DISCUSSION

David W. Chapman, National Analysts
1700 Market St., 17th Floor, Philadelphia, PA 19103

When Joe Waksberg and I agreed to be discussants for this session, we decided to split the five papers between us in terms of the major responsibility for discussant comments. My allocation was the Fahimi and Judkins paper and the Ernst paper.

I do have one comment on the paper by Gorsak, et. al. regarding the methods they compared for stratifying blocks within PSUs. I suggest that they consider the "cum $\sqrt{f}$ rule," described in Cochran's Sampling Techniques (3rd ed., pp. 129-130).

Discussion of the Fahimi and Judkins Paper

First I will discuss the paper "PSU Probabilities Given Differential Sampling at Second Stage," by Mansour Fahimi and David Judkins. They address a very important topic which arises in many survey sampling applications: defining a measure of size for PSUs which takes into account multiple domains (subpopulations) of interest in the population.

The authors define and compare four alternative PSU measures of size (MOS). The first, MOS^T, is the standard MOS: the total number of population members in the PSU. The other three measures, which they refer to as MOS^1, MOS^2, and MOS^3, are alternative measures of size that attempt to account for the sampling needs for the domains of interest.

They define the first alternative MOS for the i^{th} PSU as follows:

$$MOS^1_i = \sum_{k=1}^{D} f_k N_{ik} \quad (1)$$

where

f_k = the specified overall sampling rate for the k^{th} domain,

N_{ik} = the number of population members in k^{th} domain in the i^{th} PSU.

This MOS has good properties; it is the only MOS that allows for both a self-weighting sample for all of the domains and a constant total PSU sample size (i.e., a constant workload). Therefore, MOS^1 may be the MOS that is generally best to use in applications of this type.

There is some confusion in the paper -- at least in the first draft -- regarding the definition of MOS^1. Near the beginning of the paper the authors indicate that the definition of MOS^1 given above is equivalent to making the MOS proportional to the sum of the inclusion proportions of the D domains:

$$MOS_i \sim \sum_{k=1}^{D} N_{ik}/N_{\cdot k} \quad (2)$$

It is easy to show that equations (1) and (2) are equivalent only if the total sample sizes for the domains are equal. In the general case when the domain sample sizes differ, the MOS defined in equation (2) may be unacceptable in most applications since it is not sensitive to the different sample sizes or sampling rates assigned to the various domains.

By definition of MOS^1, one of the domains in equation (1) is that part of the population that does not fall in any of the domains of interest. (Call this "leftover" domain the D^{th} domain.) If the domains of interest collectively make up a relatively small part of the total population, as is often the case, MOS^1 will be strongly influenced by the size of the population not covered by any of the domains of interest (i.e., by the size of domain D). This suggests the following variation of MOS^1:

$$MOS^{1A}_i = \sum_{k=1}^{D-1} f_k N_{ik} + (f_D/c) N_{iD}$$

where c = a "deflation factor" that reduces the influence of the "leftover" domain.

With this alternative the equality of workloads from PSU to PSU is lost, but more emphasis is placed on the concentrations in the PSUs of the various domains of interest.

The second alternative measure, MOS^2, is to take as the PSU measure of size the maximum over all the domains of $N_{ik}/N_{\cdot k}$ (i.e., the maximum over all the domains of the proportion of a domain contained in the PSU). However, this measure does not take into account the relative importance of the various domains. A variation of MOS^2 which does take this into account is:

$$MOS^{2A}_i = \max_k \{f_k(N_{ik}/N_{\cdot k})\}$$

The authors point out (correctly) that a high correlation between PSU selection probabilities and PSU survey totals for a survey characteristic makes the between-PSU variance small. In their discussion of the effect on between-PSU variance of the use of MOS^2, they claim that it guards against a large between-PSU variance for any domain total because MOS^2 assigns a high selection probability to a PSU that has a high concentration of any of the domains. However, its unclear that this effect necessarily provides a high correlation between the PSU selection probability and a specific domain total. A PSU may have a high selection probability because of a high concentration of a specific domain, but other PSUs could have high selection probabilities even though they have a relatively low concentration of the specific domain. With MOS^2 the relationship between a domain total and the selection probability for a PSU is a rather complex and cannot easily be generalized.

In their introductory remarks to their simulation study, they state that the value of empirical research with real data is limited because of the lack of appropriate variables to evaluate between-PSU variances. It seems possible that there might be some useful variables available from the Health Interview Survey for this purpose. Perhaps additional consideration should be given to the potential value of empirical research.

