will describe an individual's usual diet for one year. The study is based on several assumptions: 1) that an individual is reasonably consistent in his or her choice of food; 2) that he or she is sufficiently aware of the foods eaten to be able to recall them; 3) that he or she can generalize about eating habits over time and in terms of groups of foods, not just single foods; and 4) lastly, that he or she is willing and able to read and write. The specific study questions are:

1. Are questions about food frequencies valid indicators of last year's diet? Validity will be measured by the degree of correspondence between food frequencies for the past year and 16 days of recall/records sampled from the same period.

2. What characteristics of respondents, and of dietary patterns, are associated with validity or with various aspects of lack of validity such as bias?

The study design requires each respondent to provide information five times over the course of one year, four times providing one-day recalls and three-day records and the fifth time providing food frequency data for the previous year. Because interviewing for each round of data collection takes three months, the field work will take about 15 months. The 24 hour recall in the first round was used primarily as a respondent training tool for the three-day records. It was retained in all rounds as a training reinforcer and to provide another day of data. Each respondent is also given measuring spoons and cups, a ruler and a set of eighth-inch thickness measures adapted from the NHANES models to help in estimating amounts.

We have completed collecting the recall/records and are in the fifth round - that is, we are now collecting the food frequency data. This is a progress report using data from the pretest of the food frequency instrument administered to all respondents prior to the first 24 hour recall/three-day record, and from two rounds of food records.

Methods

Sample

Of the 273 enrolled into the project, we hope to retain at least 200 respondents, aged 23-50, roughly equally divided among males and females, black and white. We attempted to get an equal representation of high school education and greater than high school education within each category but we could not achieve this in our age group and communities. Refusal rates ran around 70 percent in Ypsilanti and 50 percent in Ann Arbor, perhaps because we required participation for a year and because our initial interview took about two hours. Respondents were enrolled by

interviewers visiting the homes in census tracts selected because they were at least 25 percent black in the 1980 census. The first interview consisted of the food frequency interview, a 24 hour recall and a brief health and demographic questionnaire. The 3 day record was left to be picked up 3 or 4 days later. In the second, third and fourth rounds, the interviewers visited the respondents for the 24 hour recall and left the 3 day record form. They again returned after several days to pick it up, and review the record before leaving the respondent's home.

Questionnaire Development

The food frequency tool is composed of 113 foods or food groups, each on a separate slip of paper which the respondent sorts into frequency categories.

The initial work in defining food groups was completed for an earlier contract, comparing the intake of food groups for NFCS and NHANESI. The identification of these groups was based on the food group order already built into the data bases, nutrient composition and similarity of use among foods. In the next stage of the development of food groups, we used the 24 hour and 2 day record of NFCS, selecting a 30 percent subsample of the adults aged 23 to 74.

Basically, we analyzed the pattern of occurrence of the most frequently consumed foods, comparing two foods at a time, for individuals over three days. This identified whether the two foods were eaten together, eaten at different times, or eaten without affecting each other. To correct for the frequency of occurrence, standardized scores were computed. A high positive score indicated a stable complementary relation, such as bread and butter. A high negative score indicated a substitute relationship, such as skim milk and whole milk. Next in the development of food groups, we considered the nutritive value of the food groups. Substitute food groups with similar nutrient values were combined - for example, whole wheat and white bread.

Complementary foods were put into separate groups but used in the description of the complementary food. For example, CRACKERS is one group and CHEESE another, but CHEESE is described as "CHEESE, including on CRACKERS." Some foods are used only as complements, such as margarine or butter. Some foods are partially used as complements, such as milk used in in coffee or on cereal. For butter and margarine, the complementary use was asked in a probe. Milk in coffee was asked in a probe, and as an independent food group for milk as a drink.

Some foods used in the same manner are very different nutritionally, regular and light beer are examples. Beer, unspecified, is sorted first, then separated into the regular and light by a probe.

Some foods are used very differently depending upon whether they are eaten singly or in mixtures. For example, tomatoes are eaten raw, cooked and in sauces. Prevalence information determined the foods for which this occurred most often. These foods were then presented to the respondent as

separate foods.

Typical serving size was asked for each food except lettuce and an OTHER VEGETABLES group. A group of vegetables of different size and shape is difficult to generalize about in regard to serving size.

The frequency of use of food groups, whether the groups are used alone or in combination, the nutrient value, and the ease in generalizing about serving size were all considered in defining 110 food groups for the pretest in July through September, 1984. After three rounds of food records were collected, we revised the list, adding avocado, soy products and yogurt as separate foods and adding croissants to the bread group.

Food Frequency Questionnaire Administration

The basic format of the food frequency is a set of 113 slips of paper, 5 1/2 by 8 1/2, listing the name of the food or food group on the front and a partly precoded recording form on the back.

The interviewer places on the table before the respondent three envelopes marked with the frequencies NOT LAST YEAR, LESS THAN 12 TIMES, LAST YEAR, and 12 TIMES OR MORE LAST YEAR and says, "Please take these slips with the food printed on them and sort them into one or the other of three piles according to how often you ate the food last year." The NOT LAST YEAR and LESS THAN 12 TIMES piles are then placed in their respective envelopes and set aside. The slips sorted into the LESS THAN 12 TIMES pile will be used analyticallyin a diversity index but nutrients contributed by these rarely eaten foods will not be included in the determination of the whole year's intake.

The interviewer next takes the slips in the 12 TIMES OR MORE pile and hands them to the respondent saying "Now will you sort these foods which you said you ate 12 or more times last year into one of two piles according to whether you ate the food more often in CERTAIN SEASONS or about the SAME ALL YEAR ROUND." She then places the two envelopes so marked on the table. When this sort is completed, the slips in the CERTAIN SEASONS pile are placed in their envelope and temporarily set aside.

The interviewer then hands the SAME ALL YEAR ROUND slips to the respondent and says, "Now will you sort these foods which you said you ate about the same all year round into one of these piles according to how often you ate them last year." She then places envelopes marked with these frequencies on the table:" MORE THAN ONCE A DAY, ONCE A DAY, ABOUT 5-6 TIMES A WEEK, ABOUT 3-4 TIMES A WEEK, ABOUT 1-2 TIMES A WEEK, ABOUT 1-3 TIMES A MONTH. When the respondent finishes this sort, the interviewer puts the slips in the appropriate frequency envelope. Then, starting with the most frequent envelope, the interviewer asks, for each slip in turn, the amount of usual serving size for that food and any further probes indicated on the back of the slip. Figure 1 is the basic format. For example, the interviewer says," You said you ate eggs in salads and sandwiches 3-4 times a week all year round. When you eat eggs in salads

and sandwiches, about how many do you eat?" All measuring aids are on the table to help in the estimating. The interviewer records directly on the back of the slip the frequencies, amounts and all other information and files the completed slip in a COMPLETED envelope.

When all the SAME ALL YEAR ROUND foods have been probed for amount and other information, the interviewer asks the respondent to sort the MOSTLY IN CERTAIN SEASON foods into the same frequency categories but for the appropriate season. She will then say, "You said you ate strawberries mostly in certain seasons and about 5-6 times a week in that season. About how much did you eat each time you had strawberries? What is your stawberry season? Did you eat strawberries at all during the rest of the year?" If yes, "How often?" The question about the length of the season was added after the pretest disclosed variable seasons for locally produced fruits and vegetables and for chili, cake and other items not ordinarily considered "seasonal." The question about seasonal foods being eaten during the rest of the year was included to get a closer estimate of annual intake.

Other Data Collected

At the first contact, interviewers collected demographic and health information, including information on the use of special diets for each of the health conditions described to the respondent. The demographic data includes employment status of the respondent and spouse and income. At the second through fifth contacts, a short life events questionnaire is completed. The respondent is asked about any changes in health, family composition, and other events that might affect eating patterns.

