METHODS OF SELECTING SAMPLES IN MULTIPLE SURVEYS TO REDUCE RESPONDENT BURDEN

Charles R. Perry, Jameson C. Burt and William C. Iwig, National Agricultural Statistics Service
Charles Perry, USDA/NASS, Room 305, 3251 Old Lee Highway, Fairfax, VA 22030

KEY WORDS: Respondent burden, multiple surveys, stratified design

Summary

The National Agricultural Statistics Service (NASS) surveys the United States population of farm operators numerous times each year. The list components of these surveys are conducted using independent designs, each stratified differently. By chance, NASS samples some farm operators in multiple surveys, producing a respondent burden concern. Two methods are proposed that reduce this type of respondent burden. The first method uses linear integer programming to minimize the expected respondent burden. The second method samples by any current sampling scheme, then, within classes of similar farm operations, it minimizes the number of times that NASS samples a farm operation for several surveys.

The second method reduces the number of times that a respondent is contacted twice or more within a survey year by about 70 percent. The first method will reduce this type of burden even further.

Introduction

The National Agricultural Statistics Service (NASS) surveys the United States population of farm operators numerous times each year. Some surveys are conducted quarterly, others are conducted monthly and still others are conducted annually. Each major survey uses a list dominant multiple frame design and an area frame component that accounts for that part of the population not on the list frame. The list frame components of these surveys constitute a set of independent surveys, each using a stratified simple random sample design with different strata definitions. With the current procedures some individual farm operators are sampled for numerous surveys while other farm operators with similar design characteristics are hardly sampled at all. Within the list frame component, two methods of sampling are proposed that reduce this type of respondent burden.

Historically, NASS has attempted to reduce respondent burden and also reduce variance. In 1979, Tortora and Crank considered sampling with probability inversely proportional to burden. Noting a simultaneous gain in variance with a reduction in burden, NASS chose not to sample with probability inversely proportional to burden. NASS has reduced burden on the area frame component of its surveys. There, a farmer sampled on one survey might be exempt from another survey, or farmers not key to that survey might be sampled less intensely. Statistical agencies in other countries have also approached respondent burden. For example, the Netherlands Central Bureau of Statistics does some co-ordinated or collocated sampling, ingeniously conditioning samples for one survey on previous surveys (de Ree, 1983).

Formal Description of Methods I and II

Method I is formally described by four basic tasks.

(a) Cross-classify the population by the stratifications used in the individual surveys. This produces the coarsest stratification of the population that is a substratification of each individual stratification.

(b) Proportionally allocate each of the individual stratified samples to the *substrata*. Use random assignment between substrata where necessary.

(c) Apply integer linear programming within each substratum to assign the samples to the labels of units belonging to the substratum so that the respondent burden is minimized.

(d) Randomize the labels to the units of substratum. The final assignment within each substratum is a simple random sample with respect to each of the proportionally allocated samples.

Method II is formally described by four basic tasks.

(a) Using an equal probability of selection technique within a stratum, select independent stratified samples for each survey. Notice that the equal probability of selection criterion permits efficient zonal sampling techniques on each survey within strata. Currently, within strata samples are selected systematically with records essentially in random order.

(b) Substratify the population by cross-classifying the individual farm units according to the stratifications used in the individual surveys.

(c) Randomly reassign within each substratum the samples associated with units having excess respondent burden to units having less respondent burden.

(d) Iterate the reassignment process until it minimizes the number of times that NASS samples a farm operator for several surveys in the substratum.

For both methods, define respondent burden by an index that represents the comparative burden on each individual sampling unit in the population. Each survey considered is assigned a burden value. When a sampling unit is selected for multiple surveys, the burden index may be additive or some other functional form dependent on the individual survey burden values. Consequently, each sampling configuration is assigned a unique respondent burden index.

For any reasonable respondent burden index, the first method minimizes the expected respondent burden. This follows easily from the following observations, where it is assumed that for each of the original surveys an equal probability of selection mechanism (*epsm*) is used within strata. First, from the independence of the original sample designs, it follows that for each individual unit the expected burden from the original stratified samples is equal to the expected respondent burden using proportional allocation followed by epsm sampling within substrata. Since the respondent burden over any population is the sum of the respondent burden on the individuals of the population, the equality holds for the entire population or any subpopulation including the substrata. That is, the expected respondent burden over any arbitrary substratum for the proportionally allocated samples is equal to the expected respondent burden of the original stratified sample allocations over the substratum. Originally, these allocations are random to each substratum, constrained only so that the substratum sample sizes sum to their stratum sample size. Second, for the first method the respondent burden is minimized over each substrata by the linear programming step.

