THE USDA'S CONTINUING SURVEY OF FOOD INTAKES BY INDIVIDUALS: ISSUES AND APPLICATIONS (DISCUSSION)

S. M. Nusser
Department of Statistics, Iowa State University, Ames, IA 50011-1210

Key Words: CSFII, Dietary Assessment

Introduction

The 1994-96 Continuing Survey of Food Intakes by Individuals (CSFII) is the latest in a series of dietary intake surveys conducted periodically by the U.S. Department of Agriculture (USDA). These surveys are an important source of information on the food consumption patterns for various segments of the U.S. population. The data are widely used to study the impact of nutritional policies as well as the effect of diet on various health-related problems and exposure to pesticide residues via food intake.

The primary observation on each sample person is a 24-hr recall, during which the respondent is prompted to fully enumerate foods consumed during the previous 24 hours. Ideally, this includes accurate determinations of the specific type of food and the amount consumed for each food. During a subsequent coding phase, recipes are used to decompose the food intake data into component foods and nutrient intakes. It is widely recognized that numerous sources of measurement error may be present dietary recalls and subsequent translation into food and nutrient intakes, including errors induced by recall bias and translation data bases (NRC, 1986).

Even in the absence of such errors, daily intake data exhibit a variety of behaviors that affect choices for data collection procedures and statistical analyses. An individual's dietary intake varies greatly from day to day. While much of the intra-individual variation is due to variety in the diet, part of it is attributable to systematic patterns induced by factors such as season and day of the week. The within-person variance of nutrient and food intakes exceeds the between-person variance, sometimes by several times between-person variance. In addition, individual means are frequently correlated with individual standard deviations. Nutrient and food intake data are typically highly skewed, and food intakes contain numerous zeros arising from a mixture of nonconsumers who never consume the food and consumers that were not observed to consume the food on a sample day.

Many analysis objectives involve the concept of "usual intake," which is defined to be the long-run average intake of a dietary component by an individual. Statistically, the usual intake of individual i, y_i, is defined to be the conditional expectation of the daily intakes, Y_{ij}, for individual i, $y_i = E\{Y_{ij} \mid i\}$. The observed daily intake measures the usual intake with error, commonly expressed through the measurement error model $Y_{ij} = y_i + e_{ij}$, where $i = 1, 2, \ldots, n$ individuals, $j = 1, 2, \ldots, r_i$ days for individual i, y_i has mean μ_y and variance σ_y^2, and e_{ij} is an independent error term with mean 0 and variance σ_{ei}^2. The distribution of interest in population-level analyses of usual intakes is that of y_i. Using daily intakes to make inferences about usual intakes will result in erroneous conclusions. Because within-person variances are typically 1.5 - 8 times the usual intake variance and the number of daily intakes for person i is generally small, the distribution of individual means may be significantly overdispersed relative to the usual intake distribution. Since many risk assessments rely on estimated proportions of the population whose usual intake exceeds or is lower than a threshold value, using a proper estimate of the usual intake distribution is critical.

With these thoughts in mind, we now discuss a series of papers concerning methodologies used to conduct the CSFII and potential applications of the CSFII for policy analysis.

Sample Design and Data Collection Methods

The sample design described by Goldman et al. (1997) relies on well-established area sampling methods. The design combines several techniques to maximize the chances of obtaining a representative sample, and to insure that adequate sample sizes are obtained for subpopulations. The sample design also attempts to provide balance across levels of nuisance effect variables, such as day of the week and seasons, yielding data that allow these effects to be estimated and removed from the analyses if so desired. Balance across years permits interim analyses to be performed and early results to be released, which is increasingly important as legislative bodies, federal agencies, and the public demand more timely information. The design also addresses operational constraints while meeting analysis objectives.

Operational efficiency was achieved via clustering. An existing master sample was used to select as PSUs, resulting in reduced fixed costs associated with developing materials for sample selection.

Data collection methods were based on procedures designed to reduce nonsampling errors. As noted earlier, food intake data are known to contain many biases as a result of the recording process. While there is still great concern regarding reporting biases in dietary intakes, the multiple pass method with aids to quantify amounts consumed represents an important method of reducing recall and quantification biases.

Methods to estimate parameters of usual intake distributions rely on multiple days of intakes per person to adjust for within-person variation. Collecting two independent days of intake per subject enables estimation of and adjustment for within-person variances for nutrient intakes and large food groups that are typically consumed daily. However, for foods less frequently consumed, two positive intakes are required for variance estimation. Further research is required to determine whether this design will adequately support analyses that rely on usual food intake distributions, such as exposures to pesticide residues.

Food Coding and Nutrient Translation

LaComb et al. (1997) describe a computerized coding system called Survey Net, which translates food intakes into component food and nutrient intakes. When properly designed, computer-assisted survey information collection (CASIC) methods can make significant contributions to the efficiency and accuracy of survey data collection. Survey Net represents an example of how CASIC methods can lead to significant improvements in the quality of and time required to produce a survey data base. The query system helps to generate consistency in selecting recipes, while the flexible recipe option supports accurate translation of the foods into food components and nutrient intakes. Efficiency is gained by establishing an electronic communication system for notes and by the savings that Survey Net reaps in from reduced processing time. As with the multiple pass enumeration method of eliciting intakes, Survey Net is another strategy to reduce errors in developing component intakes.

While much has been gained by using Survey Net, the accuracy of nutrients and food intakes generated by Survey Net is highly dependent on the translation tables used to decompose foods consumed by the respondents. Methods used to establish these translation tables would benefit from further research. No unified sampling approach is used to gather data for nutrient densities contained in the tables. The current data base relies on a mixture of information obtained from private firms, FDA and USDA analyses. The quality of this information is quite variable and frequently can not be assessed due to inadequate documentation of results. While there has been some effort to develop "weighting" schemes based on quality ratings for each datum, there are still serious nonsampling errors present in the data. In response to a similar set of problems cited for pesticide residue concentrations in foods, NRC (1993) recommended the development of a unified approach to sampling foods for obtaining pesticide residue concentrations.

