THE VARIANCE OF HEALTH SURVEYS AT NATIONAL CENTER FOR HEALTH STATISTICS

Jai Won Choi, National Center for Health Statistics,
3700 East-West highway, Hyattsville, MD 20782

Introduction : In 1956, Congress passed the Health Acts for health surveys on the U. S. population. The health surveys include three distinct types: interview surveys, examination surveys, and record surveys. Any one type can not cover all the aspects of health events. Table 1 shows NCHS surveys. Other NCHS surveys - telephone sample survey, the master facility inventory survey (not sample survey), and the sampling of vital records to construct U. S. life table - are not included in this report. The oldest survey is the National Health Interview Surveys (NHIS) which was started in 1957. NHIS has been the model for all other surveys.

Table 1 Types of Surveys at NCHS

A. INTERVIEW
1. National Health interview survey (NHIS)
2. National Family Growth Survey (NFGS)
3. National Medical Care Utilization and Expenditure Survey (NMCUES)
B. EXAMINATION
4. National Health and Nutrition Examination Survey (NHANES)
5. National Hispanic Health and Nutrition Examination Survey (NHHANES)
C. RECORD
6. National Hospital Discharge Survey (NHDS)
7. National Ambulatory Medical Care Survey(NAMCS)
8. National Natality Survey (NNS) AND National Fetal Mortality Survey (NFMS)
9. National Nursing Home Survey (NNHS)
10 National Mortality Follow Back Survey (NMFS)
11.National Maternal and Infant Health Survey (NMIHS).

Most NCHS surveys are multistage designs; often the smapling plan includes selecting some sampling units with probability proportional to the size (pps). Such a sampling plan provides self-weighting samples from which we can obtain estimates with minimum variance among all others. However, sometimes certain subpopulations like blacks and poor people, were sampled more than others to allow for meaningful analysis of their data.

Table 2 Type of Variances uesd

Name of Survey	Design	Estimation Basic $ ratio	Variance
1. NHIS	3 stages	3 steps	BRR, SUDAAN
2. NFGS	5 stages	3-7 steps	BRR
3. NMCUES	5 stages	3 steps	BRR
4. NHANES	5 stages	2 steps	BRR, SUDAAN
6. NHDS	2 stages	7 steps	BRR
7. NAMCS	3 stages	2 steps	BRR
8. NNS/ NFMS	2 stages 2 stages	2 steps 2 steps	BRR BRR

The sample estimation of population parameters is usually done by basic weighting and ratio adjustment. Variances are usually estimated by BRR methods and presented by generalized variance function (GVF). Currently a modified SESUDAAN procedure has been used for theNHIS, replacing the BRR method. Six of the survey designs in Table 2 will be briefly described in the remainder of the paper.

1. NATIONAL HEALTH INTERVIEW SURVEY (NHIS):

Series A-2 (U.S. National Health Survey, 1958) and Technical report 7 (Bureau of the Census, 1963) and the technical report 40 (Census Bureau, 1977) describes the NHIS design from 1957 till 1984.

1.1 Sampling: The NCHS sampling design is periodically changed. New design has been used since 1985.

The questionaire consists of two parts, core questions that remaine the same and supplementary questions that are changed every year except the six chronic conditions. The core questions are on the housing, person, doctor visit, and acute and chronic conditions.

The target population is the U.S. noninstitutionalized civilian population. The NHIS survey plan is basically a three stage probability sampling of selecting primary sampling units (PSU), enumeration districts (ED), and segments. The ED and segment samplings are considered as one step and are systematically sampled. The persons in the segment are often subsampled for the supplements.

The new design was done by the Census Bureau (Moore, 1985). The U.S.A. was stratified into self- or nonself-representing strata according to clustering algorithm developed by Friedman and Rubin (1967). Each PSU is again stratified according to segment types. 52 PSUs from SR strata were automatically included in the sample, and two PSU's from each of 73 NSR strata were selected without replacement with pps, giving 146 PSUs. The three substrata within PSU i have different samplingrate of segments to permit more sampling of blacks. If PSU is not oversampled for blacks, the sampling rate is the same.