Regarding variance reduction, this means that if the original sample was selected using simple random sampling within each stratum, then the first method reduces respondent burden without any offsetting increase in variance, since proportional allocation is at least as efficient as simple random sampling. However, the first method would be less efficient for variance than zonal sampling unrestricted by burden. But the second method, by reallocating some zonal sampling units to reduce respondent burden, may only slightly increase variance over no reallocation and then only when zonal sampling is effective.

A Simple Simulation of Method I

Method I reduces respondent burden in the following simulation of two surveys. Survey I samples $n = 20$ from $N = 110$. Survey II also samples $n = 20$ from $N = 110$, though each of its strata has either a larger or smaller population size ($N_{.1} = 40$ and $N_{.2} = 70$) than the corresponding strata of survey I ($N_{1.} = 30$ and $N_{2.} = 80$). Here, the first subscript corresponds to the first survey, with its strata 1 and 2. Similarly, the second subscript corresponds to the second survey. For example, $N_{21} = 30$ corresponds to the size of the population in stratum 2 of survey I and in stratum 1 of survey II, while $\bar{n}^{(1)}_{21} = 3.75$ corresponds to the proportional allocation of survey I's stratum 2 sample, $n^{(1)}_{2.} = 10$, to the population in both stratum 2 of survey I and stratum 1 of survey II.

Survey I	Survey II: Stratum 1	Survey II: Stratum 2	
Stratum 1	$N_{11} = 10$ $\bar{n}^{(1)}_{11} = 3.33$ $\bar{n}^{(2)}_{11} = 2.5$	$N_{12} = 20$ $\bar{n}^{(1)}_{12} = 6.67$ $\bar{n}^{(2)}_{12} = 2.85$	$N_{1.} = 30$ $n^{(1)}_{1.} = 10$
Stratum 2	$N_{21} = 30$ $\bar{n}^{(1)}_{21} = 3.75$ $\bar{n}^{(2)}_{21} = 7.5$	$N_{22} = 50$ $\bar{n}^{(1)}_{22} = 6.25$ $\bar{n}^{(2)}_{22} = 7.15$	$N_{2.} = 80$ $n^{(1)}_{2.} = 10$
	$N_{.1} = 40$ $n^{(2)}_{.1} = 10$	$N_{.2} = 70$ $n^{(2)}_{.2} = 10$	

With two surveys, at most we will sample a respondent twice. For the above two surveys, without any proportional allocation, we simulated two independent stratified simple random samples 3 million times. These simulated samples produced, on average, 3.6 double hits for the whole population of 110 potential respondents, and four percent of the simulations produced 7 or more double hits. With the proportional allocations indicated in the diagram for Method I, the population exceeds the total sample for both surveys in each substratum, so no sampling unit needs to be selected for both surveys. The high respondent burdens of independent sampling are reduced to 0 double hits with Method I!

Operational Description

Basic Notation

Let $\mathcal{U} = \{u_i\}_{i=1}^{N}$ be a finite population of size N. Suppose that $\mathcal{U}$ is surveyed on K occasions and that on each occasion a different independent stratified design is used. For these K stratified designs, denote the survey occasion by $k = 1, 2, \dots K$ and let us use the following notation.

$H^{(k)}$:the number of strata for design k,

$\mathcal{U}_h^{(k)}$:the units (the set of them) in stratum h for design k,

$N_h^{(k)}$:the size of stratum h for design k,

$n_h^{(k)}$:the sample size in stratum h for design k,

$f_h^{(k)} = n_h^{(k)}/N_h^{(k)}$:the sampling fraction in stratum h for design k,

$n^{(k)} = \sum_{h=1}^{H^{(k)}} n_h^{(k)}$:the overall sample size for design k, and

$N = N^{(k)} = \sum_{h=1}^{H^{(k)}} N_h^{(k)}$:the overall population size.

