SAMPLE DESIGN FOR THE 1988 NATIONAL MATERNAL AND INFANT HEALTH SURVEY

Joe Fred Gonzalez, Jr., Kenneth G. Keppel, Steven L. Botman
National Center for Health Statistics, 6525 Belcrest Rd., Hyattsville, MD, 20782

KEY WORDS: Sampling, Estimation, Data Collection

Introduction

The 1988 National Maternal and Infant Health Survey (NMIHS) [1], which was conducted by the National Center for Health Statistics (NCHS), was a major study of women in 1988 who had a live birth, fetal death, or an infant death. The objective of the 1988 NMIHS was to produce the data needed by Federal, State, and private researchers to study factors related to fetal loss, low birth weight, and infant death. The NMIHS was conducted to extend the range of statistical information on live births, late fetal deaths, and infant deaths beyond that available on the vital record. Data were collected from three types of respondents associated with sampled deliveries: from mothers; from hospitals; and from prenatal care providers. This paper will describe the sample design, questionnaires, estimation procedures, and implications of the sampling design for analysis.

Background

The vital registration system provides data on vital events that occur in the United States [2]. Because vital records serve both legal and statistical purposes, they provide limited social, demographic, health, and medical information. The data from vital records can be augmented, however, through periodic "followback" surveys. These surveys are referred to as "followback" because they obtain additional information from sources named on the vital record. A followback survey is a cost-effective means of obtaining supplementary information for a sample of vital events. From the sample it is possible to make estimates of vital events according to characteristics not otherwise available. Periodic followback surveys respond to the changing data needs of the public health community without requiring changes in the vital record forms.

Previous National Natality Surveys were conducted by NCHS based on live births in 1963, 1964-66, 1967-69, and 1972. In 1980 a National Natality Survey and a National Fetal Mortality Survey were conducted. The 1988 NMIHS is the equivalent of a combined National Natality Survey, National Fetal Mortality, and National Infant Mortality Survey.

Sample Design

The overall sample design for the NMIHS was based on: knowledge and experience about precision of estimates and response rates obtained in The 1980 National Natality and Fetal Mortality Surveys and budgetary resources available to conduct the NMIHS nationwide. As shown in Table 1, there were three national stratified systematic samples of vital records that were selected. For live births, there were six sampling strata by race and birthweight. For fetal deaths and infant deaths, there were only 2 strata by race. Although not shown in Table 1, in order to assure a representative sample by such variables as age of mother and marital status, implicit stratification was employed. That is, after the live birth records were stratified as shown in Table 1, further sorting of vital records was done by age of mother and marital status within each of the strata shown for live births. Similar subsorting was carried out for fetal and infant death records.

Table 1 presents the distribution of the population and samples by sampling strata.

In order to make reliable comparisons by race, the expected sample sizes shown in Table 1 reflect a goal of the design which was to equally allocate the sample size by race.

The actual sample sizes given in Table 1 reflect inflated samples sizes (over the corresponding expected responding sample sizes) to compensate for projected nonresponse based on response rates in the 1980 National Natality and Fetal Mortality Surveys.

Vital records were sampled within each State throughout 1988. Provisional data were produced in 1990 from the first group of 1988 respondents. Data are still being received from certain States, and it is expected that all data will be collected by the end of 1990. The Bureau of the Census is the data collection agency under contract for the NMIHS.

Supplemental Samples

In addition to the NMIHS national samples, the following supplemental samples were selected:

- 1000 live birth records of American Indians in 35 urban areas where the Indian Health Service funds urban Indian programs
- Hispanics were oversampled in the State of Texas in order to produce reliable estimates for Hispanics in Texas

Note: It should be emphasized that the supplemental samples are not national samples and, therefore, do not imply that national estimates can be produced for American Indians or Hispanics.

