7. OPERATIONAL IMPACT OF RE-SELECTING THE MCBS PSUS

7.1 Introduction

Chapters 1-6 have presented an evaluation of the advantages and costs of re-designing the MCBS with an alternative measure of size. Chapter 7 explores the practical impact of such a re-design. Topics discussed in this chapter include the implementation of a new PSU sample, weighting and estimation, imputation, variance estimation, and data distribution.

7.2 Design and Implementation of the New PSU Sample

In this section, we discuss the design of a new PSU sample and how the new sample would be integrated with the current panel structure. In Section 7.2.1, we briefly outline the approach for designing the new PSU sample. Next, in Section 7.2.2, we describe how the new PSU sample will be implemented in the context of the existing field operations. Finally, in Section 7.2.3, we describe some approaches for selecting Zip clusters with the sampled PSUs.

7.2.1 Design of the PSU Sample

PSU Formation

In designing a new survey, there is some advantage to starting with a clean slate of newly designed PSUs. However, experience has shown that the types of PSUs defined for the MCBS and other in-person national surveys (i.e., PSUs consisting of metropolitan areas or groups of rural counties) are generally robust and efficient with regard to maximizing sampling precision and minimizing survey costs. For this reason, we propose to use the same PSU definitions that are currently being used for MCBS, with the only exception being very large certainty PSUs that might be subdivided to form more efficient units for data collection. Similarly, it is useful in a few cases to combine small PSUs to ensure an adequate workload. Note that while the definition of a PSU may remain the same, the measure of size will change to reflect the current number of Medicare beneficiaries in the PSU.

Stratification for Current MCBS Design

As described earlier in Chapter 3, the MCBS PSU sample was based on Westat's 100-PSU 1980 master sample. In the 1980 master sample, the 20 largest PSUs were included in the sample with certainty. The remaining noncertainty PSUs were then stratified geographically by the four Census regions and by metropolitan status (MSA vs. non-MSA) within region.

Within these broad groups, PSUs were placed into substrata based on socioeconomic factors such as percent black, percent other minority, per capita personal income, and percent change in population between 1975 and 1980. The substrata were constructed to be internally homogeneous with respect to the PSU-level socioeconomic characteristics, and to be of roughly equal size.

Stratification for Redesign

For the redesign, all PSUs with a measure of size exceeding a specified cutoff will be included in the sample with certainty. For example, if the PSU measure of size is based on the number of Medicare beneficiaries, the analysis in Chapter 3 suggests that the 26 PSUs with ___ or more Medicare beneficiaries would be included in the sample as certainties.

To select the noncertainty PSUs, we will start by first stratifying the PSUs by Census region and MSA status, as was done for the 1980 master sample, except that Puerto Rico will be treated as a separate "region" for sampling purposes. In addition to the types of socioeconomic variables used to stratify the 1980 master sample, we will use HCFA enrollment data to identify areas with relatively high concentrations of Medicare beneficiaries. This information will be used in conjunction with the PSU-level socioeconomic data to define the detailed strata for selecting the sample of noncertainty PSUs. The strata will be constructed to be of roughly equal size (e.g., in terms of a weighted or unweighted count of Medicare beneficiaries), with the goal of selecting two PSUs per stratum with probabilities proportionate to size.

Once the design for the new sample has been specified, selection of the noncertainty PSUs can proceed using, for example, the Ernst procedure described in Chapter 6. [**???** In strata where the number of PSUs is so large that the Ernst procedure cannot be applied, an alternative approach (e.g., using independent sampling or Keyfitz selection) may be used instead.]

For the discussion below, we are assuming an overlap of about 70 percent (or 76 PSUs) will overlap with the old sample. As discussed in Chapter 6, there are a range of options for overlap methods that would affect the degree of overlap actually achieved in a redesign of the MCBS.

