EFFICIENCY OF COMBINING TWO COMPLEX SAMPLE SURVEYS

Donna Brogan and Richard Porter, Emory University

ABSTRACT

In planning the 1983 Georgia High Blood Pressure Survey, a choice of two sample designs was available. One design, referred to as the single design, was to allocate a total of 3000 sample housing units (SHU's) so as to achieve an equal probability sample of HU's throughout Georgia. In the second design, referred to as the dual design, an additional 960 SHU's could be added to the 3000 SHU's, but these additional SHU's needed to be located in six specific small counties in south Georgia. Then, the other 3000 SHU's would be allocated to the remainder of Georgia in an equal probability sample. Although it might seem that increasing the total sample size by almost one-third, from 3000 SHU's to 3960 SHU's, would substantially increase the precision of estimates for Georgia, the increase was expected to be counterbalanced to some extent because of widely differential weights inherent in the dual design. Theoretical comparisons of the two potential designs indicated that the precision for statewide estimates would be about the same under each design. The dual design was used to do the survey. Statistical analysis of the actual survey data from the dual design were in agreement with the theoretical comparisons done prior to designing the survey.

1983 GEORGIA HIGH BLOOD PRESSURE SURVEY

Purpose of Survey

In 1983 an NHLBI funded contract (the FOCUS project) called for a statewide survey of adults (18+) in Georgia in order to assess the prevalence of elevated blood pressure, the prevalence of hypertension and the awareness, treatment and control status of hypertension. The sample size for this survey had been predetermined as 3000 sample housing units (SHU's) based on power considerations and cost constraints. At the same time a separately funded NHLBI grant (the CHIP project) called for a similar survey to be conducted in six rural small counties in middle Georgia; these six counties contained only 1.2% of the total housing units (HU's) in the state of Georgia. The sample size of the middle Georgia survey was predetermined as 960 SHU's based on budget constraints and experimental considerations

More detail on the purpose and the larger experimental context of these surveys, as well as the determination of the particular sample sizes (3000 and 960), is given in Brogan (1985a, 1985b). This paper takes as a given that the middle Georgia survey of the CHIP project will be done with 960 SHU's and that the FOCUS project budget allows a maximum of 3000 SHU's.

Single versus Dual Design

From the viewpoint of planning the survey for the FOCUS project, two sample designs were possible. In the single design, the total FOCUS sample size of 3000 SHU's would be allocated in an equal probability sample throughout Georgia. This design would be developed independently of the middle Georgia survey, thus allowing the possibility (although remote) that one or more of the six CHIP counties would be selected by the single design. In this event, some housing units could be selected for both surveys, an undesirable outcome since both surveys are quite similar in content.

A second design, called the dual design, was possible by coordinating both the FOCUS and CHIP surveys to use similar sample designs and exactly the same survey data collection forms and field work procedures. Under this dual design 960 SHU's would be allocated to the six specific counties in middle Georgia to be covered by the CHIP survey. Then the 3000 FOCUS SHU's would be allocated in an equal probability sample to the remainder of Georgia, i.e. to all of Georgia less the six specific CHIP counties. Thus, the dual design makes statewide estimates by taking the union of two distinct and mutually exclusive sample surveys, and the statewide sample size under the dual design is 3960 HU's. Of course, the dual design also meets the objectives of the CHIP survey by allowing inference to these six specific counties based on 960 SHU's.

Using the dual design would result in major cost savings on the survey design for the CHIP survey since the CHIP survey would be a component of the statewide survey and use all forms/procedures developed by the FOCUS project. In particular, the selection of SHU's for the CHIP survey would be done by FOCUS survey personnel so that the sample design would fit the objective of both the CHIP and statewide surveys.

From the viewpoint of the statewide survey, however, the dual design had some disadvantages. Besides entailing extra work, the allocation of the total sample size of 3960 SHU's across the state could not be done via an equal probability sample. Thus, there would be increased variability in the precision of statewide estimates, over an equal probability sample of the same size, via differential weighting of sample housing units. Considering precision of statewide estimates as the major concern regarding the dual survey, the question was whether the dual design (with 3960 SHU's and unequal weights) would yield significantly better precision than the single design (with 3000 SHU's in an equal probability sample).

