A COMPARISON OF PRECISION BETWEEN TWO- AND THREE-STAGE SAMPLE DESIGNS FOR NATIONAL HOSPITAL DISCHARGE SURVEY

Iris Shimizu, National Center for Health Statistics, and Emily Cole, University of Houston
Iris Shimizu, NCHS, 6525 Belcrest Road, Room 915, Hyattsville, MD 20782

KEY WORDS: Survey design

1. Introduction

Prior to 1988, the National Hospital Discharge Survey (NHDS) used a two-stage stratified sample. In 1988, the NHDS survey implemented a three-stage stratified sample. It was desired to know how use of a three-stage design instead of a two-stage design to select the NHDS sample affects the precision of that survey's estimates. This paper documents research designed to compare the precision from the implemented three-stage design with the corresponding precision from a simulated two-stage design. The design of the simulated two-stage sample approximates that which would likely have been used if NCHS had chosen to continue with two-stage designs for the NHDS.

To minimize the possibility of factors other than the number of sampling stages confounding the comparisons, a two stage sample was simulated for the comparison from the data in the 1988 three-stage NHDS. Use of the 1988 NHDS data for both designs in the comparison eliminates confounding that could result if the two samples were not from a common year or a common discharge sample. Use of the 1988 data also removes confounding that could occur due to the sampling frames and methodologies for sampling, non-response imputation, and variance computation which were not common between the 1988 and the prior NHDS design.

Section 2 outlines the three-stage sample implemented in 1988 while section 3 describes the two-stage design that was assumed for the comparisons. Section 4 outlines the study procedures and Sections 5 and 6 discuss the findings and conclusions, respectively.

2. NHDS present three-stage design

The NHDS universe consists of non-institutional, non-Federal hospitals in the 50 States and the District of Columbia which have an average length of stay for all patients of less that 30 days and have six or more beds staffed for patient use. In the redesigned NHDS the universe was expanded to include general hospitals, regardless of their lengths of stay. The sampling frame for the 1988 NHDS consisted of hospitals in the universe which were listed in the April 1987 SMG Hospital Market data tape.

The present NHDS sample includes with certainty all hospitals with 1,000 or more beds or 40,000 or more discharges annually. The remaining sample of hospitals was selected using a stratified three-stage design.

The first stage consists of 112 PSU's that comprise a probability subsample of primary sampling units (PSU's) used in the 1985-94 National Health Interview Survey (NHIS). The PSU's are counties or groups of counties or county equivalents or towns and townships (some PSU's in New England and Hawaii). The PSU strata were defined within four geographic regions by number of people in the 1980 Census of Population and NHIS stratification variables. From each stratum, the PSU's were selected with probability proportional to the projected 1985 population. For details of the NHIS PSU sample design, see NCHS, et al (1989).

The second stage consists of systematic random samples of non-certainty hospitals selected from the sample PSU's with probability proportional to their annual number of discharges. Primary or secondary strata of hospitals were defined by four geographic regions, by PSU size (1980 Census of population and number of hospitals), by whether the hospital subscribed to a commercial abstracting service, by PSU, by whether the hospital participated in the 1987 NHDS, and by hospital specialty and bed-size class. Finally, hospitals were arrayed within strata by number of discharges occurring annually at the hospitals and then sampled. The sampling rates were such that at least three hospitals were selected from every PSU containing three or more eligible hospitals. In PSU's with fewer than three hospitals, all hospitals in the PSU were selected.

The third stage consists of a systematic random sample of discharges from each hospital. For hospitals whose samples were selected by non-NCHS personnel, discharges were selected from lists in which discharges were listed in some chronological order. For hospitals whose samples were selected at NCHS, the discharges were selected from computerized discharge medical abstract files purchased from abstracting services and prior to sampling, those discharges were sorted on the first two digits of the ICD-9-CM code of the first-listed diagnosis, patient age group at time of admission (under 1 year, 1-14 years, 15-44 years, 45-64 years, 65-74 years, 75-84 years, 85 years and over, and age unknown), sex, and date of discharge. For details of the 1988 NHDS sample design, see Shimizu (1990).

The resulting sample included 542 hospitals. Of the 542 hospitals, 531 were eligible for NHDS for part or all of 1988. Of the 531 inscope (eligible) hospitals, 422 hospitals responded in 1988. From these hospitals, about 250,000 discharges records were included in the sample.