Table 3. Old (1957-84) and New Design (1985-88)

	OLD DESIGN	NEW DESIGN
frame	Address Area New Construction	Area New Construction
PSUs	Multicounty 1970 SMSAs	Multicounty 1980 MSAs
SR	156	52
NSR	220	146 (73 stratum x 2)
	376	198
PSU/strt	1	2
HUs/year	51,000	61,400
Sampling interval	1700 - 1900	1638
Segment size	4 households	old const. 8 hsehlds new const. 4 hsehlds
Panels	Not defined	4 (1 or more used)

A major change of new design is the introduction of four independent panels of approximately equal size, each representing the U.S. population. In case of a budget cut, one or

more panels may be dropped without damaging the sampling efficiency. Another point is the sampling of two PSUs from NSR strata, which makes it possible to derive variance between PSUs. The Table 3 highlights some of the changes made for the new design.

1.2 Estimation: As the sampling is basically pps design until 1985, in theory the probability of selection is more or less same for all sample persons (Kendall and Stuart, vol3, 1977, p 199).

1.2.1 Basic Weighting: The indexes l, i, h, j, and k are for

Strata $l = 1,\ldots, L$,
PSUs $i = 1,\ldots, n_l$, $(n_l = 2)$,
substrata $h = 1, 2, 3$. (used after 1985)
segments $j = 1,\ldots, m_{li}$,
and persons $k = 1,\ldots, D_{lij}$.

Denote $w_{lijk} = M_{li}/(\pi_{li}\, m_{li})$ for all k.
The basic population total with characteristic x before 1985 is given by

$$X' = \sum_{l}^{L} \sum_{i}^{2} \sum_{j}^{m_{li}} \sum_{k}^{D_{lij}} w_{lijk}\, x_{lijk} \qquad (1.1)$$

where $x_{lijk} = 1$ if the (lijk)-th person in the sample has x characteristic, and = 0 otherwise. With differential sampling rate in the PSU after 1985, X' is

$$X' = \sum_{l}^{L} \sum_{i}^{2} \sum_{h}^{3} \sum_{j}^{m_{lih}} \sum_{k}^{D_{lihj}} w_{lihjk}\, x_{lihjk} \qquad (1.2)$$

where $w_{lihjk} = 1/(\pi_{lih}\, f_{lih})$, and similarly for x_{lihjk}.

1.2.2 Poststratified Ratio Estimation: The first stage adjustment is for the 48 cells of color-residence, indexed by c = 1, ..., C (C = 48), done only for the NSR PSUs. The second stage ratio adjustment is for 60 cells of age-sex-race categories, indexed by a = i, ..., A (A = 60). The age-sex-race ratio adjustment is

$$X' = \sum_{a}^{A} \frac{x'_a}{y'_a}\, \hat{Y}_a \qquad (1.3)$$

where x'_a and y'_a are so defined as the numerator and denominator of next equation (1.4) under the summation sign. The color-residency ratio adjustment done as

$$X' = \sum_{a}^{A} \frac{x_{aSR} + \sum_{c=1}^{C} \frac{x_{acNSR}}{z_c}\, Z_{cCEN}}{y_{aSR} + \sum_{c=1}^{C} \frac{y_{acNSR}}{z_c}\, Z_{cCEN}}\, \hat{Y}_a, \qquad (1.4)$$

where $x_{aSR} = \sum_{l}^{L} \sum_{h}^{3} \sum_{j}^{m_l} \sum_{k}^{D_{ij}} w_{ljk}\, x_{aljk}$. (1.5)

$x_{alhjk} = 1$ if the (lhjk)-th sample person belongs to the a-th age-sex-race cell who has x characteristic, and $x_{alhjk} = 0$ otherwise, i=1.

$$y_{aSR} = \sum_{l}^{L} \sum_{h}^{3} \sum_{j}^{m_l} \sum_{k}^{D_{lj}} w_{lhjk}\, y_{alhjk}. \qquad (1.6)$$

$y_{alhjk} = 1$ if the (lhjk)-th sample person belongs to the a-th age-sex-race cell and = 0, otherwise.
$\hat{Y}_a$ = independent population total (CPS) of the age-sex-race category.
Z_{cCEN} = the 1970 census population in the c-th collaspsed race-residence category.