THEORETICAL COMPARISON OF TWO DESIGNS

General Approach

Under either design the sampling plan was to interview all adults (age 18 or older) at each SHU. Based on similar surveys conducted in Georgia a few years earlier (Brogan, 1985a), each selected SHU was projected to yield an average of 1.5 completed interviews after allowance for nonchargeable SHU's and nonresponse to either household enumeration or personal interview. Thus, the FOCUS sample size of 3000 SHU's was anticipated to yield 4500 completed interviews with the 960 CHIP SHU's expected to yield 1440 completed interviews. Hence, the single design for the statewide survey, which would be the FOCUS survey alone, would have an anticipated sample size of 4500 adults. Under the dual design for the statewide survey, which would be the union of the FOCUS and CHIP surveys, a sample size of 5940 completed interviews was anticipated. Further, since all adults within a SHU were to be interviewed, the probability with which an adult was chosen was equal to the probability with which the adult's SHU was chosen. Thus, the unadjusted weight (inverse of probability of selection) for any randomly chosen adult is the same as the unadjusted weight for the adult's SHU.

A theoretical comparison of the two designs can be made by calculating the anticipated variance of statewide point estimates under each design. In general, consider a point estimate of a mean or proportion in the population of inference, i.e. adults in Georgia. This estimator can be expressed as $\bar{y} = (\sum w_i y_i)/(\sum w_i)$, where $\sum$ indicates summation over all interviewed adults, y_i is the value of the variable measured on person i and w_i is the weight for person i. In actual analysis of complex sample survey data, w_i frequently is a product of several different weights which include the inverse of probability of selection, as well as adjustments for nonresponse and poststratification. In the theoretical comparison here w_i is defined to be the inverse of the probability of selection.

Assuming an infinite population where the y_i are distributed with some mean μ and variance σ^2 and further assuming the y_i to be statistically independent yields

$$V(\bar{y}) = \sigma^2 (\Sigma w_i^2)/(\Sigma w_i)^2 \quad (1)$$

However, (1) will underestimate the variance of $\bar{y}$ because of clustering inherent in multi-stage designs, and both the single and dual designs are multi-stage samples. Thus, the formula in (1) will be multiplied by the constant $C > 1$ to account for the increased variability due to clustering. Thus,

$$V(\bar{y}) = \sigma^2 C(\Sigma w_i^2)/(\Sigma w_i)^2 \quad (2)$$

Define now the theoretical relative efficiency (TRE) of the dual (D) design to the single (S) design as

$$TRE(D/S) = RE(\bar{y}_D/\bar{y}_S) = V(\bar{y}_S)/V(\bar{y}_D) \quad . \quad (3)$$

Substituting (2) into (3) yields

$$TRE(D/S) = \frac{\sigma^2 C_S \left[\sum_S w_{Si}^2\right] \Big/ \left[\sum_S w_{Si}\right]^2}{\sigma^2 C_D \left[\sum_D w_{Di}^2\right] \Big/ \left[\sum_D w_{Di}\right]^2} \quad (4)$$

The notation $\sum_S$ is the summation over the anticipated 4500 interviewed persons in the single design, where person i has weight w_{Si}. $\sum_D$ is defined similarly over the anticipated 5940 adults interviewed under the dual design. Note that $\sum_S w_{Si}$ and $\sum_D w_{Di}$ are equal, since they both estimate the number of adults in Georgia. Since the single and dual designs are quite similar with respect to the stages and methods of sampling, it is reasonable to assume that C_S and C_D are approximately equal. Thus, equation (4) reduces to

$$TRE\ (D/S) = \left[\sum_S w_{Si}^2\right] \Big/ \left[\sum_D w_{Di}^2\right] \quad . \quad (5)$$

This formula can be evaluated numerically to yield the relative efficiency of the dual design to the single design.