Remark

Requiring the population to be exactly the same for each survey may seem rather restrictive. However it is not, since, for each survey, one can easily introduce an extra stratum that contains the units not covered by that survey. Obviously the sample sizes associated with the extra noncovered strata are taken to be zero. This permits one to apply either Method I or Method II over years.

Warning: In multiyear applications, care must be taken to ensure that no information from the sample data is used to update any of the frames being considered. Failure to do so can lead to biased estimates. These are the same restrictions that apply to the permanent random number techniques discussed by Ohlsson (1993).

<u>Method I</u>

Using this notation for Method I, we next describe a sequence of simple data manipulation steps that can be used operationally to perform tasks (a) through (d) on page 1 for each of the K surveys.

Suppose that each unit, u_i, of the population $\mathcal{U}$ has been stratified for each of the K surveys. Further suppose that this information has been entered into a file containing N records, so that the ith record contains the stratification information for unit i. To be definitive, assume that the variable $S(k)$ denotes the stratum classification code for survey k and that $S(k : i)$ denotes the value of the stratum classification code for unit u_i.

For each survey k $(k = 1, 2, \dots, K)$ perform the following sequence of operations.

(a) Sort the data file by the variables $S(k), \dots, S(K), S(1), \dots, S(k-1)$. This will hierarchically arrange the records of the population, first by the stratification of survey k, by the stratification of survey $k+1$ within the stratification of survey k, then by the stratification of survey $k+2$ within the stratification of survey $k+1$, $\dots$, by the stratification of survey K within the stratification of survey $K-1$, then by the stratification of survey 1 within the stratification of survey K, $\dots$, then by the stratification of survey $k-1$ within the stratification of survey $k-2$. In terms of the *substrata* formed by the cross-classification, the records of the population are arranged sequentially after sorting as

$$\mathcal{U}_{1,1,\dots,1,1,\dots,1,1}^{(k,k+1,\dots,K,1,\dots,k-2,k-1)},$$
$$\mathcal{U}_{1,1,\dots,1,1,\dots,1,2}^{(k,k+1,\dots,K,1,\dots,k-2,k-1)},$$
$$\vdots$$
$$\mathcal{U}_{1,1,\dots,1,1,\dots,1,H^{(k-1)}}^{(k,k+1,\dots,K,1,\dots,k-2,k-1)},$$
$$\mathcal{U}_{1,1,\dots,1,1,\dots,2,1}^{(k,k+1,\dots,K,1,\dots,k-2,k-1)},$$
$$\vdots$$
$$\mathcal{U}_{H^{(k)},H^{(k+1)},\dots,H^{(K)},H^{(1)},\dots,H^{(k-2)},H^{(k-1)}-1}^{(k,k+1,\dots,K,1,\dots,k-2,k-1)},$$
$$\mathcal{U}_{H^{(k)},H^{(k+1)},\dots,H^{(K)},H^{(1)},\dots,H^{(k-2)},H^{(k-1)}}^{(k,k+1,\dots,K,1,\dots,k-2,k-1)}$$

where

$$
\begin{aligned}
& \mathcal{U}_{h_k,\ldots,h_K,h_1,\ldots,h_{k-1}}^{(\mathrm{k},\ldots,\mathrm{K},1,\ldots,k-1)} \\
= \; & \mathcal{U}_{h_k}^{(k)} \bigcap \ldots \bigcap \mathcal{U}_{h_K}^{(K)} \bigcap \mathcal{U}_{h_1}^{(1)} \bigcap \ldots \bigcap \mathcal{U}_{h_{k-1}}^{(k-1)} \\
= \; & \mathcal{U}_{h_1}^{(1)} \bigcap \ldots \bigcap \mathcal{U}_{h_{k-1}}^{(k-1)} \bigcap \mathcal{U}_{h_k}^{(k)} \bigcap \ldots \bigcap \mathcal{U}_{h_K}^{(K)} \\
= \; & \mathcal{U}_{h_1,h_2,\ldots,h_{k-1},h_k,h_{k+1},\ldots,h_K}^{(1,2,\ldots,k-1,\mathrm{k},k+1,\ldots,\mathrm{K})}
\end{aligned}
$$

