

## SAMPLE SURVEYS AT NATIONAL CENTER FOR HEALTH STATISTICS

Jai Won Choi and Lester R. Curtin, National Center for Health Statistics,  
6525 Belcrest Rd., Hyattsville, MD 20782

### I. INTRODUCTION.

The National Center for Health Statistics (NCHS) has maintained the mechanism which provides the information on the matters of health as well as vital events, health resources and costs, and related affairs. NCHS data system can be classified into two broad categories by its characters: the first category is the census data. Such data are based on the entire population, and do not require sampling. The second category is the information obtained from sample. The scope of this report is limited to the sample surveys, which require special statistical techniques of sampling, estimation, and variance. We shall discuss mainly these three aspects for the sample surveys. NCHS sample surveys are broadly divided into four categories: Population Based Surveys, Care Provider Record Based Surveys, and Vital Registrations Based Surveys, and Employer Based Health Insurance Survey. NCHS has concurrently executed these surveys. The sample design of most NCHS surveys employs multistage probability sampling designs. NCHS data may be contaminated with both random and systematic errors from the process of data collection and estimation. We will discuss mainly random errors for variance. The exact variance of NCHS data is often difficult to obtain, because most NCHS data are not based on a traditional sampling design, but on a complex multistage unequal probability sample to meet special situations. Previously, at least ten methods have been used: Keyfitz Method (Keyfitz, JASA 52(280),1957), Balanced Half Samples (McCarthy, Series 2, No.14,1963), Third Sample Replication (Bryan, 1971), Jackknife (Fay, JASA, 1980), Bootstrap (Efron, 1986), Taylor series expansion, SUDAAN (Woodruff, JASA 66, 1971), Model Based Variance (Choi and McHugh, Biometrics, 1989), Generalized Variance Function, and Conditional Variance method. The details of variance estimation will be presented in other publication (Choi, Series 2 report draft, 1994).

### II. POPULATION BASED SURVEYS.

The population based surveys are mostly accomplished through personal interview or examination, or both. Eight such surveys of different durations are briefly discussed below:

#### II.1. NATIONAL HEALTH INTERVIEW SURVEY

NHIS is the first NCHS survey, and has been done every year since 1957. HIS SAMPLE DESIGN: Most NHIS samples have been selected approximately by

probability proportional to size (pps). However, since 1985, the blacks have been oversampled from the areas where more than 5 percent of population are black. NHIS sampling designs for 1957-1972 and 1973-1984 were described in Technical Reports 7 and 40 (Bureau of Census), respectively, and these two designs were almost same except the rotating PSU or segments. The current design for 1985-1995 (NCHS, Series 2, No 110, 1989) differs from the 1973-84 design in five major points: The 376 PSUs reduced to 198 in 1985 design, while increasing the sample size within PSU in 1985. Introduction of four independent panels of approximately equal size, each representing the U.S. population. Adoption of all area sample frame replacing list frame, Sampling of two PSUs from each nonself-representing stratum to derive variance between PSUs, Over-sampling of blacks. However the principles of sampling of PSUs, EDs, segments, and sample persons remain the same. NHIS ESTIMATION: The sample units are weighted up to be consistent with population. The estimation methods remained the same for both old and new plans. First, the basic weight is the inverse of the probability of selecting final sampling units. Second, the basic weight is multiplied by the nonresponse ratio within each segment. The resulting weights are again multiplied by the ratio of the population to the sample estimate for each of 48 cells for NSR-PSUs. The fourth step is the third weight are further adjusted by the ratio of population to the sample estimate for each of the 60 cell in the age-sex-race. Finally, the weights were further adjusted by reference period. NHIS VARIANCE: Since 1973 up to 1985, the Balanced Half Sample (BHS) has been used for the NHIS variance estimation. Recently SUDAAN method is applied to estimate NHIS variance.

