A CROSS REFERENCE OF HANES I AND NFCS (1978) NUTRIENT COMPOSITION DATA: ANALYSIS AND EVALUATION

John Stanton and Bonnie Sherr
Institute for Research in Food Consumption & Nutrition
Temple University, Philadelphia, PA

Introduction:

The HANES I and NFCS surveys have given researchers the opportunity to assess American eating habits. Each study presents a complete and comprehensive picture of how much and what foods are consumed in the U.S. Since these studies were conducted by different government agencies, they were not designed with an easy mechanism to compare between studies. Two major obstacles make specific comparison difficult. First, the item codes assigned to each food consumed were different. Second, the philosophy of coding was quite different.

It is not the intent of this paper to review the data collection philosophy of both studies. However, to summarize, HANES I broke down each item into the smallest unit foods, while NFCS used standard recipes. For example, if one had fruit flavored yogurt, in HANES I plain yogurt and fruit preserves were coded, whereas in NFCS the item was coded as fruit yogurt.

For a variety of research questions it would be useful to compare across the studies. For example, did people eat more of or less of certain foods in one study than in the other? Or if one was to assign codes not currently in either data base, such as an indication to show plant or animal source of the food item, it only need be done once. Furthermore, specific nutrient composition data available to only one survey would subsequently be accessible from both (eg. B6, B12, Mg in NFCS: Na, K in HANES I).

It should be apparent that a method to compare across studies would be useful. The method we chose is referred to as a correspondence method, or developing a correspondence file.

The correspondence file stores, for each HANES I food item, the corresponding (or equivalent) NFCS food item. Since the system is not symmetric, for each NFCS item, the corresponding HANES I food item is also determined.

For example, whole fluid milk, item 1321 in HANES I, was matched to whole fluid milk in NFCS, item 1111100. However since all items were not as simple as this, criteria had to be developed to determine whether these matches were valid.

These criteria can be summarized as:

1) Do verbal descriptions correspond;
2) Is the nutrient composition of each matched item within 15% for the calories and three macronutrients;
3) Are the foods coded as equivalent by HANES I and NFCS; and
4) Are the foods of commonly accepted equivalent composition.

Procedure:

Given the above criteria, items were assigned specifically for the dairy group. Using criteria 1) on verbal descriptions, the number of dairy items matched was 113. Criteria were established in order to determine whether these matches were valid. Nutrient composition data for calories and the three macronutrients, protein, fat and carbohydrate, were compared. Those items exceeding a certain percent difference were listed, to be considered as possible non-matches. A percentage of 15% was used as the cutoff for the allowable difference after it was evaluated and compared with non-match lists determined using 10%, 15%, 20% and 25%. The objective was to maximize the number of nutrient comparisons outside of the allowed range for those food items describing obviously different foods and minimize them for those foods items whose verbal descriptions are identical, or close to it. Nutrient differences were not considered as non-matches if the nutrient was present in a quantity less than 2% of the recommended diet (Recommended Dietary Allowances, Dietary Goals). The rationale for this was the fact that caloric value, for example, could differ by 100% if one food item had 2 calories and the other had 4 calories. However, these differences are not large enough to be significant. The cutoff value of 2% was chosen because it represents the quantity considered insignificant for nutritional labeling purposes. In the process of comparing nutrient levels, it was observed that for certain food items, all four nutrients in one survey were present in amounts which were in the same multiple of the other survey. For example, for HANES I item 1324, fluid evaporated milk , undiluted, and NFCS item 1121020, evaporated milk, diluted , nutrients in the HANES I item were double those in the NFCS item. Dilution factors were assigned to these foods.

At this point most HANES I items were matched with NFCS items. NFCS items were then identified for which no matching HANES I items had yet been established. Those for which corresponding HANES I items could be found were identified. Many NFCS items are composite foods, as previously mentioned. These can only be matched up with a sum of various quantities of different HANES I items. Communication with the Department of Health and Human Services led to the coding methodology used by the HANES I interviewers when those foods were consumed. This coding schema as well as commonly accepted composition of combined foods was used to assign factors to summed items. This was the last step in the assignment of corresponding foods.