3. Simulated two-stage sample from three-stage sample

The simulated design approximates a two-stage stratified design that would have been a candidate for the redesigned NHDS in 1988 if NCHS had chosen to continue using two-stage designs for the NHDS. In that design, the first stage was assumed to be a stratified sample of the hospitals with strata defined by region, by hospital specialty and bed-size class, and by whether the hospital subscribed to a commercial abstracting service. The simulated design was assumed to include with certainty all hospitals with 1,000 or more beds or 40,000 or more discharges annually. In the non-certainty strata the hospitals were assumed to be selected using probability proportional to their annual number of discharges.

The design for the second stage of the simulated sample was assumed to be identical to that for the discharge level sample described previously for the three-stage 1988 NHDS sample.

The estimation methodologies for the two-stage design were assumed to be identical to those for the three-stage design, except PSU weights were absent and the data were inflated to the strata assumed in the two-stage sample instead of to strata that were used to select the three-stage sample.

Under these assumptions, it was sufficient to recalculate the hospital level weights for non-certainty hospitals in the 1988 NHDS data set in order to estimate precision levels likely from a two-stage design. Hence, the overall 1988 NHDS weights for discharges from certainty hospitals and the 1988 NHDS discharge level weights for discharges from non-certainty hospitals were used without modification to simulate data from a two-stage design.

The hospital level weights for the discharges from non-certainty hospitals were recalculated as the inverse of the product of two factors. The first factor is the assumed initial probability of the hospital's selection from its stratum, which was calculated as the ratio of the total discharges recorded for that hospital in the 1987 frame divided by the sampling interval for the stratum. The assumed sampling interval for the stratum was the ratio of total discharges for that stratum in the 1987 frame divided by the total number of hospitals (including out-of-scopes and non-respondents) in the 1988 three-stage sample which were contained in that stratum.

The second factor in the hospital level weights was an adjustment for non-respondent hospitals in the 1988 three-stage sample. Within each sampling stratum the adjustment consisted of a ratio, the numerator of which was the sum of 1987 discharges of the inscope sample hospitals and the denominator of which was the sum of 1987 discharges of the respondent hospitals.

4. Comparison Procedures

The investigation consisted of comparing precision in the form of the relative standard errors (RSE's) of identical selected statistics from the three-stage and the two-stage designs. RSE's (also known as coefficients of variation) were calculated as the standard error divided by the estimates.

4.1. Study statistics

The study statistics include seventeen variables which were subjectively selected for 1986-87 research leading to the NHDS redesign and a probability sample of 500 points which were selected for fitting five of the generalized RSE curves included in publications presenting the 1988 NHDS statistics. The seventeen variables used in redesign research are listed in Table 1 and include all of the types of statistics produced in the NHDS.

The five generalized RSE curves used in the study were only for estimates about "Numbers of discharges, first listed diagnoses, or all listed diagnoses." The specific statistics covered by each of the curves are listed in Table 2 together with the formula for each curve. The points used to fit each curve and included in this study were selected by first arraying the statistics relevant to that curve in the order of their magnitudes. The ten largest statistics in the array plus a systematic random sample of 90 additional statistics were then used with a weighted least squares procedure to fit the curve.

4.2. Sampling error estimation

Both sets of RSE's for the study were computed with SESUDAAN software, because that software was used to compute the 1988 NHDS variances. SESUDAAN uses a linearized Taylor series technique. The RSE's for the three-stage design were derived from the 1988 NHDS data with no alternations to the weights. The RSE's for the two-stage design were computed from the same data set but with the previously described alterations to the sampling weights and to SESUDAAN-required design parameters to reflect a two-stage sampling design.

Variances for both the three- and two-stage samples were computed using the assumption of equal selection probabilities at the first sampling stage to enable use of SESUDAAN. Theoretically, this assumption causes the resulting estimates to overstate the true errors to the degree that probability-proportional-to-size selection of hospitals improves precision over simple random sampling of those hospitals. However, the SESUDAAN error estimates are the ones used in the analysis and publication of 1988 NHDS data and, hence, the error estimates used in this study.

To compute the variance estimates, individual certainty hospitals were treated as separate sampling strata so that SESUDAAN would not erroneously compute a between-hospital sampling error for these hospitals. Otherwise, the variance computations were straightforward for the two-stage sample.