#### II.2 NATIONAL FAMILY GROWTH SURVEY

NFGS has been conducted by cycles of survey to obtain data on females of reproductive ages. The first cycle of survey was conducted in 1973, the second cycle in 1976, and the third in 1982. The fourth cycle was conducted in 1987 using the NHIS sample frame. SAMPLE DESIGN. Cycle I: Target population is the women who are less than 45 years, and who are ever married or single mothers with children. The design is based on a five-stage selection of PSUs, replicate groups of PSUs, segments, dwelling units, and one eligible person in a housing unit. Cycle II: The sampling is also the five stages selection of PSU, ED,

segment, household, and one woman of 15-44 years. (Series 2, No. 87, 1981). All interviews were conducted between January and September 1976. Cycle III: The sample design of Cycle III was also a five stage probability design that incorporated oversamples of black and teenage women and a supplementary sample of women living in college dormitories and sororities. (Series 2, No. 98, Bachrach et al, 1985). Cycle IV is different from the previous cycles because this survey used the sample frame of 1985-1987 NHIS. The selection up to segments is already presented in the Section of NHIS. The selection of households and women were further needed for the Cycle IV. In the sample households with one women eligible for NFGS, that woman is selected with certainty. In the sample households with two or more women eligible, only one was selected with equal probability. Estimation and variance for NFGS are similar to NHIS. (NCHS Series 2, No.76, 1978).

### **II.3 NATIONAL MEDICAL CARE UTILIZATION & EXPENDITURE SURVEY.**

NMCUES collects data on access to and the use of medical services, associated charges and sources of payment, and health insurance coverage. Two independent nationwide samples are pooled together. One sample is taken by Research Triangle Institute (RTI), and the other by National Opinion Research Center (NORC). SAMPLE DESIGN: Both are characterized as stratified, five-stage area probability designs. (NCHS Series A, No.4, 1988; Series A, No.1, 1983). ESTIMATION and VARIANCE are obtained along the line of NHIS.

### **II.4. LONGITUDINAL STUDY OF AGING.**

The longitudinal study of aging is based on the 1984 supplements on aging to NHIS. The 1984 base line sample includes everyone, 55 years of age and over living in a household where an NHIS interviews are completed in 1984, but only half of those ages 55-65 years are selected to participate. All participants are the self-respondent except when incapacity or absence prevented it. THE 1986 LONGITUDINAL AGING: The population for the 1986 interview sample is 7,541 persons, 70 years of age and over in 1984 when they participated in the NHIS supplement for aging. A subsample of 5,151 persons is taken for the 1986 interview. THE 1988 AND 1990 LONGITUDINAL SAMPLE OF AGING: The base line for the 1988 and 1990 follow-up is also the NHIS supplement-of-aging participants, 70 years of age and over in 1984. There was no sampling, and all persons who were 70 years of age and over when they participated in the supplement-of-aging in 1984 are included. However interview does not include those who died at the time of interview.

### **II.5. TELEPHONE SURVEY ON SMOKING.**

A national random digit dialing (RAD) telephone survey is used to obtain information on smoking and health related characteristics of the civilian non-institutionalized population in the conterminous United States in 1980. Two interviews were scheduled six months apart. All adults 17 years of age and over were eligible for the interview and self-response was required for all adults. SAMPLING: In response to unlisted problem, NCHS used two stage random digit dialing telephone survey which provided coverage of both listed and unlisted telephone households by generating telephone numbers at random from the frame of all possible telephone numbers. ESTIMATION: First, the basic weight is the reciprocal of the basic probability. Second, the multiple telephones are adjusted in the sample households. Third, the nonresponses are adjusted at the block level. Fourth, the weight is adjusted for the houses without telephone. Fifth, the adjustment for a poststratified ratio.

### **II.6. NATIONAL HEALTH AND NUTRITION EXAMINATION SURVEY.**

Health Examination Surveys (NHES) started in 1959 has been expanded to include Nutrition Examination Survey and Epidemiologic Follow-up Surveys. NHES 1: Nov.1959-Dec.1962. It focussed on the prevalence of selected chronic conditions. The target population is the civilian non-institutionalized U.S. adults aged 18-79. NHES 2: July 1963-Dec.1965. The target population is the children who were between the ages of 6 and 11 years and resided in the United States, excluding those confined to an institution or living in Indian reservation. NHES 3: March 1966-Mar.1979. The target is the development of the U.S. non-institutionalized children of 12-17 years inclusive. DESIGN: One PSU is drawn from each geographic region according to the probability proportional to its 1950 census population (pps). PSU was divided into geographically bounded segments, each containing an expected six households. From a listing of households within the segment, a random sampling procedure created a subsegment of approximately four households, each of which was interviewed. Some segments are drawn by systematic selection from the PSU, and the subsegments are randomized in sequence. The Households in the subsegment are randomly selected, and all persons in the reduced segment are eligible. Finally sample persons are systematically taken from eligible persons. NHANES 1: March 1971-March 1974. it assessed the nutritional status of persons 1-74 years, in addition to the other health examinations. NHANES 1a: 1974-1975 (14 months). To augment the size of the sample originally included in the sample of