Both the size and sequential arrangement of the substrata of stratum h for survey k are displayed schematically as

$$N_{\mathrm{h},1,\ldots,1,1,\ldots,1,1}^{(\mathrm{k},k+1,\ldots,\mathrm{K},1,\ldots,k-2,k-1)}$$

$$N_{\mathrm{h},1,\ldots,1,1,\ldots,1,2}^{(\mathrm{k},k+1,\ldots,\mathrm{K},1,\ldots,k-2,k-1)}$$

$$\vdots$$

$$N_{\mathrm{h},h_{\mathrm{k+1}},\ldots,h_K,h_1,\ldots,h_{\mathrm{k-2}},h_{\mathrm{k-1}}}^{(\mathrm{k},k+1,\ldots,\mathrm{K},1,\ldots,k-2,k-1)}$$

$$\vdots$$

$$N_{\mathrm{h},H^{(k+1)},\ldots,H^{(K)},H^{(1)},\ldots,H^{(\mathrm{k-1})}-2,H^{(\mathrm{k-1})}-1}^{(\mathrm{k},k+1,\ldots,\mathrm{K},1,\ldots,\mathrm{k-2},\mathrm{k-1})}$$

$$N_{\mathrm{h},H^{(\mathrm{k+1})},\ldots,H^{(K)},H^{(1)},\ldots,H^{(\mathrm{k-1})}-1,H^{(\mathrm{k-1})}}^{(\mathrm{k},k+1,\ldots,\mathrm{K},1,\ldots,\mathrm{k-2},\mathrm{k-1})}$$

where

$$N_{h_k,\ldots,h_K,1,\ldots,H_{k-1}}^{(\mathrm{k},\ldots,\mathrm{K},1,\ldots,k-1)}$$

denotes the number of units in

$$\mathcal{U}_{h_k,\ldots,h_K,1,\ldots,H_{k-1}}^{(\mathrm{k},\ldots,\mathrm{K},1,\ldots,k-1)} .$$

(b) To randomly proportion the sample $n_h^{(k)}$ for stratum h of survey k to the subintervals of stratum k:

(1) Divide the length of stratum h for survey k, $N_h^{(k)}$, into a sequence of $n_h^{(k)}$ subintervals of integer length that differ in length by at most 1. Do this by forming $\frac{N_h^{(k)}}{n_h^{(k)}}$ as yet unpopulated subintervals, each with the length $n_h^{(k)}$, leaving $N_h^{(k)} - \left(\left[\frac{N_h^{(k)}}{n_h^{(k)}} \right] n_h^{(k)} \right)$ imaginary population units to be assigned. Randomly distribute these remaining imaginary units (without replacement) to the $\left[\frac{N_h^{(k)}}{n_h^{(k)}} \right]$ subintervals. Now populate these subintervals by randomly selecting a starting unit from the $N_h^{(k)}$ units. This starting unit begins the first subinterval, with its size randomly determined as above, $\left[\frac{N_h^{(k)}}{n_h^{(k)}} \right]$ or $\left[\frac{N_h^{(k)}}{n_h^{(k)}} \right] + 1$. Sequentially continue to populate the above subintervals, wrapping around to the first unit for one of the subintervals. This method of forming subintervals will let us keep the same probability of selection $\frac{n_h^{(k)}}{N_h^{(k)}}$ for each unit in that subinterval. It does not choose a sample.

(2) Randomly select an integer from each subinterval [while this integer corresponds to a population unit, it is not used here to select that population unit–for that, see (d) below].

The number of these random integers falling in the interval corresponding to

$$N_{h_k,h_{\mathrm{k+1}},\ldots,h_K,h_1,\ldots,h_{\mathrm{k-2}},h_{\mathrm{k-1}}}^{(\mathrm{k},k+1,\ldots,\mathrm{K},1,\ldots,k-2,k-1)}$$

in the sequential ordering is the size of the randomly proportioned sample for survey k to be drawn from the substratum population

$$\mathcal{U}_{h_k,h_{\mathrm{k+1}},\ldots,h_K,h_1,\ldots,h_{k-2},h_{k-1}}^{(\mathrm{k},k+1,\ldots,\mathrm{K},1,\ldots,k-2,k-1)} .$$

Denote this sample size for the substratum by

$$m_{h_k,h_{\mathrm{k+1}},\ldots,h_K,h_1,\ldots,h_{k-2},h_{k-1}}^{(\mathrm{k},k+1,\ldots,\mathrm{K},1,\ldots,k-2,k-1)}$$

or

$$m_{h_1,\ldots,h_k,h_{\mathrm{k+1}},\ldots,h_{\mathrm{K}}}^{(k)}$$

where the subscripts in the last expression are understood to be in natural order.