7.2.2 Implementation of New PSU Sample

Transition between Old and New Samples

After the new sample of PSUs has been selected, all subsequent annual samples of beneficiaries will be selected from the new PSUs. However, during the three years immediately following implementation, beneficiaries selected in previous rounds will continue to be followed in the old PSUs until their respective panels are rotated out. Thus, PSUs from the original design that are not included in the new sample will be phased out gradually over a three-year period. Moreover, in those newly-selected PSUs that do not overlap with the old PSUs, the sample workload will start out at about one-third the desired level and gradually increase over three years until it reaches the desired level under the rotating panel design.

After selection of the new PSU sample, there will be five categories of PSUs (the expected number of PSUs for each category is given in parentheses):

1. Certainty in both designs (26)
2. Certainty in old design, retained as noncertainty in new sample (3)
3. Noncertainty in old design, retained for new sample (47)
4. Certainty/noncertainty in old design, not selected for new sample (31)
5. Not in original design, selected as noncertainty for new sample (31)

The 107 PSUs in groups 1, 2 , 3, and 4 constitute the original MCBS PSUs, whereas the 107 PSUs in groups 1, 2, 3, and 5 constitute the new PSU sample. As discussed in Chapter 3, we estimate that 26 of the 33 current MCBS certainty PSUs (group 1) will be retained as certainties in the new design, with a 75% certainty cut-off. Furthermore, we assume that about 50 of the remaining MCBS PSUs (groups 2 and 3) will be retained in the new sample (see Chapter 6 for more details on the expected overlap). Thus, until the workload in the old PSUs is completely phased out of the study, the MCBS will be operational in an expected 138 PSUs.

Illustration of Transition

The phasing in and out of the PSU workload is illustrated in Table 7-1. The five groups shown in the table correspond to the groups discussed above. For the purpose of illustration, we assume that a new PSU sample is selected for the fall round of year 1.

For the fall round, no new samples of beneficiaries would be selected from those original PSUs not included in the new sample (Group 4). However, beneficiaries in the three most recent panels in these PSUs will continue to be interviewed. In year 1, the workload in these PSUs will be roughly 70 percent of the full workload. In years 2 and 3, the workload will be reduced further (to roughly 43 percent and 20 percent of the full workloads, respectively) as the older panels are released from the study. Thus, each original MCBS PSU that is not included in the new sample will remain in the study for another 3 years.

At the same time, the workload in newly selected PSUs will start out at a reduced level of approximately 30 percent since it will include only the current supplemental sample (Group 5 in Table 7-1). However, with the addition of new panels in each of the following two years, the workload will gradually increase to 57 percent and 80 percent, respectively, until it achieves full capacity in its fourth year of operation.

Finally, for the estimated 76 PSUs that are included in both the new and original designs, the workload will be maintained at the desired 100 percent level since the annual supplement will replace the panel that is scheduled to be released under the rotating panel design.

Table 7-1. Illustration of phase-in of new MCBS PSUs

Status of PSU	Number of PSUs	Approximate percent of typical workload†			
		Year 1	Year 2	Year 3	Year 4
1. Certainty in both designs	26	100%	100%	100%	100%
2. Certainty in old design, retained as noncertainty for new sample	3	100%	100%	100%	100%
3. Noncertainty in old design, retained for new sample	47	100%	100%	100%	100%
4. Certainty/noncertainty in old design, not selected for new sample	31	70%	43%	20%	0%
5. Not in original design, selected as noncertainty for new sample	31	30%	57%	80%	100%

† The "typical" workload refers to the workload associated with four active panels. Due to sample attrition, the panels are not equal in size. Older panels are generally smaller in size than newer ones. The percentages shown are intended to reflect the different sample size losses in the component panels over time. They are based on an assumed workload per panel of 4,500 interviews for the first year in sample, 4,000 interviews for the second year, 3,500 interviews for the third year, and 3,000 interviews for the fourth year.

7.2.3 Selection of Zip Clusters Within PSUs

Another consideration in the transition from the old to the new samples is how the second-stage selection of Zip clusters within PSUs will be affected by the introduction of the new PSUs.

