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