I. Aggregates (Totals)

A) Estimator

$$X' = \sum_{i=1}^{\eta_D} W_i x_i = \text{weighted total}$$

B) Variance Estimator

$$Var(X') = \underset{SRS}{Var(X')} \cdot DEF$$

$$= \frac{X' (T_D - X')}{\eta_D} \cdot DEF$$

$$SE(X') = \sqrt{\frac{X' (T_D - X')}{\eta_D} \cdot DEF}$$

where X′ is as previously defined,

T_D = complement total of X′ within D

DEF = design effect = $\frac{Var(X)}{\underset{SRS}{Var(X)}}$

Imputation

Imputation for item nonresponse was carried out only for the Mother's questionnaire by using a "hot deck" imputation procedure. The following is the imputation matrix that was used within birth outcome and sampling strata:

	Marital Status	
Age of Mother	Married	Unmarried
<20 years		
20-29 years		
30+ years		

Estimation Procedures

Notation: let x_i = X-characteristic for the i^{th} sample person within a certain domain D

η_D = sample size within a certain domain D

P_i = probability of selection of i^{th} sample person

NR = nonresponse adjustment

PS = poststratification adjustment

W_i = final weight for i^{th} sample person

= $\frac{1}{P_i} \cdot NR \cdot PS$

Table 1. Numbers of events in the universe, expected responding sample sizes, selected sample sizes, and sampling fractions for the NMIHS by survey component and sampling stratum

Live Births

	Universe	Expected Responding Sample Size	Selected Sample Size	Sampling fraction
Total	3,898,922	10,000	13,898	1/280
Black:				
<1500g	18,403	750	1,315	1/14
1500-2499g	67,651	750	1,230	1/55
2500+g	579,740	3,500	5,130	1/113
Nonblack:				
<1500g	29,815	750	1,028	1/29
1500-2499g	153,379	750	959	1/160
2500+g	3,049,934	3,500	4,236	1/720

Fetal Deaths

	Universe	Expected Responding Sample Size	Selected Sample Size	Sampling fraction
Total	15,256	4,000	5,400	1/2.83
Black	3,847	2,000	2,317	1/1.66
Nonblack	11,409	2,000	3,083	1/3.70

Infant Deaths

	Universe	Expected Responding Sample Size	Selected Sample Size	Sampling fraction
Total	38,917	6,000	8,304	1/4.69
Black	11,840	3,000	4,625	1/2.56
Nonblack	27,077	3,000	3,679	1/7.36

II. Proportions

A) Estimator

$$\hat{p} = \frac{X'}{N_D} = \frac{\sum_{i=1}^{\eta_D} W_i \, x_i}{\sum_{i=1}^{\eta_D} W_i}$$

B) Variance Estimator

$$Var(\hat{p}) = Var_{SRS}(\hat{p}) \cdot DEF = \frac{\hat{P}(1-\hat{P})}{\eta_D} \cdot DEF$$

$$SE(\hat{p}) = \sqrt{\frac{\hat{P}(1-\hat{P})}{\eta_D} \cdot DEF}$$

III. Rates

A) Estimator

$$\hat{R} = \frac{X}{Y}$$

B) Variance Estimator

Case 1: X is a subclass of denominator Y

$$Var(\hat{R}) = \hat{R}^2 [RSE^2(X) - RSE^2(Y)]$$

$$SE(\hat{R}) = \hat{R}\sqrt{RSE^2(X) - RSE^2(Y)}$$

where $RSE^2(X) = \text{Rel-variance of } X = \frac{Var(X)}{X^2}$

and $RSE^2(Y) = \text{Rel-variance of } Y = \frac{Var(Y)}{Y^2}$

Case 2: X is not a subclass of denominator Y

$$Var(\hat{R}) = \hat{R}^2 [RSE^2(X) - RSE^2(Y)]$$

$$SE(\hat{R}) = \hat{R}\sqrt{RSE^2(X) - RSE^2(Y)}$$

where all quantities are as previously defined

The variance estimators presented above are suggested variance estimators only if the design effects are known for the specific variables of interest. Of course, variances are more accurately computed by using an appropriate computer software package such as SUDAAN [3].

Implications of the Sample Design on Analysis

Referring to Table 1 under total for live births we see that the overall sampling fraction for live births is 1/280. This means that the corresponding overall sampling weight, which is the reciprocal of the sampling fraction, is 280. However, since Blacks were oversampled, in order to achieve equal sample allocation with Nonblacks, there is a large difference in sampling fractions by race and birthweight. The sampling fractions for Blacks by birthweight were: 1/14, 1/55, and 1/113 as compared to Nonblacks 1/29, 1/160, and 1/720, respectively. This means that sampling weights ranged from 14 to 720 (a ratio of 1:52) for

Table 2. Design Effects for Percentage Estimates for Sampling Strata in the 1988 NMIHS by Survey Component

Live Births (6 strata)

- Black (3 strata)
  - <1500g, 1500-2499g, 2500+g } 1.18
- Nonblack (3 strata)
  - <1500g, 1500-2499g, 2500+g } 1.31

} 2.42

Fetal Deaths (2 strata)

- Black, Nonblack } 1.13

Infant Deaths (2 strata)

- Black, Nonblack } 1.32

live births in the United States. It should be mentioned that this range of sampling weights is prior to nonresponse and poststratification adjustments which could further increase the variability of the final weights. This differential weighting increases the variance of estimates of characteristics.