NHANES 1, an additional sample is drawn from the 25-74 age group for detailed examinations. The target population is the civilian non-institutionalized U.S. population of 25-74 years of age (excluding Alaska and Hawaii, Indian reservations). NHANES 2: Feb. 1976-Feb. 1980. The target population is the civilian non-institutionalized U.S. population of 6 months-74 years. NHANES 3: 1988-1994. The target population is the civilian non-institutionalized U.S. population of 2 months and older. The similarity of these NHANES to NHESs is that both use multistage, stratified, probability sample of clusters of persons in land based segments. The successive elements in sampling process are PSU, segments, households, eligible persons, and finally sample persons. NHANES design was further complicated by the fact that unlike the NHES cycles it had two distinct examination components -nutrition and detailed - instead of only one. The age range of NHANES covers more than one specific age group, and emphasis is placed on the low income group, preschool children, women of child bearing age, and the elderly, because these are the most effected by malnutrition, and for which detailed information is most needed. Therefore, the design has to take into consideration the sample size requirements for the population subgroups to obtain reliable estimates. The first three surveys are done by pps design, while the last three are done with unequal probability in order to select more sample persons from the poor, the old, young age, or other groups. ESTIMATION and Variance are similar as NHIS. (NCHS Series 2, No.1971; Series 2, No.92, 1982).

#### **II.7 NATIONAL HISPANIC HEALTH AND NUTRITION EXAMINATION SURVEY.**

NHHANES is quite similar to that of the NHANES, involving a specific ethnic population instead. Specifically this population includes three Hispanic national origin groups of Mexican-Americans, Puerto-Ricans, and Cuban-Americans. (NCHS Series 1, No.19) DESIGN: The sampling design is also similar to that of the NHANES. Only difference is that NHHANES covers only three special groups in three selected areas of the United States rather than whole population. The target population is the civilian non-institutionalized eligible Hispanics of Mexican-Americans in the three regions, who are older than or equal to 6 months but less than 75 years of age. All households are included at least one member of Hispanic origin in the 1980 census. The age groups, 6 months to 19 years and 45 to 74 years, are oversampled.

**II.8 NHANES I EPIDEMIOLOGIC FOLLOW-UP** NHEFS is a longitudinal study that uses as its baseline those sample persons who were examined in the previous NHANES I. The first wave of data collection,

the 1982-84 NHEFS, include all persons 25-74 years of age at their NHANES I (n=14,407). The second wave, the 1986 NHEFS, is conducted for the members of the cohort who are 55-74 years of age at their baseline examination, and not known to be deceased at the time of 1982-1984 NHEFS (n=3,980). The third wave, the 1987 NHEFS, is to recontact the entire non-deceased NHEFS cohorts (n=11,750), who are contacted again in 1992.

#### **III. PROVIDERS' RECORD SURVEYS.**

This survey depends not only on the records but partly on the interviews to obtain some socio-demographic information; but the main body of the survey is done by sample records.

##### **III.1. NATIONAL HOSPITAL DISCHARGE SURVEY.**

NHDS has been done ever since 1965, (NCHS Series 2, No.39, 1970), and the same sampling design has been used except for minor changes and revision in 1988. 1985 NHDS DESIGN: The NHDS universe is the hospital discharges from noninstitutional hospitals, located in 50 States and the District of Columbia. Only hospitals with six beds or more for patient use and those in which the average length of stay for all patients is less than 30 days are included in the survey. Discharges of all patients from Federal hospitals are excluded. The universe in 1984 Master Facility Inventory of Hospitals (MFI) includes 6,023 hospitals of which 553 were selected for the sample of which 86 hospitals refused to participate and 60 were out of scope, either out of business or failed to meet the definition of a short-stay hospital. Remaining 407 hospitals participated in the survey in 1984 and provided about 192,000 abstracts of medical records. The plan follows a basic two-stage stratified sample, the first stage being hospital chosen from the NCHS MFI with stratification by geography, size, population concentration, type of hospital, and type of ownership. Large hospital with 1,000 beds or more were selected with certainty and the hospitals less than 1,000 beds were stratified into 24 size-by-region strata. Within each stratum, sample hospitals were drawn with probability ranging from one to 1/40. The second stage sampling is the systematic sampling of discharges from sample hospitals. The within hospital sampling ratio for selecting sample discharges varied inversely with the probability of hospital selection. The smallest sampling fraction of discharges was taken in the largest hospitals, and the largest fraction was taken from the smallest hospitals. This sampling was done to ensure that overall probability of selecting a discharge would be about the same in each size strata. NHDS ESTIMATION also follows the similar steps of NHIS.

