A COMPARISON OF INDIRECT VARIANCE ESTIMATION

DH Freeman, Dartmouth Medical School,
M Livingston, L Leo, P Leaf, Yale University School of Medicine

Taylor series (TS) and balanced repeated replication (BRR) are the two most common sampling variance estimation techniques used in design based analysis of complex sample survey data. They have been compared on a variety of artificial samples with mixed results. Frankel (1971) used the 1967 Current Population Survey as a target population from which he drew a Monte Carlo simulation of three sample designs. These had 6, 12, and 30 strata each with two PSU's per stratum. He found both methods (as well as the jack-knife) provided reasonable variance estimates for ratios and differences of ratios, if the variance estimates are regarded as estimates of mean square error. BRR was superior to TS for correlation coefficients and the reverse was true for simple regression coefficients. When considering t-tests for the two, BRR consistently performed better than TS, producing more nearly correct p-values. Increasing the number of strata did not appear to improve the approximate p-values.

Bean (1975) compared the two methods for estimating the variance of a post-stratified ratio estimator with a more complex design. She concluded both methods provided generally similar and valid results. The major qualification concerned the poor performance of the procedures in estimating one-sided confidence intervals for ratios.

It should be noted that a number of authors have considered other techniques such as "jack--knifing" and more recently "boot-strapping." These are also reasonable approaches but there is no available evidence of their superiority to the TS and BRR procedures which are available in package computer programs such as SAS and OSIRIS. The availability of convenient software has made BRR and TS available to most data analysts.

In this paper the procedures are compared using data from two medium size surveys. The Epidemiologic Catchment Area Project (ECA) had the adult population of the New Haven MSA as its target population with an "elderly" over sample. The target period was July 1980 through December 1981. The Yale Health and Aging Project (YHAP) had, as its target population, the 65 and over population of the city of New Haven. Its target period was February to December 1982. Both surveys used the same frame but the sampling strategies differ significantly. Hence the underlying population parameters should be similar for appropriate sub-populations. Attention focuses on differences in the estimation procedures using actual surveys rather than the simulated surveys previously reported.

In addition to comparing the estimation procedures, alternative strategies for constructing "sampling error computing units (SECU)" are considered. It is concluded that differences between the procedures are generally smaller than previously reported. Moreover among the strategies considered for constructing SECU's all seem to yield similar sampling error variance estimates. Finally, problems with each procedure suggest that no one approach is applicable in every situation.

The Surveys

A statistical summary of the ECA sample selection design is shown in table 1. The target population of the ECA project is non-institutionalized adult residents of the city of New Haven and of the surrounding 12 towns (the New Haven MSA 1970 census definition). The frame consists of residential electrical utility connections with a supplemental list for bulk meters. From this list a systematic sample of households was drawn for each of the towns.

Clusters of eight housing units were selected at an interval of every 61-st unit. The first and fifth household of each cluster were selected for a general community sample. Persons 18 and over were eligible for interview. From each of these households one "community" respondent was selected using random sub-sampling techniques (Kish 1965).

An elderly over sample was obtained by screening the remaining six households for persons 65 and over, all of whom were selected for interview. Note that an elderly person could be included in the community sample or in the "elderly over sample." Thus the ECA selection design is a stratified cluster sample with disproportionate over sampling of the elderly.

A total of 5034 respondents were obtained in the ECA sample. Because the ECA is a probability sample respondents can be "up-weighted" to the complete target population. Sampling weights depend in part on whether an individual fell in the over sample. The survey weights shown in Table 1 reflect not only the selection design but also a non-response adjustment and post-stratification to the 1980 census totals according to the age-race-sex distribution of the New Haven MSA. The relatively low mean weight for the 65+ group reflects the over sampling of this group.

The target population of the YHAP is more narrowly defined to be non-institutionalized persons 65 and over resident in the City of New Haven. A statistical summary of the YHAP sample selection design is shown in Table 1. The frame is the same as for the ECA but geographically more limited. The focus on the elderly led to several important design differences. The predominance of females over males and the fact that large numbers of elderly live in age restricted housing meant that over-sampling of males and housing projects was desirable for statistical, economic and substantive reasons.