Results

Table 1 contains a preliminary analysis for only 172 participants, comparing the nutrient values from the pretest food frequency administered at the first round with eight recall/records collected from the first two rounds. The food frequency questionnaire was administered last July through September, asking about the previous year. The correct comparison, which is the heart of the study, will be between the food frequency currently being administered and the past year's 16 recall/records. We also look forward to seeing whether last year's food frequency bears any resemblance to the one being collected this year. That is, for those who say their diets did not change, what is the reliability of the instrument?

Total mean energy intake from the 24 hour recall, day 1, day 2 and day 3 records bear a close resemblance to each other, but are significantly different from results obtained by the food frequency. The frequency value is 20-25 percent greater than the means of the 24 hour and the records.

Is there any food group that is overestimated by the food frequency instrument? Since the 24 hour recall and records appear, at this stage of the analysis, to be similar, we combined the four days in order to compare information from the records and the food frequency.

We classified the foods from the food frequency and records into eleven food groups. For each method, the food frequency and the records, we computed the percentage of the total calories that each of these eleven food groups contributed.

Comparing the baseline food frequency and six months of record data for each food group, there is a moderate correspondence between percentages of calories from each food groups, especially for sugar by probe, beverages, fruits and dairy products, as shown here by correlation coefficients in Table 2. If these results hold up in the final analysis, it will indicate that most respondents were able to generalize fairly effectively about their intake of some foods. The paired t-test indicate that there is a significant difference between the two instruments for most food groups, excepting vegetables and fats.

The percentages of total calories contributed by grains, meats, desserts, sugar and beverages were systematically higher from the food frequency than the recall records. The percentages of total calories contributed by dairy products, eggs, fruits, and fats were systematically lower from the food frequency than the recall records.

In an attempt to understand the overestimation of meat and the underestimation of fruit, we have added a few questions to the food frequency currently being used. We asked about how many times a week meat as a group and fruit as a group are usually eaten, after the food frequency is completed. The problem of over and under estimation may be partially due to the number of foods or slips of paper in each food group. For example, there are 21 different meat items in the food frequency. Perhaps there are too many, so that small errors in each item add up to a substantial error. However, 24 different vegetable items are asked in the food frequency yet results from this instrument are quite similar to the recall/records. The egg group is the smallest, just two slips. The difference in results here may be because eggs are used in such a variety of foods that respondents may have trouble estimating them in a food frequency type format.

The length of time respondents took to complete the food frequency questionnaire is similar across race, age, and sex categories. The range of minutes in Table 3 is fairly large for some groups. Based on our observation of some respondents, the concept of how often food is eaten, and how much, is new to some people. The first part of the sort, whether or not food is eaten and whether or not it is a seasonal pattern, moves along fairly rapidly. Decisions on how often, and especially on how much, took a longer period of time. We plan to get an estimate of the distribution of time required for the different portions of the food frequency tool. We also plan to analyze whether questioning for the usual serving size is useful, or whether one can estimate intakes just as well using sex/age standards.

During the initial interview, about 30 percent of the respondents indicated that they were on a "special" diet, as listed in Table 4. About 25 percent of this group said they were on a weight loss diet. We do not

know the degree to which they altered their usual diet, or how long they followed the "special" diet.

Table 5 includes the foods that the respondents said were most affected by these special diets included meat, dairy products, sweets and vegetables. Meat was mentioned more often in this circumstance. The meat group was one of the groups that showed the greatest disparity between the food frequency and record recall measures of intake. There may be an ongoing process in the pattern of meat consumption.

When the three day record is picked up by the interviewer the record is reviewed and the respondent asked how typical these days have been. For various reasons, the kind or amount of food, how often, or the time the food was eaten, the records were not typical for a number of respondents. Table 6 lists the percentage of respondents reporting a typical diet for rounds one and two.

The use of a special diet, change in the use of a specific food and typicality of the diet may or may not affect the degree to which the food frequency and the recall record agree. Our analysis will explore these relationships.

Conclusion

A final caution against relying on any of the figures in the latter tables. Preliminary data are presented to illustrate trends in the analysis and the methods. There is also the more basic question of validity. We really don't know how valid the series of recall/records are. It could be that a series of recall/records is a good indicator of a individual's average diet, and that a food frequency is a good estimate of an individual's usual diet. We are assuming that the terms are interchangeable. By this time next year we should be able to give you some indication whether or not that assumption is valid.

Figure 1. Basic Format Food Frequency Questionnaire

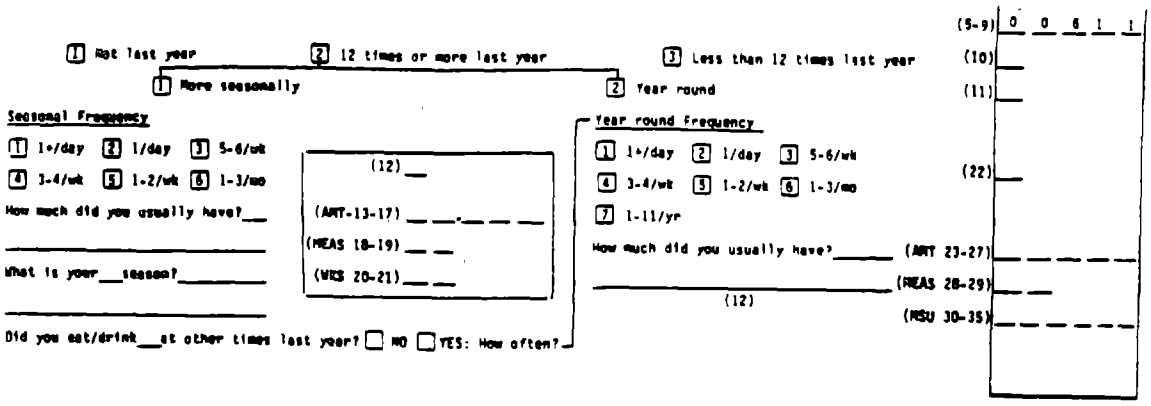


Table 1. **Mean Nutrient Intake Estimated from Food Frequency, Recall and Records, Preliminary Results**

	Food Frequency I (N=172)	Rounds I, II Records (N=234)				
		24 Hr	Day 1	Day 2	Day 3	Mean
Energy (kcal)	2808	2201	2111	2196	2223	2183
Protein (gm)	103	81	81	81	81	81
Fat (gm)	127	96	93	95	97	95
Carbohydrate (gm)	305	242	229	244	239	239

Food Frequency significantly different ($p < 0.1$) from records by paired t-test

Table 2. **Comparison of Percentage of Total Energy Contributed by Eleven Food Groups for Food Frequency and Recall plus Records: Preliminary Results**

Food Group	Percent of Total Calories			
	Food Frequency I	Recall and Records I & II	r	Paired t-test Significance
Grains	23.8	19.3	.37	.0000
Meats	21.0	15.5	.40	.0000
Desserts	7.3	4.5	.44	.0000
Sugar (probe)	1.7	1.1	.69	.0001
Dairy	10.8	14.1	.56	.0000
Beverages	9.0	7.0	.66	.0000
Eggs	2.1	2.4	.42	.0402
Fruits	6.9	10.7	.61	.0000
Vegetables	7.6	8.2	.35	.1535
Fats (butter, margarine, probe)	4.5	4.3	.31	.6571
Other fats	6.3	7.5	.25	.0116

Correlations are all significant at $p<.01$

Table 3 Time (in Minutes) to Complete Baseline Food Frequency by Race, Sex and Age, Continuing Participants As of Round Four Only (N=234): Preliminary Results

Demographic Group	N	Time (in minutes)		
		Minimum	Maximum	Mean
White				
Men				
23-34	36	34	160	75
35-50	29	40	120	72
	65			
Women				
23-34	37	48	120	73
35-50	36	15	150	73
	73			
Black				
Men				
23-34	21	50	145	78
35-50	24	45	119	71
	45			
Women				
23-34	25	47	120	78
35-50	24	30	110	75
	49			
Missing information for time	2			
Total	234			

Table 4. **Frequency of Special Diets Mentioned from Health Information Questions:**

Special Diets	Numbers of Times Mentioned
Weight loss	17
Pregnant/lactating	8
Low salt	7
Low cholesterol/fat	5
Increased calories	4
Vegetarian	4
Lactose avoidance	4
Low salt and low fat	3
Diabetic	3
Hypoglycemia	2
Gallbladder	2
Migraine	2
High fiber	1
Low fat and high fiber	1
Other	7
Total Special Diets	70
Respondents	234

Table 5. **Number of Continuing Participants as of Round Four Reporting Foods Eaten Less, More, or Not at All for Special Diets**

Meat	27
Dairy Products	25
Sweets/Sugar	20
Vegetables	19
Fruits/Juices	15
Non-Sweet Grain Products	12
Fish	11

Twelve other food groups were mentioned by fewer than ten respondents.