Similarly, x_{acNSR} and y_{acNSR} can be written for the NSR-PSUs, and

$$z_c = \sum_{l}^{L} \sum_{i}^{2} \sum_{h}^{3} \sum_{j}^{m_{li}} \sum_{k}^{D_{lij}} w_{lihjk}\, z_{c\,lihjk}, \qquad (1.7)$$

where $z_{clihjk} = 1$ if the (lihjk)-th person in all sample PSUs belongs to the c-th color-residence category, and = 0 otherwise for all l, i, h, j, and k. The ratio adjustments remained the same after the new design took place in 1985.

1.3 $\hat{var}(X')$: After 1973, the combination of Balanced Repeated Replication (BRR) and Generalized Variance Function (GVF) are used in order to present relative standard errors.

Since 1985, the same GVF has also been used to present sample variances, but the sample variances are estimated differently. The ratio estimate X' is linearized , and then the usual formula for two stage sampling variance formula is applied to these linearized variables instead. The variances so derived are now used to estimate the parameters, a and b of GVF curve. The GVF is also used, but instead of the curve, theestimator of parameters a and b are given so that one can easily calculate the relative standard error of x by $(a + b/x)^{1/2}$.

1.3.1 Variance by BRR: 149 pseudo-strata are used to generater 152 balanced half sample replications in order to estimate the variance. NHIS classified estimates into about 75 domians. Each domain uses about 100 points of varying sizes to produce the GVF curve of each domain.

1.3.2 Variance for two stage sampling: The sampling is done only in the i-th and j-th levels, the l-th level indicates the strata, the h-th level indicates different sampling rate, and the k-th level indicates the units in the segement.

The variance is now obtained by two steps. The first step is to find the variance formula of two stage sampling. The second step is to find the linear form of two stage ratio estimation.

We are interested in the variance of the basic estimate, which is the form of Brewer or Durbin estimator, and it was extended to two stage sampling as shown in Cochran (1977, p308, 11.41)

Since strata are independent, first we find the variance of each stratum and sum the variances for all the strata.

$$t = \sum_{l}^{L} \sum_{i}^{2} \sum_{h}^{3} \sum_{j}^{m_{lih}} \sum_{k}^{D_{lihj}} \frac{1}{\pi_{li}}\, \frac{M_{lih}}{m_{lih}}\, t_{lihjk}$$

$$= \sum_{l}^{L} \sum_{i}^{2} \sum_{h}^{3} \sum_{j}^{m_{lih}} \frac{1}{\pi_{li}} \frac{M_{lih}}{m_{lih}} t_{lihj+}$$

$$= \sum_{l}^{L} \sum_{i}^{2} \frac{1}{\pi_{li}} \sum_{h}^{3} M_{lih} \bar{t}_{lih++}$$

$$= \sum_{l}^{L} \sum_{i=1}^{2} \frac{\bar{t}_{li}}{\pi_{li}} \qquad (1.8)$$

where $t_{lihjk} = 1$ if the (ihjk) person has the t characteristic, and = 0 otherwise. The variance and its estimator are shown in (11.42) and (11.44) (Cochran, 1977), respectively, when two PSUs are sampled in the ith level.