Using a weighted least squares approach, a generalized error curve of the form:

$$RSE(\hat{X}) = \sqrt{Var(\hat{X})/\hat{X}^2}$$

was then fit to each of the five sets of 100 points for the two-stage design. Such curves are used to approximate RSE's in publications which include so many estimates that inclusion of the actual RSE for every estimate is impractical in the publication. More detail on the process for fitting generalized RSE curves in general and the computations of error estimates for the three-stage NHDS design are given in Shimizu (1990).

5. Findings

Tables 1 - 3 and Figure A present results computed from the 1988 NHDS data. In these tables and figure, positive differences mean that the three stage sample RSE's are greater than those from the two stage sample. "Small" differences may be assumed to indicate that the two designs give about the same precision. While the definition of "small" is arbitrary, it will be assumed in the following that the precision for a statistic from the two designs is approximately the same if the absolute value of the RSE difference for that statistic is less than one percentage point.

Table 1 presents the comparison of RSE's for the 17 variables used in redesign research. For all statistic types (discharges and diagnoses, days of care, and procedures), the RSE differences are positive except those for the diagnoses of psychoses or alcohol dependence. Also, the alcohol dependence variables have differences of about 10 percentage points while the absolute differences for the remaining variables are all less than 3.5 percentage points. Three of the differences have absolute values of less than one percentage point.

Tables 2 and 3 and Figure A display results derived from the probability sample of 500 points selected to fit generalized RSE curves. Table 2 gives the formulas for the RSE curves fitted to those points while Figure A displays the differences between the corresponding fitted RSE's from the two designs. Based on the curves in Figure A, the fitted RSE's for the three-stage design were greater than or equal those for the two-stage design for all estimates of 30,000 or more discharges and for all estimates for the discharge population cells represented by error curves D and E. It appears the three-stage design increased the fitted RSE's by at most five percentage points and that was for statistics about non-whites or about discharges in three of the four geographic regions. For the remaining curves, the increase in fitted RSE's appears limited to three percentage points.

Because the approximate RSE's here are fitted to samples of points rather than to all the points published, the fitted RSE's are subject to sampling error. That is, use of a different sample of points could yield a different set of curves. Resources for the study, however, did not permit the examination of RSE differences for all possible estimates. It is believed unlikely, however, that the overall magnitude of the maximum differences between the curves would vary significantly from those experienced with the observed curves.

Table 3 displays the frequency distribution of the RSE differences for the individual sample points. As expected, the three-stage RSE's exceed (by one percentage point or more) the two-stage RSE's for a majority (about 60 percent) of these points. On the other hand, the three-stage design appears to yield better precision than the two-stage design for almost 20 percent of the points. An examination of the individual points revealed no particular discharge characteristics which consistently explain which variables are more likely than others to result in better or worse precision from the three-stage design than the two-stage design. One may note, though, that the number of differences with absolute values greater than 5 appears to vary inversely with the size of the discharge population cell for the statistic. Indeed, the largest absolute differences in the observed sample occurred only to statistics for some of the smallest discharge population cells considered for curves, that is, for discharges from hospitals having fewer than 100 beds or discharges to persons of unstated race or age 15 years or less. Plots (not shown) of the RSE differences by the estimated numbers of discharges revealed that large differences (10 or more) in either direction occurred for estimates ranging from 5,000 to around one million. Hence, size of the estimate offered little explanation for the size or direction of the RSE differences.

6. Conclusions

Based on the findings from the study, it appears that use of a three-stage sample does decrease the precision for many, but not all, of the statistics produced from the NHDS. No clear patterns emerged from the study that would help in determining which statistics have better or worse results from a three-stage sample.

It does appear that the three-stage sample design increased generalized RSE's by at most three percentage points for national statistics, other than those for non-whites, and by at most five percentage points for regional statistics and statistics for nonwhites. However, it appears that the three-stage design also afforded some precision gain (at most six percentage points) for the smallest statistics for larger hospitals, for regions (not Northeast), and for nonwhite race.

All errors used in this study were produced assuming equal selection probabilities within strata due to the limitations of the SESUDAAN software which was available for computing the 1988 errors. The study was not repeated after SUDAAN (which allows for the unequal selection probabilities present in the hospital sample) became available. It is believed that use of the new feature in SUDAAN would not change the basic relationship between errors for the three and two-stage sampling design--the change in results, if any, would likely be in the magnitude of the errors themselves.

REFERENCES

Shah, B.V.; LaVange, Lisa M.; Barnwell, Beth G.; Killinger, Joyce E.; and Wheeless, Sara C. 1989. SUDAAN: Procedures for Descriptive Statistics User's Guide. Research Triangle Institute, P.O. Box 12194, Research Triangle Park, NC 27708-2194.