Table 6. **Percent Reporting Three-Day Record was Typical for Rounds One and Two, Continuing Participants as of Round Four Only (N=234)**

Typicality of 3-Day Record for	Percent Reporting Typical	
	Round 1	Round 2
Kinds of foods	75	75
Amounts of foods	78	78
Frequency of eating	83	86
When food eaten	82	88

Data-based Questionnaire Design: the NCI Survey

Gladys Block, Ph.D.

We at the National Cancer Institute have developed a dietary assessment questionnaire, making use of a large national survey and database as an integral part of the development approach. My purpose here is not to make a case for a particular instrument, but to demonstrate the value of the data-based approach (1). This approach is not unique. Others have used the percent contribution of a nutrient to guide food item selection, notably Jean Hankin (2), Pickle and Hartman (3), and Buzzard (4).

Several purposes guided the choice and development of the dietary assessment method. First and foremost, it must be a valid reflection of an individual's usual diet, since we want to make associations with individual health outcomes. This clearly means that 24-hour recalls are inappropriate, and in my view also means that dietary intake over a few consecutive days is inadequate, as well. And it must be brief, since only then will a dietary assessment be acceptable to many investigators. These requirements made a quantified frequency questionnaire the appropriate assessment method.

There are two broad aspects of a diet questionnaire: 1) the components of the questionnaire itself -- the food list and quantitations; and 2) the human response to it -- the ability to estimate frequency, for example. I will address the second briefly at the end of the paper. Most of what I will present will concern the first aspect, the development, improvement and testing of the components of the instrument itself. When these two aspects are kept separate and are tested separately, it is possible to maximize the potential performance of the instrument itself, before the vagaries of the human response are introduced. In this way it is possible to identify the sources of error, and target improvements more appropriately.

Within aspect #1 above, the instrument itself, there are three main components the food list, the nutrient composition associated with each food on the list, and the portion size assumption associated with each food on the list. Each of these three components was developed in a data-based manner.

The food list was selected by identifying those foods which were the main contributors to population intake, in the NHANES II survey. That survey collected 24-hour recalls from approximately 12,000 adults. For calories and each of the 17 nutrients on the NHANES II database, we summed the total nutrient consumed by the entire adult population, and then determined what percent of the total each food contributed. This work has been published in detail elsewhere (5,6). We then identified a list of foods which represents at least 90 percent of each of the 17 nutrients and energy on the database. For example, the list contains foods which represent 95 percent of the vitamin C, 97 percent of dietary cholesterol, and 93 percent of total calories consumed by the U.S. adult population. By capturing calories and 17 other major nutrients adequately, and by keeping the food items reasonably distinct, it is likely that other nutrients not now calculated by the program (or perhaps not even discovered) will be able to be assessed by this food list.

As indicated above, the list is comprehensive, that is, represents a high percentage of each of the nutrients. However, comprehensiveness is not enough. In order to make accurate nutrient estimates, precision is also necessary. For example, an item, "pasta", may be comprehensive but is not precise enough to enable you to distinguish between retinol (in pasta with cheese sauces), carotenes (in pasta with tomato sauce) or neither (in plain pasta). To as great a degree as possible consistent with time constraints, foods were kept separate and were grouped only if conceptually similar and similar in nutrient content per usual serving.

The nutrient composition of each listed food item was selected as follows. All of the NHANES II foodcodes which comprised a food item were grouped together. For example, there were 11 different codes of green beans, each of which could have a different content of, e.g., protein per 100 grams. How is one to choose among the different protein contents, in order to determine a single protein value to be used with the questionnaire line item, "green beans"? Clearly it is not appropriate simply to take the mean or the median of the 11 values, since that is greatly influenced by what happens to have been analyzed. The value must take into account the frequency of consumption of each of the items. We used the frequency-weighted median, that nutrient composition for which half of the consumers ate green beans with a higher protein content, and half ate green beans with a lower protein content. This approach minimizes the error more than does a weighted mean, and frequently chooses the value of the food eaten by the greatest number of respondents.

In addition to the food list and the nutrient composition of each food, the third component of a dietary assessment instrument is the portion size assumptions for each food. These were also developed in a data-based way, by examining the actual reported portion sizes among the 12,000 adults in NHANES II. The NHANES II survey is an invaluable source for such data, since three-dimensional models were used by the respondents to estimate portion sizes. We thus have large-population data using three-dimensional models, a unique resource. In addition, because the NHANES II population is so large we were also able to develop age- and sex-specific portion sizes for the foods on our questionnaire. This greatly improves the precision of our nutrient estimates, particularly for men and for older women whose usual portions differ quite substantially from "standard" portions.

To take account of individual food preferences and variability, the questionnaire also asks the respondent to indicate whether his usual portion is small,

medium or large with respect to a stated medium portion. This feature, unique to this instrument, adds additional precision in nutrient estimates.

The resulting questionnaire consists of a 98-item food list (printed on the front and back of a single page), and the nutrient composition and portion size arrays as discussed above. There is also an open-ended question which permits the respondent to identify other commonly-eaten foods which are not on the list; and questions on restaurant foods and on the frequency and types of fat added in cooking and at the table, all of which are used in the nutrient calculations.

We evaluated and perfected this list and its quantitations by testing how well it could approximate the values resulting from a detailed diet record. For this evaluation we used 50 one-day diet records calculated in the usual way, and then used our food list and quantitations as a brief scoring system. Nutrients were calculated by both methods, and the two estimates compared.

This approach is a test strictly of the instrument itself, before the addition of the uncertainty of the human ability to estimate frequency. It enabled us to evaluate the adequacy of the list, since if foods could not be coded on the food list the nutrient estimates would be low and the correlations poor. It also enabled us to evaluate the nutrient composition and portion size assumptions; if our quantitations were not adequate to reflect the diversity of foods and portion sizes in actual diets, again the correlations would be poor.

The results indicate that the list, nutrient composition and portion size assumptions are robust in the face of the diversity of the normal free-living human diet. When used as a brief coding system for the 50 diet records, the nutrient estimates correlated with those of the more detailed method at $r = 0.7$ to 0.9. Figure 1 shows scatterplots of the relationship between the record and food-list methods, for calories and vitamin A.

As indicated previously, the above evaluations test the degree to which the food list and quantitations could produce accurate estimates, if respondents could respond accurately about their diets. The performance of the questionnaire when actually administered to respondents has been tested in three field validations. In one, the reference method was a detailed diet history; in another, it was a seven-day diet record; and in the third, questionnaire results were compared with a known diet actually administered to the subjects. Results of these validations will be reported elsewhere (7-9). Preliminary results, however, indicate correlations of 0.7 or better in each of these validations, for factors as diverse as vitamin A, calcium, and percent of calories from fat.