There are three strata: community housing which utilized the ECA frame, age restricted housing (Private), and age and income restricted housing (Public). For the community stratum clusters of twelve consecutive housing units were selected using the same sampling interval as the ECA but with a different starting point, one which ensured the YHAP and ECA clusters would be as widely separated as possible. All elderly males were included and females were sub-sampled at a rate of 1 in 1.5.

The housing projects were censused, excluding those housing units selected for the ECA sample. Artificial clusters of eight units existed but these were used only when estimating variances. In public housing all persons were selected and in private housing 1 in 2.5 females were selected. Thus the YHAP is a stratified cluster sample where the strata are type of housing and the cluster sizes differ depending on the stratum.

A total of 2811 elderly respondents were obtained in the YHAP sample. The weights shown in Table 1 were constructed similarly to those of the ECA. They incorporated an adjustment for non-response and post-stratification to the 1980 census totals by age and sex.

Two variables common to both surveys are the proportion of persons with at least one overnight hospitalization and the average number of hospitalizations for these persons. These variables will serve to illustrate our comparison of variance estimating procedures across the two samples.

<u>The Variance Estimation Procedures: Computation</u>

The programs used to estimate variances were SURREGR (Holt 1977) and a special purpose SAS program written using PROC MATRIX (SAS 1982). The former is based on the Taylor series linearization approach and is reasonably convenient for use with SAS data sets. With appropriate data manipulation it can be used for the estimation of the standard errors of means and proportions at slightly lower cost than the more commonly employed SESUDAAN (Shah 1979). The estimates are identical.

A small problem with SURREGR is that the statistic for testing hypotheses about several parameters is incorrectly referred to as approximately F in distribution. As noted by Koch, Freeman, and Freeman (1975) the correct test statistic has a chi-squared distribution. This can be readily obtained from the SURREGR output. For a test of d parameters, multiply the reported F value by the number of parameters: d. The resulting statistic is distributed as a chi-square with d degrees of freedom.

Software for computing variances by the balanced repeated replication method was written using SAS PROC MATRIX. The program uses sixty artificially created strata. For the ECA these were based on aggregating the clusters over the entire sample; for YHAP the aggregation was based on twenty for each of the 3 housing type strata. These are then sub-divided into half-samples yielding 120 pseudo-replicates for which totals and means can be generated. From these, 60 pseudo-replicates are selected according to the Plackett and Burman orthogonal matrix. Deviations of each half-sample estimate from the entire sample estimate are computed. The usual formula for estimating the variance based on 60 observations can then be employed to generate an indirect estimate of variance. As noted by Koch, Freeman, and Freeman (1975) if several domains are to be compared vectors of the domain estimates for the half samples can be used. The usual variance-covariance matrix estimator is then appropriate.

Subsequently, OSIRIS IV (ISR 1981) became available. The results for the regression models using its BRR program were quite similar to the SURREGR TS results. However, the computing time was about eight times longer than the other procedures examined. Consequently this package was not pursued farther. The BRR program using PROC MATRIX in SAS used CPU time comparable to SURREGR.

<u>Variance Estimation: Standard Error Computing Units</u>

The second issue is deciding on the appropriate standard error computing units (SECU's). In theory the surveys are designed with a fixed number of primary sampling units (PSU's) per strata and fixed cluster sizes within each PSU. In practice things are not so neat. Moreover, for BRR there should be precisely two PSU's per stratum. Hence in practice there is often some ambiguity in defining the SECU's. As noted above for YHAP the segments were combined to produce 20 pseudo-strata per housing stratum. The ECA design allowed more choices.

The most obvious approach was to treat the 13 towns as strata and the clusters as PSU's. Initially, this was thought to be expensive and potentially subject to an unacceptable degree of sampling variation. As an alternative, the clusters were combined to produce 60 pseudo-replicates in a manner similar to the YHAP procedure. It was thought that the large number of pseudo-strata might cause either computational or estimation problems, hence a set of 15 collapsed pseudo-replicates were also created.

<u>Results</u>

The comparisons across methods, samples, and SECU's are shown in Tables 2-7. The first three show the percentage of persons hospitalized in the year prior to the date of interview, while the last three show the mean number of hospitalizations among persons who have been hospitalized. By considering different age groups a range of percentages can be considered, from under 10 to over 20. Similarly by limiting the means to persons who have been hospitalized, the effect of varying sample sizes can be evaluated.