Repeating steps (a) and (b) above for each of the K surveys, we have randomly proportioned the K original stratified sample sizes to the substrata.

(c) Next we describe how to use integer linear programming to assign *within a substratum* the above proportioned samples to the substratum unit labels–not specific population units yet. We do this so that the respondent burden is minimized for an *arbitrary* positive linear respondent burden function (index).

Suppose that $m^{(1)}$, $m^{(2)}$, ..., $m^{(K)}$ samples have been randomly proportioned to a substratum of size M. Clearly the random proportioning procedure described above insures that $m^{(k)} \leq M$ for $k = 1, 2, \ldots, K$.

Moreover, if the size of the total sample $m = m^{(1)}+m^{(2)}+\ldots+m^{(K)}$ randomly proportioned to the substratum is less than or equal to M, then any positive linear respondent burden index is minimized by selecting the total sample m by simple random sampling (SRS) without replacement (WOR) where the first m_1 units selected are associated with survey I, the second m_2 units selected are associated with survey II, etc.

If the size of the total sample $m = m^{(1)} + m^{(2)} + \cdots + m^{(K)}$ is greater than M, then linear integer programming can be used to find an assignment of the total sample to the (unspecified) *labels* of the stratum that minimizes the respondent burden. Reiterating, we are working with labels here, so we are considering the burden of an arbitrary unit in the substratum, not the population units themselves, though we will use the natural terminology "population unit." When assigning samples from K surveys to the population units, there are 2^K possible ways of assigning the samples to any one population unit. These possible assignments can be represented by the 2^K K-dimensional vectors, call them *assignment configurations*,

$$\vec{v}_1 = \begin{pmatrix}0\\0\\0\\\vdots\\0\end{pmatrix}, \quad \vec{v}_2 = \begin{pmatrix}1\\0\\0\\\vdots\\0\end{pmatrix}, \quad \vec{v}_3 = \begin{pmatrix}0\\1\\0\\\vdots\\0\end{pmatrix},$$

$$\vdots$$

$$\vec{v}_{K+1} = \begin{pmatrix}0\\0\\0\\0\\\vdots\\1\end{pmatrix}, \vec{v}_{K+2} = \begin{pmatrix}1\\1\\0\\0\\\vdots\\0\end{pmatrix}, \vec{v}_{K+3} = \begin{pmatrix}1\\0\\1\\0\\\vdots\\0\end{pmatrix},$$

$$\vdots$$

$$\vec{v}_{2^K-1} = \begin{pmatrix}0\\1\\1\\1\\\vdots\\1\end{pmatrix}, \quad \vec{v}_{2^K} = \begin{pmatrix}1\\1\\1\\1\\\vdots\\1\end{pmatrix} .$$

where component k of the vector is 1 if the unit is sampled for the kth survey and 0 otherwise. Now we must determine the number x_1 of the population units to assign the configuration $\vec{v}_1$, the number x_2 to assign the configuration $\vec{v}_2$, ..., the number x_{2^K} to assign the configuration $\vec{v}_{2^K}$.

Suppose the ith assignment configuration, represented by the ith assignment configuration vector $\vec{v}_i$, produces a respondent burden of $b_i \geq 0$. Then the problem of assigning the $m^{(1)}, m^{(2)}, \ldots, m^{(K)}$ samples to the M (unspecified) unit labels such that the total respondent burden over the substratum is minimized is equivalent to minimizing the linear objective function (respondent burden index)

$$\begin{aligned} & f(x_1, x_2, \ldots, x_{2^K}) \\ = \; & b_1x_1 + b_2x_2 + \ldots + b_{2^K}x_{2^K} \\ = \; & \vec{b}\,\vec{x}^T \end{aligned}$$

subject to the $K+1$ linear constraints

$$\begin{cases} \vec{v}_1x_1 + \vec{v}_2x_2 + \cdots + \vec{v}_{2^K}x_{2^K} = \vec{m} = \\ \quad (m^{(1)}, m^{(2)}, \ldots, m^{(K)})' \quad \leftarrow \text{K constraints} \\ x_1 + x_2 + \cdots x_{2^K} = M\,, \end{cases}$$

where $x_1, x_2, \ldots, x_{2^K}$ are non-negative integers.