These results suggest that a systematic data-based approach to questionnaire development may permit us to enhance the accuracy of nutrient assessments in some situations.

Bibliography

1. Block G, Hartman A, Dresser CM, Gardner L, Gannon J. Data-based diet questionnaire design and testing. Submitted for publication.

2. Hankin JH, Rawlings V, Nomura A. Assessment of a short dietary method for a prospective study on cancer. Am J Clin Nutr 1978; 31:355-9.

3. Pickle LW and Hartman AM. Indicator Foods for vitamin A assessment. Nutr Cancer 1985; 7:3-23.

4. Buzzard M, Personal communication; 1983.

5. Block G, Dresser CM, Hartman AM et al. Nutrient Sources in the American Diet: Quantitative data from the NHANES II Survey. I, Vitamins and Minerals. Am J Epidemiol 1985; 122:13-26.

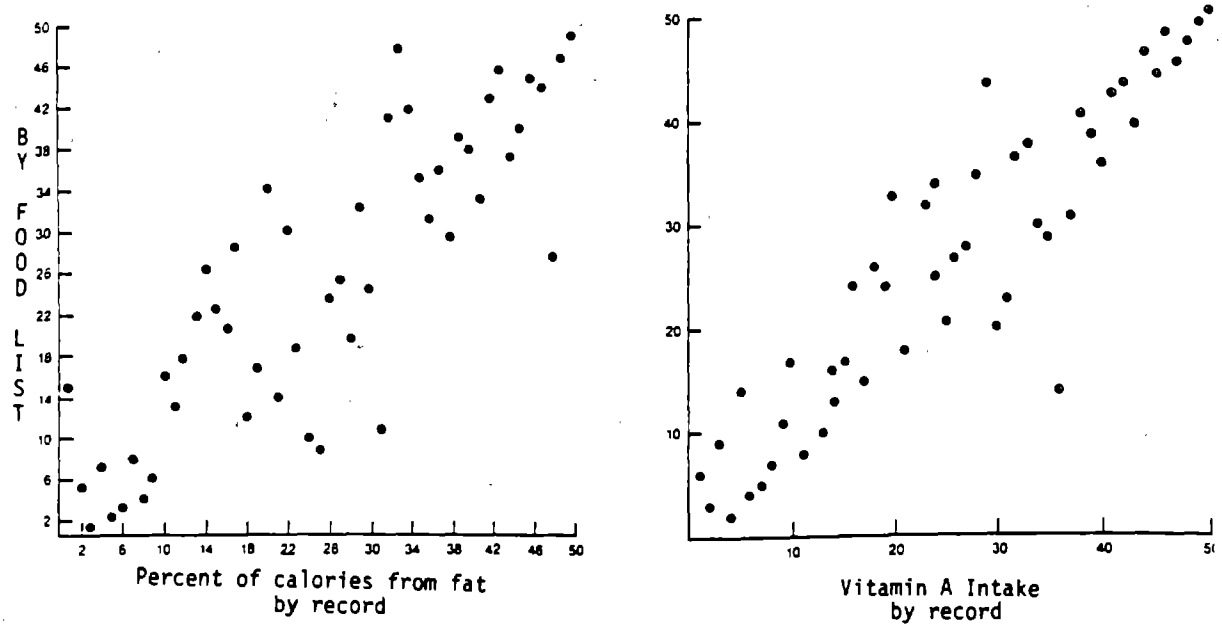
6. Block G, Dresser CM, Hartman AM et al. Nutrient Sources in the American Diet: Quantitative data from the NHANES II Survey. II, Macronutrients and Fats. Am J Epidemiol 1985; 122:27-40

7. Cummings SR, McHenry K, Block G, et al. Validation of two check-list methods for measuring dietary calcium intake of elderly women. Am J Clin Nutr, Submitted.

8. Block G, Jones Y, Hartman AH. Agreement of a quantified frequency questionnaire with known dietary intake. In preparation.

9. Burgess M, Block G, Bowen P, Hartman AH. A comparison of serum, diet interview and self-administered questionnaire for carotenoids and total vitamin A. In preparation.

Figure 1



Data-based Questionnaire Design: the NCI Survey

Gladys Block, Ph.D.

We at the National Cancer Institute have developed a dietary assessment questionnaire, making use of a large national survey and database as an integral part of the development approach. My purpose here is not to make a case for a particular instrument, but to demonstrate the value of the data-based approach (1). This approach is not unique. Others have used the percent contribution of a nutrient to guide food item selection, notably Jean Hankin (2), Pickle and Hartman (3), and Buzzard (4).

Several purposes guided the choice and development of the dietary assessment method. First and foremost, it must be a valid reflection of an individual's usual diet, since we want to make associations with individual health outcomes. This clearly means that 24-hour recalls are inappropriate, and in my view also means that dietary intake over a few consecutive days is inadequate, as well. And it must be brief, since only then will a dietary assessment be acceptable to many investigators. These requirements made a quantified frequency questionnaire the appropriate assessment method.

There are two broad aspects of a diet questionnaire: 1) the components of the questionnaire itself -- the food list and quantitations; and 2) the human response to it -- the ability to estimate frequency, for example. I will address the second briefly at the end of the paper. Most of what I will present will concern the first aspect, the development, improvement and testing of the components of the instrument itself. When these two aspects are kept separate and are tested separately, it is possible to maximize the potential performance of the instrument itself, before the vagaries of the human response are introduced. In this way it is possible to identify the sources of error, and target improvements more appropriately.

Within aspect #1 above, the instrument itself, there are three main components the food list, the nutrient composition associated with each food on the list, and the portion size assumption associated with each food on the list. Each of these three components was developed in a data-based manner.

The food list was selected by identifying those foods which were the main contributors to population intake, in the NHANES II survey. That survey collected 24-hour recalls from approximately 12,000 adults. For calories and each of the 17 nutrients on the NHANES II database, we summed the total nutrient consumed by the entire adult population, and then determined what percent of the total each food contributed. This work has been published in detail elsewhere (5,6). We then identified a list of foods which represents at least 90 percent of each of the 17 nutrients and energy on the database. For example, the list contains foods which represent 95 percent of the vitamin C, 97 percent of dietary cholesterol, and 93 percent of total calories consumed by the U.S. adult population. By capturing calories and 17 other major nutrients adequately, and by keeping the food items reasonably distinct, it is likely that other nutrients not now calculated by the program (or perhaps not even discovered) will be able to be assessed by this food list.

As indicated above, the list is comprehensive, that is, represents a high percentage of each of the nutrients. However, comprehensiveness is not enough. In order to make accurate nutrient estimates, precision is also necessary. For example, an item, "pasta", may be comprehensive but is not precise enough to enable you to distinguish between retinol (in pasta with cheese sauces), carotenes (in pasta with tomato sauce) or neither (in plain pasta). To as great a degree as possible consistent with time constraints, foods were kept separate and were grouped only if conceptually similar and similar in nutrient content per usual serving.

The nutrient composition of each listed food item was selected as follows. All of the NHANES II foodcodes which comprised a food item were grouped together. For example, there were 11 different codes of green beans, each of which could have a different content of, e.g., protein per 100 grams. How is one to choose among the different protein contents, in order to determine a single protein value to be used with the questionnaire line item, "green beans"? Clearly it is not appropriate simply to take the mean or the median of the 11 values, since that is greatly influenced by what happens to have been analyzed. The value must take into account the frequency of consumption of each of the items. We used the frequency-weighted median, that nutrient composition for which half of the consumers ate green beans with a higher protein content, and half ate green beans with a lower protein content. This approach minimizes the error more than does a weighted mean, and frequently chooses the value of the food eaten by the greatest number of respondents.

