Vital and Health Statistics

Covariances for Estimated Totals When Comparing Between Years

Series 2: Data Evaluation and Methods Research No. 114

This report presents the results of a study to measure the degree of correlation between yearly estimates of totals for selected characteristics from the National Hospital Discharge Survey and from the National Health Interview Survey. The report also gives an idea of the magnitude of the covariances of estimated totals and the effect of assuming independence when making year-to-year comparisons of totals.

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
Centers for Disease Control
National Center for Health Statistics

Hyattsville, Maryland
April 1992
DHHS Publication No. (PHS) 92-1388

Suggested Citation

Bean JA, Hoffman KL. Covariances for estimated totals when comparing between years. National Center for Health Statistics. Vital Health Stat 2(114). 1992.

Library of Congress Cataloging-in-Publication Data

Bean, Judy A.
Covariances for estimated totals when comparing between years.
p. cm. – (Vital and health statistics. Series 2, Data evaluation and methods research ; no. 114) (DHHS publication no (PHS) 92-1388)
"Results of a study to measure the degree of correlation between yearly estimates of totals for selected characteristics from the National Hospital Discharge Survey and the National Health Interview Survey."
By Judy A. Bean and Keith L. Hoffman.
Includes bibliographical references.
ISBN 084.6.4548
1. Medical Care – United States – Utilization – Statistics. 2. National Hospital Discharge Survey (U.S.) – Evaluation. 3. National Health Interview Survey (U.S.) – Evaluation. 4. United States – Statistics, Medical. 5. Analysis of covariance. I. Hoffman, Keith L. II. National Center for Health Statistics (U.S.) III. National Hospital Discharge Survey (U.S.) IV. National Health Interview Survey (U.S.) V. Title. VI. Series. VII. Series: DHHS publication no (PHS) 92-1388. VIII. Series: DHHS publications ; no. (PHS) 92-1388.
[DNLM: 1. Analysis of Variance. 2. Health Status – United States – statistics. 3. Health Surveys – United States. 4. Patient Discharge – United States – statistics. W2 A N148vb no. 114] RA409.U45 no. 114
[RA 407.3]
362.1'0723 s – dc20
[362.1'0973'021]
DNLM/DLC
for Library of Congress 91-36074
CIP

Contents

List of text tables

Covariances for Estimated Totals When Comparing Between Years

by Judy A. Bean, Ph.D., Department of Epidemiology and Public Health, University of Miami, and Keith L. Hoffman, M.S., Office of Research and Methodology

Introduction

The National Center for Health Statistics (NCHS) conducts several large-scale surveys whereby data are obtained through complex designs including sampling in two or three stages, clustering, and stratification. If the design is multistage the first stage consists of drawing a sample of primary sampling units (PSU's). These PSU's can be hospitals, counties, or other units, depending upon the specific survey. In several surveys sponsored by NCHS the primary sampling units employed are often sampled on more than one occasion.

Two examples are the National Hospital Discharge Survey (NHDS) and the National Health Interview Survey (NHIS). In NHDS, the design used from 1965 to 1987 was a two-stage highly stratified one. The PSU's selected in the first stage were hospitals, and within the chosen hospitals, discharges were sampled. Although some hospitals have closed, new ones have been added, and a few do not participate, the data collected annually from 1965 to 1987 came from the same hospitals originally selected. NHIS is an annual survey of households located in the 50 States and the District of Columbia. Each year interviewers contact approximately 41,000 new households; however, these households are selected from the same primary sampling units, which are groups of counties and metropolitan areas. From one year to the next the households are chosen from adjacent segments of housing units within PSU's.

When sampling inquiries are conducted at successive intervals of time on a continuing basis and covering the same universe, such as in NHDS and NHIS, three different types of parameters for totals may be estimated. These estimators are the change in a total from one time period to the next, the average total for all the time periods, and the average total for each time period. The best design is not the same for these estimators and depends upon the correlation of the estimators from one time period to another. This report focuses only on the effect of the correlation on estimated changes between time periods.

This procedure of gathering data in a survey from the same basic units from one year to the next means that the estimates produced each year are not independent. This lack of independence is an important feature that, for a variety of reasons, is often ignored by analysts.

Statistics produced from NHDS and NHIS are published annually in reports prepared by NCHS. In the appendix to reports presenting the NHIS statistics, the reader can find relative variances curves from which variances can be calculated. If the reader wishes to compute the variance of a difference between two yearly estimates, a formula is provided. This variance formula is the sum of the variances of the two estimates. In using this formula one is assuming that the covariance between the two estimates is negligible. Similarly, the variances of NHDS estimates are given in reports, but nothing is explicitly stated about the covariances of estimated totals when comparing between years. This report is intended to give NCHS users an idea of the magnitude of the covariances of estimated totals and the effect of assuming independence when making year-to-year comparisons of totals.

To determine the degree of correlation between the estimates of totals, an empirical investigation was conducted using data from two NCHS multistage, complex surveys, NHDS and NHIS.

The main objective of this investigation was to measure the degree of correlation between yearly estimates of totals for selected characteristics produced from the National Hospital Discharge Survey and the National Health Interview Survey. To achieve this overall objective the specific aims were to:

- Determine a method for applying the balanced repeated replication (BRR) technique to the NHDS design.
- Estimate variances and covariances of totals for selected variables from NHDS and NHIS using the BRR method.
- Compute the variances of the year-to-year differences for totals using the BRR technique.
- Calculate the year-to-year correlations between the statistics using the BRR variances and covariances.

This report shows the results of the study. Specifically, correlations of estimated totals for selected variables and a measure of the effect of independence are examined.

Description of NHDS and NHIS

In order to understand the methods and results of this study it is crucial that a person have an understanding of the sample designs for the two surveys. The two designs are briefly described here.