The basic pattern for the procedures is shown in Table 2. Here the percentage with a hospital visit in the year prior to the date of interview varies from 9.47 to 20.95. The percentages for the 65+ age groups did not differ materially for the two surveys. In addition there is a clear increase with age for both surveys. The first observation with respect to the standard errors is that the TS and BRR methods of estimation give nearly identical results when the same definition of SECU is used (60 pseudo-replicates).

A second point is that no single method and choice of SECU yields estimates which are higher or lower than the other methods. The TS 15 pseudo-replicate estimate is most frequently the lowest but there are enough exceptions to preclude any general speculation. Finally for

Table 2 it appears there is a negative design effect in the ECA over 65 domains. In these cases the estimates under simple random sampling are greater than the design based estimates.

The remaining tables follow the same general pattern as Table 2 except the negative design effect is either small or absent. These results suggest several points. First the choice between BRR and TS seems less important than the definition of SECU's. Presumably those closest to the underlying survey design are to be preferred.

Second if the SECU's are defined on a post-sampling basis then one should probably opt for a larger rather than smaller number of SECU's or pseudo-replicates. This is because the TS and BRR estimates are nearly identical when more SECU's are defined. It will be important to know whether this result holds in general.

Third in terms of computational efficiency (speed) BRR programs can be comparable to TS programs. In the present study for tables 2-7, SESUDAAN took from 9.0 to 10.9 cpu seconds, SURREGR 7.1 to 8.9 cpu seconds, and BRR about 12 cpu seconds. As noted earlier the OSIRIS version of BRR took about eight times as long as the TS or PROC MATRIX BRR.

Finally, one point came up that was unanticipated. When 60 pseudo-replicates were employed for small domains which led to empty pseudo-strata it was frequently impossible to obtain the estimates of standard error. It was for this reason the 15 replicates were developed. Hence while either TS or BRR is acceptable as a methodology, no available single piece of software meets all requirements.

References

Bean JA (1975). Distribution and properties of variance estimators for complex multistage probability samples, Vital and Health Statistics, Series 2:65.

Frankel MR (1971). Inference from Survey Samples: An Empirical Investigation, Ann Arbor, MI: Institute for Social Research.

Institute for Social Research (1981). OSIRIS IV User's Manual: Fifth Edition, Ann Arbor, MI: Institute for Social Research.

Holt MM (1977). SURREGR: Standard Errors of Regression Coefficients from Sample Survey Data, Research Triangle Park, NC: Research Triangle Institute.

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

Koch, GG, Freeman, DH, Freeman, JL (1975). Strategies in the multivariate analysis of data from complex surveys, Int Statist Rev, 43: 59-78.

SAS Institute (1982). SAS User's Guide: Statistics, 1982 Ed, Cary, NC: SAS Institute.

Shah BV (1979). SESUDAAN: Standard Errors Program for Computing of Standardized Rates from Sample Survey Data. Research Triangle Park, NC: Research Triangle Institute.

Table 1. Summary of selection designs

Design Characteristic	ECA <65	ECA 65+	Community	YHAP Public	YHAP Private
Sampling Weight (including post-stratification)					
n	2458	2576	1215	728	868
Mean	102.54	18.66	9.75	1.85	2.54
Standard Deviation	45.65	7.83	3.26	0.36	1.34
Minimum	12.70	10.19	5.00	1.00	1.00
Maximum	354.43	98.46	62.00	2.00	5.00
Respondents within strata					
Strata	13	13	20	20	20
Mean	189.08	198.15	60.75	36.40	43.40
Standard Deviation	196.22	223.41	4.30	2.37	3.33
Minimum	23	14	47	33	39
Maximum	749	815	68	43	53
Respondents within clusters					
Clusters	609	581	494	132	214
Mean	4.03	4.43	2.46	5.52	4.06
Standard Deviation	3.64	3.60	1.65	1.75	2.17
Minimum	1	1	1	1	1
Maximum	24	18	10	10	11

Table 2. Percentages and standard errors for any hospital visit by age and sample

	ECA <65	ECA 65-74	ECA 75+	YHAP 65-74	YHAP 75+
Percentage Any Visit	9.47	15.10	20.19	15.53	20.95
Sample size	2447	1613	970	1552	1234
Simple Random Sample	0.47	1.38	1.76	0.92	1.16
Taylor Series					
True Strata (13)	0.63	0.98	1.41	NA	NA
60 Pseudoreps	0.66	0.90	1.18	1.23	1.44
15 Pseudoreps	0.67	0.91	1.02	NA	NA
BRR					
60 Pseudoreps	0.67	0.91	1.19	1.23	1.44

Table 3. Percentages and standard errors for any hospital visit by sex and sample.