For the existing MCBS design, the initial sample of Zip clusters was selected in 1991, using measures of size based on HCFA enrollment data. Once selected, the intention was to retain these sampled Zip clusters for all future sampling activities, in much the same way as the sample of PSUs is retained from year to year. However, unlike PSUs, Zip codes frequently change over time. Therefore, to ensure proper coverage of the newly formed ZIP codes, a sample of new Zip clusters is selected each year under the current MCBS design. These new Zip clusters are simply added to the previous year’s sample of Zip clusters. Then, for each annual supplement under the rotating panel scheme, beneficiaries are selected from both old and new Zip clusters. Over the nine years in which the MCBS has been in operation, the number of sampled Zip clusters has increased from the 1,500 originally selected in 1991 to over ___ active Zip clusters in 1999.

With the introduction of the new PSU sample, the method of selecting Zip clusters will depend on whether or not the PSU was included in the original sample.. For example, for the set of PSUs in the original MCBS sample that are not included in the new sample (Group 4 of Table 7-1), there would be no need to augment the Zip cluster sample. The existing sample of Zip clusters in these PSUs still provides an unbiased sample of beneficiaries who were eligible for selection for the older panels. On the other hand, a completely new sample of Zip clusters will be selected from the corresponding set of newly selected PSUs (Group 5 of Table 7-1). The sample of Zip clusters in these PSUs will be designed and selected “from scratch” using the most up-to-date beneficiary counts available in HCFA’s enrollment data base.

However, for the PSUs in the new design that overlap with the original PSUs (Groups 1-3 of Table 7-1), there are at least three alternative methods for selecting the Zip cluster sample.

- The first would be to simply continue the practice of augmenting the existing sample of Zip clusters in these PSUs with a sample of new Zip clusters. Although this approach is unbiased, it will lead to reduced sampling precision and possibly increased costs because the older Zip clusters will be retained with their original (and outdated) selection probabilities.
- A second alternative would be to select a “new” sample of Zip clusters using Keyfitz-type procedures to maximize the overlap with the existing Zip cluster sample. The advantage of this approach is that the Zip cluster selection probabilities can be updated to be consistent with the current size of the Zip cluster. On the other hand, it is likely that over time the Zip cluster sample will have achieved such a high level of dispersion throughout the PSU that the benefits of retaining the same clusters is significantly reduced. Under these conditions, the costs associated with fielding a completely new and independently selected sample of Zip clusters in addition to the existing Zip cluster sample may not be significantly higher than either of the first two alternatives.
- A third alternative, therefore, would be to design and select a completely new sample of Zip clusters in these PSUs that would eventually replace the existing Zip clusters.

Like the phasing out of old PSUs described in Table 7-1, the old Zip clusters will be eventually be phased out of the sample as the existing sample of beneficiaries in the old Zip clusters are rotated out of the study. While more research needs to be done to fully understand the implications of the various alternatives, the third approach described above seems to be a promising one.

7.3 Impact on Weighting and Estimation

Sampling weights can be thought of as indicating the number of persons that a particular sample observation represents. Sampling weights vary across members of the sample for three main reasons: (1) to compensate for unequal selection probabilities, (2) to attempt to compensate for differential nonresponse and undercoverage, and (3) to attain greater precision for the survey estimates through poststratification. Like many complex surveys, MCBS uses sampling weights for all three purposes.

The weighting procedures that would be used in a redesigned MCBS would be almost identical to those currently in use. Moreover, since most of the weighting steps are internal to each panel, very few modifications would be necessary to incorporate the new sample.

For example, under the current weighting procedures the initial weighting step for each new panel is to assign a baseweight for each sampled beneficiary. The second step uses a raking algorithm to adjust the baseweights so that weighted counts correspond to administrative counts. Sampling weights for each panel are then created by adjusting the raked weights for nonresponse. Since all of these steps are applied to each panel separately, it does not matter whether the different panels are based on the same or a different sample design.

With the introduction of the new PSU sample, estimates for the newest panels will be based on a different set of PSUs than those used in the older panels (i.e., until the old panels are "phased out" of the study). The samples for each weighting delivery are comprised of sampled beneficiaries from several different MCBS panels. For these samples, the panel-specific weights will be appropriately adjusted using composite estimation to account for overlapping coverage. As described in detail below, even this final compositing step will be unaffected by the introduction of the new sample.