National Hospital Discharge Survey

This survey is a source of data on the short-stay hospital experience of the civilian noninstitutionalized population. The data are obtained through a probability sample of all discharges, both alive and deceased, from short-stay non-Federal hospitals. Estimates produced from the survey include the number, rate, and average length of stay of patients discharged from short-stay hospitals by selected patient and hospital characteristics.

The NHDS sample plan for 1965–87 was a two-stage stratified design. Each hospital with 1,000 or more beds in the universe of short-stay hospitals was considered to be a stratum; the other hospitals, called noncertainty hospitals, were stratified by bed size and geographic region. The noncertainty hospitals were classified into 24 size-by-region primary strata. If a stratum contains only one hospital, that hospital came into the sample with a probability of one. Controlled-selection techniques were employed to choose hospitals within each of the 24 size-by-region classes (1). Primarily because of cost, the sample hospitals were allocated to a series of subsamples, referred to as panels, and the subsamples were implemented in the field over time.

The second stage of selection consisted of sampling discharges from within each of the hospitals. The daily listing of discharges was the sampling frame. The within-hospital sampling ratio for choosing the discharges varied inversely with the probability of selection of the hospital.

In 1988 the sampling plan changed from a two-stage design to a three-stage design. The first stage is geographic area, with hospitals selected from sampled areas in the second stage. Discharges are then sampled from the selected hospitals. A more indepth description is given by Shimizu (2).

After the abstracted data are extensively edited, an estimating procedure is used to produce national estimates from the survey. The procedure has three components: Inflation by reciprocals of the probabilities of sample selection, adjustment for nonresponse, and ratio adjustment.

National Health Interview Survey

This survey is designed to produce national estimates for the civilian noninstitutionalized population residing in the United States. Members of the Armed Forces are not included in the target population. Each year data are collected on personal and demographic characteristics, illnesses, injuries, impairments, chronic conditions, and other health topics. Estimates produced from this survey include morbidity rates and measures of utilization of health services.

The sampling design produces a weekly national probability sample of households that are representative of the target population. The weekly sampling permits continuous measurement of characteristics and, through consolidation of data over time, detailed analysis of rare health-related items.

The first stage in the design used from 1973 to 1984 consisted of drawing a sample of 376 primary sampling units (PSU's) from approximately 1,900 geographically designed PSU's. (A PSU is either a county or a group of contiguous counties, metropolitan areas, or minor civil divisions.) Then the housing units within each sample PSU were geographically clustered into segments, and each year a cluster of four expected households was selected. Prior to 1985, the annual NHIS sample consisted of approximately 41,000 eligible occupied households, which yielded a sample of about 106,000 individuals (3).

For several reasons a redesign of NHIS was implemented in 1985. Besides reducing the number of PSU's from 376 to 198 and increasing the number of interviewed households to 49,000, four other significant changes were made. Instead of using a list sampling frame, an area sampling frame was employed. The number of PSU's selected in certain strata was changed from one to two. Furthermore, national subsamples of PSU's were created. The fourth change was to oversample the black population (4).

In order to produce estimates the basic data are extensively manipulated. The four steps are: Inflation by the reciprocals of the probability of selection, household nonresponse adjustment, first-stage ratio adjustment, and poststratification by age, sex, and race. Statistical theory indicates that a ratio estimator for a statistic generally has smaller variance than an inflation estimator has. The first-stage ratio adjustment and poststratification are carried out for this reason. These adjustments also bring the sample data into close conformity with U.S. Bureau of the Census population totals.

Study design

Data from NHDS

Data collected on sampled discharges of inpatients from short-stay hospitals during the years 1982, 1983, and 1984 were used in the study. The original frame of the survey consisted of the hospitals listed in the 1963 National Master Facility Inventory. To reflect the universe of short-stay hospitals for a given year, hospitals that have come into existence since 1965 were also sampled. These discharges represent three years of data collection in NHDS, which has been conducted since 1965. Data used in the study were on diagnostic and surgical procedures for 1982, 1983, and 1984. There was one record for each discharge. The variables chosen for this investigation were a subset of the variables NCHS employed for computing variances for NHDS statistics.

Data from NHIS

As described previously, NHIS uses an annual probability sample of households throughout the 50 States and the District of Columbia. Data collected each year are available on five different types of record formats. In this study, tapes for the years 1982, 1983, and 1984 containing a record per individual or a record per condition were used. The 60 statistics examined in this study were used in generating the relative variance curves formed in NCHS Current Estimates Series 10 reports.

Variance estimator

The variances of NHDS sample statistics are calculated directly using the estimating equations discussed by Simmons and Schnack (1). In NHIS the balanced repeated replication (BRR) procedure for general variance estimation is employed. Further discussion of the variances calculated for the survey estimates is presented in an appendix to NCHS Current Estimates Series 10 reports, which give NHIS estimates, and in Series 13 reports on NHDS estimates. For ease of computation the BRR method was used to estimate variances and covariances in this study for both data sets.

The basic premise of BRR is that the variability of a statistic based on the entire sample data set can be estimated by the variability of that statistic based on subsamples that reproduce the complex design of the total sample. Several investigators—McCarthy (5,6), Kish and Frankel (7,8), Frankel (9), Koch and Lemeshow (10), and Bean (11)—have studied this procedure extensively. The basic technique is described below.

Consider a finite population of N primary units classified into L strata, each containing N_h units ($h = 1, 2, \ldots, L$), with

$$N = \sum_{h=1}^{L} N_h$$

The sample design is a stratified simple random sample with two units chosen from each stratum with replacement. An unbiased estimator of the population mean $\bar{Y}$ is

$$\bar{y} = \sum_{h=1}^{L} W_h \bar{y}_h$$

where

W_h = weight for stratum h ($h = 1, 2, \ldots, L$)
$= N_h/N$
$\bar{y}_h$ = the sample mean for stratum h

The usual estimator of the variance of $\bar{y}$ is

$$V(\bar{y}) = \sum_{h=1}^{L} W_h^2 s_h^2/2$$

where s_h^2 is the sample variance for stratum h.