Shimizu, Iris M. 1990. American Statistical Association 1990 Proceedings of the Section on Survey Research Methods. American Statistical Association. pp.702-707.

National Center for Health Statistics; Massey, J.T.; Moore, T.F.; Parsons, V.L.; and Tadros, W. 1989. Design and Estimation for the National Health Interview Survey, 1985-94. Vital and Health Statistics. Series 2, No. 110. DHHS Pub. No. (PHS) 89-1384. Government Printing Office, Washington, D.C. 20402.

TABLE 1. Relative standard errors (RSE's) for three-stage sample and comparison with RSE's for two-stage sample for redesign research variables: 1988 National Hospital Discharge Survey

Variable (ICD-9-CM code)	Estimate	RSE3[1]	RSE3/ RSE2[1]	RSE3 - RSE2
	(1,000's)	%		%
DISCHARGES AND FIRST LISTED DIAGNOSES				
1. All discharges excluding births	31,721	3.13	2.71	1.97
2. Diseases and disorders of the Circulatory system (390 - 459)	5,379	3.88	1.61	1.47
3. Psychoses (290 - 299)	784	13.58	0.94	-0.90
4. Alcohol Dependence (303)	245	9.38	0.50	-9.56
5. Fractures of femur and Hip+Femur Procedures (820 - 821)	259	6.38	1.39	1.80
DAYS OF CARE FOR FIRST LISTED DIAGNOSES:				
6. All discharges excluding births	206,744	3.48	1.46	1.10
7. Diseases and disorders of the Circulatory system (390 - 459)	40,424	4.04	1.69	1.64
8. Psychoses (290 - 299)	11,830	15.92	0.85	-2.74
9. Alcohol Dependence (303)	2,753	12.01	0.52	-10.93
10. Fractures of femur and Hip+Femur Procedures (820 - 821)	3,473	5.80	1.17	0.85
DISCHARGES WITH PROCEDURES (LISTED IN ANY ORDER):				
11. All Procedures	39,529	4.02	2.13	2.13
12. Cesarean Sections (74.0 - 74.2, 74.4, & 74.99)	947	5.29	1.31	1.25
13. zzzLens Procedures (13.1 - 13.7)	219	17.09	1.02	0.33
14. Transurethral Prostatectomy (60.2)	328	5.82	1.34	1.46
15. Permanent Cardiac Pacemaker Procedures (37.80)	38	14.81	1.27	3.19
16. Coronary Bypass (36.1)	347	12.10	1.38	3.34
17. Circumcision (64.0)	32	13.06	1.09	1.11

[1] RSE3 and RSE2 are the relative standard errors from the three- and two-stage sample designs, respectively.

Table 2: Coefficients for selected relative standard error (RSE) curves by number of sampling stages and by error curve: 1988 National Hospital Discharge Survey

{Curves of the form $\mathrm{RSE}(\hat{X}) \approx \sqrt{A + B/\hat{X}}$ are fitted by a weighted least squares procedure to points whose coordinates are the estimate $\hat{X}$ and $\mathrm{RSE}(\hat{X})$.}

	Three stage sample		Two stage sample	
Curve and the statistics class covered by the curve	A	B	A	B
A. Region other than Northeast/Nonwhite race	0.005106	337.01830	0.000524	484.59569
B. Bedsize greater than or equal 100	0.013719	154.82287	0.007900	211.39703
C. Age greater than 15/White race/Northeast region/Sex	0.001591	403.12398	0.000193	411.93267
D. Bedsize less than 100/Unstated race	0.029699	539.56300	0.024706	476.81766
E. Age less than or equal 15	0.012551	393.91963	0.012104	329.43459