7.3.1 Baseweights

The baseweight for each sample person is the reciprocal of the overall probability of being selected for the sample. Currently, equi-probability samples are selected from each age domain for each panel, and this will also be true for each panel in the new sample.

The overall probability of selection of any person is the product of the probabilities at the various stages of selection; that is, the product of the probability of selection for the PSU, the conditional probability of selection for the ZIP cluster given that the PSU containing the ZIP cluster was selected, and the conditional probability of selection for the beneficiary given that the beneficiary's ZIP cluster and PSU were selected. To account for the fact that the HISKEW file from which the beneficiary samples are drawn is a 5 percent sample, the baseweight also includes a factor of 20.

7.3.2 Poststratification

Poststratification adjustments will be made using a raking adjustment to adjust the baseweights so that weighted sample counts correspond to administrative counts from the HISKEW sampling frame. Adjustment cells will be defined based on age category by sex by race, region by age category, metropolitan status by age category, and accretion year.

Each eligible beneficiary will be assigned to a cell, using data from the sampling frame to determine eligibility status and to identify the appropriate cell. Within each adjustment cell, the baseweights for each eligible sampled beneficiary are adjusted by a factor that is equal to the ratio of the control total for the cell and the weighted estimate for the cell. The resulting weighted count, using the poststratified weights, is then the same as the control total for each cell.

Since poststratification is done only to each new panel, no modifications of this procedure are necessary for the new sample.

7.3.3 Nonresponse Adjustments

Following the initial poststratification for each new panel, we will adjust the poststratified weights for nonresponse in the initial interviewing round. For nonresponse in subsequent rounds, we will make a single adjustment for each year that accounts for nonresponse over the three interview rounds (winter, spring, fall).

A separate set of adjustment cells will be created for each panel for each nonresponse adjustment. Special adjustment cells will be created for groups such as recent deaths and residents of facilities, for whom response propensities are different from the general beneficiary population. For the remaining sampled beneficiaries, adjustment cells will be based on modeled response propensity within panel.

Within each cell, weights of respondents will be adjusted to account for weights of nonrespondents. Response propensity will be modeled using logistic regression. Within each adjustment cell, a ratio adjustment is applied so that the resulting sum of the adjusted weights for respondents is equal to the sum of incoming weights for both respondents and nonrespondents. The adjustment factor is computed from the sum of the incoming weights for all beneficiaries in the cell divided by the sum of weights for responding beneficiaries in the cell.

Again, since the adjustment is done within each panel, no modification of existing procedures is needed for the new sample.

7.3.4 Sample Combination

Each annual MCBS supplement (panel) is selected as a nationally representative sample that represents the population of beneficiaries who are alive and eligible as of January 1 of the given year. Samples for MCBS public use files include beneficiaries from several different MCBS panels. Weighting adjustments for the combined sample are needed to account for overlapping coverage in the panels. This adjustment involves the use of "combination factors" that are applied to the previously calculated panel-specific nonresponse-adjusted weights. In general, these combination factors are proportional to the effective sample sizes for the panels being combined. Separate adjustment factors will be computed for each age stratum.

Generally, this weighting adjustment is applied in three steps. The first step is to combine respondents in the panel that is three years old with the portion of respondents in the two-year-old panel that overlaps in coverage (e.g., beneficiaries who became eligible on or before January 1 of earlier year). This is known as the "two-year backward longitudinal" sample. The second step combines these combined panels with the portion of respondents in the year old panel that overlap in coverage. This is known as the "one-year backward longitudinal" sample.

The final step is to combine the three panels with the portion of the current panel respondents that overlap in coverage. At each combination step, combination factors will be based on the proportion of the effective sample size in each sample. Combination factors will be determined and applied separately within each age stratum and accretion status. This procedure results in weights that are adjusted for multiple chances of being sampled; and it produces weighted estimates for the combined sample that are substantially unbiased. In addition, under certain circumstances, the factors have the property that they yield combined estimates with approximately minimum variance.