One general method for estimating variances is random groups, but for this situation there are only two independent random groups. An alternative approach is to use half-samples comprised of one unit from each stratum. This results in half-samples that have overlapping units, which means that they are correlated. This estimator can be expressed as

$$V_B(\bar{y}) = \sum_{i=1}^{K} (\bar{y}_i - \bar{y})^2/K$$

where

$$K = 2^L$$

$$\bar{y}_i = \sum_{h=1}^{L} W_h (\delta_{h1i}\, y_{h1} + \delta_{h2i}\, y_{h2})$$

$$\delta_{h1i} = \begin{cases} 1 \text{ if unit } (h,1) \text{ is selected for the } i\text{th half-sample,} \\ 0 \text{ otherwise} \end{cases}$$

$$\delta_{h2i} = 1 - \delta_{h1i}$$

There are 2^L possible half-samples. For this case of a linear statistic, $V_B(\bar{y}) = V(\bar{y})$. When L is large the computation of 2^L half-samples is not feasible. McCarthy (5) derived a method for choosing a subset of the half-samples in a specific fashion so that $V_B(\bar{y})$ is reproduced. The set of replicates generated is said to be orthogonally balanced.

The subsets of replicates are formed by deleting one unit in each stratum. The set of A half-samples is defined by an $A \times L$ matrix with elements a_{ih}, where

$$a_{ih} = \begin{cases} +1 \text{ if unit 1 in the } h\text{th stratum is in the } i\text{th half-sample where } i = 1, 2, \ldots, A, \\ -1 \text{ if unit 2 in the } h\text{th stratum is in the } i\text{th half-sample.} \end{cases}$$

This set of A replicates satisfies the property

$$\sum_{i=1}^{A} a_{ih}\, a_{ih'} = 0$$

for all $h < h' = 1, \ldots, L$. The BRR variance estimator is

$$V_r(\bar{y}) = \sum_{i=1}^{A} (\bar{y}_i - \bar{y})^2/A$$

where

$\bar{y}_i$ = the estimate of $\bar{y}$ utilizing half of the total sample

The approach of balanced repeated replications has been described for a linear statistic; however, as indicated earlier, the method is utilized for nonlinear statistics produced in NHIS. By analogy with the linear problem, an estimator of the variance of $\hat{\theta}$, a nonlinear statistic, based on A balanced replicates is

$$V_r(\hat{\theta}) = \sum_{i=1}^{A} (\hat{\theta}_i - \hat{\theta})^2/A$$

Krewski and Rao (12) show that both linear and nonlinear statistics are asymptotically normal and that the BRR variance estimators are consistent under certain conditions.

If more than two units are selected from a stratum but n_h is an even integer, investigators suggested forming pairs within the stratum randomly (8,13). However, research indicates that as the number of units within a stratum increases, the BRR variance estimator becomes increasingly unstable (14). This finding played a role in determining the methodology used in this study.

Besides being useful for estimating the variances of ratio and nonlinear estimators, BRR is applicable in the area of inferential statistics. Several studies (15–19) have demonstrated that the BRR approach could be used to estimate covariances as well as variances.

Logistics of the investigation

The various steps taken in conducting this investigation are as follows.

Step 1: Creating a working file

For the NHDS data the working file consisted of one record per hospital from the more than 500,000 discharges for the study. A computer program using the statistical package SAS (20) was written to select the variables needed from each record; then records within a hospital were combined. For a diagnosis the output consisted of panel number (denoting the particular subsample the hospital was in), hospital number, bed size, region, and 19 variables. For each surgical tape, the output consisted of 23 surgical variables, along with the panel number, hospital number, bed size, and region. When the statistic was the number of diagnoses or surgical discharges for selected patient characteristics, the value for a hospital equaled the sum of the final weights across all the individual records having the specified diagnosis-related group (DRG) code and the selected characteristic. To compute a hospital's total days of care for patients by selected characteristics, the number of days of care was multiplied by the final weight and then accumulated across all the records having the specific DRG code and characteristic. (Discharges for newborn infants were excluded from the diagnostic, days-of-care, and surgical statistics.) A total of 61 statistics were produced for NHDS.

This same general procedure was used to prepare the data from NHIS. For this survey the working file consisted of one record per primary sampling unit for a specific year. The statistical package SAS was utilized to select the variables needed from each record and then to combine the records within a PSU. The output consisted of PSU number, demographic variables, and the variables of interest. The value of a variable for a PSU equaled the sum of the final weights across all individual records having the selected characteristics and the specific variable. A total of 60 statistics were produced for NHIS.

Step 2: Certainty strata

Because 17 strata in NHDS consist of only one hospital, these hospitals enter the sample with a probability of one. Each of these hospitals was treated as a pseudostratum by randomly dividing the records into two groups. (These hospitals are in panel 0.) Using the random number function available in SAS, the program used to create the hospital records was modified to accomplish this step.

NCHS has a procedure for NHIS whereby PSU's consisting of an entire stratum are allocated to a pseudostratum. Accordingly, these pseudostrata were employed.

Step 3: Nonresponse problem

One program in SAS was written to perform all the remaining tasks, but these various tasks will be explained in steps. In NHDS not all of the sampled noncertainty hospitals in the size-by-region classes responded for all three years—1982, 1983, and 1984. The hospitals not responding for all three years were dropped from the data file used in estimating the variances and covariances. Because the number of hospitals not responding for all three years was small—making up less than 5 percent of the total—this nonresponse is not thought to invalidate the results. NHIS does not have this type of nonresponse, because a household is included in only one year of sampling.

Step 4: Noncertainty strata

As discussed earlier in the section "Variance estimator," research indicates that having two units per stratum is the most efficient design in terms of BRR variance estimator stability. However, if the NHDS noncertainty hospitals were simply paired, the number of pseudostrata would be more than 200, which means that an extremely large orthogonal matrix would be required for variance estimation. To avoid the problem of constructing a large orthogonal matrix, a pseudostratum was defined to be four hospitals. The pseudostrata formed were independent of the sampled hospitals.

