Adjusting for Intra-Individual Variability when Estimating Nutrient Intakes

Patricia Guenther, USDA-ARS-HNIS

This paper discusses a method of estimating the distribution of usual daily intakes for a dietary component, where "usual" is defined as "long-run average," which is effectively a year. Our approach is based on the assumption that an individual can more accurately recall and describe the types and amounts of foods eaten yesterday than the types and amounts of foods eaten over any longer period of time. We also assume that the nutrient data base used can also reflect the nutrient content of the foods eaten at that time reasonably well.

Dietary data contain both within- and between-person variation. When analyzing such data, nutritionists typically are interested only in the between-person variation. That is, they want to remove the within-person variation; or otherwise said, they are interested in the variation of "usual" intakes.

Nutritionists have sought to remove or minimize this within-person variation by lengthening the observation period from 1 day to as many as 365 days. The method we will discuss takes much of the burden of estimating usual intakes away from the subjects. They need only provide as little as 2 independent days of food intakes or 3 consecutive days of such information.

The Human Nutrition Information Service (HNIS) has been working cooperatively with statisticians at Iowa State University, who have developed what we call the ISU method for estimating usual intake distributions (Nusser et al. in press.) It is important to point out that the ISU method does not assume that the data under investigation are normally distributed or that they come from simple random samples. Neither assumption, although often made, is appropriate for most dietary intake data.

The ISU method assumes that a reported 1-day nutrient intake for an individual found in a food intake survey data set can be represented by the equation:

$$y_{its} = x_i + c_t + b_s + e_{it},$$

where y_{its} is the reported nutrient intake by individual i for date t, and s is the sequence number of the day for which the individual has provided intake information. The value we are interested in estimating is x_i, the usual intake of individual i. Also in the equation are c_t--the temporal effect on nutrient intake caused by the particular day of the week or season of the year--and b_s--the bias associated with intakes on a particular reporting day of the survey. The last term, e_{it}, is simply the difference between the reported intake, y_{its}, and the other three terms.

The ISU method assumes that the first day of reported nutrient intake values represent the truth. This means not only that the individual reports his or her food intakes correctly for that day, but also that the nutrient values assigned to those intakes represent the truth.

Whether or not there is bias on the first day of reported intake should not obscure one of the important attributes of the ISU method; namely, it removes the well-known biases of subsequent reporting days compared to the first. In addition to that, temporal effects, such as day-of-the-week and seasonal effects, are also easily removable from data sets in which such temporal factors are recorded.

Let us now explore the assumptions that individuals are correctly reporting their food intakes on the first day of the survey and that they are coded correctly and focus our attention on the quantity of a particular nutrient in 100 grams of a particular food eaten, which we will call the "nutrient density" of the food. The "within" variation we are looking at here is the within-food variation rather than the within-individual variation that the ISU method successfully deals with.

It is helpful to express the true nutrient density of a food eaten and then recorded in a survey data set as--

$$d_{it} = u + f_i + g_t + e_{it},$$

where d is the density; the subscript "it," again refers to individual i and survey data t; u is the true mean nutrient density of the food consumed by the population; f, is the potential variation across different individuals because they prefer and habitually consume different varieties or brands of the food, for example, because one uses mostly Heinz ketchup while another prefers Hunt's; and g_t reflects the potential variation of the nutrient density of the food over different seasons of the year. Finally, e_{it} is simply what remains after u, g_t, and f, are subtracted from d_{it}, the true nutrient density of the particular food eaten.

Ideally, if our goal is to estimate the distribution of usual nutrient intakes for a population correctly, we would like to remove the effects of season and purely random variation from our determination of the correct nutrient densities to use when translating reported food intakes into nutrient intakes. Unfortunately, this means that even if the nutrient data base we use gives us good estimates for u, the average nutrient density of a food consumed by the population, we will fail to incorporate the tendency for individual i to be different from his or her fellows, the f_i, in our equation. We recognize this problem, but we feel that the relative difference between, for example, the average vitamin A content in the brand of ketchup one person prefers to that another favors is very small compared to the relative differences in their total vitamin A intakes--small enough to be safely ignored. This leaves the question of estimating u, the average nutrient density of the food. USDA presently does this for the National Nutrient Data Bank (NNDB).

The great statistician George Box is credited with the saying "all models are wrong, but some models are useful." This should be the guiding principle in estimating the average nutrient content of foods in a data base. Estimating the average nutrient density of a food is, unfortunately, difficult and thankless work that will require the extensive use of models to be effective--models linking the foods analyzed to what is actually eaten and models reflecting the measurement errors inherent in lab work. The results may not be completely satisfying, but they will be generated using the best methods at our disposal.

We feel that error in the NNDB's model-based estimate for u, the average density of the nutrient content of food, is small compared to the variation in nutrient composition across foods of different types--small enough to be ignored in most cases. Nevertheless, USDA is engaged in strengthening the statistical underpinnings of the NNDB. It is investigating more sophisticated methods of combining data from various sources that will allow us to consider quality factors in the redesigned nutrient data bank. USDA will continue to consider what the appropriate mix of varieties and brand information should be.

Finally, we turn to the assumption that the nutrient values reported for the first survey day represent the truth. This means that the respondent actually reports correctly his or her food intakes on the first survey day and that we correctly translate these food intakes into nutrient intakes. USDA is working towards making this assumption more tenable. With researchers at the Census Bureau's Center for Survey Methods Research, it is investigating the cognitive aspects of the 24-hour dietary recall task. A multiple pass approach has been developed, which gives the respondents more time to think and focus on the recall task, and work is continuing to improve it. We believe that research aimed at reducing reporting error has the

most to offer in terms of improving the quality of the estimates of usual nutrient intakes produced from dietary surveys.

Reference

S.M. Nusser, A.L. Carriquiry, K.W. Dodd, and W.A. Fuller. A Semiparametric Transformation Approach to Estimating Usual Intake Distributions, Journal of the American Statistical Association, in press.