In the discussion of their simulation results, the authors express surprise that MOS^1 leads to smaller between-PSU variances than does MOS^T, the standard measure of size, for estimating prevalence for the entire population. But this is not surprising in light of the fact that the various domains have different sampling rates. With MOS^1 being effected by these rates, the chances are increased that PSUs with higher concentrations of the domains of interest will be selected, providing a more stable estimate, from PSU to PSU, of these domain totals. This is important in the models (1 and 2) for which the disease prevalence rates are higher for the two special domains. For the other model it seems reasonable that the two measures come out about the same.

With respect to the simulation research that was conducted, some variations might be worth considering:

(1) The simulation provided PSUs that averaged 7700 persons. It might be useful to create variation in the expected PSU sizes.
(2) The simulation assigned the number of blacks and Hispanics independently in each PSU. However, it might be more realistic to allow the number of blacks and Hispanics in a PSU to be correlated.
(3) The sampling rates and disease incidence rates were basically fixed for the three domains. It might be of general interest to see what would happen if these rates were varied.

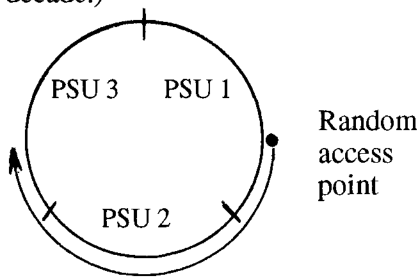
Finally, in spite of the advantages of a self-weighting sample, it might be more efficient to oversample domains in PSUs that have higher concentrations of these domains. This would reduce the amount of screening that would have to be done to achieve a given total sample size for a domain.

Discussion of the Ernst Paper

I worked with Larry Ernst at the Census Bureau for many years. I have always been impressed with his innovative skill in addressing applied survey design problems. His work in this paper is another example of this ability.

This topic, rotating PSUs for current household surveys, is of special interest to me because of a cross-division committee on PSU definitions for current surveys that I served on for several months, prior to leaving the Census Bureau. One committee member, Jim Roebuck, mentioned on several occasions his preference to minimize the number of rotating PSUs because of the practical difficulty and expense of dealing with them.

The main problem with the random access method (RAM) for dealing with rotating PSUs is that a small PSU can be selected for the sample for a period of time considerably less than the maximum amount of time that the small PSU could participate. This "partial" selection of small PSUs allows for the possibility of needing to select an "extra" PSU to complete the decade. For example, suppose that a stratum has three small PSUs and that each one has enough households to cover 2/3 of the sample needs for the decade. Then, with RAM, if the random access point for selecting small PSUs falls in the latter third of any of the three PSUs, all three will be selected for the sample. (This is illustrated in the accompanying chart where each PSU is represented by a third of the circle and the outside arc illustrates the selection of small PSUs for the decade.)



For the case in which any small PSU has enough housing units to provide sample for at least half the decade, Dr. Ernst has designed a clever method -- his Half-interval Method (HIM) -- that minimizes the number of rotations required to provide adequate sample for the decade. One weakness of this approach, relative to the RAM method, is that there is no control over the location of the two small PSUs selected; they could be very far apart. With the RAM method, the two or three small PSUs chosen would be relatively close together which may have some field operations advantages.

An alternative which might be considered is one which actually defines the problem away. The small PSUs, typically sparsely-populated counties, could be paired off prior to PSU selection. (If there were an odd number of PSUs, the odd one could be grouped with one of the other pairs or with a large PSU.) The pairing would be done primarily on a geographical basis. The resulting pairs (or groups) of small PSUs would serve as large PSUs so that there wouldn't be any small PSUs to worry about rotating. The main concern with this approach would be the problems in the field that would occur whenever any of these geographically large PSUs were selected.

Dr. Ernst's method of restricting the number of surveys for which a small PSU can be selected has some useful potential. Currently it is difficult to determine the most appropriate criterion for a small PSU -- i.e., one that would have to rotate if selected. For a PSU to be non-rotating, should it have to contain enough housing units to supply all current surveys for the decade? This criterion seems to be too conservative since the probability of selecting a relatively small PSU for multiple surveys is very small. So, what should the criterion for a small PSU be? With an approach like that of Dr. Ernst, in which PSUs are only selected for at most the number of surveys for which they could supply sample, the definition of a small PSU would be straightforward.

However, there appear to be some difficulties with the application of his approach. First, PSUs for current surveys are not identical, though they are quite close. In particular, PSUs defined for the CPS do not cross state boundaries; whereas PSUs for other surveys do. Allowing for some variation in PSU definitions could add considerable complexity to the procedure. In addition, the procedure is already very complex due to the need of establishing initial PSU selection probabilities, α_{ik}'s, for the various surveys. It appears that the choice of these probabilities has to be customized for the specific surveys involved.