System Drift and Dietary Data Analysis

by Alison L. Eldridge, PhD, RD, Nutrition Coordinating Center, Division of Epidemiology, University of Minnesota

Introduction

It is widely known that there are many potential sources of error in dietary data. In fact, it is unlikely that dietary intake can be estimated without error (Beaton, 1994). It is important, therefore, to understand the nature and magnitude of the error so that we are able to improve our analysis and interpretation of dietary intake data. Error in dietary data can be partitioned into three major potential sources: participants, interviewers, and the database itself. Participant errors include such pitfalls as memory problems, inability to accurately estimate portion size or frequency of consumption, and unwillingness to disclose actual food consumption (Smiciklas-Wright, 1994; Thompson and Byers, 1994; Willett, 1990). Interviewer errors may be introduced if the interviewers are not properly trained, if they use leading questions during the interview, or if they do not use standardized procedures for probing and coding dietary intake information (Willett, 1990). In addition to errors introduced by participants and interviewers, there are potential sources of error in the databases themselves. Sampling practices, analytical techniques, laboratory quality control, seasonal variability, estimation, and missing data all play a part in the accuracy of nutrient values in food composition databases.

This paper is devoted to a discussion of another type of database error, namely system drift. Researchers need to be able to assess dietary changes without confounding due to artifactual database changes (Buzzard, 1991). Therefore, system drift is of importance to researchers who use the Minnesota Nutrition Data System (NDS) and other dietary analysis programs. System drift refers to the underlying database changes that occur with each new database release. These changes are due to improvements that are made in the underlying databases. Since improvements are made with each new release of a food composition database, significant differences may be detectable over time.

Food composition databases are used by a variety of researchers for a diverse group of research objectives. System drift is a potential factor in only a few types of studies, namely surveys interested in monitoring changes in food consumption patterns over time and epidemiological studies that follow subjects longitudinally. System drift is not a factor in case-control studies, where dietary measures at a point in time in one group are compared to dietary measures at the same point in time in another group. Likewise, system drift is not really an issue for ecological or correlational studies where disease rates in populations are compared with per capita consumption of foods based on disappearance data.

Sources of System Drift in Dietary Data

There are two broad categories of "improvements" in food composition databases that have the potential to impact system drift: actual changes in the marketplace and better data for foods that have not really changed (Table 1). Let's look at each of these in more detail.

The food marketplace is very dynamic. New foods are constantly being introduced. Other products are reformulated, package sizes are changed, serving sizes are altered, items are removed from the market, and consumers change their preferences in food preparation and cooking methods. All of these marketplace changes have the potential for affecting the information contained in our food composition databases.

Addition of new foods to the marketplace poses an interesting problem for maintenance of food composition databases and monitoring system drift. The reason new foods pose a problem in analyzing system drift is that people who collect the dietary records may come up with their own creative solutions when entering new foods into the computer system. For example, they may make a judgment about the new food that may or may not accurately reflect its nutrient composition. Recently, Post introduced a new cereal called Blueberry Morning. If Post Blueberry Morning cereal was reported on a dietary intake record or recall, the interviewer may make an incorrect judgment about it. Perhaps the interviewer decides that Blueberry Morning Cereal is similar to Raisin Bran and calculates nutrients for the dietary record using Raisin Bran. Later, when Blueberry Morning Cereal is available in the database, the dietary record is edited and updated, and a new nutrient profile is generated. The difference between the two records could erroneously be interpreted as system drift.

Many food products are reformulated without changing the food name. Some of these changes can significantly affect the nutrient profile of the food. Fleischmann's Egg Beaters is a good example. In 1992, 1/4 c of Egg Beaters had 25 kcal, 5 g protein, 1 g carbohydrate, and no fat. Egg Beaters supplied 10% of the US RDA for riboflavin. In 1994, the product was reformulated with more protein and, therefore, more kcal per serving. In addition, more riboflavin was added so that the product then provided 60% of the US RDA for riboflavin. Without making these changes in the database, riboflavin, protein and total calories could be underestimated in dietary intake records containing Egg Beaters.