After the original sample of hospitals was selected, other short-stay hospitals came into existence. The NHDS survey design was altered so that a sample of these noncertainty hospitals was chosen; these sample units are designated as panel B. The original noncertainty hospitals sampled belong to panels 1–6. The hospitals in panels 1–6 were placed into one group and collapsed into strata. However, because weighting is different within panel B, the hospitals within it were collapsed to form strata.

Again a function of the statistical package SAS was used to form the strata and pairs within them. The hospitals within panel B were sorted by region, bed size, and hospital. This was done so that hospitals with similar characteristics, especially size, were paired together. Each group of four was considered a stratum. Because there were only three hospitals for the last stratum, the records for the final hospital were duplicated. This procedure was also followed for the hospitals in panels 1–6. However, the last pseudostratum had only two hospitals, which meant that the records for two hospitals were duplicated.

Prior research into the behavior of BRR estimators indicates that if the number of units sampled from a stratum is even, pairs within the stratum should be formed randomly (8,13). Because each NHDS pseudostratum contained four noncertainty hospitals, random pairs were formed.

As stated above, the pairing used by NCHS to produce BRR variances for NHIS statistics was employed in this study. Because there were noncertainty PSU's, pseudostrata were formed.

The creation of pseudostrata raises questions concerning the bias introduced in the estimates of variances through this procedure. Stanek (21) investigated the effect on variance estimates of pairing strata into pseudostrata. The findings indicate that, as the pairing became heterogeneous, the bias of the BRR variance estimator increased. In addition, subjective pairing was studied using data from the National Health Examination Survey (another NCHS survey). The results show that the BRR estimates of variance of mean weights and heights of children were not highly dependent on the arrangement of pseudostrata.

For NHDS noncertainty hospitals, the pairing for this investigation was done by combining adjacent geographic-size classes. The pairing utilized in NHIS is subjective. Although no estimates of bias were made, the variance estimates for NHDS estimates should be only slightly overestimated.

Step 5: Estimation of variances, correlation coefficients, and effect of independence assumption

The next step was to generate an estimate of the statistic utilizing the entire sample. SAS data statements were employed to write a program to construct a 120 × 111 orthogonal matrix consisting of 120 half-samples and 111 pseudostrata for NHDS. For NHIS, a 160 × 149 matrix was constructed.

The final component was calculation of the half-samples of the statistic. There were 120 half-sample estimates in NHDS and 160 in NHIS. This step was accomplished by taking advantage of SAS features. One example is the use of a pointer to read both of the records in a stratum at one time. This process also allows the location in the data file where the program is reading to be maintained.

The theory of BRR indicates that data from a half-sample should be subjected to the same estimation procedure as data from the total sample are. This would mean, for example, recomputing the first-stage ratio adjustment and poststratification for each half-sample for NHIS data—a task that would require considerable work. Investigations by Simmons and Baird (22) and Kish and Frankel (7,8) found that the adjustment factors based on the parent sample can be applied without seriously biasing the estimate. Other studies (11,23) indicate that these adjustment factors should be calculated for each specific half-sample. For this study, the decision was made to use the factors based on all the survey data.

After the half-sample estimates were produced, the variance was calculated. This procedure was repeated for each of the 38 diagnostic and days-of-care statistics and each of the 23 surgical statistics from NHDS and for each of the 60 statistics from NHIS.

The basic program was modified to yield covariances between the estimates across years. The BRR formula for the covariances computed is

$$CV_r(Y_{82}Y_{83}) = \sum_{i=1}^{A} (y'_{82,i} - y''_{82})(y'_{83,i} - y''_{83})/A$$

where

A = the number of half-sample estimates produced
$y'_{82,i}$ = the estimate of Y_{82} utilizing half of the total sample
y''_{82} = the estimate of Y_{82} based on the entire sample
$y'_{83,i}$ and y''_{83} are defined similarly

In addition to the variances and covariances, the variance estimates of the year-to-year differences were directly computed. For example, the estimator for the years 1982 and 1983 is

$$V_r(Y_{82} - Y_{83}) = \sum_{i=1}^{A} [(y'_{82,i} - y'_{83,i}) - (y''_{82} - y''_{83})]^2/A$$

All the output from the programs was passed to a SAS data file so that additional statistics could be computed.

Two other statistics were also calculated.

- *Correlation coefficients for the estimates*—Correlation coefficients were calculated for the estimates of 1982 and 1983 and for the estimates of 1983 and 1984. The formula is:

$$\hat{\rho} = CV_r(Y_{82}Y_{83})/\sqrt{V_r(Y_{82})\,V_r(Y_{83})}$$

- *Effect of independence assumption*—To determine the effect that making the assumption of no correlation has on the variance of the difference between two variables, the ratio

$$E = [V_r(Y_1) + V_r(Y_2)]/V_r(Y_1 - Y_2)$$

was computed for all the statistics.

Verification of the NHDS variance estimator

Table 1 gives the number of patients discharged, standard error, and standard error of the differences for 1982, 1983, and 1984 from NHDS. (The variables in the tables are shown in the order of magnitude of the number of patients with the characteristic discharged in 1982.) Examination of the relative standard errors (RSE's) for several characteristics indicates that the BRR estimates are reasonable. (The standard error divided by the estimate = RSE.) For example, consider these relative standard errors:

Females:
RSE = (635/21,554) × 100 = 2.9 percent
Persons 15–44 years:
RSE = (520/14,515) × 100 = 3.6 percent
Atherosclerotic heart disease:
RSE = (25/465) × 100 = 5.4 percent
Malignant neoplasm of lung, male:
RSE = (11/188) × 100 = 5.9 percent

As the size of the statistic decreases, the RSE increases.