When units were selected with unequal probabilities with replacement. Unbiased estimator of the population total of pps sample, its variance, and the unbiased estimator of the variance are also shown in Cochran (1977, 11.31, 11.32, and 11.35).

$$\hat{var}(t) = \sum_{l=1}^{L} \frac{(\pi_{11}\pi_{12} - \pi_{112})}{\pi_{112}} \left(\frac{t_{11}}{\pi_{11}} - \frac{t_{12})^2}{\pi_{12}}\right.$$

$$+ \sum_{l=1}^{L} \sum_{i=1}^{2} \frac{1}{\pi_{li}} \sum_{h}^{3} M_{lih}^2 (1-f_{lih}) \frac{s_{lih}^2}{m_{lih}} \qquad (1.9)$$

where s_{lih}^2 is the usual sample estimator of S_{lih}^2. π_{11}, π_{12} and π_{112} is also known (Cochran, 1977, p 308, 11.44).

1.3.3 Linearization of X': (Woodruff, 1971).

The first step is the linearization of $(x'_a / y'_a)\, \hat{Y}_a$ as

$$X' = \sum_{a}^{A} \gamma_a \left(x'_a - \frac{X_a}{Y_a} y'_a \right), \qquad (1.10)$$

where $\gamma_a = \hat{Y}_a / Y_a$, $X_a = E(x'_a)$, $Y_a = E(y'_a)$.

Replacing x'_a and y'_a, the numerator and denominator of (1.4), (1.10) can be written as

$$X' = \sum_{a}^{A} \gamma_a \left[\left(x_{aSR} + \sum_{c}^{C} \frac{x_{acNSR}}{z_c} Z_{cCEN}\right) - \frac{X_a}{Y_a} \left(y_{aSR} + \sum_{c}^{C} \frac{y_{acNSR}}{z_c} Z_{cCEN}\right) \right]. \qquad (1.11)$$

Second linearization under the summation over c is

$$X' = \sum_{a}^{A} \gamma_a \left(x_{aSR} - \frac{X_a}{Y_a} y_{aSR}\right) + \sum_{a}^{A} \gamma_a \sum_{c}^{C} \frac{Z_{cEN}}{\tilde{Z}_c} \left[x_{acNSR} - \frac{\tilde{X}_{acNSR}}{\tilde{Z}_c} z_c - \frac{X_a}{Y_a} \left(y_{acNSR} - \frac{\tilde{Y}_{acNSR}}{\tilde{Z}_c} z_c\right)\right] \qquad (1.12)$$

where $\tilde{Z}_c = E(z_c)$, $\tilde{Y}_{acNSR} = E(y_{acNSR})$, $\tilde{X}_a = E(x_{acNSR})$.

Replacing the five variables x_{aSR}, y_{aSR}, and z_c with (1.4), (1.5), and (1.6), and y_{acNSR} and x_{acNSR} with similar definitions, and exchanging the summations signs of the levels l, i, h, j, and k with the summation signs of a and c, the equation (1.12) can be expressed as

$$X' = \sum_{l}^{L} \sum_{h}^{3} \sum_{j}^{m_{li}} \sum_{k}^{D_{lij}} \left[\sum_{a}^{A} \gamma_a \left(w_{lhjk} x_{alhjk} - \frac{X_a}{Y_a} w_{lhjk} y_{alhjk}\right)\right] \qquad (1.13)$$

$$+ \sum_{l}^{L} \sum_{i}^{2} \sum_{h}^{3} \sum_{j}^{m_{lih}} \sum_{k}^{D_{lihj}} \sum_{a}^{A} \gamma_a \sum_{c}^{C} \frac{Z_{cCEN}}{\tilde{Z}_c} \left[w_{lihjk} x_{aclihjk} - \frac{\tilde{X}_{acNSR}}{\tilde{Z}_c} w_{lihjk} z_{c\,lihjk} - \frac{X_a}{Y_a} \left(w_{lihjk} y_{aclihjk} - \frac{\tilde{Y}_{acNSR}}{\tilde{Z}_c} w_{lihjk} z_{clihjk}\right)\right]$$

