THE USE OF UNEQUAL PROBABILITY SAMPLING TO REDUCE RESPONDENT BURDEN

Robert Tortora, USDA
Keith Crank, USDA

INTRODUCTION

The Economics, Statitics, and Cooperatives Service conducts many surveys during the year. A large number of these surveys are done year after year and often more than once during the year. In some states a list frame is used in conjunction with an area frame, and because of stratification, many operators are contacted many times a year, year in and year out.

The purpose of our study was to find out if we could reduce the amount of respondent burden for these operators who are contacted so often without removing them from the frame. Our plan was to use an unequal probability sampling scheme with respondent burden used as an inverse measure of size.

BURDEN

In order to use burden in assigning probabilities it is first necessary to have some manner in which to assign values of burden to each unit on our frame. For our study we used two methods of determining burden. One method, which we called response burden based on number of contacts, was to count the number of times that an operator was in a survey during the year. Table 1 shows some of our surveys and the number of times they are conducted each year. So if an operator were in the samples for hogs and for cattle his burden would be six, since that is the number of times he would be contacted during the year.

The second method of assigning burden we called response burden index (RBI). For this method we assigned weights to different aspects of each survey. For example the length of the questionnaire and the amount of time for which information was to be recalled, as well as the number of times the survey would be conducted during the year, were included in the formula for assigning burden. Table 1 has the value of RBI in the last column. So using this method the operator who was chosen for the hog survey and the cattle survey would have a response burden of 9.12.

Because of the problems involved in finding out which names had been chosen for each of the surveys for the year, we used expected response burden instead of actual response burden. Also we limited ourselves to seven surveys since we this would be sufficient to test the feasibility of this sampling plan.

The total expected burden for an individual is then calculated using the formula:

$$B_{ik} = \sum_{j=1}^{7} m_{ij} \Pi_{jk}; \quad i=1, 2; \; k=1, 2;$$

where

B = Total burden for an individual based on one of the two sampling methods and one of the two methods of computing response burden

M = Burden for a given survey for one of the two methods of computing response burden

Π = Probability of inclusion in a given survey for one of the two sampling plans.

The formula used in assigning probabilities is: (Tortora, 1976)

$$P_i = c^{1-i}/(M_1 + M_2 c^{-1} + \ldots M_d c^{1-d})$$

where

P_i = Probability of selection in class i

M_j = Number of units in class j

c = Constant (>1)

d = Number of classes

In order to use this formula the population must be divided into classes so that all units in one class will have the same probability of selection. It is not necessary that each class contain the same number of units. In the formula d denotes the number of classes and M_j denotes the number of units within class j.

The constant c is necessary to give unequal probabilities. If c = 1, we have equal probability sampling. If c > 1, then a class with a high index will have a smaller probability than a class with a low index. If c < 1, then the opposite happens. We chose values of c > 1. This meant that by assigning units with larger burdens to higher numbered classes, these units would be given a smaller probability of selection. For this study we used three values of c namely c = 1.1, c = 1.25 and c = 1.5.

For consistency we limited the number of classes to 5. This was done so that we would be better able to judge the differences between the values of c that we used. We also had the criterion that not more than one class would have less than 5 percent of the population.

Initially an expected response burden is calculated for each individual for both measures of burden based on a stratified simple random sample. The stratified simple random sample used for this calculation is the same survey design currently being used by ESCS for this survey. Then each individual is assigned a selection probability for the first survey. This is done separately for each value of c. From this an expected response burden is calculated and substituted in the formula for total expected burden.

This process is continued, one survey at a time, for all seven surveys. Then we have a total expected burden based on unequal probabilities of selection, which we can compare with the total expected burden based on equal probabilities of selection. This is done in the graphs.

RESULTS

All three of these graphs use a response burden based on number of contacts. The y - axis represents the difference between the burden using equal probabilities of selection and the burden using unequal probabilities of selection. The

x - axis represents the total expected burden using equal selection probabilities. Zero on the y - axis indicates no change in expected burden, and a positive value on the y - axis indicates an improvement from using unequal probabilities of selection.

The first graph is for burdens computed using a value of c = 1.1 in the formula for selection probabilities. There is some change in burden; however, the change is not large in either direction, and those operators with a large burden have not been helped much at all.

The second graph is for c = 1.25. Here a lot more changes have occurred and to a greater degree. Also there appears to be a positively sloped trend. This means that for those operators with larger burdens, the improvement in expected burden is more substantial than for those with small burden.

The final graph, which is for a value of c = 1.5, shows approximately the same information as the previous one. There is the wide spread in the changes in burden and there is a trend with a positive slope.