(NCHS Series 13, No. 2, 1967; NCHS Series 13, No. 3, 1967). NHDS VARIANCE: The variance of certainty and noncertainty strata are separately calculated, and are added later. Each part used approximately exact formula of variance. BHS method or SUDAAN is recently used for variance estimation. (NCHS Shimizu, 1989).

### III.2 NATIONAL AMBULATORY MEDICAL CARE SURVEY.

The medical records are sampled to explore the provision and utilization of ambulatory care in the physician's offices. The survey has been conducted yearly since 1973 over the conterminous United States. The physicians whose specialty is anesthesiology, pathology, or radiology, or physicians in the Government services are also excluded. The NAMCS collected the data on the visits by 13 specialty, age-sex-race of patients, diagnostic and therapeutic services, prior visit, and disposition and duration of visits. NAMCS DESIGN: This sample survey design consists of three stage selections. First pps selection of PSUs out of 1924 PSUs. Second systematic selection of office based physicians from the master files maintained by the American Medical Association and the American Osteopathic Association. Third stage was the selection of patient visits within the annual practices of the sample physicians. First sampled physicians were randomly assigned to one of the 52 week periods in a survey year. A systematic sample of visits in each office of sampled physicians were selected each day during the week. The sample size depended on the clinic size. NAMCS ESTIMATION: The basic weights were obtained by the inverse of the three selection probabilities of individual sample visits for selecting the PSUs, physicians in the PSU, and office visits within the physician's practice. All weekly estimates were inflated by a factor of 52 weeks to derive annual estimates. (NCHS Series 2, No. 90; Series 13, No. 94; NCHS Advancedata No. 88, 77, 66, 60, 48, and 30). NAMCS VARIANCE: Sample variance was estimated by BHS with 48 balanced half-samples from 48 pseudo-strata (Shimizu, June 1, 1976).

### III.3. NATIONAL HOSPITAL AMBULATORY MEDICAL SURVEY.

In December, 1991, NCHS started NHAMS to gather information on the health care provided by hospital emergency and outpatient departments to the population of the United States, exclusive of Federal, Military, and Veterans Hospital. NHAMS used four stage probability design with sample PSUs, hospitals within PSUs, clinic within hospitals, patient visits within clinics. (NCHS Series 1, No. 34, 1994).

### III.4. NATIONAL NURSING HOME SURVEY.

NCHS conducted three nursing homes surveys to obtain their expenditures, the number of residents of nursing homes, functional status of residents, staff and other characteristics of nursing homes, and, in recent surveys, people discharged from nursing homes. 1973-1974 SURVEY: the universe included only those nursing homes that provided some level of nursing care. Thus, homes providing only personal or domiciliary care were excluded. 1977 NNH SURVEY: This survey included all types of nursing homes, including personal care and domiciliary care homes. 1985 NNH SURVEY: As in other nursing home surveys, the 1985 survey included all types of nursing homes, and produced estimates about the nursing and related care homes, their current residents, registered nurses for staff, and their current discharges, admission events in a 12 month period, and the location in a Metropolitan Statistical Area. NNHS DESIGN: The sampling design for the NNHS had been a stratified two stage probability design. Facilities were selected at the first stage, and current residents, discharge, and staff members were sampled at the second stage. In 1985 the staff sample was limited to the registered nurses. The overall probabilities of selecting residents would be same across facilities within each primary sampling stratum because the number of residents tends to be correlated to the number of beds and facilities were selected with probability proportional to the facility bed size. NNHS ESTIMATION: As it was a two-stage sampling of (1) facilities and (2) residents, discharges, and staff members, the two-stage weighing are used with other adjustments. The numbers of facilities were estimated by the reciprocals of probabilities of the facility selection, adjusted for non-responding facilities within bed size, certification status, and metropolitan status. The estimates for current resident, discharge, and RN characteristics based on the second stage samples would be produced by multiplying the adjusted facility sampling weight by the inverse of the probability of selection within the facility and by an adjustment for questionnaire non-response unique to the facility. NNHS VARIANCE: The sample variance was calculated by the BHS, SESUDAAN or Taylor Expansion or exact method.