Charts providing general relative standard errors for a wide variety of estimates are available in the publications concerning the utilization of short-stay hospitals—for example, (24). Comparing the RSE's produced by the BRR method with the values of RSE's from the figures in the publications shows that the BRR estimates are larger. The ratios of the BRR estimate of RSE divided by the value derived from the figures are: Females—1.16; persons 15–44 years—1.44; atherosclerotic heart disease—1.35; and malignant neoplasm of lung, male—1.18. Note that the denominators were obtained from the figures by rounding the estimates. Because empirical evidence (8,9,11) shows that in general the BRR method overestimates the variance, this is the result one would expect. The standard errors are consistent across the years.

Associated with each diagnosis is a days-of-care estimate. These values are examined next. The estimates for days of care are presented in table 2. Most of the patterns found in table 1 are seen again in table 2. The BRR relative standard errors are usually larger than the values obtained from the chart of general RSE's for NHDS estimates shown in NCHS publications.

The estimates for the surgical procedures and their standard errors are presented in table 3. The size of the estimates varies from 33,519,000 to 55,000 for 1982 data. The estimates for many variables increase from 1982 to 1983 but decrease in 1984. The RSE's for four variables in 1982 are:

All procedures:
Total:
RSE = (1,186/33,519) × 100 = 3.5 percent
Persons 15–44 years:
RSE = (647/14,281) × 100 = 4.5 percent
Bilateral occlusion of fallopian tube, 15–44 years:
RSE = (54/558) × 100 = 9.7 percent
Hysterectomy, 65 years and over:
RSE = (4/55) × 100 = 7.9 percent

To compare these RSE's with the ones obtained from the published NCHS charts, the ratio of the BRR RSE divided by the published RSE was calculated. The ratios are: All procedures, total—1.09; all procedures, persons 15–44 years—1.12; bilateral occlusion of fallopian tube, 15–44 years—1.61; and hysterectomy, 65 years and over—0.88. Again the BRR estimates are generally larger, but in some cases the two RSE's are essentially the same.

These comparisons indicate that, in general, the BRR estimate of variance is larger than the estimate produced by the design variance formula. Based on BRR research findings, this overestimation is to be expected. Although no one has compared BRR estimates of covariance with exact covariances, it is reasonable (using the results from variance comparisons) to assume that covariances are also overestimated. However, assuming that the magnitude of the overestimation is the same for variances and covariances, the correlation coefficients calculated will be approximately the same as those calculated using the correct design formulas.

Results

NHDS diagnoses

The correlation coefficients in table A for the number of discharges for the 1982 and 1983 statistics range from 0.96 for two statistics—the number of discharges for mental disorders for persons ages 15–44 years and the number of discharges for persons under 15 years of age—to a low of 0.45 for the number of discharges for malignant neoplasm of the lung among males. (The variables in the tables are shown in the order of the magnitude of the estimates in 1982.) Of the 19 variables, only 7 have values below the value of 0.80 and none has a negative value. The average value of the correlations is 0.80. No clear pattern related to the magnitude of the estimate was observed. If the statistic has an estimate of more than 800,000, the correlation is always larger than 0.88. The other statistics have smaller correlations, with one exception: The variable congenital anomaly for persons under 15 years of age, which has the smallest estimate, has a correlation of 0.95 for 1982–83. The three malignant neoplasm diagnoses and inguinal hernia for males have the smallest correlations, ranging from 0.45 to 0.58.

For the correlations for 1983 and 1984 (table A), the values vary from a high of 0.98 to a low of 0.41, with an average of 0.81. Again the statistics with estimates larger than 800,000 have the higher correlations; they are all bigger than 0.93. The rest of the variables—with the one exception, congenital anomaly for persons under 15 years of age—range from 0.41 to 0.83. The estimated number of discharges for the age group under 15 years has a correlation of 0.98, whereas malignant neoplasm of the lung among males has the value of 0.41. Although 11 of the 19 variables have larger correlation coefficients for 1983–84 than for 1982–83, generally the correlations do not differ greatly from one year to the next.

Table A also shows the ratios (E = independence effect) showing the effect on the variances when the assumption is made that the correlation between estimated totals across years is zero. (The formula for E is presented earlier in the section "Logistics of the investigation.") For 1982 and 1983 the ratios range from a minimum of 1.8 to a maximum of 22.4. For example, for persons ages 15–44 years with mental disorders, the variance of the difference (assuming the correlation is zero) is 22.4 times larger than the variance taking into consideration the correlation. The values for 1983 and 1984 range from 1.7 to 40.4. The mean value is 9.6 for 1982–83 and 13.3 for 1983–84.

These statistics indicate that making the assumption that the correlation is zero results in the variance of the difference for diagnoses always being an overestimate.

Table A. Correlation coefficients and effect of independence assumption on variances for number of patients discharged from short-stay hospitals, by selected characteristics: United States, 1982–84

Characteristic	Correlation coefficient		Independence effect (E)	
	1982–83	1983–84	1982–83	1983–84
Female	0.91	0.96	10.8	28.2
15–44 years	0.92	0.96	12.2	22.4
Male	0.91	0.95	11.2	19.9
65 years and over	0.91	0.95	10.2	18.5
45–64 years	0.88	0.95	8.6	18.9
Females with deliveries	0.95	0.96	18.7	24.3
Under 15 years	0.96	0.98	19.1	40.4
Mental disorders	0.95	0.94	19.0	16.7
Mental disorders, 15–44 years	0.96	0.93	22.4	14.5
Acute myocardial infarction	0.72	0.75	3.5	4.0
Cataract	0.89	0.77	8.4	3.7
Atherosclerotic heart disease	0.79	0.83	4.7	5.1
Inguinal hernia, male	0.58	0.57	2.4	2.3
Cataract, 65 years and over	0.88	0.76	7.3	3.6
Malignant neoplasm of lung	0.49	0.61	1.9	2.5
Malignant neoplasm of breast, female	0.52	0.49	2.1	2.0
Malignant neoplasm of lung, male	0.45	0.41	1.8	1.7
Fracture of neck of femur, 65 years and over	0.67	0.60	2.9	2.5
Congenital anomaly, under 15 years	0.95	0.96	15.5	21.4

SOURCE: National Center for Health Statistics. National Hospital Discharge Survey, 1982–84.

