DISCUSSION

Graham Kalton, University of Michigan
Survey Research Center, University of Michigan, Ann Arbor, MI 48106-1248

I should like to thank Sue Ahmed for organizing an excellent session. The large attendance attests to the widespread interest in the weighting of survey data. I should also like to commend the authors for their thought-provoking and valuable papers. I will comment on each of the papers in turn.

Kish Paper

As Leslie Kish notes, there is no readily available, simple yet comprehensive, treatment of weighting for survey practitioners. His paper is designed to fill this gap. In a short space he has successfully addressed a multitude of issues relating to weighting. In reading the paper I was stimulated to think of other aspects of weighting that I might have included. I will outline some of these below.

First, I think it would be useful to discuss the data needed for weighting to compensate for nonresponse and noncoverage. In one type of adjustment, which I term sample weighting, respondents are weighted up to represent similar nonrespondents. The data requirements for this form of weighting are information for both the respondents and nonrespondents on the auxiliary variables used in the adjustment. Since often little is known about the nonrespondents apart from information from the sampling frame (e.g., their PSUs and strata), there is usually a limited choice for creating weighting cells for this type of adjustment. Sample weighting attempts to compensate for nonresponse but not for noncoverage.

Another type of adjustment is generally known as poststratification, but I prefer to term it a population weighting adjustment when it is used to compensate for nonresponse and noncoverage. Population weighting adjustments require information on the auxiliary variables for the respondents (but not for the nonrespondents) and knowledge of the population joint distribution across the weighting cells from some external source. Care is needed to ensure that the auxiliary variables are measured in a comparable way in the survey and in the external source.

Raking is a form of weighting adjustment that has less detailed data requirements than those of population weighting adjustments. In addition to information on the auxiliary variables for the respondents, it requires external information on only the marginal population distributions for the auxiliary variables, not on their joint population distribution.

Apart from its less detailed data requirements, another reason for using raking is to employ several auxiliary variables without resorting to separate weighting adjustments in a large number of cells. An alternative procedure for this situation is some form of response propensity weighting (Little, 1986).

Another possible topic for the paper is the avoidance of weights. For instance, nonresponse adjustment weights may sometimes be avoided by making substitutions for the nonrespondents (Chapman, 1983). Similarly, noncoverage adjustment weights may sometimes be avoided in sample selection. For example, a sample of hospital patients may have to be drawn from an incomplete list of hospitals because the other hospitals do not have the requisite records (perhaps from abstracting services). The sample of hospitals may be drawn as a disproportionate stratified sample in order to make the resulting sample reflect the distribution of hospitals across the strata that would have occurred had the total population of hospitals been sampled.

The issue of norming the weights often arises. Since government statistical agencies frequently want to estimate population totals, they may norm the weights to make their sum correspond to the population size. Researchers concerned with means and proportions and other standardized estimates, not totals, often norm the weights to sum to the sample size. The rationale here is that the sum of the weights can then be treated as comparable to the sample size in gauging the precision of the survey estimates. This, however, fails to take account of the fact that the use of weights for sampling frame inequalities and nonresponse and noncoverage adjustments leads to a loss of precision; under certain assumptions, the effective sample size is in fact reduced to $(\Sigma w_i)^2/\Sigma w_i^2$. This suggests the possibility of norming the weights to sum to the effective sample size, but even this is imperfect because it will not produce the appropriate effective sample sizes for subclass analyses.

It is useful to note that imputation schemes for item nonresponse that assign values from respondents to item nonrespondents are closely related to a weighting procedure that replicates the weights of the respondents (Kalton, 1983). For univariate analyses, assigning the value of a respondent to a nonrespondent is equivalent to adding the nonrespondent's weight to that of the respondent.

Kish notes that the use of weights for analytical statistics, like regression analysis, poses philosophical problems. This issue is a complex one, and as such may fall outside the scope of his paper. However, in view of the considerable importance of this issue, I would have liked to see some discussion of it.

Cowan and Ahmed Paper

Chuck Cowan and Sue Ahmed present an interesting approach to noncoverage based on maximum likelihood. In order to examine their approach, I will review the classical survey sampling approach for the same problems that they address.