Summing over a and c, we obtain

$$X' = \sum_{l}^{L} \sum_{h}^{3} \sum_{j}^{m_l} \sum_{k}^{D_{lj}} t_{lhjk+} + \sum_{l}^{L} \sum_{i}^{2} \sum_{h}^{3} \sum_{j}^{m_{li}} \sum_{k}^{D_{lij}} t'_{lihjk++} \qquad (1.14)$$

where $t'_{lihjk++} = \sum_{a}^{A} \gamma_a \sum_{c}^{C} \frac{Z_{cCEN}}{\tilde{Z}_c} \left[w_{lihjk} x_{aclihjk} - \frac{\tilde{X}_{acNSR}}{\tilde{Z}_c} w_{lihjk} z_{c\,lihjk} - \frac{X_a}{Y_a} \left(w_{lihjk} y_{aclihjk} - \frac{\tilde{Y}_{acNSR}}{\tilde{Z}_c} w_{lihjk} z_{clihjk}\right)\right]$

for the NSR-PSUs and

$$t_{lhjk+} = \sum_{a}^{A} \gamma_a \left(w_{lhjk} x_{alhjk} + \frac{X_a}{Y_a} w_{lhjk} y_{alhjk}\right)$$

for self-representing PSUs.

Since the variables $t'_{lihjk++}$ of NSR-PSUs in the equation (1.14), are the weighted variables, which included the weights $M_{lih} / (\pi_{li}\, m_{lih})$, the variance formula (1.9) can be used by replacing X'_{li}/π_{li} with t'_{li++++} in the first term, and $M_{lih}^2 / (\pi_{li}\, m_{lih})$ with $\pi_{li}\, m_{lih}$ in the second term.

Similarly the t_{lhik+} of SR-PSUs of the first term in (1.14), as the sampling is done only in the j-th level in the SR-PSUs, the second term of the variance formula (1.9) can be adjusted, dropping

the first term.

New software is developed Research Triangle Institute for the two stage sampling when simple random sampling is used within each stage sampling They also show how to use this formula for the pps sampling case.

Thus, the total variance is estimated by

$$\hat{var}(X') = s^2_{SR} + s^2_{NSR}. \qquad (1.15)$$

The use of this method instead of BRR method may reflect the actual survey design better.

2. NATIONAL FAMILY GROWTH SURVEY (NFGS)

There have been four cycles of NFGS survey since NFGS was established in 1972. The first cycle of survey was conducted in 1973, the second cycle in 1976, the third in 1982, and the fourth in 1987.

In the cycle 1, the sample plan is based on a five-stage selection of 203 PSUs, 2 replicate groups of PSUs out of 4, EDs, segments, dwelling units, and one eligible person.

The resulting data are stratified into the 12 age-race classes, and ratio-adjusted. Nonresponses were also ratio-adjusted corresponding to each stage. The weights of the inverse of selection probability in each stage were also mutiplied to get the weights for each individual.

Since it may be impossible to find exact formula of the variance of an estimate from this sample, the balanced half sample replication is used to find variances under certain assumptions.

The 103 FGS PSUs were grouped into 48 strata, seven of self-representing and 41 nonself-representing. The self-representing PSU's were randomly divided into two pseudo PSUs, and the nonself-representing PSUs are paired. Taking one member of the pair, a 48 x 48 replication matrix was formed. Following the usual procedures, the variance of an estimate was obtained. The details are prevously given in NHIS section.

In the cycle 2, the sampling is a five stage selection of PSUs, EDs, segments, households, and finally women of 15-44 years. Each sample person is inflated by the inverse of the selection probabilities in each of the five stages. The basic weight was adjusted by multiplying nonresponses at the segment level, and for the 12 age-race ratios only of ever married women. The 79 cycle 2 PSUs were regrouped into 37 super-strata, 18 self-representing strata and 19 nonself-representing strata. The 18 self-representing strata were randomly divided into two pseudo-strata, and two PSUs were selected from each of nonself-representing superstrata. Taking one of the pair, the 37 strata gives 40 replications for the application of BRR method.