The procedures for the final combination step are slightly modified to accommodate special one-time supplements (e.g., the ORD managed care supplements). The resulting sample from the steps already described is known as the "classical" MCBS cross-sectional sample for current year access to care. The final step is to combine the special supplement with the "classical" MCBS cross-sectional sample. Adjustment factors for the final combination step are computed from the effective sample size in each sample by age stratum, supplement area, and the strata used to select the special sample.

Table 7-2 illustrates the sample combination while phasing in the new sample. For example in weighting year 1, steps 1 and 2 combine the three continuing panels selected under the old design. The third step combines these three continuing panels with TIS-1 panel selected under the new design. Each year, an additional panel selected under the new design is included in the combinations until the fourth year when all panels are from the new design.

Similar procedures will be used for each combination step involving a panel selected from the new sample. We will evaluate the methods for computing the effective sample size for each combination to incorporate the increased precision of the panels selected from the new sample. Currently, the effective sample size is computed based on the coefficient of variation of the weights. It will be useful to compute the effective sample size for a number of variables of interest from the survey, in order to examine the extent to which the effective sample sizes varies depending on the variable chosen. The results of this analysis may suggest more nearly optimal combination factors.

Table 7-2. Illustration of sample combination

	Weighting Year			
Panel	1	2	3	4
Current year minus 3	Old Design Continuing			
Current year minus 2	Old Design Continuing	Old Design Continuing		
Current year minus 1	Old Design Continuing	Old Design Continuing	Old Design Continuing	
Current year				New Design Continuing
Current year + 1				New Design Continuing
Current year + 2				New Design Continuing
Current year + 3				New Design first year

7.4 Impact on Imputation and Variance Estimation

7.4.1 Introduction

One of the most important products of the MCBS is the annual sourcebook series. These publications provide a broad range of estimates from the MCBS, furnishing health analysts with a longitudinal time series that can be used to evaluate trends over time in health and health care among Medicare Beneficiaries.

This section discusses the potential impact of an MCBS redesign on the imputation, survey estimates, and variance calculations that are required to produce the sourcebook series.

7.4.2 Impact of PSU Redesign on Imputation

Imputation

Imputation is used to correct “item nonresponse,” questions that are left blank because a survey respondent has incomplete knowledge or refuses to answer. Item nonresponse occurs in virtually all large surveys. Sample persons in the MCBS, for example, often are unable or unwilling to provide complete information on their income and expenditures for medical care. Westat uses logical edits and a hot-deck imputation to produce complete information on income and medical care expenditures for the MCBS cost and use files.

In the MCBS, “hot deck” imputation is used. In this procedure, a value from one respondent (the donor) is used to “fill in” the missing value for another respondent (the recipient). Donors and recipients are matched in the imputations by using auxiliary variables to identify persons or events with common characteristics. In the MCBS hot-deck procedure, the auxiliary variables are called "boundary" variables. They are used to partition the sampled units into homogeneous classes with similar characteristics. The goal is to explain as much of the variance in the imputation variable as possible by the boundary variables rather than other random factors.

In the income and asset (IA) imputation, both demographic characteristics and IA data collected in previous rounds are used as boundary variables. Currently, PSU is not a boundary variable in our hot-decking procedures. For some time, we have considered the addition of PSU as either a hard or soft boundary variable in imputation procedures (both in IA and ghost-donor matching). This addition should control for geographic variations that explain differences between survey responses. It will also help to smooth the transition between the new and old samples in the first three years after the PSU redesign, when the sample will contain a mix of PSUs from both designs.

PSU redesign should not affect the time or level of effort required to impute for missing data in the MCBS. The edit and imputation programs have been completely developed at this point, and they will not change as PSUs are added to or dropped from the MCBS.