Consider, first, a simple random sample (SRS) from a frame with complete coverage of teachers to estimate the total number of mathematics teachers. The probability of t mathematics teachers in the sample is given by the hypergeometric distribution

$$p(t) = \binom{T}{t} \cdot \binom{N-T}{n-t} \Big/ \binom{N}{n}$$

in the notation of Cowan and Ahmed. The sample proportion of mathematics teachers $p = t/n$ is unbiased for $P = T/N$ and $\hat{T} = Np$ is unbiased for T.

Now consider the problem of noncoverage, with M teachers on the frame. Then the total number of mathematics teachers in the population is $T = MP+(N-M)P_0$, where P and P_0 are the proportions of mathematics teachers among the M teachers on the frame and among the $(N-M)$ teachers not on the frame, respectively. As above, p is an unbiased estimator of P. In order to estimate T, an assumption is needed about P_0. A simple assumption is $P_0 = P$, in which case P_0 may be estimated by p, and hence T may be estimated by $\hat{T} = Np$.

Next consider a stratified sample with noncoverage, with M_i listed teachers and N_i total teachers in stratum i. Extending the SRS case, the total number of mathematics teachers is $T = \Sigma T_i = \Sigma M_i P_i + (N_i - M_i) P_{0i}$. If the N_i are known, and if the assumptions that $P_{0i} = P_i$ are made, T may be estimated by $\hat{T} = \Sigma N_i p_i$. This is akin to a two-phase sample: first, a SRS of M_i of the N_i teachers is selected for the frame in stratum i, and then a SRS of m_i teachers is selected from the M_i teachers on the frame. Under the assumptions $P_{0i} = P_i$, the poststratified estimator $\hat{T} = \Sigma N_i p_i$ follows in a straightforward manner.

Suppose now in the stratified example that $N = \Sigma N_i$ is known, but that the N_i are unknown. In this case T may be estimated by $\hat{T} = \Sigma \hat{N}_i p_i$, again assuming that $P_{0i} = P_i$. In this situation, estimators of N_i are also needed, and this requires a further assumption. A simple assumption is that the rate of noncoverage is the same for all strata, in which case N_i may be estimated by $\hat{N}_i = NM_i/M$. This leads to the estimator $\hat{T} = N\Sigma M_i p_i / M$.

Finally, consider the situation with marginal controls for two control variables. In this case $\hat{T} = \Sigma\Sigma \hat{N}_{ij} p_{ij}$, under the assumptions that $P_{0ij} = P_{ij}$. The N_{ij} may be estimated by the raking algorithm under the assumption that the noncoverage rate in cell (ij) can be expressed as $\phi_{ij} = \rho_i \sigma_j$.

The above estimators, based on a criterion of unbiasedness, are very similar to those derived by maximum likelihood by Cowan and Ahmed. For these simple situations, the forms of estimator to use with the classical approach are obvious. The assumptions involved under the classical approach are clearly identified and somewhat less stringent than those needed for the maximum likelihood approach, and the derivations are perhaps more straightforward. Moreover, the results for the classical approach generalize readily to the complex cluster sample designs that are widely used in practice. The extension of the maximum likelihood approach to complex sample designs is much more difficult.

I feel that the strength of maximum likelihood methods lies in complex situations for which no obvious estimator can be identified under the classical approach. The systematic application of maximum likelihood procedures, with a clear specification of the assumptions involved, may then lead to estimators that otherwise would not have emerged.

Cohen Paper

Panel surveys like the National Medical Care Survey (NMES) and its predecessors, the Survey of Income and Program Participation, and the Panel Study of Income Dynamics face a problem of how to take account of changes in families over time in conducting longitudinal family analyses. Families are formed and dissolve, and change in composition, as the result of births, deaths, marriages, divorces and separations. By the nature of a panel survey the problems of family dynamics have to be squarely faced. It is worth noting that these problems also occur with cross-sectional surveys that collect retrospective data on family members, but they are often side-stepped in the analysis. Steve Cohen's paper provides a clear discussion of the issues involved in dealing with family dynamics in the NMES.