In the cycle 3, the sample design is similar to that of the cycle 2. It also uses a five stage probability sampling that incorporated oversamples of black and teenage women, but the cycle 3 takes a supplementary sample of women living in college dormitories and sororities.

Cycle 4 is entirely different from the previous designs, and linked to the NHIS sample persons of the previous years. It uses the 4th quarter sample of 1985 design (3 panels system), from which the national sample of black women 15-44 years was obtained, all 4 quarters of 1986 (3 panels system), from which the national sample of women 15-44 years was taken; the 1st quarter of 1987 (4 panels system), from which the national sample of black women 15-44 years was used. At most one woman was sampled from each household. The sample size was 10,694 women of which about 1/3 were black. The estimation, variance, and other details are not yet available.

3. NATIONAL MEDICAL CARE UTILIZATION AND EXPENDITURE SURVEY (NMCUES)

3.1 Sampling: NMCUES was designed to collect data about the U.S. civilian noninstitutionalized population during 1980. During the 1980, information was obtained on health, access to and use of medical services, associated charges and sources of payment, and health insurence coverage.

Two independent nationwide samples were pooled together. One was the sample taken by Research Triangle Institute (RTI), and the other by National Opinion Research Center (NORC). Both are characterized as stratified, five-stage area probability designs. The differences between two designs are mainly the type of stratification variables and the specific definitions of sampling units.

3.2 Estimation: The basic method used in NHIS estimation was applied to the NMCUES sample data. The details arepresented in other publication.

3.3 Variance: The balanced half sample replication was used to find the variance of the sample estimates.

4 NATIONAL HEALTH AND NUTRITION EXAMINATION SURVEY

There have been six surveys since 1959. The first three of them are the health examination surveys, the two of them (NHANES I and II) are the health and nutrition examination surveys, and one of them is the detailed examination survey (NHANES IA). These are:

1. National Health Examination Survey (HES I). Nov., 1959 - Dec., 1962. It focussed on the prevalence of selected chronic conditions.
2. National Health Examination Survey (HES II). July, 1963 - December, 1965. It was devoted to the growth and development of children 6 - 11 years.
3. National Health Examination Survey (HES III). March, 1966 - March, 1979. It was devoted to the growth and development of children 12 - 17 years.
4. National Health and Nutrition Examination Survey (HANES I). March, 1971 - March, 1974. It assessed the nutritional status of persons 1 - 74 years, and other health examinations.
5. National Health and Nutrition Examination Survey (HANES IA). 1974 - 1975 (14 months). Persons 25 - 74 years. To augment the size of the sample originally included in the sample of NHANES I, an additional sample was drawn from the same age group for detailed examinations.
6. National Health and Nutrition Examination Survey (HANES II). 2/1976 - 2/1980. Nutrition status and Examinations of persons of 6 months - 74 years.

4.1 Sampling: All these surveys are using the similar plan of multistage probability sampling design. Minor modifications of the self-weighting features were introduced to handle particular situations. The last three surveys took more sample from the poor, and young and older age groups. The probability of inclusion of every sample person was known over each stage of

sampling.

Table 4 Six Surveys

	Strata	PSU	ED per PSU	SEGMT per ED	HU per SEG	AGE OF PERSON
HES I	42	42	-	60	6	18-79
HES II	40	40	10	2	6	6-11
HES III	40	40	20	1	6	12-17
HANES I	15(sr)	15	*	sys. sple of SEGs	6	1-74 # (28k)
	25(ns)	50			8	
	$ 20% detailed Examination					25-74 (6k)
$ HANES IA	&10(sr) 25(ns)	10	**	syst. smple	8	@ 25-74#
HANES II	64(ss)	64	***	1	8	6m-74# (28k)

sample rate of segments from EDs is * for 1/8, ** for 1/2, and *** for 2.3.
different sampling rate by age-sex groups
@ one of every two eligible persons sampled.
$ Combined for augment survey.
& 5 of largest SR-superstrata, 10 of other strata grouped into five superstrata of two and select one of two.