7.4.3 Impact of PSU Redesign on Longitudinal Data Analysis

The MCBS provides a valuable data source for policy makers and researcher to follow trends in health care cost and financing, and utilization by this population because of its continuous rotating panel design. The MCBS annual sourcebook series illustrates the use of this survey data to describe trends of this dynamic population. Estimates of net change between years are obtained by computing the difference between two cross-sectional estimates (O'Connell, J., A. Chu, and R.C.Bailey. (1997). Considerations for analysis of the Medicare Current Beneficiary Survey (MCBS) across time. *1997 Proceedings of the Survey Methods Research Section, American Statistical Association*: Anaheim, CA.). Annual rates of change are also calculated based on net changes. Other types of longitudinal analysis are also possible, using the MCBS data. It is important to keep the consistency of the data so that comparisons made across years are valid, and irregularities reflected in the data can be attributed to factors other than sampling design.

A question might be raised as to whether the new PSU design may have confounding effects on the trend data, especially on utilization and expenditure data. The MCBS samples are drawn, stratified by basic demographic characteristics rather than by beneficiaries' utilization patterns. Providers in different geographic areas tend to vary significantly in medical practices. This fact is in turn directly associated with how and how much beneficiaries utilize medical services. Furthermore, fee schedules of providers tend to vary significantly across geographic areas. Switching to new PSUs may propagate these geographic differences, and cause discontinuity in the trend data across the years.

However, we would argue that significant disruptions in trend data are unlikely to happen. First of all, the impact of PSU redesign on the trend data, if any, is determined by the proportion of new PSUs phased in each year's supplement. As discussed in an earlier section, the complete redesign may achieve a 75 percent of overlap of PSUs between the current and the new design. The resulting 25 percent of new PSUs, phased in over a three-year period, amounts to slightly over 8 percent of new PSUs in a single year. Therefore, the impact of the redesign is gradually introduced in our data. Secondly, the majority of certainty PSUs (n=26) and the PSUs retained from the old design (n=50) tend to be larger metropolitan/urban areas where there are larger concentrations of beneficiaries. The uncertain PSUs (n=31) in the new design tend to be PSUs in more rural areas. Since sample sizes in uncertain PSUs are likely to be smaller than those of certainty and more urban PSUs, the actually impact of the redesign on the MCBS sample should be less than 8 percent. Third, given the fact that the redesign is going to retain the nationally representative characteristics of the sample, we should be able to assume that geographic variations of high use and low use are evened out in the sampling process. Therefore, it is safe to assume that PSU redesign is expected to have minimum impact on trend data analysis and on longitudinal data presented in the sourcebook.

7.4.4 Sourcebook production

The MCBS annual sourcebook series presents cross-sectional as well as trend data on a dynamic population, ever-changing with incoming newly eligible beneficiaries and exiting beneficiaries who died. Estimates of trend data in the sourcebook are using cross-sectional weights, because longitudinal weights can not account for natural changes of the Medicare population over time. Estimates of net change between years are obtained by computing the difference between two cross-sectional estimates (O'Connell, J., A. Chu, and R.C.Bailey. (1997). Considerations for analysis of the Medicare Current Beneficiary Survey (MCBS) across time. *1997 Proceedings of the Survey Methods Research Section, American Statistical Association*: Anaheim, CA.).

PSU redesign is expected to have minimum impact on trends data presented in the sourcebook, since each year's sample is theoretically nationally representative in nature. These samples are wighted up to national totals regardless of PSU mix. Nevertheless, it is possible that irregularities might show up in the trend data over the time period when the new PSU design is phased in. However, as discussed earlier, these irregularities are expected to be minor. (Is there any statistical method to tease out the differences caused by PSU redesign?)

7.4.5 Variance Estimation

For the MCBS, standard error estimates are calculated using a modified version of balanced repeated replication (BRR) known as "Fay's method." This method is a form of pseudo replication in which a predetermined number of subsamples (referred to as "replicates") are initially generated from the full sample. The replicates are formed in such a way that they resemble the full sample with respect to the stratification, clustering, and other relevant features used in the MCBS sample design. Each replicate is then reweighted using the procedures developed for the full sample. Each reweighting produces a set of "replicate weights" that not only reflect the stratification and clustering used in the MCBS, but also features of the estimation process such as the nonresponse and poststratification weighting adjustments described in Section 7.3.