Manufacturer changes to packages or portion sizes may also affect nutrient values for a food. Orville Redenbacher's Gourmet Light Butter Flavored Microwave Popping Corn has changed in serving size in recent years. One serving in 1992 contained 70 kcal, 3 g fat and 3 g fiber. In 1994, the serving size was changed so that one serving contained 114 kcal, 4 g fat and 5 g fiber. A more confusing example is that of McDonald's french fries. In 1991, the available package sizes were small, medium and large. These package sizes now go by the names small, large and supersize. Since the word large has different meanings depending on the year, re-analysis of older dietary records to track system drift could produce different gram weights and nutrient values for so-called large fries.

Another marketplace change that affects food composition databases is removal of foods from the market. Examples of discontinued products include Space Sticks, Tofulicious non-dairy imitation ice cream, Sara Lee Lights carrot cake and Libby's fruit float. Databases should be "improved" so that these products cannot be chosen once they are no longer available in the marketplace. However, they should not be removed from the database so that nutrients can be assigned to them, even if the record is re-analyzed years later.

Consumers have also changed preferences in food preparation and cooking methods. Less fat is left on meat cuts after trimming in response to consumer demand. Consumers are using less fat in recipes, and asking for options to lower fat in commonly prepared foods. Also, the use of non-stick cookware has increased. All of these factors can contribute to reductions in dietary fat that may or may not be reflected in food composition databases in a timely manner. These changes may also affect interpretation of system drift in dietary data.

In addition to the marketplace changes discussed above, better data for foods may result in changes in food composition databases. These changes can be of several types. New information may be provided for recipe yields, densities and conversion factors. Analytical methods may be improved. Coding rules for entering and calculating nutrient values may change over time. The specificity of the database may change. Finally, new nutrient values may be added to food composition databases.

Changes in recipes, yields, density data and conversion factors may affect interpretation of system drift analysis of dietary data. For example, in the Nutrition Coordinating Center's databases, the recipe for chicken salad made with mayonnaise was changed in 1994. The new recipe resulted in a change in fat from 21.2 g per cup to 16.3 g per cup. A change in the recipe for French toast resulted in a reduction in fat from 10.4 g per serving in 1991 to 3.8 g per serving in 1994. This change was the result of a recipe that used less fat in frying. An example of a density change occurred in 1994 when fish and seafood densities were updated to reflect new density data from the USDA survey database.

Examples of improved analytical methods that may affect interpretation of system drift include revised nutrient values for eggs (American Dietetic Association, 1989) and individual carotenoids. Most food composition tables quantify total carotenoids rather than individual carotenoid values. Most of these values were obtained using methods similar to the official Association of Official Analytical Chemists (AOAC) methods, which tended overestimate total carotenoid values. It is now possible to analyze individual carotenoids in foods using high-performance liquid chromatography (Khachik et al., 1992). Improvement in analytical methods could result in identical dietary records yielding different nutrient values when analyzed using old and updated databases.

Another potential source for system drift is coding rule changes. At the Nutrition Coordinating Center, fast food entrees were formerly coded by individual constituents, such as hamburger plus bun plus dressing, etc., in amounts reported by the study participant. Now they are coded by name (e.g., McDonald's Big Mac) using quantities provided by the manufacturer. Similarly, zucchini bread and other quick breads were formerly coded by individual ingredient (e.g., flour plus sugar plus egg, etc.) in amounts reported by study participants. Now they are coded by recipe name using standard amounts from commonly used recipes. Medium rich ice cream was formerly coded as containing 16% fat; now it is coded as containing 10% fat because of coding rule changes from USDA.