Numerous items from both surveys were matched with more than one item. These duplicate items were listed and a primary key was assigned to the closest match in each direction (i.e. HANES I to NFCS, and NFCS to HANES I). For example, HANES I item 948 (custard,baked) is matched with NFCS items 1321030 (custard, NFS) and 1321031 (custard, homemade). Homemade custard is listed as the primary correspondence because of the verbal description even though all three items have the same values assigned for calories, protein, fat and carbohydrate.

The final correspondence file resulting from this procedure used the following four criteria for matching:

1) Foods with same verbal description;
2) Foods with close to equivalent nutrient composition;
3) Foods coded as equivalent by HANES I and NFCS; and
4) Foods of commonly accepted equivalent composition.

Results:

In the matchup of dairy items, which has been completed thus far, 197 separate correspondences were formed. This included 115 distinct HANES I items (some of which were not dairy items, but were used to form composite NFCS items) and 178 distinct NFCS items. Forty-six (46) HANES I items had more than one NFCS item assigned to them. Thirteen (13) NFCS items occurred in multiple correspondences. Of all the HANES I items in government group 01 (dairy), only two were not matched with corresponding NFCS codes. These two items both were ingested fewer than five times in the HANES I study. 103 NFCS items were left unmatched. Of these, 52 were ingested more frequently than 10 times, 17 more than 50 times, 9 more than 100 times and 2 more than 200 times. The most frequently consumed unmatched NFCS item was item 1410701, "cheese, mozzarella" at 217 ingestions. NFCS foods which remained unmatched were basically of two types:

(1) those for which no comparable HANES I item could be identified, such as acidophilus milk (item 1111212), and mozzarella cheese (item 1410701); and
(2) those composite foods for which we had no way of deciding on their exact composition, such as cheese cake with fruit (item 1461011).

Analysis:

In order to evaluate the cross-reference file formed by the procedure just described, two criteria were employed. We wanted to

(1) evaluate the accuracy of the correspondence formed, and
(2) evaluate the effect of missing values.

The usefulness of the correspondence file in representing an individual's daily dietary intake depends on how accurately the corresponding items represent the person's intake of certain nutrients. Since all items were not matched, it was necessary to

weigh how crucial the missing values would be in a total analysis of a person's daily intake. Separate analyses were performed for the purpose of examining these two evaluative criteria.

Nutrients for which the dairy group provides more than 10% of the total intake of the U. S. diet are as follows (Briggs & Calloway, 1979, Nutrition and Physical Fitness, WB Saunders Co., p398):

Calories	11.1%
Protein	22.0
Fat	12.5
Calcium	74.6
Phosphorus	35.0
Vitamin A	13.0
Riboflavin	39.0
Magnesium	21.7
Vitamin B-6	10.6
Vitamin B-12	20.1

The last three were omitted in the analysis because they are not available in the HANES I nutrient composition data base.

The NFCS Survey of Food Intake of Individuals was used for this analysis. Record types 52, 53, and 54, which contain nutrient intake information for each food consumed by the first 1000 persons for three days in the survey, were combined. For each food consumed, the corresponding HANES I item or items were determined, their nutrient composition retrieved, and the specific nutrient intake computed using the amount consumed in grams on the NFCS file.

The total intake from the dairy group for each of the three days was calculated using both the NFCS values and the corresponding HANES I items. The differences between these two values for each of the major nutrients supplied by the dairy group were calculated along with the percentage difference. It was decided to use this percentage difference as a measure of the closeness of matching. In cases where no corresponding food existed or where individual nutrient data was missing (negative values on HANES I file), no comparison was made in this analysis.