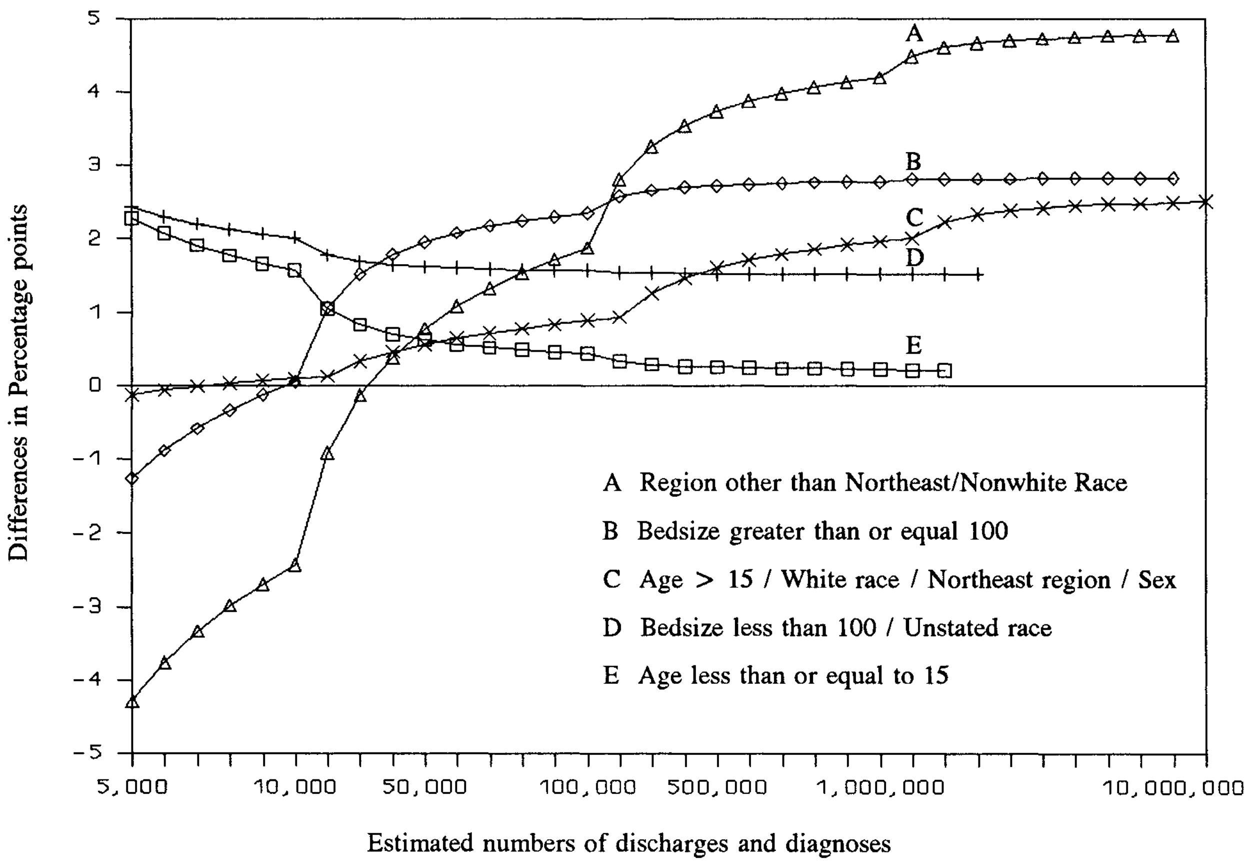

Figure A: Differences between fitted relative standard errors (three-stage minus two-stage): 1988 National Hospital Discharge Survey

Table 3: Frequency distribution of differences between the three- and the two-stage design relative standard errors (RSE's) for sample estimates: 1988 National Hospital Discharge Survey

Range	RSE curve					
	All	A[1]	B	C	D	E
	Number of points					
All sample estimates	500	100	100	100	100	100
<-20	3	0	0	0	2	1
[-20, -15)[2]	3	1	1	0	0	1
[-15, -10)	6	2	1	0	1	2
[-10, -5)	9	2	1	1	2	3
[-5, -3)	19	3	2	3	4	4
[-3, -1)	38	7	4	6	10	11
[1, 1]	110	23	19	23	21	24
(1, 3]	143	23	35	49	16	20
(3, 5]	79	23	24	10	14	8
(5, 10]	48	12	7	5	14	10
(10, 15]	22	0	5	2	6	9
(15, 20]	12	2	1	1	4	4
>20	8	0	0	0	4	4
Mean of differences	2.35	1.5	2.4	1.7	3	3.1
S.E. of differences	6.42	4.4	4.4	3.2	9.4	8.4

[1] Each sample of 100 points is used to fit a curve of the form $RSE(\hat{X}) = \sqrt{A + B/\hat{X}}$ to approximate the RSE's for one of the following sets of estimates $\hat{X}$:

A Region other than Northeast or Nonwhite race,
B Bedsize greater than or equal 100,
C Age greater than 15 or White race or Northeast region or sex
D Bedsize less than 100 or Unstated race, and
E Age less than or equal 15.

[2] [x,y) is set of those points greater than or equal to x but less than y.