The price of the assumption is a variance that is at least double the true one. Using the overestimated variance will have an impact on trends of the year-to-year differences and on hypothesis testing done on these statistics. Examination of the ratios reveals that for all statistics with an estimate larger than 800,000, the effect is huge, ranging from 8.6 to 22.4 for 1982–83 and from 16.7 to 40.4 for 1983–84.

NHDS days of care

The correlation coefficients for NHDS days-of-care statistics are displayed in table B. The total number of days of care among persons under 15 years has the largest correlation, 0.94, for 1982–83 and again, 0.96, for 1983–84. The number of days of care for malignant neoplasm of the lung among males has the lowest correlation for 1982–83, 0.34, and again for 1983–84, 0.29. All the correlations were positive. The ranges of the correlation coefficients are 0.94 to 0.34 for 1982–83 and 0.96 to 0.29 for 1983–84. The means are 0.74 and 0.73, respectively. The pattern by magnitude of the estimate is similar to that seen for number of discharges. For statistics with an estimate larger than 10 million, the correlations are above 0.84. Most of the other statistics have correlations ranging from 0.29 to 0.82. Exceptions are a correlation of 0.85 for cataract procedures for 1982–83 and a correlation of 0.93 for congenital anomaly among children under 15 years for 1982–83.

Examination of tables A and B reveals that the correlations are always smaller for days of care than for number of discharges. One possible explanation is the introduction of the concept of diagnosis-related groups (DRG's) in hospitals. DRG's were put into use during the years 1982 and 1983. The use of DRG's would affect the number of days of care more than the number of discharges for these selected 19 variables.

Because the correlations are all positive and most are greater than 0.8, making the assumption of independence results in relatively large overestimates of the variance of the difference. The ratio for days-of-care statistics reaches a high of 26.3 for persons under 15 years for 1983–84 and a low of 1.4 for malignant neoplasm of the lung among males for 1983–84. The value of 26.3 means that when the assumption of independence is made, the variance used is 26 times larger than a more accurate estimate of variance. The averages across the statistics are 6.4 for 1982–83 and 8.4 for 1983–84. In general, as the correlation coefficient approaches one, the effect of the independence assumption increases.

NHDS surgical procedures

The impression obtained from studying the correlation coefficients in table C is that these surgical statistics are as highly correlated from year to year as the number of discharge statistics are. For 1982–83, the correlation coefficients vary from a high of 0.93 for females ages 15–44 with bilateral occlusion of fallopian tube to a low of 0.43 for females ages 65 and over with a hysterectomy. The correlation coefficients for 1983–84 range from 0.96 for all procedures among females to 0.20 for females ages 65 and over with a hysterectomy. The average values are 0.80 for both time periods. In general, the statistics with larger estimates have larger correlations. The statistics with the three smallest estimates have the smallest correlations, ranging from 0.20 for 1983–84 to 0.55 for 1982–83.

Table C gives the effect of the independence assumption (E) for the surgical procedures. The ratios range from 1.8 to 14.2 in 1982–83 and from 1.3 to 23.4 in 1983–84. The statistic for hysterectomy among females 65 years and over has the smallest ratio for both sets of values. The average ratio is 6.7 for 1982–83 and 10.1 for 1983–84. The variance assuming the correlation is zero is two times larger than the true variance when the correlation is about 0.43.

Table B. Correlation coefficients and effect of independence assumption on variances for number of days of care for patients discharged from short-stay hospitals, by selected characteristics: United States, 1982–84

Characteristic	Correlation coefficient		Independence effect (E)	
	1982–83	1983–84	1982–83	1983–84
Female	0.85	0.93	6.6	13.0
Male	0.88	0.91	8.3	11.0
65 years and over	0.87	0.91	7.7	11.1
15–44 years	0.81	0.92	8.2	12.0
45–64 years	0.84	0.89	6.3	8.8
Mental disorders	0.92	0.93	12.1	14.4
Under 15 years	0.94	0.96	12.4	26.3
Females with deliveries	0.90	0.95	9.8	18.7
Mental disorders, 15–44 years	0.92	0.93	11.3	14.5
Acute myocardial infarction	0 53	0.63	2.2	2.7
Atherosclerotic heart disease	0.69	0.79	3.1	4.1
Fracture of neck of femur, 65 years and over	0.53	0.41	2.1	1.7
Malignant neoplasm of lung	0.44	0.46	1.8	1.9
Malignant neoplasm of breast, female	0.47	0.32	1.9	1.5
Malignant neoplasm of lung, male	0.34	0.29	1.5	1.4
Inguinal hernia, male	0.40	0.32	1.6	1.5
Cataract	0.85	0.74	6.6	3.1
Cataract, 65 years and over	0.82	0.70	5.5	2.8
Congenital anomaly, under 15 years	0.93	0.89	13.1	9.1

SOURCE: National Center for Health Statistics. National Hospital Discharge Survey, 1982–84.