Foods formerly grouped together because of similar characteristics are separated as other nutrients of interest are identified. For example, Rice Krispies, Corn Flakes, Kicks, Cheerios and Wheaties once shared a database code at the Nutrition Coordinating Center because they did not differ substantially in nutrients that were of interest: calories, fat and cholesterol. When more researchers began to look at fortification levels of vitamins and minerals, it became clear that these cereals needed to be separated. However, dietary records analyzed when these foods shared a code would get the nutrients for Rice Krispies, even though the subject may have eaten one of the other cereals. Changes in the specificity of the databases also affect foods such as cookies, crackers, cheeses and vegetables. For example, vegetables used to be grouped together regardless of whether they were frozen or canned. Now they are separated because of differences in sodium levels.

Adding new nutrient values has the potential to affect drift in the system because nutrient values for new nutrients would not be available for dietary data collected at an earlier point in time. The Nutrition Coordinating Center is in the process of adding the following nutrient values to the database in response to requests from researchers: individual carotenoids, *trans* fatty acids, oxalates and phytates.

The Minnesota Heart Survey

Two potential sources of system drift in dietary data, marketplace changes and changes that reflect better data about foods, have now been examined. An example of the effects of system drift on dietary data can be seen using data from a longitudinal study, the Minnesota Heart Survey.

The purpose of the Minnesota Heart Survey (MHS) is to examine trends in risk factors, hospitalization, and mortality for coronary heart disease and stroke in Minnesota populations. Surveys have been conducted in 1980-82, 1985-87, and 1990-92. A fourth survey period is underway for 1995-97. Detailed descriptions of the methods have been published elsewhere (Luepker et al., 1985; Sprafka et al., 1990), but a brief description follows.

To identify possible participants, the seven county Twin Cities metropolitan area was divided into 704 clusters of approximately 1,000 households each. Of these, forty clusters were randomly selected for surveillance, and approximately 5 to 10 percent of the households within these clusters were randomly selected for participation. Eligible individuals were asked to complete a 20-minute home interview to determine various socio-demographic characteristics, health attitudes and beliefs, medication use, and a brief medical history. Home interviews were followed by a field clinic visit where dietary and physiological assessments were made.

Dietary information was collected using 24-hour recalls from 50% of all clinic visit participants. Food models were used to assist with estimation of portion size. Non-dietitian interviewers conducted the interviews and were trained and certified by the Nutrition Coordinating Center. Every six months, interviewers were recertified to insure data quality and standardization of methods. Particular emphasis was paid to information regarding foods and food preparation methods that would affect fat, cholesterol, and sodium intake. Coding of the recall, quality control, and maintenance of the nutrient database were the responsibility of the Nutrition Coordinating Center.

MHS chose to monitor quality control using 74 records randomly selected from the first survey period. There was still some variation when these records were blindly sent back to Nutrition Coordinating Center for recoding. This variation was due to problems with poorly documented foods in the original records. To overcome this limitation, the original records were unambiguously documented to eliminate the potential for coder variation when the records were subsequently recoded.

Examples of some of the changes that were made to standardize the records can be seen in Table 2. For example, the original record documented a 5½ to 6 inch hot dog, 1 inch in diameter. The standard further specified the hot dog to be a pork and beef hot dog 5¾ inch long and 1 inch in diameter. A 1¼ by 5 inch wedge of watermelon is somewhat difficult to quantify, so the standard specified watermelon to be 1 cup of watermelon chunks. An average size frozen chicken wing from Banquet was further specified as 1 ounce of frozen, pre-breaded and fried, commercially prepared chicken wing in which the person ate skin and breading. It is not important that the standard reflect exactly what the subject ate. It is important that the standard be unambiguously documented so that judgment error in coding is removed from the investigation of the effect of system drift on dietary data in this long-term study.

The mean values for total kcal from the 74 quality control samples in 1983, '85, '86, '88, '90 and '93 can be seen in Figure 1. The upper line represents the original, unaltered dietary data. The lower line represents the dietary record standards (unambiguously documented records) coded in 1983 and 1993 only. The original records were analyzed using ANOVA, and were not found to be statistically different in the different years (mean = 2121 kcal ± 27.5; range = 2074 to 2149 kcal). The standards were analyzed using mixed model regression to take into account changes over time in repeated analysis, and were also found to have no drift over the last decade (2047 kcal in 1983 and 2051 kcal in 1993).