4.2 Estimation: The subscripts i, j, k, t, and l are
for the PSU (i = 1, ..., n),
for ED (j = 1,..., J_i),
for the segment (k = 1,..., K_{ij})
for poor (t = 1) and for nonpoor (t = 2),
for the person (l = 1)

Define $1/\pi_{ijktl} = w_{ijktl}$, and $Y_{aijktl} = 1$ if the aijktl-person included in the sample and = 0 otherwise. π_{ijktl} is the probability of the ijktl-th sample unit being in the sample.

If this person belongs to the a-th age-cell, and the poststratified ratio of this cell is

$$Y_a / \hat{Y}_a$$

Y_a is the census cell counts of a-th age-sex and

$$\hat{Y}_a = \sum_i\sum_j \sum_k \sum_t \sum_l w_{ijktl} Y_{aijktl},$$

The final weight of this person is adjusted as

$$w_{ijktl} Y_a / \hat{Y}_a.$$

The final estimator of the population total is

$$Y = \sum_i\sum_j \sum_k \sum_t \sum_l \sum_a w_{ijktl} Y_a / \hat{Y}_a.$$

Actually w_{aijktl} is the inverses of the selection probabilities. Note that, if it is a self-representing PSU, w_{ijktl} do not include the inverse of the probability of the first stage selection for $\pi_i = 1$.

Define $x_{ijktl} = 1$ if the ijktl-th sample person has x characteristic, and = 0 otherwise.

The final estimator of the population total with x characteristic is

$$X' = \sum_i\sum_j \sum_k \sum_t \sum_l \sum_a w_{ijktl} x_{ijktla} Y_a / \hat{Y}_a$$

4.3 Variance: Table 5 shows the method used for variance estimations.

Table 5 Variance Estimation for 6 Surveys.

HES I	BRR (19 pseudo-strata x 20 repl.)
HES II	BRR (same)
HES III	BRR (same)
HANES I	BRR (40 pseudo-strata x 40 repl.)
HANES IA	
1-65 stds	BRR (40 pseudo-strata x 40 repl.)
66-100stds	BRR (19 pseudo-strata x 20 repl.)
1-100 stds	TSR (35 pseudo-strata x 81 repl.)
HANES II	BRR (32 pseudp-strata x 40 repl.)

All the surveys used the BRR except the two cases, that is, third sample replications (Bryant, 1975), using 100 stands (or strata) in HANES IA and SESUDAN in HANES II. The formula (1.9) presented in Section 1 can be adapted to this situation.

6 NATIONAL HOSPITAL DISCHARGE SURVEYS

The plan is the two-stage stratified sample design. The first stage being hospital chosen from the NCHS Master Facility Inventory (NCHS, 1971), with stratification by geography, size, population concentration, type of hospital, and type of ownership. Large hospital should have a higher probability of being in the sample than small hospitals. The second stage sampling is the systematic sampling of discharges from sample hospitals. If possible, the overall design should be nearly self-weighting.

The estimation of patient statistics is carried out through seven stages of adjustment, three of the adjustments bring sample observations to hospital levels, three others bring the hospital levels up to the primary stratum level, and then a ratio estimation takes the primary stratum level to the final estimates.

The statistics produced by the NHDS pass through two phases: first, monthly estimates based on data collected from all responding hospitals are computed, and, secondly, these estimates are summed over the months to produce the published statistics. The variance of certainty and noncertainty strata are separatly calculated, and are added later. Each part was approximately exact formula of variance. Later the BRR method or SESUDDAN (Shah, 1981) was used for variance estimation. A cutomized program was proposed for use after 1988.

7. NATIONAL AMBULATORY MEDICAL CARE SURVEY (NAMCS)

Target population is all the visits to the non-federal office based physicians offices of certain practices. Sample survey design consists of three stage selections. First pps selection of 87

psu's out of 1924 PSU's from 3141 counties, county equivalents and independent cities of the United States, Second systematic selection of office based physicians from the list of MD's from the selected PSU's such that we obtain a self-weighting sample of 4,681 physicians. They were randomly assigned to one of the 52 week periods. Third systematic selection of only ten visits each day, giving about 30 patient records during the week in each office sampled physicians.