Tables 2 and 3 show quantitatively what happens to the expected burden of large operators. EO's or extreme operators are operators that have a large number of a given specie of livestock. Using an expected burden based on number of contacts, the average expected burden decreased using unequal probability sampling with the size of the change ranging from 6 to 17 percent (Table 2). For non-EO's the average burden increased but the increase was less than two percent.

Table 3 shows what happens to operators who have a large burden when the unequal probability sampling scheme is used. For those operators whose expected burden was greater than ten using equal probability sampling, the expected burden was reduced by almost 25 percent, while those with a burden less than or equal to 10 showed an increase of slightly more than one percent. Operators with burdens greater than 15 and 20 showed decreases of 15 percent and 16 percent respectively.

ESTIMATION

We used a modified PPS estimator to compare the efficiency of this design with that of a stratified simple random sample. Using simulated data the standard errors were seven to nine percent higher for the PPS estimator when response burden was based on number of contacts. For the response burden index, standard errors for the PPS estimator were thirteen to twenty-four percent higher than for the stratified simple random sample.

SUMMARY

We have designed a sampling plan which will help reduce respondent burden. This plan uses unequal probabilities of selection and could prove to be useful when a number of surveys are to be conducted using the same sampling frame. For the seven surveys in our study we were able to reduce respondent burden by almost 25 percent for some large operators.

REFERENCES

Arends, W. (1976), "Sample Selection System," Proceedings of SRS National Conference, USDA, pp. 135 - 141.

Graham, B. (1976), "Long Range Plans," Proceedings of SRS National Conference, USDA, pp. 17 - 27.

Raj, D. (1968), Sampling Theory, McGraw Hill, New York

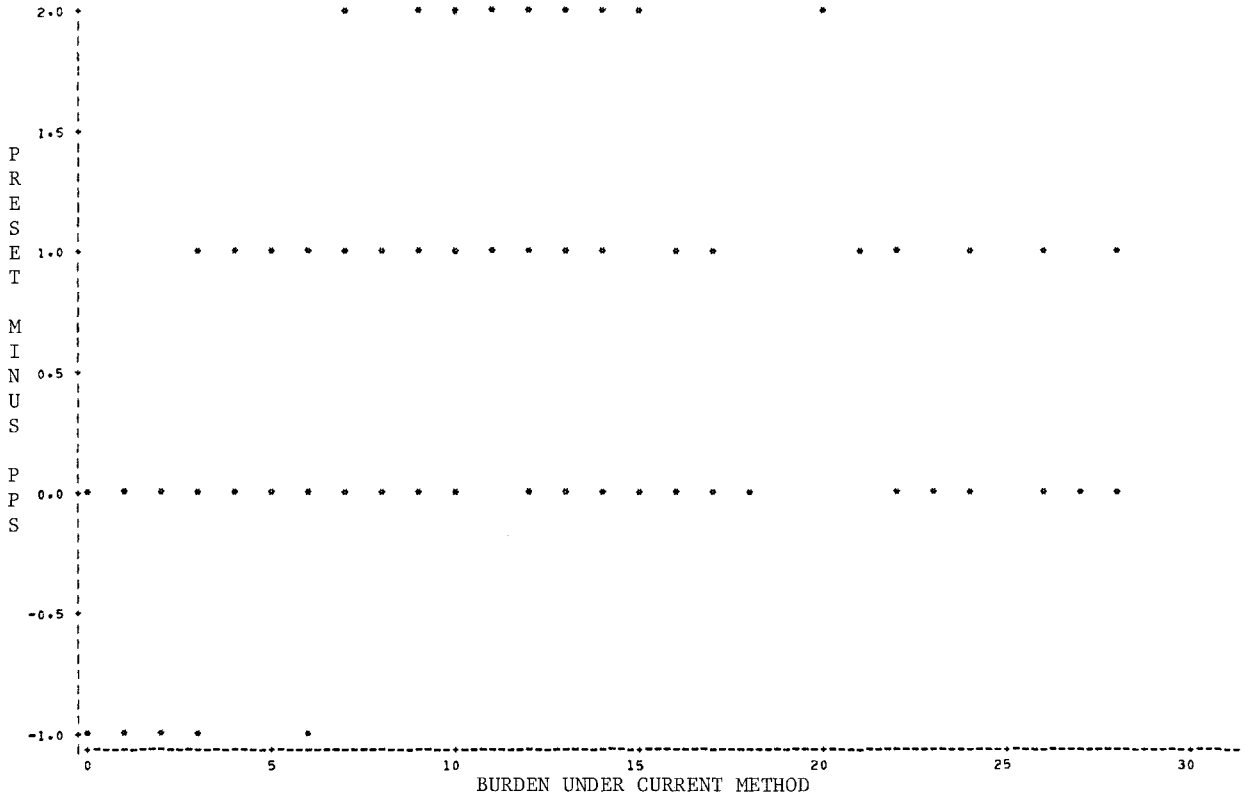
Rao, J. N. K. (1966), "Alternative Estimators in PPS Sampling for Multiple Characteristics," Sankhya, Series A, 29, pp. 47 - 60.