<u>Relative Efficiency for Statewide Estimates</u>

Under the single design a sample of 3000 HU's is allocated in an equal probability sample to 2,028,664 HU's in Georgia. Thus, each of the 4500 interviewed adults has an expected weight of (2,028,664)/3000 = 676.2. Hence,

$$\sum_S w_i^2 = 4500\ (676.2)^2 = 2.0576 \times 10^9.$$

Under the dual design the CHIP survey allocates 960 SHU's to 23,701 HU's, although not in an equal probability sample. In the context of the CHIP experimental design explained in detail in Brogan (1985a, 1985b), the six counties comprise an "experimental" area of four counties and a "control" area of two counties. The predetermined sample allocation (Brogan, 1985b) was 576 SHU's in an equal probability sample to the 16,755 HU's in the "experimental" area and 384 SHU's in an equal probability sample to the 6946 HU's in the "control" area, yielding a weight of (16755)/(576) = 29.1 for adults interviewed in the experimental area and a weight of (6946)/(384) = 18.1 for adults interviewed in the control area. The 4500 adults interviewed in the FOCUS survey will have a weight of (2,004,963)/(3000) = 668.3. Thus,

$$\sum_D w_i^2 = (4500)(668.3)^2 + (576)(1.5)(29.1)^2 + (384)(1.5)(18.1)^2 = 2.0107 \times 10^9$$

The relative efficiency of the dual design compared to the single design is given by TRE(D/S) = (2.0576)/2.0107 = 1.023. Thus, although the dual design increases total sample size by 32% (1440/4500), the precision is increased only 2% over the single design. Hence, using the dual design will not impact negatively upon statewide estimates; however, the dual design will not result in significantly increased precision for statewide estimates.

<u>Relative Efficiency for Domains</u>

In addition to making statewide estimates, a secondary objective is to make estimates for major demographic domains based on race, age, and sex. One domain of major interest is blacks. Although the percentage of the adult population which is black is 25.5% statewide, it is 43.1% in the six CHIP counties. In estimating the anticipated relative efficiency of the dual design to the single design for making statewide estimates for the domain of black adults in Georgia, formula (5) was used where now the summation is over black interviewed adults.

Using the single design 25.5% of the 4500 interviewed adults are anticipated to be black. Thus, for the single design for blacks only,

$(\Sigma w_i^2) = (4500)(.255)(676.2)^2 = 5.24690x10^8$.

For the dual design, 43.1% of the CHIP interviewed adults are expected to be black with 25.3% of the FOCUS interviewed adults. Thus, there will be 1139 black adults with an expected weight of 668.3, 248 with an expected weight of 18.1 and 372 with an expected weight of 29.1. The quantity (Σw_i^2) for this sample of 1759 black adults is $5.09102x10^8$. Thus, the relative efficiency of the dual design to the single design for making statewide estimates for blacks is 5.24690/5.09102 = 1.031. Again, the precision is better for the dual design but not dramatically so.

Based on the above relative precision findings and the concern with efficient use of limited resources for planning of surveys, the decision was made to implement the dual design. The next section uses the survey data collected under the dual design to estimate empirically the relative efficiency of the two designs.

EMPIRICAL COMPARISON OF TWO DESIGNS

Population Parameters

In order to investigate empirically the relative efficiency of these two designs, the estimation of five population parameters is considered:

(1) Percentage of adult population with an elevated blood pressure (EBP), where EBP is defined as mean diastolic blood pressure (DBP) of 95 mg Hg or higher OR mean systolic blood pressure (SBP) of 160 mm Hg or higher.

(2) Percentage of adult population who are hypertensive, where hypertensive is defined as having an EBP or taking antihypertensive medication.

(3) Percentage of adult population who have been told by a doctor that they are hypertensive.

(4) Percentage of adult population who never drink alcohol.

(5) Percentage of adult population who are employed.

Each parameter is estimated for all adults and for each of the following domains: blacks, whites (defined as nonblack), black females, black males, white females and white males.

Variance Estimation Procedure

The standard error for each point estimate is calculated using software developed by B.V. Shah (1981a, 1981b) at Research Triangle Institute. This approach is a linearization of the point estimate (which is typically a ratio estimator) by using a Taylor Series expansion around the relevant population parameters. The ultimate cluster approach was used where only stratum and PSU (primary sampling unit) are identified for each interviewee. Once the standard error (s.e.) is obtained for a given point estimate under each design, the squared ratio of the two standard errors is taken as the empirical relative efficiency, i.e.

$ERE(D/S) = [s.e (S)]^2/[s.e.(D)]^2$.

Weighting the Cases

The dual survey included 4101 SHU's rather than the planned 3960 SHU`s and had an overall response rate of 86% for a total of 6083 interviewed adults rather than the 5940 cases anticipated. The total sample of 6083 includes 4507 FOCUS cases (1226 blacks) and 1576 CHIP cases (750 blacks). The percentage of blacks among FOCUS and CHIP interviewees was 27% and 48%, respectively; it was estimated to be 25% and 43%, respectively. These differences are due primarily to the slightly higher interview response rate among blacks.