The basic weights were obtained by the inverse of the selection probability of individual sample units. These basic weights were further adjusted by multistage post-stratified ratios to estimate the total of the target population.

The variables most frequently used in the report were general-family-practitioner, internal medicine, pediatrics, general surgical, ob-gyn, psychiatry on one side of two way table, and age, sex, and race on the other side in the publication. Sample variance was estimated by BRR method with 48 balanced half-samples from 48 pseudo-strata.

8. THE NATIONAL NATAILTY SURVEY (NNS) AND THE 1980 NATIONAL FETAL MORTALITY SURVEY (NFMS)

The NNSs were conducted by NCHS based on live births in 1963, 1964-66, 1967-69, 1972, and 1980.

The 1980 NNS is based on 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. The certificates were registered for an estimated 99.3 percent of all live births.

In each area, a sequential file number is assigned to each birth certificate received from the beginning to the end of each year. 105 of every 1,000 certificates were randomly selected starting from the most recent month completed. Next 25 of every 105 certificates were randomly selected. From the remaining 80, only those less than 2,500 grams were taken as sample. Finaly, it was necessary to exclude from the sample some additional certificates in the State of Washington where it was required to obtain the permission of married mothers , and Idaho where it was needed to get the consent of unmarried mothers.

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, and ages.

The variance was estimated by the BRR method from the 20 half sample repicates, based 20 strata. The relative standard error was equated to the square root of $(A + B/x)$, where A and B are the weighted least square estimators.

REFERENCES

Bryant, E. (1976). Memorandum from Bryant to Director, Division of Health Examination Statistics. April 12, 1976.

Cochran, W.G. (1977). Sampling Techniques, 3rd ed. John Wiley and Sons.

Cypress B.K.(1983). Patterns of Ambulatory Care in General and Family Practice: The National Ambulatory Medical Care Survey, NCHS Series 13, No. 73

Hollis G.G.(1971). Design and Methodology of the 1967 Master Facility Inventory Survey, NCHS Series 1, No. 9

Kendall, M.G. and Stuart, A.(1968). The Advance Theory of Statistics, Vol 3. Hafner Publishing Co., New York.

Keppel K.G. et al (1986). Method and Response Characteristics, 1980 National Natality and Fetal Mortality Surveys. NCHS Series 2, No. 100.

Moore, T. (1985). Statistical Methods Division Methodological Memorandum Series. Redesign of the National Health Interview Survey. U.S. Bureau of the Census, Unpublised Document, Report No.:CB/SMD/MM/85/02.

Shah, B. B. (1981). SESUDAAN: Standard Error Program for Computing of Standardized Rates from Sample Survey Dtata, Research Triangle Institute.

Simmons W. R. and Schnack G. (1970). Development of the Design of George A. Hospital Discharge Survey. NCHS Series 2, No. 39.

U.S. National Health Survey (1958). Health Statistics - the Statistical Design of the HEALTH HOUSEHOLD-INTERVIEW SURVEY, Series A2, U.S. Department of Health Education and Welfare.

U.S. Bureau of the Census (1963). The Current Population Survey - A Report on Methodology by Joseph Steinberg, Thomas B. Jabine, and Leon Pritzker.

U.S. Bureau of the Census (1978). The Current Population Survey - Design and Methodology. Technical paper 40. Washington D.C. Government Printing

Friedman, H. P. and Rubin, J (1967). "On some variance criteria for grouping data." JASA, 62, pp. 1159-78.

Woodruff, R.S. (1971). "A Simple Method for approximating the variance of a complicated estimate. JASA 66, 411-14.

Acknowledgement

Author thanks to Dr Lester R. Curtin whose comments made this paper more readable.