In addition to the food list and the nutrient composition of each food, the third component of a dietary assessment instrument is the portion size assumptions for each food. These were also developed in a data-based way, by examining the actual reported portion sizes among the 12,000 adults in NHANES II. The NHANES II survey is an invaluable source for such data, since three-dimensional models were used by the respondents to estimate portion sizes. We thus have large-population data using three-dimensional models, a unique resource. In addition, because the NHANES II population is so large we were also able to develop age- and sex-specific portion sizes for the foods on our questionnaire. This greatly improves the precision of our nutrient estimates, particularly for men and for older women whose usual portions differ quite substantially from "standard" portions.

To take account of individual food preferences and variability, the questionnaire also asks the respondent to indicate whether his usual portion is small,

medium or large with respect to a stated medium portion. This feature, unique to this instrument, adds additional precision in nutrient estimates.

The resulting questionnaire consists of a 98-item food list (printed on the front and back of a single page), and the nutrient composition and portion size arrays as discussed above. There is also an open-ended question which permits the respondent to identify other commonly-eaten foods which are not on the list; and questions on restaurant foods and on the frequency and types of fat added in cooking and at the table, all of which are used in the nutrient calculations.

We evaluated and perfected this list and its quantitations by testing how well it could approximate the values resulting from a detailed diet record. For this evaluation we used 50 one-day diet records calculated in the usual way, and then used our food list and quantitations as a brief scoring system. Nutrients were calculated by both methods, and the two estimates compared.

This approach is a test strictly of the instrument itself, before the addition of the uncertainty of the human ability to estimate frequency. It enabled us to evaluate the adequacy of the list, since if foods could not be coded on the food list the nutrient estimates would be low and the correlations poor. It also enabled us to evaluate the nutrient composition and portion size assumptions; if our quantitations were not adequate to reflect the diversity of foods and portion sizes in actual diets, again the correlations would be poor.

The results indicate that the list, nutrient composition and portion size assumptions are robust in the face of the diversity of the normal free-living human diet. When used as a brief coding system for the 50 diet records, the nutrient estimates correlated with those of the more detailed method at $r = 0.7$ to 0.9. Figure 1 shows scatterplots of the relationship between the record and food-list methods, for calories and vitamin A.

As indicated previously, the above evaluations test the degree to which the food list and quantitations could produce accurate estimates, if respondents could respond accurately about their diets. The performance of the questionnaire when actually administered to respondents has been tested in three field validations. In one, the reference method was a detailed diet history; in another, it was a seven-day diet record; and in the third, questionnaire results were compared with a known diet actually administered to the subjects. Results of these validations will be reported elsewhere (7-9). Preliminary results, however, indicate correlations of 0.7 or better in each of these validations, for factors as diverse as vitamin A, calcium, and percent of calories from fat.

These results suggest that a systematic data-based approach to questionnaire development may permit us to enhance the accuracy of nutrient assessments in some situations.

Bibliography

1. Block G, Hartman A, Dresser CM, Gardner L, Gannon J. Data-based diet questionnaire design and testing. Submitted for publication.

2. Hankin JH, Rawlings V, Nomura A. Assessment of a short dietary method for a prospective study on cancer. Am J Clin Nutr 1978; 31:355-9.

3. Pickle LW and Hartman AM. Indicator Foods for vitamin A assessment. Nutr Cancer 1985; 7:3-23.

4. Buzzard M, Personal communication; 1983.

5. Block G, Dresser CM, Hartman AM et al. Nutrient Sources in the American Diet: Quantitative data from the NHANES II Survey. I, Vitamins and Minerals. Am J Epidemiol 1985; 122:13-26.

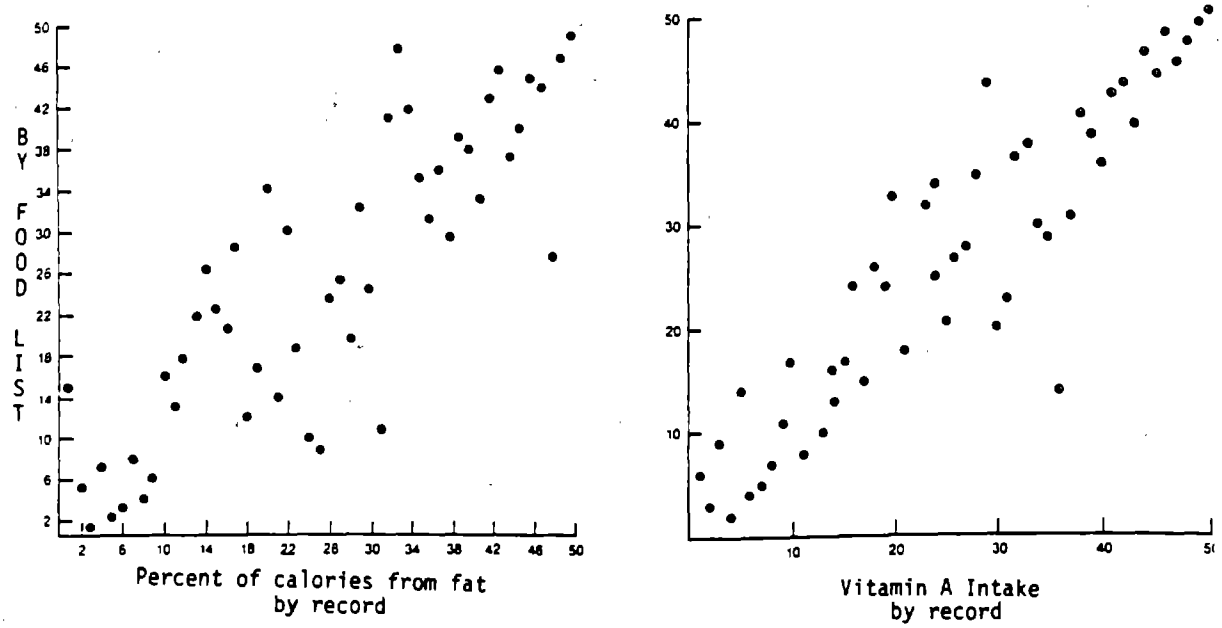
6. Block G, Dresser CM, Hartman AM et al. Nutrient Sources in the American Diet: Quantitative data from the NHANES II Survey. II, Macronutrients and Fats. Am J Epidemiol 1985; 122:27-40

7. Cummings SR, McHenry K, Block G, et al. Validation of two check-list methods for measuring dietary calcium intake of elderly women. Am J Clin Nutr, Submitted.

8. Block G, Jones Y, Hartman AH. Agreement of a quantified frequency questionnaire with known dietary intake. In preparation.

9. Burgess M, Block G, Bowen P, Hartman AH. A comparison of serum, diet interview and self-administered questionnaire for carotenoids and total vitamin A. In preparation.

Figure 1



USE OF A SHORTENED NUTRIENT DATA BANK TO ANALYZE NATIONAL SURVEY DATA

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

Today I'd like to share with you some of the results of a project I undertook to examine relationships between dietary intake and trace element status in the United States. Due to time constraints, I will present only the first portion, which concerns the use of a shortened nutrient data base to analyze national survey data. The most recent nationwide survey to collect both dietary and biochemical measures of trace element status is the Second National Health and Nutrition Examination Survey (NHANES II). This survey was conducted between 1976 and 1980 by the National Center for Health Statistics (NCHS) (1). Approximately 20,000 individuals were examined, with a response rate of 73%. The age range was 6 months to 74 years, from 64 geographic locations, with certain types of individuals oversampled. I chose to work with a subset of the NHANES II population, due to limitations of time and funding, and selected women 18-24 years of age, since this group is thought to be at risk for trace element deficiency (2-3).

There were 1066 women 18-24 years old in NHANES II. Of these, 56 were pregnant and 14 were lactating; 20 had partially incomplete dietary records; 127 were Black and 26 were other non-White races. Approximately one-third were oral contraceptive users, which did not affect intake, but was important when looking at biochemical parameters.