Unfortunately, this is not the case with dietary fat (Figure 2). Again, there were no statistically significant differences in total fat calculated from the original, unaltered records (mean = 92.3 g ± 2.0; range = 89.6 to 94.4 g), but the standards told a different story. Mixed model regression showed a small but significant decrease in total fat calculated from the standard dietary records over the last decade from 89.8 g to 86.9 g.

Looking back through the list of potential sources of system drift, it is difficult to pinpoint the exact cause of the drift (refer to Table 1). At least the list of possibilities can be narrowed. Because the same records were used in 1983 and 1993, addition and removal of foods from the marketplace are not factors. Also, package sizes should not be an issue because the standards carefully documented package sizes and amounts. Therefore, the marketplace changes that could affect the MHS data include product reformulations and changes in the database that reflect changes in food preparation and cooking methods.

It is not so easy to eliminate changes to the database that reflect better data for foods. Updates or "improvements" made to reflect changes in recipe yields, densities and conversion factors, coding rule changes and changes in the specificity of the database may all play a part in the system drift for total fat. In this case, improved analytical methods did not affect the data because the same nutrient database was used in both time periods.

Now that system drift has been identified, what can be done with this information? MHS plans to use results from the system drift analysis to more accurately interpret dietary data collected in recent years of the survey. They will expand their investigation to evaluate system drift in other nutrients of interest. They also plan to continue to use the standard, unambiguously documented records as quality control standards in the future. Finally, they are planning to move away from hand coding to the NDS automated interview system for the next survey period and will use the standards to monitor any affect due to that transition.

Summary and Conclusions

To summarize, system drift can be caused by many factors, including marketplace changes and better data for foods that have not changed. System drift can have significant impact on interpretation of dietary data from longitudinal studies. Finally, system drift can and should be monitored for proper interpretation of dietary data.

These results indicate the necessity for monitoring and characterizing system drift in long-term studies of dietary intake. Several research opportunities present themselves. First, additional research needs to be done to explore our ability to partition system drift. Specifically, is there a way to separate system drift due to changes in the marketplace from drift due to improvements in analytical methods? Also, within each of these broader categories, can we attribute drift to any specific change? Secondly, what impact does a time-related database have on system drift? With the National Health and Examination Survey, NHANES III, finishing data collection in October, 1995, we are in a unique position to use data from this survey to further explore the issue of system drift in dietary data. Finally, what statistical methods are most appropriate for assessing or monitoring system drift in dietary data? Obviously, we get very different results using mixed model regression versus ANOVA. Is mixed model regression really the best tool? Hopefully, we can continue to explore causes and effects of system drift in dietary data until we are able to answer these questions.

References

American Dietetic Association. Updated nutrient composition of eggs. J Am Diet Assoc. 1989;89:1049-1050.

Beaton GH. Approaches to analysis of dietary data: relationship between planned analyses and choice of methodology. Am J Clin Nutr. 1994;59(suppl):253S-261S.

Buzzard IM. Maintaining time-related databases for dietary data collection and nutrient calculation. In: Proceedings, 16th National Nutrient Databank Conference, June 17-19, 1991, San Francisco, CA. p. 73-77.

Khachik F, MB Goli, GR Beecher, J Holden, WR Lusby, MD Tenorio, MR Barrera. Effect of food preparation on qualitative and quantitative distribution of major carotenoid constituents of tomatoes and several green vegetables. J Agric Food Chem. 1992;40:390-398.

Luepker RV, DR Jacobs Jr, RF Gillum, AR Folsom, RJ Prineas, H Blackburn. Population risk of cardiovascular disease: the Minnesota Heart Survey. J Chron Dis. 1985; 38:671-682.