Weighting Procedures

Sample weights are used to adjust for unequal inclusion probabilities and nonresponse. As with the sample design, the methods used to generate weights described by Chu and Goldman (1997) represent standard procedures in survey sampling. Demographic data from the Current Population Survey were used to control weights using numerous classification variables and a ratio-raking procedure. The authors note that the large number of variables can act to increase variance of estimates from the survey. However, the benefit of such procedures is a robustness against bias when analyses are performed for variables or subpopulations other than those used in controlling the weights. The CSFII data base is used for a wide variety of purposes, and thus it seems the potential increase in variance is worth the robustness gained in using many demographic classes to control weights.

Applications

As noted earlier, the CSFII and other USDA food consumption surveys are used for a wide variety of analytic purposes. Basiotis et al. provide an example of how CSFII data can be used to explore the effects of policy changes on the nutritional status of affected subpopulations. The timeliness of the data base release is underscored by the topic of their analyses, the impact of new welfare policies on the nutritional status of able-bodied adults without dependents. Though their results are preliminary, they point to particular conditions under which the 1996 welfare reforms may lead to reduced food security for individuals.

The USDA also produces summary statistics on intakes of a variety of dietary components for specific age/sex populations as well as for high risk populations such as low-income households. These summaries include estimates of mean intakes and

percentiles for usual intake distributions. Over the past several years, USDA has sponsored cooperative research that enables them to publish estimates of percentiles for subpopulations based on usual intake distributions that are adjusted for various sources of measurement error (Fuller 1995, Nusser et al., 1996, and Nusser et al., 1997a). These methods are now being employed by the National Academy of Sciences in determining reference values for nutrient intakes, and are being applied to the National Health and Nutrient Examination Survey, conducted by the National Center for Health Statistics. The USDA-sponsored cooperative research represents a positive contribution by USDA to the field of dietary assessment.

USDA food consumption data are also widely used by the food safety community. Food safety is one of the most important future uses of the CSFII, and cooperative interagency efforts will be required to address problems presented in collecting and analyzing data for food safety purposes. While food residue risk assessment benefits from the existence of a comprehensive food intake data base, many statistical and measurement error problems remain that are similar to those present in dietary intakes. Extensions of methods to estimate usual nutrient and food intake distributions are currently being explored as a framework for estimating usual residue exposure distributions for chronic health effects (Jensen et al., 1997, Nusser et al., 1997b). One problem that remains is whether CSFII's design of two recalls per person will be sufficient for estimating usual residue intake distributions, especially for narrow pesticide or food groupings that are infrequently consumed.

Concluding Remarks

Dietary intake data are difficult to collect and to analyze in part because of numerous potential sources of measurement error. The role of food consumption data in monitoring and improving the nation's health underscores the importance of addressing these challenges. The USDA and their partners at Westat have used survey methodologies for conducting the CSFII that represent a reasonable balance among survey objectives, the need to minimize nonsampling errors, and operational constraints.

Literature

Basiotis, P. P., Kramer-LeBlanc, C. S., and Kennedy, E. T. (1997). Using USDA's Continuing Survey of Food Intakes by Individuals to assess welfare reform nutritional impacts on selected low-income groups. *1997 Proceedings of the Section on Survey Research Methods.* American Statistical Association.

Chu, A. and Goldman, J. (1997). Weighting procedures for USDA's Continuing Survey of Food Intakes by Individuals 1994-96. *1997 Proceedings of the Section on Survey Research Methods.* American Statistical Association.

Fuller, W. A. (1995). Estimation in the presence of measurement error. *International Statistical Review,* 63(2):121-147.

Goldman, J. D. and Borrud, L. G. (1997). An overview of USDA's 1994-96 Continuing Survey of Food Intakes by Individuals and the Diet and Health Knowledge Survey. *1997 Proceedings of the Section on Survey Research Methods.* American Statistical Association.

Jensen, H. H., Chen, C. F., and Ralston, K. (1997). Estimating usual intake distributions for pesticide residues intake. Final Report submitted to USDA Economic Research Service under USDA Cooperative Agreement No. 43-3AEM-5-80090. 30 pp.

LaComb, R., Green, A., and Ingwersen, L. (1997). Survey Net: coding and management of CSFII food intake data. *1997 Proceedings of the Section on Survey Research Methods.* American Statistical Association.

National Research Council (1986). *Nutrient Adequacy.* National Academy Press, Washington, DC.

National Research Council (1993). *Pesticides in the Diets of Infants and Children.* National Academy Press, Washington, DC. 386 pp.

Nusser, S. M., Carriquiry, A. L., Dodd, K. W., and Fuller, W. A. (1996). A semiparametric transformation approach to estimating usual intake distributions. *Journal of the American Statistical Association,* 91:1440-1449.

Nusser, S. M., Fuller, W. A., and Guenther, P. M. (1997). Estimating usual dietary intake distributions: adjusting for measurement error and nonnormality in 24-hour food intake data. In L. Lyberg, P. Biemer, M. Collins, E. DeLeeuw, C. Dippo, N. Schwartz, and D. Trewin (editors), *Survey Measurement and Process Quality.* Wiley, New York. pp. 689-709.

Nusser, S. M., Carriquiry, A. L., and Jensen, H. H. (1997). Usual intake distributions in risk assessment for food residues. *1997 Proceedings of the Section on Statistics and the Environment.* American Statistical Association.