Each adult (case) in the dual survey has a case weight WT which is the product of three weights: WT1, the inverse of the probability of selection; WT2, an internal adjustment for nonresponse to SHU household enumeration; and WT3, a poststratification weight based on geography, age, race and sex which also adjusts for total nonresponse to the interview. The dual survey design anticipated that the FOCUS WT1 values would be around 668; in fact 87% of the calculated WT1 values for FOCUS cases were between 650 and 670. The dual design anticipated that 40% of the CHIP cases would have WT1 values around 18; in fact 37% of the calculated WT1 values for CHIP cases were between 18.0 and 18.9. The dual design also anticipated that 60% of the CHIP cases would have WT1 values around 29; in fact 51% of CHIP cases have a calculated WT1 value between 28 and 30. The product weight variable WT ranges from 10.7 to 84.4 for the 1576 CHIP interviewees, with 80% of the weights between 17 and 40. The variable WT for the 4507 FOCUS interviewees ranges from 104 to 2128 with 83% of the weights between 500 and 1000. See Brogan (1985b) for more detailed discussion on weighting.

Estimation Procedure for Each Design

In order to calculate the point estimates, their standard errors and design effects (deffs) under the dual design, the complex sample survey software was used on the database of all 6083 cases with the weight factor WT as described above.

Since the single design was not actually implemented, estimating the point estimates, standard error and deffs that would have been obtained under this design requires an approximation. This was done by using only the 4507 FOCUS interviewees and multiplying each case's WT1 value by an adjustment factor. This adjustment factor will be greater > 1 so that the newly weighted 4507 cases will estimate the number of adults in Georgia rather than in Georgia less the six CHIP counties.

An a priori adjustment is (676.2)/(668.3) = 1.0118, where 676.2 and 668.3 are the anticipated WT1 values for FOCUS interviewees under the single and dual designs, respectively. An adjustmenmt factor based on the actual WT values for the survey data is (3964900)/(3916670) = 1.012314, where the numerator is the sum of the WT values for all 6083 interviewees and the denominator is the sum of the WT values for the 4507 FOCUS interviewees only. The a priori and the empirical adjustment factors are quite close; the empirical adjustment factor 1.012314 is used in estimating standard errors for the single design.

Results

Table 1 presents the results of the empirical work for all adults, for estimating all five population parameters under each design.

Under the single design, the point estimate of the percentage of adults with elevated blood pressure (EBP) is 7.81% with standard error 0.51%. Under the dual design, the point estimate and standard error for estimating the same parameter are 7.83% and 0.50%, respectively. The relative efficiency of the dual design to the single design for estimating percentage with EBP is $(.0051)^2/(.0050)^2 = 1.024$.

Looking across the row ERE(D/S) of TABLE 1 for all five population parameters indicates that the relative efficiency of the dual design to the single design is 1.02, exactly as predicted by the earlier theoretical approximation. Note in Table 1 that the relative efficiency is independent of the population parameter being estimated.

Also included in TABLE 1 is the design effect or deff (Kish, 1965) for each design. The deff is expected to be larger under the dual design because of the larger variability of sampling weights; the data in Table 1 indicate that the design effect for the dual design is about 30% higher than for the single design. Although the design effects vary quite a bit over the population parameters, the magnitude of the design effect for a given variable has no significant impact on the relative efficiency of the single to the dual design since the component of the deff unique to the variables cancels out when the ratio of standard errors is calculated. This is illustrated in Table 1 by the empirical relative efficiency being 1.02 for all variables and the relative design effect being 1.3 for all variables.

Thus, in Table 1 the sample size is increased 35% (6083/4507) by using the dual design, but the design effect is increased 31%-32%. Hence, the relative precision of the dual design to the single design is only 1.02.

Tables 2, 3 and 4 (available from the authors) give similar information for three domains: all blacks, black females and black males. By using the dual design the sample size for blacks was increased 61% (1976/1226), but the design effect was increased by about 55%. The relative efficiency of the dual to the single design for these three domains is 1.04 for all five variables, close to the theoretical prediction of 1.03.