Since $\vec{v}_{K+2}, \ldots, \vec{v}_{2^K}$ can each be written as a nonnegative integer combination of $\vec{v}_2, \ldots, \vec{v}_{K+1}$ and since $m^{(k)} \leq M$ for each k, it is easy to see that

$$\begin{aligned} & \vec{v}_2x_2 + \vec{v}_3x_3 + \cdots + \vec{v}_{2^K}x_{2^K} \\ = \; & (m^{(2)}, m^{(3)}, \ldots, m^{(K)})' \end{aligned}$$

has a solution over the nonnegative integers, say $x_2, \ldots, x_{2^K}$. Setting

$$x_1 = M - x_2 - x_3 - \cdots - x_{2^K}$$

then provides a feasible solution to the integer linear programming problem. So there exists a solution and hence there exists an optimal solution.

(d) Finally, select specific sampling units u_i from the population. Consider a specific substratum and treat other substrata similarly. From the results of (c) above, we now randomly choose x_2 farmers from the M substratum farmers for the configuration $\vec{v}_2$, randomly choose x_3 farmers for the configuration $\vec{v}_3, \ldots$, randomly choose x_{2^K} farmers for the configuration $\vec{v}_{2^K}$. This sample of farmers reduces burden, yet within each stratum of each survey, this approach selects farmers with equal probability.

Note that this sample is not a type of systematic sample–the randomness in (b)-(2) reveals this.

Method II

In Method II, a sample is selected by some preferred technique. That sample might be selected by some equal probability of selection technique using zonal sampling to reduce variance, eg, by Chromy's Procedure, Chromy (1981). Method II largely retains that sample, but alters it to reduce burden. Thus Method II alters the sample by redistributing it within the substrata.

Since this Method II is no more complicated than Method I and has many similarities to it, the following description is brief.

(a) Within each stratum of each of the K surveys, independently select a sample with equal probability.

(b) Cross-classify the population as in (a) of Method I. This not only cross-classifies the population, it also cross-classifies the sample chosen in (a) of Method II. From the

$$N_{h_k, h_{k+1}, \ldots, h_K, h_1, \ldots, h_{k-2}, h_{k-1}}^{(k, k+1, \ldots, K, 1, \ldots, k-2, k-1)}$$

units in the substratum population

$$\mathcal{U}_{h_k, h_{k+1}, \ldots, h_K, h_1, \ldots, h_{k-2}, h_{k-1}}^{(k, k+1, \ldots, K, 1, \ldots, k-2, k-1)},$$

denote the number sampled by

$$m_{h_k, h_{k+1}, \ldots, h_K, h_1, \ldots, h_{k-2}, h_{k-1}}^{(k, k+1, \ldots, K, 1, \ldots, k-2, k-1)}.$$

This subsample size will not be changed, but it will be distributed among the substratum's population in (c) below.

(c) Within a substratum, reassign or swap some of the surveys associated with a sampling unit having excess respondent burden to a sampling unit have less respondent burden. If the respondent burden index is linear, then only one survey for one sampling unit need be reassigned to reduce burden. For example, when we measure respondent burden by the number of times we hit a farmer with a survey. Then we would move one survey from the farmer who got 4 hits to the farmer who got 0 hits, or to the farmer who got 2 hits if no farmer got 0 or 1 hit.

If the respondent burden is non-linear, then sometimes more than one survey must be reassigned to reduce burden. And when respondent burden in non-linear, then sometimes three sampling units (not two) must swap to ever reduce burden.

(d) Repeat (c) above until no reassignments can be made. Then respondent burden has been minimized.

With this method, one might want to retain most of the original sample selection for the first survey but not necessarily for the other surveys. Then, in (c), try to reassign other surveys before reassigning the first survey. Sequential application of Method II is justified since each survey uses equal probability of selection in each stratum which implies that all units of a substratum have the same selection probability for any given assignment configuration.

