DESIGNING OPTIMAL INTEGRATED SAMPLES USING NONLINEAR PROGRAMMING

Rudolph E. Schwartz, Social Security Administration

I. INTRODUCTION

At present, there are in existence at least three quality control programs which the States are required to carry out under Federal guidelines. They are:

Aid to Families with Dependent Children (AFDC)
Food Stamps (FS)
Medicaid (Md)

In most States, all of these programs are administered by a single agency.

The purpose of each of the quality control programs is to measure and by corrective action reduce the frequency of errors.

The three programs have many similarities. Each consists of monthly simple random samples of beneficiaries. Each operates independently with its own specifications and separate reviewers.

Since, in many instances, there is substantial overlap between recipients and programs (for example, an AFDC recipient might be eligible for both Food Stamps and Medicaid, or a Food Stamp recipient might receive Medicaid, etc.). Many States have begun to see the advantage of interviewing for one program and receiving information that can be used by several programs. This process is known as integrated sampling.

II. FACTORS FOR CONSIDERATION

However, before this process can be considered, there are several issues with which the State must concern itself. They are:

1. Maintenance of statistical validity of the sample. This requires proper frames and definitions of the populations of interest. Development of adequate estimators of central tendency and variance since many of the sampling procedures introduced require complex variance estimators.

2. Quality of Generic Reviews (a generic review is one in which the interviewer is trained in reviews for all programs as opposed to being specialized in one program). This requires cross-training of reviewers in the nuances of each of the several programs.

   There is evidence to indicate that with proper cross-training generic program review quality is as acceptable as individual specialized review quality.

III. COST EFFECTIVENESS

Increased efficiency in integrated sampling comes about for several reasons; the most important are: (a) reduced travel, (b) increased efficiency, (c) overlap between beneficiaries and programs.

a. Since a reviewer can do several reviews in one household as opposed to separate individuals doing reviews in the same household, there can be saving as much as 30 percent in travel costs as well as a reduced respondent burden.

b. Efficiency increases due to concentration of reviewers in smaller geographic areas. This arises because several collateral contacts can be done at one time and often at the same institutions. Also waiting time for contacting call backs can be reduced since other beneficiaries in the same area can be reviewed while waiting for call backs. This gain can be as great as 30 percent of review time cost.

c. The number of sample cases can be reduced because of overlap between beneficiaries and programs. For example, a review of a member of a group receiving multiple benefits such as AFDC, Food Stamps, and Medicaid can be used by each of the programs as a sample case. Thus the overall sample case count is reduced with only nominal increase in interviewing costs.

In a few States that are making efforts at integration, it has been shown that it is possible to achieve cost savings of about 30 percent by integrating as opposed to running separate programs.

IV. POSSIBLE METHODS OF SAMPLING

1. One can continue to choose independent samples from each of the program files and after selection merge sample cases on individuals that belong to more than one program. There are various approaches using this method. Schneider (3) has developed a post stratification scheme which maximizes the use of sample overlap. However, this method only assures that the sample size requirements of the various programs can be satisfied. It does not meet the variance requirements of each of the individual programs when the complex estimation procedure required is considered.

   An example using the Schneider procedure and the same data in table I indicated that while a 34 percent reduction in sample size could be achieved, in no instance would the variance requirements of each of the separate programs be met. Also, the variance shortfall in each instance was substantial.

2. One can merge all three program files and construct mutually exclusive strata consisting of a maximum of seven (7) strata as follows:

   (1) AFDC only; (2) AFDC-FS; (3) AFDC-FS-Md; (4) AFDC-Md; (5) FS only; (6) FS-Md; (7) Md only.

   Having done this the problem becomes one of designing a sample that is optimal for all three programs combined. This paper describes the second approach using nonlinear programming.

V. OPTIMAL SAMPLE DESIGN

An optimal sample takes into consideration 3 elements:

(a) Reliability or allowable sampling error.

(b) Variability of the universe or sub-universes from which samples are drawn.

(c) Cost of sampling which considers set-up, travel, interview, and verification costs.

The difficulty with designing an optimal sample is that each of the various programs has its own values for the three elements mentioned above. Thus the sample that is optimal for one program may not be optimal for another.

In order to solve such a problem for stratified samples and several programs one must solve a nonlinear programming problem.

Mathematical Formulation of an Optimal Stratified Integrated Sampling Problem

$$\text{MIN } C = \Sigma\, a_h n_h \qquad \begin{matrix} j = 1 \ldots J \text{ programs} \\ h = 1 \ldots L \text{ strata} \end{matrix}$$

$$V_j \geq C_j \quad \text{(neglecting finite correction factor)}$$

$$V_j = \sum_{h=1} \frac{W_h^2 S_{hj}^2}{n_h} \qquad 0 \leq n_h \leq N_h$$

where N = Total units of population
N_h = Total units in h stratum
n_h = Sample units in hth stratum
$W_h = N_h/N$ = Stratum weight
S_h^2 = estimate of population variance in h stratum
a_h = review costs of a case

Below is a contrived example of a typical optimal solution which has been solved by a non-linear programming package.

Example - Design of an optimal integrated sample.

Assume the present three programs have the following sample size requirements.

Program	Sample	SRS Variance* (10^{-4})
Aid to Families with Dependent Children (AFDC)	800	1.10
Food Stamps (FS)	1,200	0.61
Medicaid (MED)	275	3.12

*SRS Variance = Simple random sample variance equivalent for sample sizes shown if P = .095.

Construct mutually exclusive strata by merging the files of the three programs into seven mutually exclusive strata.