In order to measure the increase in variance for percentage estimates due to differential sampling rates, a method outlined in Kish [4] and in a Westat memo written by Joseph Waksberg [5] was used. We then approximated the design effects (DEF) resulting from disproportionate sampling.

The results of the calculated DEFF's are shown in Table 2. From Table 2, we observe that for the Black stratum and the Nonblack stratum the DEFF = 1.18 and 1.31, respectively. These DEFF's represent an 18% and 31% increase in the variances of estimates for Black and Nonblack live births, respectively. However, the overall DEFF = 2.42 for the combined races represents an increase of 142% in the variances of estimates. This large increase in the variance is supported by the previous discussion about the range (1:52 ratio) of the sampling weights across Black and Nonblack strata.

For fetal deaths and infant deaths, the DEFF = 1.13 and 1.32 for the combined races, respectively.

Before conducting an in-depth analysis a data user should be fully aware of the NMIHS sample design as well as the differential sampling that was used.

Questionnaires

Approximately 20,000 mothers, the hospitals where their births and infant deaths occurred, and their providers of prenatal care will be followed back with mail questionnaires and interviews which will be linked with the vital records. The following are the types of data that will be collected with each of the three mail questionnaires:

From Mothers
- Barriers to prenatal care
- Source of payment
- Women, Infants and Children (WIC) use patterns
- Smoking, drinking, and marijuana use
- Work patterns before and after delivery
- Infant feeding practices
- Infant health and medical care up to 6 months
- Sociodemographic characteristics

From Hospitals
- Maternal hospitalizations-prenatal and up to 6 months postpartum
- Maternal and infant diagnoses and procedures (in ICD-9)
- Charges for care and DRG's
- Cesarean delivery and trial of labor
- Fetal monitoring
- Medical devices, apnea monitors, and respirators
- Infant resuscitation and neonatal intensive care
- Infant hospitalizations up to 6 months

From Prenatal Care Providers
- At each prenatal visit: weight, blood pressure, hematocrit, urine glucose, urine protein, hemoglobin
- Patient education, advice, and referral
- AIDS and sexually-transmitted disease testing
- Sonograms and X-rays
- Prescribed medications and vitamins
- Amniocentesis and chorionic villus sampling
- Charges for care

Cosponsors

The NMIHS was conducted by the NCHS in collaboration with the following cosponsors:

National Institute for Child Health and Human Development/NIH
Food and Nutrition Service, U.S. Department of Agriculture
Bureau of Maternal and Child Health and Resources Development/HRSA
Office of Minority Health/OASH
National Institute on Drug Abuse/ADAMHA
Center for Food Safety and Applied Nutrition/FDA
Office of Planning and Evaluation/PHS
Assistant Secretary for Planning and Evaluation/OASH
Center for Prevention Services/CDC
National Institute on Alcohol Abuse and Alcoholism/ADAMHA
Center for Devices and Radiological Health/FDA
Agency for Toxic Substances and Disease Registry
Center for Health Promotion and Education/CDC
Office of Minority Health/CDC
Indian Health Service
National Institute of Mental Health

Acknowledgement

The authors would like to express their gratitude to Ingrid Amos and Janice Melvin for preparing this manuscript.

References

1) National Center for Health Statistics, 1988 National Maternal and Infant Health Survey Brochure.

2) National Center for Health Statistics, K.G. Keppel, R.L. Heuser, P.J. Placek, et al: Methods and Response Characteristics, 1980 National Natality and Fetal Mortality Surveys. Vital and Health. Series 2 No. 100. DHHS Pub. No. (PHS) 86-1374. Public Health Service. Washington. U.S. Government Printing Office, Sept. 1986.

3) SUDAAN (1989), Profession Software for Survey Data Analysis, Research Triangle Institute.

4) Kish, L. (1965), Survey Sampling. John Wiley and Sons, New York.

5) Memorandum by Joseph Waksberg, Westat, December 15, 1977.