Some NASS Examples

NASS administers many surveys with a large number of strata. For example, the Farm Costs and Returns Survey (FCRS/COPS) may have 18 strata, the Agriculture Survey may have 17 strata, and the Labor Survey may have 8 strata. This many strata over many surveys brings skepticism to any use of Methods I or II. One would expect many combinations of strata to contain but one individual, even for three surveys. Methods I and II could never reduce burden on such a sparsely (one individual) populated combination of strata. Fortunately, most stratum combinations are empty while other combinations are well populated.

Indeed, not only are many substratum combinations empty, many survey sampling combinations are empty. In some initial testing over nine major surveys, only 58 of the $2^9 = 512$ possible survey combinations occurred in Kansas and only 62 in Arkansas based on 1991 data. This fortuitous limitation on survey combinations gives some optimism that many combinations of strata will be well populated. A look at the number of population units selected for multiple surveys provides further optimism (see Table 1).

No burden exceeds five surveys. No sampling unit was selected for more than five surveys, indicating that the possible number of substrata with only one unit is limited somewhat.

There is some optimal combination of surveys to consider when reducing respondent burden by either Methods I or II. More surveys result in too few farmers being classified for any of the many substrata combinations. Fewer surveys prevent

Table 1: Number of Survey Hits over Nine Surveys in 1991

Hits	Arkansas Frequency	Kansas Frequency
0	3491	21474
1	21125	40900
2	6136	8638
3	846	938
4	60	74
5	1	7

Methods I and II from reducing any large burdens on some farmers; eg, when NASS surveys one farmer for five different surveys.

In 1991, for the four major surveys – FCRS, Labor, Quarterly AG, and Cattle/Sheep – NASS initially sampled the following numbers of farmers.

Survey	Arkansas	Iowa	Kansas
FCRS	666	1836	1356
Labor	576	728	440
Quarterly AG	4442	6477	5881
Cattle/Sheep	1727	5507	3204

Method II reduced burden by about 70 percent over the three states Arkansas, Iowa and Kansas in 1991 and 1992. Table 2 below summarizes these reductions of burden. Since the NASS samples were essentially random within strata, a huge reduction can be made in burden with no cost (increase) in variance.

Table 2: Reduction in Multiple Sample Selections Using Method II for the FCRS, Labor, Quarterly AG, and Cattle/Sheep Surveys

Number	1991		
Selections	Current	Method II	% Reduction
4	0	0	–
3	159	50	69
2	2620	782	70
Total	2779	832	70
Arkansas	733	205	72
Iowa	1105	252	77
Kansas	941	375	60

Number	1992		
Selections	Current	Method II	% Reduction
4	6	4	33
3	112	28	75
2	2371	749	68
Total	2489	781	69
Arkansas	735	124	83
Iowa	801	204	75
Kansas	953	453	52

Acknowledgements

The authors would like to express their appreciation to George Hanuschak for assistance in initiating and supporting this project early on. They would also like to thank Ron Bosecker and Jim Davies for continued support.

State Statistical Offices are quite concerned about individual farmer respondent burden and want the Agency to pursue methods to reduce it while maintaining an acceptable level of statistical integrity. This report is the first such statistical attempt in recent years in the Agency. Special thanks go to State Statisticians T. J. Byran, Ben Klugh, Dave Frank and Howard Holden for providing their substantial insights into the definition and potential relief of individual farmer respondent burden.

References

Chromy, James R. (1981). "Variance Estimators for a Sequential Sample Selection Procedure," *Current Topics in Survey Sampling*, D. Krewski, R. Platek, and J.N.K. Rao (Eds). Academic.

de Ree, S.J.M. (1983). *A System of Co-ordinated Sampling to Spread Response Burden of Enterprises.* Netherlands Central Bureau of Statistics.

Ohlsson, Esbjörn (1993). "Coordination of Samples Using Permanent Random Numbers," ASA International Conference on Establishment Survey Proceedings.

Tortora, Robert D. and Crank, Keith N. (1978). *The Use of Unequal Probability Sampling to Reduce Respondent Burden.* USDA/National Agricultural Research Service.