The dietary portion of NHANES II included two general components. First, a 24-hour recall of all foods consumed the previous day, including an estimated portion size. These recalls were collected for weekdays only. Supplement use was not quantified. In addition, a food frequency questionnaire was administered, using 26 food groups. Although I have done some work quantifying nutrient intake based on the food frequency, I will have time today to talk about the recall portion only.

For the 24-hour recall, there were approximately 2600 different foods reported by the entire survey population, and young women reported 1267 of these foods (or about half). The staff at NCHS compiled a nutrient data base for these foods, using data from Handbook 8, the food industry, and other nutrient data bases (4).

The 24-hour recall diets were analyzed for their content of the 18 nutrients shown in Table 1, by the staff at NCHS. I have also shown the range of the percent of missing values for each group of nutrients. As you can see, this percent is quite high for some nutrients, especially the fat components, and could lead to an under-estimate of the intake of these nutrients. You will notice that there is no information on the trace elements zinc and copper. Although iron is present, other nutrients that might affect iron status measures, such as folacin and vitamin B12, are not present. Finally, there is no information on the dietary factors that might affect trace element bioavailability (such as fiber and phytate). As Judy Turnlund pointed out, we are realizing that the total intake of a trace

element may be less important than its form, and its interactions with other food components.

Therefore, I decided to expand the nutrient data base for these diets, using the UCB Minilist (Table 2). The Minilist is a nutrient data base with relatively few foods (235), but a large number of nutrients (53). In addition, there are no missing values - when analytical values are not available, the value is either calculated or imputed. It was developed by Jean Pennington in the early 1970's (5). Over the last few years, I've updated the nutrient values, using USDA's Nutrient Data Base for Standard Reference (release 3) and current literature values, and added a few new foods. In addition, I've added five nutrients (two of which are actually food components, but I will use nutrients and food components as interchangable terms) - dietary fiber (which is total dietary fiber, using Southgate's method (6)), phytate (using the ion exchange method of Oberleas and Harland (7)), meat/fish/poultry iron (which has higher availability than other forms of iron), MFP protein (which is an enhancer of trace element absorption), and added (fortification) iron (which may have different availability than organic iron). Since I chose not to use the amino acid values, the total number of nutrients on my version of the Minilist was 36 (shown in Table 2).

Now the problem was to substitute Minilist foods for those 1267 foods actually reported by the NHANES II young women. Since it was not pratical to do the substitutions manually for over 16000 items (each of the women reported an average of about 14 food items), I developed a cross-reference index that could be used by a program to automatically substitute codes. Table 3, for example, shows three beef stews and a lamb stew, all of which are considered Minilist code # 371 (homemade beef stew).

Several guidelines were followed when matching foods:

1. Foods that were frequently consumed, consumed in large quantities, or with high nutrient densities, were matched with particular care. In some cases, the foods on the Minilist were modified.

2. Of the nutrients on both data bases, nine were selected as being of particular interest (iron, protein, vitamin C, carbohydrate, fat, calcium, vitamin A, thiamin, and cholesterol). Reports were generated showing differences in these nutrients between the NHANES II food, and the substituted Minilist food.

3. The 18 nutrients that were not on the NHANES II nutrient data base had to be considered when devising substitutions. For example, corn, oat, rice, and wheat products were never substituted for each other, due to their differing fiber and phytate contents. Similarly, whole grain products were not substituted for their refined equivalents, or vice versa.

4. Since the Minilist had been updated more recently than the NHANES II data base, some differences in nutrient values were to be expected. Often, these differences were due to better methodology in determining nutrient values. In other cases, however, differences were due to changes in product formulations, and the Minilist values were changed to reflect nutrient values appropriate for the time in which the survey was conducted.

Direct substitutions of this type often were not satifactory, however. I addressed this problem in two ways. First, I directed the program to adjust the serving size based on the caloric content of the original food. For example, if tomato sauce was the reported food, and canned tomatoes was the Minilist food, the serving size would be doubled, since the nutrient content of canned tomatoes is approximately half that of tomato sauce. These caloric conversions had to be checked, of course, but 99% of the time, the result was more satisfactory than a direct substitution. Table 3 shows that both types of canned beef stew had slightly more calories per 100 grams than the Minilist equivalent (homemade beef stew), so the serving size will be adjusted downward by approximately 10% (the factor shown).

The second method of increasing the precision of substitutions was to add a recipe file to the Minilist. For example, beef cabbage rolls were nutritionally different enough from beef stew, that a direct subsititution was not satisfactory, even when adjusted for caloric differences. Thus, a recipe was developed, and whenever the code for cabbage rolls was encountered, the ingredients shown in Table 4 (all of which were on the Minilist) were substituted instead. The proportions add up to more than one due to evaporative loss during cooking. I developed 163 recipes, which were substituted for 490 food items in these young women's diets.

There were 16 nutrients in common on the two nutrient data bases. When 1066 24-hour recall diets were analyzed using the two data bases, differences in mean nutrient intakes were small for most nutrients - under 2% for the proximate (energy, protein, fat, and carbohydrate), and under 10% for everything else except sodium. For most nutrients, the Minilist was higher, which is probably due to the missing values on the NHANES II data base - implying that the Minilist may be more accurate. Sodium was the worst case, differing by over 15%. None of the recipes contained sodium, so I did not expect the values to match those from NHANES II. Furthermore, NHANES II did not measure discretionary salt intake so it is not possible to estimate total sodium intake.

The nutrients of particular interest for this project were iron, vitamin C, and protein, all of which differed by less than 5%.

Although the differences in the mean values were small, they were often statistically significant, due to the large sample size. A paired t-test found significant differences (a p-value of less than .05) for all nutrients except energy, fat, and thiamin. However, I think it is important to examine the power of a test of this type - in this case, the likelihood of detecting a 5% difference is over 98% for all nutrients. Given the other errors inherent in any diet analysis methodology, I believe that a 5% error is well within the acceptable range. Thus, statistical significance may not mean pratical significance in this case.

Correlation coefficients between the nutrients on the two data bases were obtained as a measure of the similarity of the nutrient values in individual diets, as opposed to the similarity of the overall mean. All correlations were above .76, and most (all but 4 of the 16) were above .9.

I investigated some of the larger differences in individual diets. Since I was particularly interested in trace elements, I examined two diets for which the iron content differed by more than 10 mg. One of the diets contained 56 grams of hog

liver, which has an iron content of 17 mg on the NHANES II data base. The substituted food item, beef liver, contains only 4.5 mg, giving a difference of over 11 mg. Obviously, if the study population consumed large quantities of hog liver, it would be necessary to add it to the data base. It is worth noting that the latest USDA value for hog liver is 10 mg per 56 grams (8), so neither data base is correct.

The other difference of over 10 mg appeared in a diet containing two commercial diet bars. The iron value for this food item is missing on the NHANES II data base, while the Minilist substitution is a recipe which contains 25% of the USRDA for iron, as specified on the product label. A total difference of 9 mg iron was attributable to this food item, and other minor differences caused the total difference to exceed 10 mg. Thus, in this case, the Minilist substitution was more accurate than the NHANES II data base value.

Mean intakes of 18 other nutrients, not on the NHANES II nutrient data base were also calculated. At this point, pregnant and lactating women were dropped, since their intakes were significantly higher than that of the non-pregnant, non-lactating women. Of particular interest for this project were the trace elements zinc and copper, vitamins which might affect measures of trace element status (folacin, vitamin B12, and vitamin B6), and factors which might affect bioavailability (dietary fiber, phytate, MFP iron, MFP protein, and added iron). Zinc intake was 54% of the RDA (9), while copper intake was 58% of the minimum safe and adequate recommendation. Iron intake was also 58% of the recommendation, leading to the prediction that marginal trace element status should be common in this population of young women.