### III.5. THE 1985 NATIONAL NURSING HOME SURVEY FOLLOWUP.

NNHSF was designed to collect information on the patterns of nursing home and hospital use, particularly among the elderly. The followups are built on the information collected in the 1985 NNHS. WAVE 1 was conducted from August to December, 1987, WAVE 2 from July to November, 1988, and WAVE 3 from

January to April 1990. The followup data include information on the characteristics and nursing staffs of a representative sample of long term care facilities, on the residents who were living in the facilities at the time of contact, and on those who had been discharged during the 12 months prior to the facility contact. The followup information on the facilities and residents was then compared to the information obtained in 1985. (Series 1, No. 30, 1993).

#### IV. REGISTRATION SURVEYS.

NCHS collects and publishes data on births, deaths, marriages, and divorces in the United States through the National Vital Statistics System. The division of Vital Statistics obtains these vital information from the registration offices of all 50 states, and the District of Columbia. The samples are taken from these registration records to study certain aspects of the population. If person is dead, a family member or acquaintance responded the questions on behalf of the deceased person.

##### IV.1 THE NATIONAL NATALITY AND FETAL MORTALITY SURVEYS (NNS/NFMS).

The focuses of the 1963 survey were on the exposure of married mothers to radiation during their pregnancy. The surveys in 1964-66 and 1967-69 sought information on mothers social and demographic characteristics, and various health and prenatal characteristics. The 1972 surveys obtained information on social, demographic, health, prenatal, labor, and delivery status from married mothers, hospitals, and attendants at delivery. The 1980 NNS was composed of information from birth certificates and questionnaire set to married mothers, hospitals, attendants at delivery, and providers of radiation examinations and treatments. This survey provides the information on specific maternal and child health conditions and obstetric practices for live births. The 1980 NFMS used the same questionnaire as the national natality survey. Some questions in the questionnaire asked about live births, while others asked about fetal deaths. This was the first time when fetal deaths were surveyed. Married mothers, hospitals, attendants, and providers of examinations and treatments were surveyed under the same conditions as those described in NNS so that NFMS permits the comparisons between live births in NNS and fetal deaths in the NFMS. NNS/NFMS DESIGN: The NNSs were conducted by NCHS based on live births in 1963, 1964-66, 1967-69, 1972, and 1980. All the NNS/NFMSs designs are similar. The 1980 NNS used the probability sample of 9,941 live birth certificates that occurred in the U.S.A. during 1980. The files of birth certificates in the 50 states, the district of Columbia, and the independent registration area of New

York City constituted the target areas. In each area, a sequential file number is assigned to each birth certificate received from the beginning to the end of each year. Files are randomly selected. Finally, it was necessary to exclude from the sample some additional certificates in the State of Washington and Idaho where it was required to obtain the permission of married mothers. NNS/NFMS ESTIMATION: The total of national live births was the weighted sum of the post-stratified ratios of the number of births to U.S. residents in 1980 to the number of sample births in NNS according to the 50 cells of birth weights, marital status, race, ages, and birth order. Similarly the weighing of NFMS used the weights of 22 cells of marital status, race of fetus, and age of mother. (NCHS Series 2, No. 100). NNS/NFMS VARIANCE: The variance was estimated by the BHS method from the 20 half sample replicates, based 20 strata.

##### IV.2. NATIONAL MATERNAL AND INFANT HEALTH SURVEY-1988.

The 1988 NMIHS surveys the maternal health and the three birth outcome live birth, fetal death, and infant mortality. The 1988 NMIHS will be instrumental in producing information on Low Birth-Weight, Infant Mortality, adolescent pregnancy, pregnant women from smoking, and pregnancy and infancy. The live births are based on the 10,000 certificates, which are equivalent of a national natality survey (NNS). The sample of live births is taken monthly from the 1988 certificates supplied by states. The American Indian Births are somewhat restricted see memo March 30, 1989 from Gonzalez). All certificates excluded from the sampling are those from mothers under 15 years of age, from nonresidents of U.S., and from the live births occurred before Jun 30, 1988; some certificates from the States of Michigan and Oklahoma are also excluded from the sampling. Sort each month's certificates of live birth by state, location within state, mother's marital status within location, mothers age within marital status, and certificate number within mother's age group. Within the state, a stratified systematic sample is taken as discussed before in NMFS. The fetal deaths are based on the 6,000 reports of fetal deaths which are equivalent of a national fetal mortality survey (NFMS). The infant mortalities are based on the 4,000 death certificates of infants which are equivalent of a national infant mortality survey (NIMS). Data collection instruments will be the improved versions of those used in the NNS/NFMS (Series 2. No. 100), discussed previously. Mailed questionnaires and telephone interviews with mothers, hospitals of deliveries, attendants-at-delivery, and certificates of death will be used to expand information about vital events. The

documentation of the estimation and variance are not available, but the procedures would be very similar as those used in NNS/NFMS.