It is useful to separate the problem of family dynamics into two parts. First, what are the population parameters that provide meaningful summaries of family-level information? That is, what quantities would be computed if the whole of the U.S. population were included in the NMES for its one year duration? Second, how are the sample data collected in the NMES used most effectively to estimate these population parameters? This includes the question of how the sample data should be weighted.

Cohen indicates the need for two types of population parameters, means and distributional estimates, from the NMES. An example of the former is the mean annual expenditure for ambulatory physician contacts per family. An example of the latter is the percentage of families with expenditures for physician contacts above $2000 per year. Given the problem of family dynamics, one solution is to define such parameters only for stable families that do not change in composition at all during the reference period. The clear limitation of this solution is that it excludes a sizeable proportion of the population, and those excluded may well have different medical expenditures than those in stable families.

Cohen describes three other strategies for handling family dynamics in family-level analyses: (1) to define a family as a set of individuals at a specific point of time, and to define the longitudinal family to comprise that set of individuals for the survey reference period; (2) to define longitudinal families by a set of rules that allow for some changes in composition; and (3) to deal with families that exist for only part of the reference period by weighting them in the parameter according to the proportion of the reference period that they existed. I will comment briefly on each in turn.

The first strategy corresponds to what is done in cross-sectional surveys like the Current Population Survey, where retrospective data are collected on current family members. An aggregation of these data to the family level combines data of individuals who may have lived in other families during the year. For the NMES, Cohen suggests that families might be defined at the beginning or end of the panel. If this strategy is adopted, I would opt for the beginning of the panel. It is simpler from the weighting prospective, and data are intended to be collected for all family members under this definition throughout the reference period (although panel attrition will cause some missing data). Defining the family as it exists at the end of the panel presents the problem that data are not collected for nonkey persons in the NMES until they join the panel. It would, of course, be possible to impute their missing earlier data, but that seems to constitute an unnecessary amount of imputation. Whatever point of time is used, however, the basic objection to this strategy is that it does not truly reflect a family's experience.

The second approach of defining longitudinal families to allow for some change of membership aims to reduce the losses from the strict no change definition. It seems to me that it would be hard to give meaningful interpretation to the results obtained. For example, a change by the addition of a birth or the loss from a death during the course of the reference period is likely to markedly affect medical expenditures.

The third approach weights part-period families by the proportion of the period they were in existence. This is similar to the exposure to risk approach used in demographic and actuarial studies. The usual basis of this approach is that a person has a constant exposure to risk, and that the chance of becoming sick is therefore directly proportional to the length of time observed. I question whether the approach is appropriate for the NMES where family formation and dissolution are likely to be associated with atypical medical expenditures. In particular, how is the need for distributional estimates handled under this approach?

In summary, I think that all the approaches have serious limitations. While I recognize the attraction of the concept of a family to analysts, it is perhaps a misleading one. In view of the difficulties created by family dynamics, the best approach may well be to follow Duncan and Hill (1985) and Ruggles (1990) in rejecting the family as the unit of analysis, and instead conduct analyses at the individual level.

References

Chapman, D. (1983). The impact of substitution on survey estimates. In *Incomplete Data in Sample Surveys, Volume 2* (W.G. Madow, J. Olkin and D.B. Rubin, eds.), pp. 45-61. Academic Press, New York.

Duncan, G.J. and Hill, M.S. (1985). Conceptions of longitudinal households, fertile or futile? *Journal of Economic and Social Measurement*, **13**, 361-375.

Kalton, G. (1983). *Compensating for Missing Survey Data.* Survey Research Center, University of Michigan, Ann Arbor, Michigan.

Little, R.J.A. (1986). Survey nonresponse adjustments for estimates of means. *International Statistical Review*, **54**, 139-157.

Ruggles, P. (1990). Longitudinal analysis of federal survey data. Survey of Income and Program Participation Working Paper No. 9007. U.S. Bureau of the Census, Washington, D.C.