Since these nutrient values could not be validated by comparison to NHANES II, comparison to other reported values was the alternative. There have not been any national surveys that reported zinc and copper values, but several smaller surveys obtained estimates in the same range as the Minilist (2, 10-11). The energy content of the diets in these other surveys was usually higher than the 1687 kcal found for the NHANES II young women, and the trace element content is correspondingly somewhat higher. However, the Nationwide Food Consumption Survey (NFCS) reported similar energy intakes (1621 kcal for women 19-22), and their estimates of vitamins B12 and B6 were virtually identical to those obtained using the Minilist (12). I consider these numbers an excellent validation of the Minilist concept, especially since I was not particularly attempting to match these nutrients. The NFCS data base contains over 3700 foods, yet I was able to obtain similar estimates of mean intakes of these vitamins with only 235 foods.

There are very few estimates in the literature of the dietary fiber and phytate intakes of individuals, and none that I found specifically for young women. The phytate content is in the range of those reported for hospital diets (13). If the caloric content is adjusted, the dietary fiber estimate closely matches an estimate based on a simulated American diet (14), but is lower than an estimate obtained from the diets of Canadian university women (2) (who very likely consume more whole grains, fresh vegetables, etc., than a cross-section of American young women).

Estimates of animal sources of iron and protein are also hard to find, but the percents reported recently by NCHS (15), based on NHANES I data are very similar to those obtained using the Minilist.

In conclusion, I found that the precision of a shortened nutrient data base could be substantially increased by the use of a caloric adjustment factor, and recipes. The mean intakes of nutrients calculated using both a large and a shortened data base were similar for all nutrients except sodium. Differences in fat components were probably due to missing values on the large data base. High correlations indicate that the small differences in the means are not due to cancelling plus and minus differences, and that the calculated nutrient intakes agree for individuals as well as for the population. For nutrients and food components not present on the NHANES II data base, there was good agreement with published values for similar populations. Thus, I felt confident that the Minilist was giving a good estimate of the nutrient and food component intakes, based on the 24-hour recall foods reported by the young women.

REFERENCES

1. National Center for Health Statistics: Plan and operation of the second National Health and Nutrition Examination Survey 1976-80. Hyattsville, MD: U.S. Government Printing Office, 1981. (Vital and health statistics. Series 1, no 15.)

2. Gibson, R.S., and Scythes, C.A.: Trace element intakes of women. Br J Nutr 48:241, 1982.

3. National Center for Health Statistics: Dietary intake source data: United States 1976-80. Washington, DC: U.S. Government Printing Office, 1983. (Vital and health statistics. Series 11, no 231).

4. National Center for Health Statistics: Public use data tape documentation: NHANES II A. Model gram and B. Nutrient composition (Tapes no 5702 and 5703). Hyattsville, MD: National Center for Health Statistics, 1982.

5. Pennington, J.A.: Dietary nutrient guide. Westport, CT: Avi Publishing Co, 1976.

6. Southgate, D.A.T., Bailey, B., Collinson, E., and Walker, A.F.: A guide to calculating intakes of dietary fibre. J Hum Nutr 30:303, 1976.

7. Harland, B.F., and Oberleas, D.: A modified method for phytate analysis using an ion-exchange procedure: application to textured vegetable proteins. Cereal Chem 54:827, 1977.

8. United States Department of Agriculture: Composition of foods. Pork and pork products. Agriculture Handbook No. 8-10. Washington, DC: U.S. Government Printing Office, 1983.

9. National Academy of Sciences: Recommended dietary allowances. 9th ed. Washington, DC: National Academy of Sciences, 1980.

10. Holden, J.M., Wolf, W.R., and Mertz, W.: Zinc and copper in self-selected diets. J Am Diet Assoc 75:23, 1977.

11. Sempos, C.T., Johnson, N.E., Smith, E.L., and Gilligan, C.: A two-year dietary survey of middle-aged women: Repeated dietary records as a measure of usual intake. J Am Diet Assoc 84:1008, 1984.

12. Pao, E.M., Mickle, S.J., and Burk, M.C.: One-day and 3-day nutrient intakes by individuals--Nationwide Food Consumption Survey findings, spring 1977. J Am Diet Assoc 85:313, 1985.

13. Ellis, R., Morris, E.R., Hill, A.D., and Smith, J.C.: Phytate:zinc molar ratio, mineral, and fiber content of three hospital diets. J Am Diet Assoc 81:26, 1982.

14. Ahrens, E.H., and Boucher, C.A.: The composition of a simulated American diet. J Am Diet Assoc 73:613, 1978.

15. National Center for Health Statistics: Diet and iron status, a study of relationships: United States, 1971-74. Washington, DC: U.S. Government Printing Office, 1982. (Vital and health statistics. Series 11, no 229).

TABLE 1

NHANES II NUTRIENT DATA BASE

PROXIMATE:
Energy, protein, fat, carbohydrate.
Less than 1% missing values.

MINERALS:
Calcium, phosphorus, iron, sodium, potassium.
Missing values range from .7% to 6.8%.

VITAMINS:
Thiamin, riboflavin, niacin, vitamin A, vitamin C.
Missing values range from 2.0 to 4.1%.

FAT COMPONENTS:
Saturated fat, oleic acid, linoleic acid, cholesterol.
Missing values range from 13.1 to 14.8%.

TABLE 2

MINILIST NUTRIENT DATA BASE

PROXIMATE:
Energy, Protein, Fat, Carbohydrate, Ash.

MINERALS:
Iron, Zinc, Copper, Calcium, Phosphorus,
Magnesium, Potassium, Sodium, Iodine.

VITAMINS:
Thiamin, Riboflavin, Niacin, Pantothenic acid,
Folacin, Biotin, Vitamins A, C, E, D, B6, B12.

FAT COMPONENTS:
Saturated fat, Polyunsaturated fat, Cholesterol.

OTHER COMPONENTS:
Crude fiber, Dietary fiber, Phytate, Sucrose, Added iron,
Meat/fish/poultry protein, Meat/fish/poultry iron.

TABLE 3

CROSSREFERENCE INDEX
NHANES II FOODS TO MINILIST FOODS

NHANES II CODE	MINILIST CODE	FACTOR	NAME
371	371	1.00	Beef Stew, Homemade
372	371	0.89	Beef Stew, Canned
92500	371	0.91	Beef Stew, Canned
35007	371	1.00	Lamb Stew
92509	9101		Beef Cabbage Rolls

TABLE 4

SAMPLE RECIPE
BEEF CABBAGE ROLLS
RECIPE # 101

MINILIST CODE	FACTOR	FOOD
370	.2	Hamburger
512	.2	Cabbage
1872	.3	Rice, white
2284	.6	Tomatoes, canned

SELENIUM CONTENT OF SELECTED FOODS BASED ON AN EVALUATION OF THE QUALITY OF PUBLISHED DATA. A.S. Greenland, J.M. Holden, and W.R. Wolf (SPON: G.R. Beecher). USDA-ARS, BHNRC, Nutrient Composition Laboratory, Beltsville, MD 20705.

References published since 1960 which report analyses of selenium (Se) in foods were collected and evaluated according to five criteria: number of samples, analytical method, sample handling procedures, sampling plan, and quality control. Studies were grouped by food item and scored for each of the criteria which had been developed specifically for evaluating the quality of published Se data. Scores assigned to each study were averaged to yield an index. The various indices for acceptable studies were summed to determine a "confidence code," intended to indicate the relative degree of confidence the user can have in the mean value for each food item. Mean, minimum and maximum Se values and references have been compiled for more than 120 food items. Food items were selected for evaluation based upon ranked contributions to Se intake in American diets. Foods were ranked by multiplying Se concentration in the food times the amount consumed in the 1977-78 Nationwide Food Consumption Survey. The five highest ranked foods (beef, white bread, eggs, chicken, and pork) contribute one-half of the Se in American diets.

The data for two-third of the evaluated foods are of good quality. The evaluation of nutrient data by objective criteria reveals those frequently consumed foods for which poor quality or no data have been reported.