Tortora, R. D. (1977), "Reducing Respondent Burden for Repeated Samples," Agricultural Economics Research, 30, 2, pp. 41 - 44.

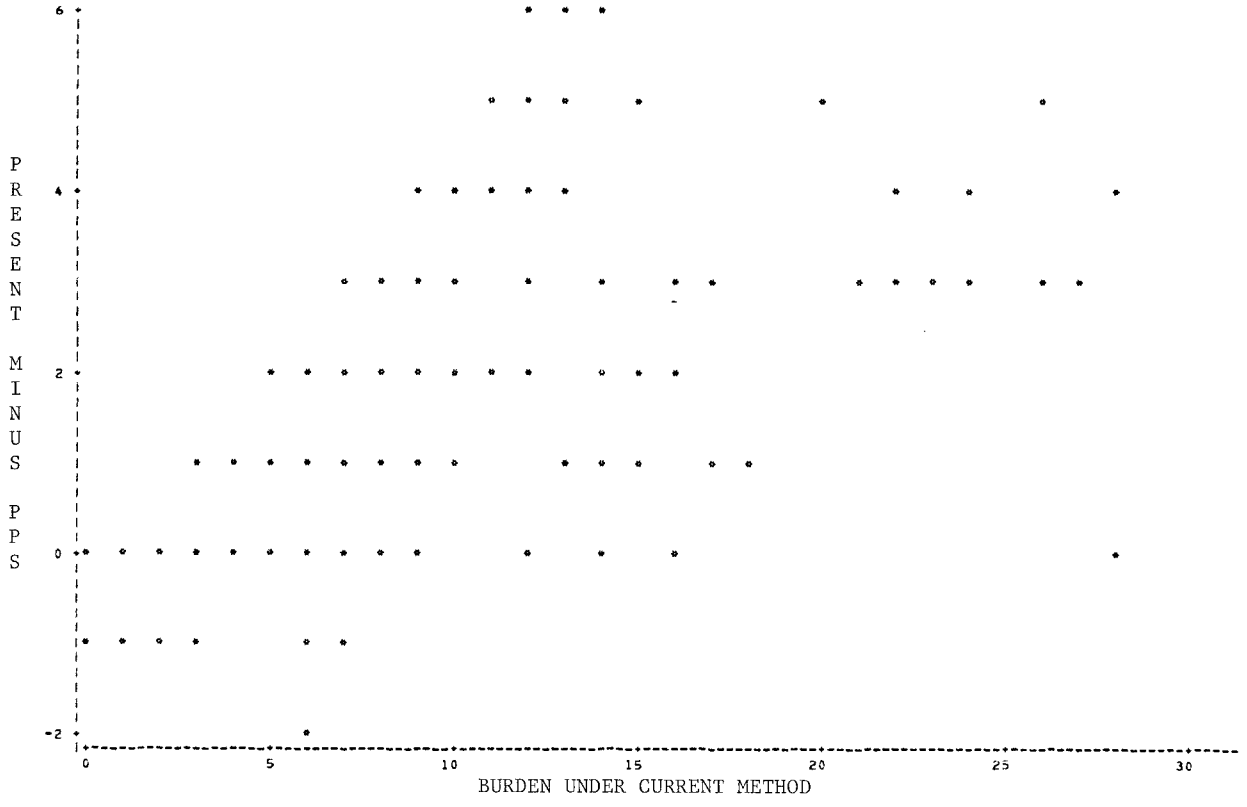
Table 1: Some periodic surveys conducted by ESCS, the number of times they are conducted in a year, and a response burden for each.

Survey	Number of Contacts	Response Burden Index (RBI)
Hogs	4	4.20
Cattle	2	4.92
Cattle on Feed	4	3.54
Sheep	2	2.40
Sheep on Feed	3	2.34
Dairy	12	4.80
Chickens	12	6.72

Graph 1: Plot of difference between expected burden under present sampling system and expected burden under proposed PPS system for c = 1.1.