Tables 5, 6 and 7 (available from the authors) give the same information for three additional domains: whites, white females and white males. The dual design increased the sample size of whites 25% (4107/3281) over the single design, but the design effect was also increased about 23%. The relative efficiency of the dual design to the single design is 1.02 for the three white domains over all five variables, the same as when all adults are the population of inference.

CONCLUSION AND RECOMMENDATIONS

The empirical comparisons shown in TABLES 1 thru 7 agree remarkably well with the theoretical comparisons made before the survey was done. This illustrates that theoretical/empirical investigation can aid in choosing the best design for a survey. The relative design effect and the relative precision of the two designs are independent of the five variables considered in this investigation. This conclusion most likely is true for all of the variables in the given survey, although this conclusion may not hold in other combined surveys where the two surveys are more equally weighted and/or have different sample designs. This investigation illustrates that careful consideration needs to be given to the potential benefit of increasing total sample size in a complex sample survey when the increased sample size also results in significant increased variability of the case weights. In some instances there may be no or little precision payoff in increasing sample size coupled with significant additional cost. In such instances the dual design may be worse than a single design.

REFERENCES

Brogan, Donna (1985a). 1981 Georgia High Blood Pressure Survey: Description and Documentation. Technical Report. Dept. of Statistics & Biometry, Emory Univ., Atlanta, Ga. (67 pages).

Brogan, Donna (1985b). 1983 Georgia High Blood Pressure Survey: Description & Documentation. Technical Report. Dept. of Statistics & Biometry, Emory Univ., Atlanta, Ga. (67 pages).

Kish, Leslie (1965). Survey Sampling, John Wiley, New York.

Shah, B.V. (1981a). RATIOEST: Standard Errors Program for Computing of Ratio Estimates from Sample Survey Data. Research Triangle Institute, Research Triangle Park, N.C.

Shah, B.V. (1981b). SESUDAAN: Standard Errors Program for Computing of Standardized Rates from Sample Survey Data Research Triangle Institute, Research Triangle Park, N.C.

ACKNOWLEDGEMENTS

The authors' time on this paper was supported by NHLBI/NIH via contract N01-HV-92912 awarded to Joseph Wilber, M.D. of Georgia Department of Human Resources and grant HL 24004 awarded to Madelyn Chenault, Ph.D. of Fort Valley State College. Appreciation for aid in data collection and data management for the survey is extended to Harriette Hance, Delores Harlow, Judy Wess, Mary Pease, Jane Ginn, Gail McCray, B.K. Mishra, Carolyn Miles, Cecile Cate, Sherald Glover, Easter White, Sheena McGraw Scott, Ruth Moody, Patricia Brown, Sharon Kurtz, Carolyn Miles, Alan Boda, Patrick Whitaker, Phyllis Newman, Beverly Martinez, Nancy Levy, Anita Baldwin, Elaine Wells, Roz Perry, Robert Wortham, Dean Duncan, Peter Day, Michael Lynn, about 20 employees of Georgia Department of Human Resources who participated in listing and about 115 temporary employees who conducted field interviews. Appreciation for review and comments on earlier drafts of the paper is extended to Joseph Wilber and Beverly Martinez. Special appreciation is given to Phyllis Newman and Michael Lynn for manuscript preparation.

TABLE 1

Point Estimate, Estimated Standard Error and Estimated Design Effect for Five Population Parameters under Two Survey Designs and Relative Efficiency of Dual Design to Single Design

Adults in Georgia, 1983

DESIGN and RELATIVE EFFICIENCY	Percent with EBP	Percent Hypertensive	Percent Told Hypertensive	Percent Never Drink Alcohol	Percent Employed
SINGLE DESIGN					
PT EST	.0781	.1952	.2709	.4087	.6257
STD ER	.0051	.0087	.0081	.0142	.0108
Deff	1.60	2.14	1.49	3.72	2.22
Sample Size	4458	4467	4493	4486	4493
DUAL DESIGN					
PT EST	.0783	.1955	.2707	.4102	.6264
STD ER	.0050	.0086	.0080	.0140	.0106
Deff	2.10	2.81	1.96	4.90	2.93
Sample Size	6008	6021	6069	6060	6068
ERE(D/S)	1.024	1.024	1.025	1.024	1.023
DEFF(D)/DEFF(S)	1.31	1.31	1.32	1.32	1.32