Stratum No.		Pop. Size (000)	Cost Review (Hours)	Factor	Unit Cost (Hours)
1	AFDC	9	9.50	1	9.50
2	AFDC, FS	18	11.50	1/2	5.75
3	AFDC, FS,MED	2	13.50	1/3	4.50
4	AFDC, MED	1	11.50	1/2	5.75
5	FS	23	9.50	1	9.50
6	FS,MED	7	11.50	1/2	5.75
7	MED	45	9.50	1	9.50

MINIMIZE:

$$C = 9.50n_1 + 11.50n_2 + 13.50n_3 + 11.50n_4 + 9.50n_5 + 11.50n_6 + 9.50n_7$$

SUBJECT TO:

$$0 \leq n_1 \leq N_1 \ldots; \quad 0 \leq n_7 \leq N_7$$

$$\left(\frac{9}{30}\right)^2 \frac{(P_{11})(1-P_{11})}{n_1} + \left(\frac{18}{30}\right)^2 \frac{(P_{21})(1-P_{21})}{n_2} + \left(\frac{2}{30}\right)^2 \frac{(P_{31})(1-P_{31})}{n_3} + \left(\frac{1}{30}\right)^2 \frac{(P_{41})(1-P_{41})}{n_4} \leq 1.10 \times 10^{-4}$$

$$\left(\frac{18}{50}\right)^2 \frac{(P_{22})(1-P_{22})}{n_2} + \left(\frac{2}{50}\right)^2 \frac{(P_{32})(1-P_{32})}{n_3} + \left(\frac{23}{50}\right)^2 \frac{(P_{52})(1-P_{52})}{n_5} + \left(\frac{7}{50}\right)^2 \frac{(P_{62})(1-P_{62})}{n_6} \leq 0.61 \times 10^{-4}$$

$$\left(\frac{2}{55}\right)^2 \frac{(P_{33})(1-P_{33})}{n_3} + \left(\frac{1}{55}\right)^2 \frac{(P_{43})(1-P_{43})}{n_4} + \left(\frac{7}{55}\right)^2 \frac{(P_{63})(1-P_{63})}{n_6} + \left(\frac{45}{55}\right)^2 \frac{(P_{73})(1-P_{73})}{n_7} \leq 3.12 \times 10^{-4}$$

Where $P_{11} = .09$, $P_{21} = .10$, $P_{31} = .08$, $P_{41} = .07$
$P_{22} = .06$, $P_{32} = .09$, $P_{52} = .10$, $P_{62} = .08$
$P_{33} = .08$, $P_{43} = .10$, $P_{63} = .07$, $P_{73} = .10$

Solution:

$n_1 = 205$, $n_2 = 504$, $n_3 = 54$, $n_4 = 19$, $n_5 = 564$, $n_6 = 144$, $n_7 = 226$

Cost = 17852 Hrs.

Description of Table I

Column

1 Program name and strata used by the program.

2 Population size of each stratum used by each program.

3 Weight = N_i/N_1.

4 P = estimate of proportion of error in each stratum for the various programs.

5 Theoretical proportional sample allocation (this sample would be self-weighting).

6 Actual proportional sample used.
Example - FS stratum 3 requires only 48 cases, however, since 54 cases have been finished to satisfy the requirements of AFDC stratum 3 it would not be sensible to throw away the 4 cases already completed unless self-weighting is crucial.

7 Optimal sample allocation.

Referring to table I we can see that optimal sample when compared with the proportional sample is cheaper in the AFDC and MED strata but slightly more expensive in the FS strata.

However, compared to independent simple random samples, both AFDC and FS are substantially cheaper while Medicaid is slightly more expensive. The overall cost, however, is much cheaper than separate simple random samples.

Bibliography

(1) W. G. Cochran; Sampling Techniques; 2nd edition, Wiley, New York, 1963.

(2) Jerome Bracken and Garth McCormick; Selected Applications of Nonlinear Programming, Wiley, 1968.

(3) Robert Schneider; How to Do It - Integrated Sample Selection for AFDC, Food Stamps and Medicaid Quality Control, unpublished paper, 2/78.

Table I - Contrast Between Proportional Sample and Optimal Sample

	(2) Pop. (000)	(3) WT	(4) P	(5) Theor. Prop.	(6) Actual Prop.	(7) Optimal
AFDC						
1	9	.300	.09	239	239	204
2	18	600	10	480	480	483
3	2	067	08	54	54	42
4	1	033	07	27	27	18
Total	30			800	800	757
$Var(10^{-4})$				1.10	1.10	1.10
Cost -					5,429	5,065
SRS*						7,600
FS						
2	18	.36	.06	432	480	493
3	2	04	09	48	54	42
5	23	46	10	552	552	578
6	7	14	08	168	168	147
Total	50			1,200	1,254	1,260
$Var(10^{-4})$				0.63	0.61	0.61
Cost -					9,213	9,860
- SRS						11,400
MED						
3	2	.036	.08	10	54	42
4	1	018	10	5	27	18
6	7	127	07	35	168	147
7	45	819	10	225	225	200
Total	55			275	474	407
$Var(10^{-4})$				3.12	2.76	3.12
Cost -					3,502	3,037
- SRS						2,613
Size (1) Net Size				1,505	1,559	1,716
(2) Gross Size					2,528	2,491

*SRS - Simple Random Sample Cost in Hours Expended

Stratum No.	Program Name
1	AFDC
2	AFDC, FS
3	AFDC, FS, MED
4	AFDC, MED
5	FS
6	FS, MED
7	MED