Graph 2: Plot of difference between expected burden under present sampling system and expected burden under proposed PPS system for c = 1.25.



Graph 3: Plot of difference between expected burden under present sampling system and expected burden under proposed PPS system for c = 1.5.

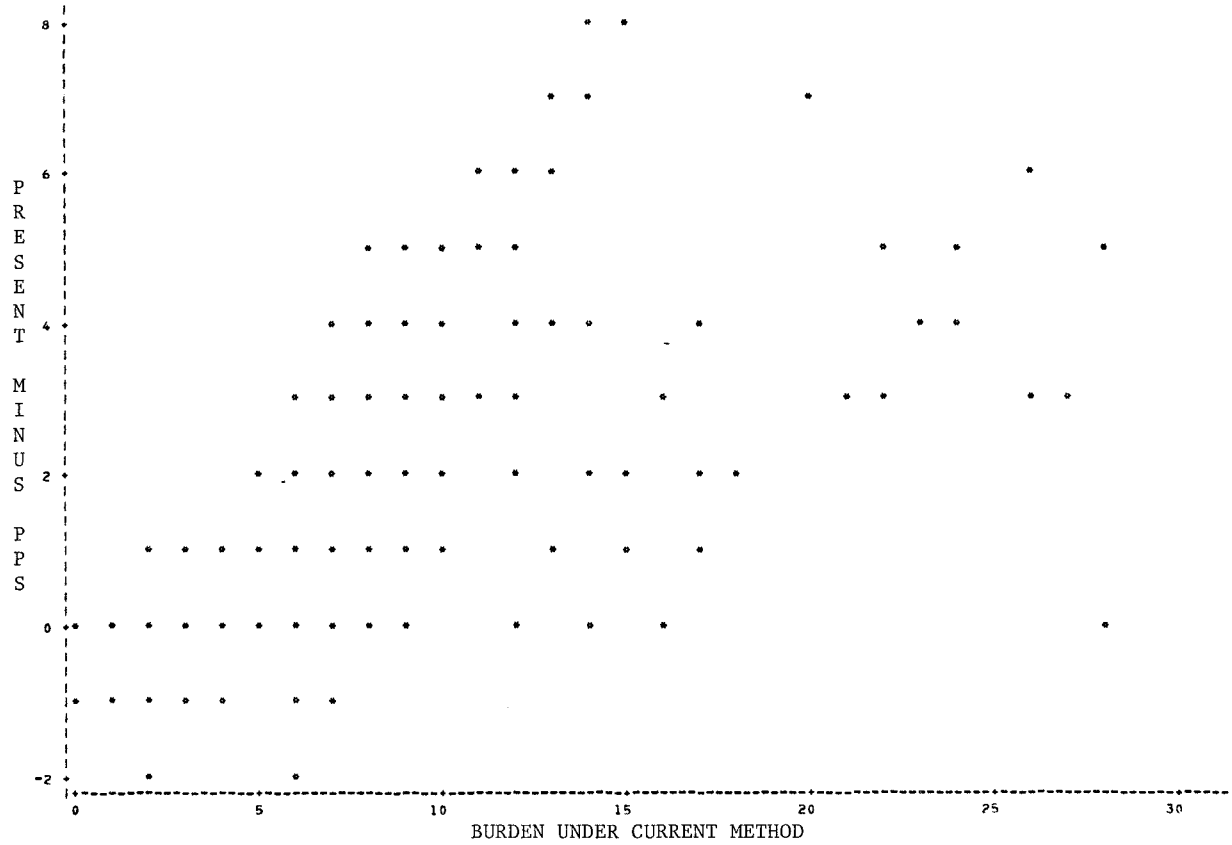


Table 2: Average expected burden based on number of contacts.

Group	#	Current	PPS	Change (%)
Cattle on Feed EO's	177	6.040	5.513	- 8.7
Cattle EO's	193	4.093	3.754	- 8.3
Sheep EO's	83	5.266	4.544	-13.7
Sheep on Feed EO's	25	7.150	5.912	-17.3
Hog EO's	415	5.167	4.472	-13.4
Dairy EO's	12	15.295	14.358	- 6.1
Chicken EO's	95	16.326	14.722	- 9.8
Non-O's	37666	1.803	1.837	+ 1.9

Table 3: Average expected burden based on number of contacts.

Group	#	Current	PPS	Change (%)
Current Burden > 10	231	13.447	10.100	-24.9
> 15	42	21.184	17.924	-15.4
> 20	26	23.998	20.058	-16.4
$\leq$ 10	38271	1.813	1.836	+ 1.2