Table C. Correlation coefficients and effect of independence assumption on variances for number of surgical procedures for patients discharged from short-stay hospitals, by selected characteristics: United States, 1982–84

Characteristic	Correlation coefficient		Independence effect (E)	
	1982–83	1983–84	1982–83	1983–84
All procedures:				
Total	0.89	0.95	8.4	20.6
Female	0.90	0.96	9.2	23.4
15–44 years	0.91	0.95	11.6	18.0
Male	0.86	0.92	7.1	12.7
65 years and over	0.86	0.93	5.5	14.5
45–64 years	0.87	0.93	7.3	13.6
Under 15 years	0.93	0.94	12.3	15.9
Circumcision, under 15 years	0.85	0.91	6.5	10.5
Cesarean section:				
All ages	0.91	0.94	8.7	15.2
15–44 years	0.91	0.94	8.6	15.2
Hysterectomy	0.88	0.92	8.5	13.1
Bilateral occlusion of fallopian tube, 15–44 years	0.93	0.93	14.2	13.4
Extraction of lens, 65 years and over	0.86	0.79	6.8	4.5
Cardiac catheterization	0.90	0.92	8.9	10.7
Prostatectomy	0.68	0.71	3.2	3.4
Tonsillectomy, under 15 years	0.74	0.75	3.8	3.9
Prostatectomy, 65 years and over	0.64	0.67	2.8	3.0
Hysterectomy, 45–64 years	0.63	0.65	3.0	2.7
Direct heart revascularization	0.84	0.85	6.4	6.6
Myringotomy, under 15 years	0.82	0.83	4.6	5.5
Arthroplasty and replacement of hip:				
65 years and over	0.53	0.47	2.0	1.8
Female	0.55	0.43	2.1	1.7
Hysterectomy, 65 years and over	0.43	0.20	1.8	1.3

SOURCE: National Center for Health Statistics. National Hospital Discharge Survey, 1982–84.

NHIS acute and chronic conditions

Table D shows the correlations and the effect of the independence assumptions on the variances for 1982–83 and 1983–84. (The statistics, standard errors, and standard errors of differences are presented in table 4.) The correlation coefficients for 1982 and 1983 range from a value of –0.12 for acute conditions among females ages 6–16 years to a value of 0.20 for chronic conditions among black persons ages 35–44 years. The correlations are mostly in the range of the value of zero; of the 20 variables, 6 are negative. The mean is 0.02, which is not different from zero. The correlations for the three statistics with an estimate greater than 100,000 are 0.08 for total number of acute conditions in the population, 0.02 for total number of chronic conditions in the population, and –0.09 for the number of chronic conditions among females. Unlike the

Table D. Correlation coefficients and effect of independence assumption on variances for number of acute and chronic conditions, by selected characteristics: United States, 1982–84

Characteristic	Correlation coefficient		Independence effect (E)	
	1982–83	1983–84	1982–83	1983–84
Acute conditions				
Total	0.08	0.12	1.1	1.1
White, 25–34 years	–0.11	0.07	0.9	1.1
Female, 6–16 years	–0.12	0.09	0.9	1.1
Male, 6–16 years	0.14	–0.15	1.0	0.9
White, 55–64 years	0.01	–0.07	1.0	0.9
Married, 45–54 years	0.03	–0.08	1.0	0.9
Female, 55–64 years	0.00	0.05	1.0	1.0
75 years and over	0.08	0.13	1.1	1.1
Separated, 17–24 years	–0.03	0.00	1.0	1.0
Black, 65–74 years	0.10	–0.01	1.1	1.0
Chronic conditions				
Total	0.02	0.35	1.0	1.5
Female	–0.09	0.39	0.9	1.6
25–34 years	–0.10	0.20	0.9	1.2
Married, 25–34 years	0.08	0.21	1.1	1.3
White, 45–54 years	0.02	–0.14	1.0	0.9
Male, 6–16 years	0.12	–0.01	1.1	1.0
Female, 45–54 years	0.04	–0.08	1.0	0.9
Income less than $5,000, 17–24 years	–0.03	0.01	1.0	1.0
Black, 35–44 years	0.20	0.11	1.2	1.1
Income less than $5,000, 65–74 years	0.06	–0.03	1.1	1.0

SOURCE: National Center for Health Statistics. National Health Interview Survey, 1982–84.

statistics for NHDS, the statistics with the largest estimates do not have the largest correlations, nor do the statistics with the smallest estimates have the smallest correlations.

The values for 1983 and 1984 also are in the vicinity of zero, with the smallest value being –0.15 for acute conditions among males 6–16 years; the largest correlation is 0.39 for chronic conditions among females. For 1983–84 the average correlation is 0.06. The correlations for the two years do not show extreme variation with two exceptions. For example, the correlation for acute conditions for the total population is 0.08 for 1982–83 and is 0.12 for 1983–84.

Because the correlation coefficients are all in the neighborhood of the value of zero, the effect of the independence assumption is not great. The ratio measuring the independence effect for 1982–83 varies from a low of 0.9 for acute conditions among females ages 6–16 to a high of 1.2 for chronic conditions among black persons ages 35–44. For 1983–84 the ratio goes from a high of 1.6 for chronic conditions among females to a low of 0.9 for acute conditions among males ages 6–16 and three other statistics. The average ratio is 1.0 for 1982–83 and 1.1 for 1983–84. The conclusion one would make from these ratios is that when one compares estimates between years by assuming that the statistics are independent, the variance estimate is not an overestimation.

NHIS physician visits and short-stay hospital episodes

The statistics, standard errors, and standard errors of differences for the three years are displayed in table 5. Examination of the correlations for 1982–83 displayed in table E shows that they vary from a low of –0.16 for short-stay hospital episodes among divorced individuals to a maximum value of 0.39 for short-stay hospital episodes among white persons. In this set of variables, no very clear pattern is seen between the size of the estimate and the size of the correlation. Two of the 20 variables have a negative correlation. The average value of the correlations for these 20 variables for 1982–83 is 0.15.

Of the 20 correlations for 1983–84, 7 are negative; the negative values range from –0.01 for physician visits among black persons and physician visits among divorced persons to –0.33 for short-stay hospital episodes among black persons ages 45–54. The range for the positive values is from a low of 0.06 for physician visits among males to a high of 0.36 for short-stay hospital episodes among white persons. The mean is 0.10 for 1983–84; therefore, the averages of the correlations for number of physician visits and short-stay hospital episodes are bigger than the means for the number of acute and chronic conditions. Although the correlations for these statistics are slightly larger, they are still relatively low correlation coefficients.