Upon examination of the first 1000 people (3000 people-days) in the NFCS file, the following statistics were gathered: 5915 dairy items were consumed. This represents an average of 1.97 dairy items per day. A total of 41,246 foods were ingested on these people-days, giving an average ingestion of 13.7 foods per person per day. 14.3% of these foods were dairy items. For the seven nutrients supplied by the dairy group (at an average of more than 10% of the US diet) and present in both studies, comparisons were made of the nutrient intake on each people-day computed using the NFCS items and also using the corresponding HANES I items. Differences between the HANES I and NFCS values were taken as a measure of the accuracy of the correspondence. Table 1 gives the results for calories, calcium and riboflavin.

TABLE 1

Evaluation of accuracy of correspondence file
between HANES I and NFCS food items

Percent difference in nutrient intake from dairy items

Percent Difference	Number of People-days	Cumulative Frequency *
	CALORIES	
less than 5%	1051	44.2%
less than 10%	2013	84.6%
less that 15%	2167	91.1%
less than 20%	2341	98.4%
	CALCIUM	
less than 5%	1436	61.1%
less than 10%	1675	71.3%
less than 15%	2137	91.0%
less than 20%	2183	92.9%
	RIBOFLAVIN	
less than 5%	547	23.4%
less than 10%	1353	57.8%
less than 15%	1573	67.3%
less than 20%	1802	77.0%
less than 25%	2074	88.7%
less than 30%	2212	94.6%

* includes only those people-days on which dairy items were ~~very~~ consumed. contained the nutrient

The second part of the evaluation analysis looked at missing data. In order to assess the effect of a lack of corresponding foods on the total dietary intake, calories supplied by items in the dairy group were computed for each person-day. The three-day dietary records of the first 1000 people in the survey contained a total of 127 dairy items ingested that were missing from the correspondence file. These represented 30 distinct foods. In examining the list of ingestions of missing foods, it was observed that for many foods, multiple ingestions take place either by the same person on different days or within the same household by different people. In tallying the ingestions, it was determined that the 127 missing ingestions represented 102 distinct people and only 68 distinct households.

The diets on 118 people-days contained missing dairy items; while 2755 people-days had no missing foods. On the remainder of the 3000 people-days, no dairy items were consumed. Thus, 4.11% of the people-days had ingestions of foods which were missing from the correspondence file. The average daily caloric intake from dairy was 252 calories. The difference in total caloric intake computed for those without missing items was 14.57 calories (or 5.8% of the total caloric intake from dairy), while those with missing items differed by a mean value of 191.63 calories (or 76.0% of the total caloric intake from dairy).

Conclusions:

The accuracy of the nutrient intake from the dairy items in the daily diet computed using HANES I and NFCS food items varied with the nutrients considered. Calories, protein, fat, calcium and phosphorus were the most accurately represented by the correspondence file. More than 70% of the sample examined had a difference of less than 10% for these nutrients when computed using each of the two studies. Riboflavin and vitamin A varied to a greater extent. With riboflavin, 77.0% of the sample had a difference of less than 20% while with vitamin A, 69.2% had a difference of less than 20%. This greater difference in the vitamin content reported by the two surveys may reflect variation in the nutrient analysis methods used. In some cases, different levels of fortification of dairy products with vitamin A may be involved. Further investigation is needed in order to determine the source of this large difference.

The error due to consumption of dairy items missing from the correspondence file is significant when missing items are consumed. However, this represented only 4.11% of the total sample of people-days examined. The error would presumably not be as great if computed on the total daily diet (including all food groups), unless by coincidence, the same people consumed foods missing from all groups included.

In conclusion, the task of matching food items from the two major government surveys in which dietary data were collected is not a simple one. Item by item matching requires the establishment of criteria to use as a measure of the validity of the match. It is hoped that further work will provide a mechanism to allow for the use of all of this valuable data with ease by investigators wishing to study dietary habits and nutrient composition of foods across surveys.