	ECA <65 men	ECA <65 women	ECA 65+ men	ECA 65+ women	YHAP 65+ men	YHAP 65+ women
Percentage Any Visit	7.30	11.45	20.25	14.89	21.40	15.68
Sample size	1055	1392	1007	1576	1156	1635
Simple Random Sample	0.85	0.82	1.18	0.95	1.21	0.90
Taylor Series						
True Strata (13)	0.87	0.89	1.38	0.96	NA	NA
60 Pseudoreps	1.00	0.93	1.22	0.99	1.41	0.98
15 Pseudoreps	0.86	0.77	1.18	1.10	NA	NA
BRR						
60 Pseudoreps	1.03	0.92	1.21	0.99	1.40	0.98

Table 4. Percentage and standard errors for any hospital visit by education (years completed) and sample.

	ECA <65 <12	ECA <65 12+	ECA 65+ <12	ECA 65+ 12+	YHAP 65+ <12	YHAP 65+ 12+
Percentage Any Visit	11.12	8.96	17.02	16.99	19.40	14.47
Sample size	477	1968	1440	1130	1853	867
Simple Random Sample	1.34	0.66	0.99	1.10	0.92	1.20
Taylor Series						
True Strata (13)	1.52	0.67	1.10	1.26	NA	NA
60 Pseudoreps	1.64	0.63	0.85	1.30	1.28	1.01
15 Pseudoreps	1.67	0.59	0.81	1.34	NA	NA
BRR						
60 Pseudoreps	1.65	0.63	0.83	1.32	1.28	0.99

Table 5. Mean number of hospital stays among those hospitalized and standard errors by age and sample

	ECA <65	ECA 65-74	ECA 75+	YHAP 65-74	YHAP 75+
Mean stays	1.32	1.46	1.54	1.55	1.52
Sample size	251	245	191	281	262
Simple Random Sample	0.06	0.14	0.16	0.09	0.09
Taylor Series					
True Strata (13)	0.08	0.11	0.15	NA	NA
60 Pseudoreps	0.08	0.10	0.14	0.06	0.17
15 Pseudoreps	0.10	0.07	0.14	NA	NA
BRR					
60 Pseudoreps	0.08	0.10	0.14	0.07	0.17

Table 6. Mean number of hospital stays among those hospitalized and standard errors by sex age and sample

	ECA <65 Men	ECA <65 Women	ECA 65+ Men	ECA 65+ Women	YHAP 65+ Men	YHAP 65+ Women
Mean stays	1.47	1.23	1.44	1.55	1.50	1.56
Sample size	80	171	201	235	272	273
Simple Random Sample	0.13	0.10	0.12	0.12	0.10	0.08
Taylor Series						
True Strata (13)	0.19	0.05	0.11	0.14	NA	NA
60 Pseudoreps	0.21	0.05	0.11	0.14	0.11	0.16
15 Pseudoreps	0.22	0.05	0.10	0.13	NA	NA
BRR						
60 Pseudoreps	0.22	0.05	0.11	0.14	0.11	0.16

Table 7. Mean number of hospital stays among those hospitalized and standard errors by education (years completed) age and sample

	ECA <65 <12	ECA <65 12+	ECA 65+ <12	ECA 65+ 12+	YHAP 65+ <12	YHAP 65+ 12+
Mean stays	1.16	1.38	1.62	1.36	1.48	1.70
Sample size	60	189	249	183	377	149
Simple Random Sample	0.16	0.09	0.11	0.13	0.08	0.12
Taylor Series						
True Strata (13)	0.07	0.11	0.15	0.07	NA	NA
60 Pseudoreps	0.07	0.10	0.14	0.08	0.07	0.25
15 Pseudoreps	0.07	0.12	0.12	0.08	NA	NA
BRR						
60 Pseudoreps	0.07	0.11	0.14	0.08	0.07	0.26