#### IV.3. NATIONAL MORTALITY FOLLOW BACK SURVEY OF 1986.

The universe is all death certificates of 1986 in the United States. The sample of ten percent of deceased who were 25 years or over is systematically taken by the 49 states and District of Columbia except Oregon. The death certificates are assembled monthly by NCHS and the current mortality sample (CMS) is also taken monthly from these certificates. The death certificates are stratified into 18 strata. All the certificates falling into the 1st, 2nd, 3rd, or 4th stratum are taken; Those certificates are stratified into (1) American Indians, Aleut, and Eskimos, (2) specified heart disease, (3) specified Asthma, and (4) specified cancer. The rest of the strata are sampled. The certificates of blacks are stratified into one of the strata 5 through 11 according to the 7 age groups. The certificates of non-blacks also stratified into one of strata 12 through 18 according to the same age classes. Note that one certificate may belong to one or more strata. The death certificates are available to NCHS after three months after the date of death. No certificates of 1985 will be included in 1986 sampling. After sorting the deaths by the ascending order of strata, the deaths are ordered by the reported month within the stratum. Thirdly we order them by the states within the month. Within the state, we order them by the sex. Within sex, they are ordered by the reported cause of death. Lastly they are ordered by the reported age within the sex. Each sample unit is identified by two numbers, one sample ID number is assigned to a sample units continuously across the months, and the other is the number assigned within the stratum. Hence a sample unit is identified by the sample ID number and stratum ID number. For every stratum, one random starting number and another number indicating sampling intervals for the systematic sample are attached to each sample unit. NMFBS VARIANCE: The NMFBS sample variance is approximated by BHS (McCarthy, 1963). The states are divided into 20 pseudo- strata, and 20 replications are established by the 20 x 20 orthogonal matrix (Blackburn and Burman, 1954). Other procedures are the same as those discussed in NHIS section.

#### IV.4. DECENNIAL LIFE TABLES.

A ten percent sample of the monthly death reports to NCHS is used to construct life table or death rates rather than using entire reports. Ten percent sample records for each month are weighted to represent the total records during the twelve months. The sample is selected systematically. Variance of death rates based

on the sample has been proposed recently (Choi, 1992, 1993).

#### V EMPLOYER HEALTH INSURANCE SURVEY

The purpose of the employer based survey is to gather information on the number of employers offering health insurance, their costs, the coverage and characteristics of their health plans, and also employee-related estimates. Estimates are desired at both the state and federal levels for public and private sector employers. The public sector employers have been stratified by state and by level of government, while the private sector employers have been stratified by state and establishment size within firm size. Three sampling frames are used for NEHIS sampling: The frame used for the Federal, state, and local governments, for private sector establishments with employees, and for private sector self-employed with no-employees. The target population of three frames are mutually exclusive. All employees of the federal, state and the District of Columbia are included with certainty, excluding Puerto Rico, the Virgin Islands, or other territories. The frame of the public sector local governments is the 1987 Census of Governments (COG). to obtain measures of size for each government. The frame for the private sector establishments with employees is the Dune Market Identifiers (DMI) frame compiled by Dun and Bradstreet as of 1993, excluding all governmental entities in order to avoid duplication.

The frame for the selfemployed without employees is the last two quarters of the 1993 National Health Insurance Survey Supplement. The stratification of the establishments are made by the size of number of employees, small, interim, and large. The larger establishment is more likely selected than the smaller ones. WESTAT won the contract, and made a report to NCHS on May 5, 1994, which included the details of the NEHIS survey design.

#### REFERENCES

- Choi, J. W. and Casady, R. J. (1984). Variance in the National Health Interview Survey Data. ASA 1983 Proceeding in the Section on Survey Research Methods.
- Choi, J. W. (1986). A Variance when data arise from multistage sampling and ratio-estimation. ASA 1985 Proceeding in the Section on Survey Research Methods.
- Choi, J.W. and McHugh, R.B.(1989). Chi-square test of goodness and independence for clustered and weighted data. *Biometrics*. 45, pp979-996.
- Groves, R.M., Biemer, P.P., Lyberg, L.E., Massey J.T., Nicholls II, W.L., and Waksberg J. (1988). *Telephone Survey Methodology*. Edited. John Wiley and Sons.