Fay's estimate of the variance of a sample-based estimate, $\hat{q}$, is given by:

$$V(\hat{q}) = \frac{1}{(1-k)^2}\left(\frac{1}{G}\sum_{g=1}^{G}(\hat{q}_{(g)} - \hat{q})^2\right)$$

where $\hat{q}$ is the sample estimate based on the full-sample weights, $\hat{q}_{(g)}$ is the gth replicate estimate based on the observations included in the gth replicate and corresponding replicate weights, G is the total number of replicates formed, and 100(1-k)% is a constant known as "Fay's perturbation factor." Based on an analysis described in Judkins (1990), a value of k = 0.3 was chosen to compute the required perturbation factor for MCBS variance estimates.

For the MCBS, 100 variance-estimation strata were used to create the required replicates. Thirty-seven of these variance strata coincided with the first-stage noncertainty strata defined for the MCBS in which two PSUs were selected from each stratum. Within each variance stratum, the first sampled PSU defined what is referred to as the first "variance unit," while the second PSU defined the second "variance unit."

The remaining 63 variance strata were created from the 33 certainty PSUs. In general, the variance units within the certainty PSUs were formed at the ZIP cluster level, except for ZIP clusters selected with certainty within the PSU. For the certainty ZIP clusters, variance units were formed at the beneficiary level. Each resulting variance stratum either contained two or three variance units. The procedures used to group the noncertainty ZIP clusters in certainty PSUs into variance strata are summarized below:

(1) Certainty PSUs with more than three sampled (noncertainty) ZIP clusters are split into two or more preliminary variance strata, each of which contains two or three ZIP clusters;

(2) Certainty PSUs with one sampled (noncertainty) ZIP cluster are paired or tripled into a single preliminary variance stratum, within which each ZIP cluster constitutes a variance unit;

(3) For preliminary strata with two or three clusters, each ZIP cluster is assigned to a separate variance unit;

(4) The resulting preliminary variance strata are combined into 63 final variance strata, with two or three variance units in each stratum.

Variance strata/units for ZIP clusters selected with certainty in the certainty PSUs are formed at the beneficiary level. Beneficiaries are paired based on variables used in selecting the beneficiary sample. Each pair of beneficiaries constitutes a preliminary variance stratum; and each beneficiary is assigned to a different variance unit. These preliminary strata are then combined into the final strata.

For the current MCBS design, the required variance strata and variance units were defined for the original 1991 panel at the time of sampling. As each new panel was added to the sample, the newly sampled beneficiaries were assigned to appropriate variance strata and variance units using the procedures described above. The variance strata and units are then used to create 100 “balanced” replicates by choosing one variance unit from each variance stratum. Additional details about the formation of the replicates using Fay’s modified BRR method are given in Judkins (1990).

Notice that new panels are integrated into the existing variance replicate structure. Because a redesigned sample would be rotated into the current sample by forming a new panel each year, there should be minimal impact on variance estimation. New panels would be rotated in, just like in previous years.

With the use of replication methods such as BRR, features of the sample design that affect sampling precision such as stratification, clustering, and use of weighting adjustments, are appropriately reflected in the resulting variance estimates. The same procedures used to create 100 variance strata and associated variance units for the current MCBS sample can also be applied to the new PSU sample. The result will be a set of replicates for the new sample that can be merged with the corresponding replicates from the original sample for variance estimation.

7.5 Impact on Data Distribution

Another operational aspect of the MCBS is data distribution, which occurs primarily through the release of public use files. The redesign of the MCBS will result in an increase in the number of PSUs for the transitional years, which could confuse users of public release files. However, weighted tabulations would not be affected (since they go across PSUs); also, as discussed earlier, standard error calculations should not be affected. The change in survey design can be documented in accompanying text files, using an illustration like Table 7-1, with new and old PSUs being indicated by flag variables, if necessary. Thus we anticipate minimal impact of a survey redesign on data distribution.