Development of a Food Hierarchy to Capture Food Intake Data for Nutrient Analyses

Diane Feskanich and I. Marilyn Buzzard
Nutrition Coding Center, University of Minnesota

An interactive approach of food specification through a progressively detailed selection of food descriptions is utilized in the Dietary Data Collection system, a micro-computer based food coding system for dietary analysis being developed at the Nutrition Coding Center. Foods are being structured into a hierarchical tree design using common food names and categories. The food hierarchy is not a transcription of any one nutrient data bank; rather, it will contai most foods as eaten in North America and will incorporate much of the food detail found in the USDA Standard Reference and NCC data bases. A pathway on the food hierarchy progresses from a first level root node through from one to six levels of increasing food specificity. Pathway choices consider such detail as the processing method, source of the food, added sweeteners or salt, and brand name of the product. Each node in the pathway is represented by a one to four letter key, and a pathway is described by a series of keys. All four letter keys, which consist of the first four letters of the food name, will be maintained in an index to permit direct entry to these foods on any level of the hierarchy. Food selection is further simplified by the cross-referencing of foods under various names and in multiple pathways. Enhancement of the food description is possible through the specification of a preparation method or modification of recipe ingredients. Choice of a preparation method may add foods, such as fat and salt, to the basic food selected. The foods used in preparation or those modified in a recipe are accessed through key lists which permit transfer between food pathways in the hierarchy. It is anticipated that use of the hierarchy for coding foods from dietary intake data will enhance standardization of data collection and increase the accuracy of the data.

NUTRIENT DATA BANK APPLICATIONS IN DENTAL EDUCATION

Lois E. Miller, R.D.
Department of Biochemistry and Nutrition
University of Missouri-Kansas City
School of Dentistry

It is understood that intake of sucrose-containing foods and beverages promotes the development of dental plaque and the organic acids that cause demineralization of tooth enamel. Ingestion of the more retentive forms of sweets and an increased frequency of ingestion, especially between meals, have been shown to be the primary factors contributing to higher rates of caries formation. Due to their role in plaque development, the sucrose-containing foods have also been implicated as a contributing factor to gingivitis. In addition, recent animal research has revealed that certain nutrient deficiencies will compromise immune responses and the integrity of periodontal structures. A computer program was developed to analyze the diets of patients of our university dental clinics, the primary sources of nutrient data being USDA publications and manufacturers' communications. The goal was to promote a better understanding of the role of diet in oral health and disease. Five-day diet records of patients are analyzed. The resulting printouts summarize the intakes of sweets, categorizing them with respect to retentive properties and timing of consumption. They also give an estimate of the teaspoon amounts of sucrose contained in the designated sweets. An evaluation of the intake of twelve nutrients is made with comparisons to the RDA's. A second evaluation of adequacy of nutrient intake is based on food group guidelines. Feedback from patients and student counselors indicates that the computer-assisted analysis is a useful tool for assessing individual eating habits and focusing on ways to alter them to achieve better oral health.

Title: The Agriculture Canada Nutrient Assessment Program

Authors: Joan Smith, Linda Robbins, Lina Robichon-Hunt
Organization: Agriculture Canada, Ottawa, Ontario

In response to Agriculture Canada's recent increased commitment to promoting the nutritional aspects of Canadian foods, the Agriculture Canada Nutrient Assessment Program (AGNAP) was developed to consolidate the food and nutrition data used within the Department. This computer program integrates Health and Welfare Canada's Canadian Nutrient File (CNF) and Recommended Nutrient Intake (RNI) data with Statistics Canada's Family Food Expenditure, Apparent Food Consumption and Food Price data. AGNAP enables Agriculture Canada to generate the following kinds of information: summaries of the apparent nutrient intake of various groups in the Canadian population; nutrient profiles of recipes and menus together with comparisons of these profiles with the RNI; rankings of foods and food groups on the basis of nutrient content or cost; and the Nutritious Food Basket, a food costing plan for Canadians. Agriculture Canada interprets this information and uses it in research and consumer publications.
The Department also uses AGNAP to provide information to other government departments and outside organizations.

THE FDA'S FACTORED FOOD VOCABULARY (FFV)

E.C. Smith*, J.H. Holden**, A. McCann*, D. Soergel*** and P.C. Wiley***, *CFSAN, Food and Drug Administration, Washington, D.C. 20204, **Nutrient Composition Lab, ARS, USDA, Beltsville, Md. 20705, ***University of Maryland, College Park, Md. 20742.

The lack of a standardized vocabulary for describing food products is a problem which faces the Center for Food Safety and Applied Nutrition (formerly Bureau of Foods) of the FDA, which deals with research, regulatory, and policy problems related to the food we eat.

Scientists and information specialists have developed a standardized vocabulary for the description of food products to be used by the FDA-CFSAN to address the problem. The Factored Food Vocabulary (FFV) is currently in the pilot-test stage. It will provide a standardized but flexible retrieval-oriented language for describing foods and food products. This system provides the capability for coding eleven aspects or "factors" of foods, including product type, food source, preservation method and food contact surface. Each factor is hierarchically organized. The FFV provides the backbone which can be overlaid with various types of data pertinent to foods, e.g., nutrient data, toxicology data or demographic data. The information needs of the FDA are diverse; the Factored Food Vocabulary will permit the integration of information from various data bases, such as the National Food Consumption Survey, the Food Labeling and Product Surveillance (FLAPS) files and NHANES II, to satisfy these diverse needs and to analyze data more effectively in the future.

Calculated sodium and potassium values compared to chemical analysis and urinary excretion.

Phyllis J. Stumbo, Ph.D., R.D. and Catherine A. Chenard, R.D., General Clinical Research Center, University of Iowa, Iowa City, IA

To provide subjects with a constant intake, research dietitians must furnish a diet containing a specified amount of one or more nutrients over a period of time and verify the diet's nutrient content. We compared thirty-one diets' sodium and potassium content as calculated from USDA food composition tables and manufacturer's data with analyzed values. Sodium content ranged from 37-124 mEq for calculated values and 42-136 mEq for analyzed values. Calculated potassium content ranged from 46-157 mEq and analyzed potassium from 35-170 mEq. To evaluate the variability of identical diets we compared the sodium and potassium analyses for 20 duplicate diets prepared on two different study days.

In addition to evaluating dietary data, research dietitians must monitor subjects' compliance to a constant diet. Urinary sodium and potassium excretion is sometimes used to assess dietary adherence. To determine whether intake correlated with excretion, we compared 31 subjects urinary sodium and potassium excretion on the third day of a constant diet with sodium and potassium intake for this day.

DEVELOPMENT OF A COMPUTER PROGRAM TO ANALYZE CHINESE DIETS. Christine S. Wilson and Linda C. Koo, Departments of Epidemiology & International Health, University of California, San Francisco, and Community Medicine, University of Hong Kong.

For epidemiologic purposes nutrient intakes of Chinese in Taiwan and San Francisco were determined from self-reported dietary data collected by one of us (LCK). An existing computer program utilizing a minilist of 230 representative U.S. foods and 48 of their nutrients was modified and expanded to analyze these data. Two-hundred thirty-one foods recorded by Chinese subjects were coded separately and added to the program, for a total of 461 foods. All nutrient data for the added foods were obtained from published sources, chiefly the FAO/WHO Food Composition Table for Use in East Asia. The nutrients selenium, oleic, linoleic and linolenic acid were added to data on all the foods. Twelve amino acids, biotin and sucrose, in the original program, were omitted, as were total saturated and unsaturated fatty acids. A 35-nutrient data base for all foods in the expanded program resulted, which has been used to analyze nutrient intake differences between Taiwan and San Francisco.

Cooperation of the Quantitative Anthropology Laboratory, University of California, is gratefully acknowledged.