Smiciklas-Wright H. Dietary intake: data collection and processing. In: Hoover LW, BP Perloff (eds.) 19th National Nutrient Databank Conference Proceedings, May 22-24, 1994, St. Louis, MO. p.99-103.

Sprafka JM, GL Burke, AR Folsom, RV Luepker, H Blackburn. Continued decline in cardiovascular disease risk factors: results of the Minnesota Heart Survey, 1980-1982 and 1985-1987. Am J Epidemiol. 1990;132;489-500.

Thompson FE, T Byers. Dietary Assessment Resource Manual, Section V. Selected issues in dietary assessment methods. J Nutr. 1994;124:2255S-2261S.

Willett W. Nutritional Epidemiology. Oxford University Press; New York. 1990. p. 54-57.

Table 1: Potential Sources of System Drift in Dietary Data

Potential Source of System Drift	Examples
Changes in the Marketplace	•Addition of new foods •Product reformulations •Package or size changes •Removal of items from the market • Changes in food preparation and cooking methods
Better Data for Existing Foods	• Changes in recipe yields, density data, or conversion factors •Improved analytical methods •Coding rule changes •Changes in the specificity of the database •Addition of new nutrient values

Table 2: Examples of Differences Between Original MHS Quality Control Records and Unambiguously Documented Standards.

Food	Original	Standard
Hot dog	5 1/2 - 6 inch long 1 inch diameter	pork and beef hot dog 5 3/4 inch long 1 inch diameter
Watermelon	1 1/4 inch by 5 inch wedge	1 cup chunks
Chicken wing	average size, frozen, Banquet; person ate skin	1 ounce, frozen, pre-breaded and fried, commercially prepared; person ate skin and breading

Figure 1. System Drift in Total Kcal, 1983-1993

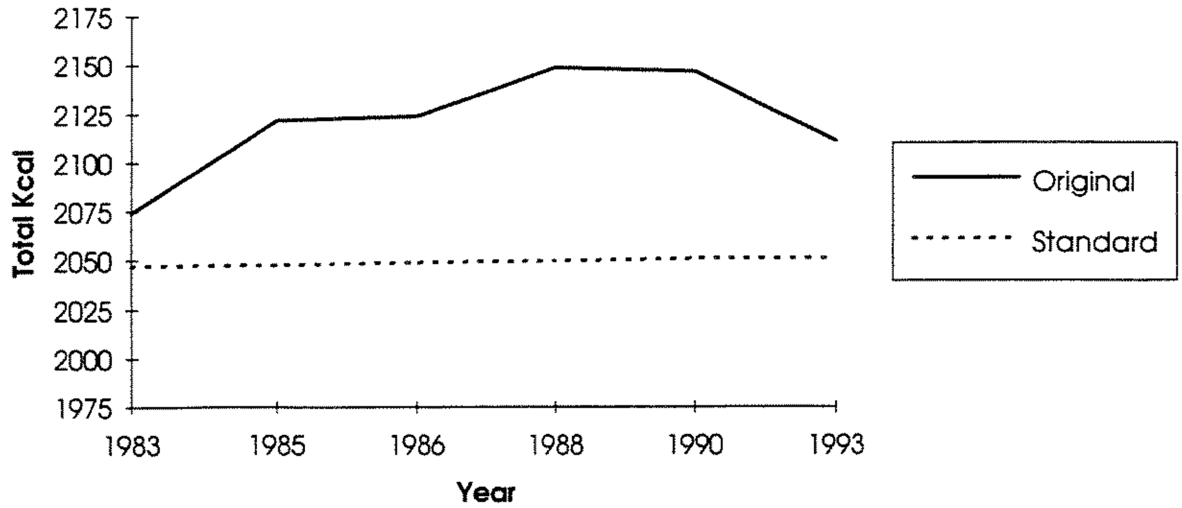

Figure 2. System Drift in Total Fat, 1983-1993