The last two columns in table E give the effect of independence assumption for the year-to-year differences. Only 2 of the 20 values for 1982–83 and 7 of the 20 values for 1983–84 are less than one. As an example of the effect of assuming independence, consider the correlations for the statistic short-stay hospital episodes among white persons—0.39 for 1982–83 and 0.36 for 1983–84. Although the correlations are not large, the effect on the independence assumption approach is a ratio of slightly more than 1½ for each time period. Thus, by making the assumption that the variables are independent, the variance is 1½ times larger than the true variance. The average ratios are 1.2 for 1982–83 and 1.2 for 1983–84.

Table E. Correlation coefficients and effect of independence assumption on variances for number of physician visits and short-stay hospital episodes, by selected characteristics: United States, 1982–84

Characteristic	Correlation coefficient		Independence effect (E)	
	1982–83	*1983–84*	*1982–83*	*1983–84*
Physician visits				
Total	0.30	0.28	1.4	1.4
Male	0.10	0.06	1.1	1.1
Black	0.14	–0.01	1.2	1.0
White, 17–24 years	0.13	0.34	1.1	1.5
White, 55–64 years	0.18	0.29	1.2	1.4
Under 6 years	0.12	0.09	1.1	1.1
Married, 35–44 years	0.11	–0.08	1.1	0.9
Divorced	0.02	–0.01	1.0	1.0
Female, 65–74 years	0.05	–0.09	1.1	0.9
Female, 75 years and over	0.37	0.11	1.6	1.1
Short-stay hospital episodes				
Total	0.27	0.29	1.4	1.4
White	0.39	0.36	1.6	1.6
Female	0.33	0.10	1.5	1.1
Male	0.04	0.15	1.0	1.2
75 years and over	0.21	0.24	1.3	1.3
Married, 55–64 years	0.17	–0.12	1.2	0.9
6–16 years	–0.01	0.09	1.0	1.1
Divorced	–0.16	–0.07	0.9	0.9
Male, 17–24 years	0.08	0.25	1.1	1.3
Black, 45–54 years	0.10	–0.33	1.1	0.8

SOURCE: National Center for Health Statistics. National Health Interview Survey, 1982–84.

NHIS restricted-activity days, bed days, and work-loss days

The correlations for restricted-activity, bed, and work-loss days are similar to those seen for the other two sets of variables from NHIS. (See table 6 for the statistics, standard errors, and standard errors of differences.) Table F reveals that the lowest correlation coefficient for 1982–83 is –0.14 for restricted-activity days among black persons 45–54 years and the lowest correlation coefficient for 1983–84 is –0.20 for bed days among persons 6–16 years. The highest correlation for 1982–83 is 0.39 for restricted-activity days among white persons. The maximum correlation seen for 1983–84 is 0.29 for restricted-activity days among the total population. The averages are 0.09 and 0.07, which fall between the averages for the other two sets of NHIS variables.

The effect of the independence assumption as measured by the ratio varies from 0.9 to 1.6 for 1982–83. For 1983–84 the minimum ratio is 0.8, whereas the maximum is 1.4. The means are 1.1 and 1.1.

Table F. Correlation coefficients and effect of independence assumption on variances for number of restricted-activity days, bed days, and work-loss days, by selected characteristics: United States, 1982–84

Characteristic	*Correlation coefficient*		*Independence effect (E)*	
	1982–83	*1983–84*	*1982–83*	*1983–84*
Restricted-activity days				
Total	0.16	0.29	1.2	1.4
White	0.39	0.28	1.6	1.4
Male	0.01	0.20	1.0	1.3
55–64 years	0.04	–0.07	1.0	0.9
Black	–0.06	0.11	0.9	1.1
Male, 75 years and over	–0.07	0.17	1.0	1.2
Black, 45–54 years	–0.14	0.04	0.9	1.0
Bed days				
Total	0.13	0.04	1.1	1.0
White	0.34	0.05	1.5	1.0
Female	0.20	0.06	1.2	1.1
Male	0.00	0.08	1.0	1.1
Male, married	0.11	–0.06	1.1	0.9
6–16 years	0.17	–0.20	1.2	0.8
Female, black	–0.08	0.15	0.9	1.2
Work-loss days				
Total	0.05	–0.01	1.0	1.0
White, married	0.32	–0.05	1.5	1.0
Female	0.10	0.02	1.1	1.0
25–34 years	–0.04	0.02	1.0	1.0
Black	–0.11	0.07	1.0	1.1
Separated	0.06	0.13	1.1	1.2

SOURCE: National Center for Health Statistics. National Health Interview Survey, 1982–84.

Impact of inference

One of the consequences when there are correlations among the sample units is an impact upon the inferences analysts draw from the findings. Kish (25) discussed the magnitude of the effect that clustering has upon the construction of confidence intervals. Using his strategy and this body of empirical evidence, one can see the effects on confidence intervals of assuming the correlation is zero.

Krewski (26) and Krewski and Rao (12) showed that under certain conditions the quantity

$$t = \frac{\hat{\theta} - \theta}{\sqrt{V_r(\hat{\theta})}}$$

is asymptotically distributed as a standard normal random variable $N(0,1)$ when L (the number of strata) goes to ∞. This theory was derived for linear and nonlinear statistics and is valid for any stratified multistage design in which the PSU's are selected with replacement and in which independent subsamples are taken for those PSU's selected more than once. The PSU's in NHDS and NHIS are not selected with replacement, and the effect of this difference on the result is not known.

To observe the magnitude of the effects of independence upon the confidence intervals, correlations from each of the six sets of variables for NHDS and NHIS that vary across the observed range were used. The variables, correlations, and effects are displayed in table G.

The column labeled E shows the ratio of the variance assuming the correlation is zero to the true variance, whereas the next column gives the ratio of the standard errors. The last column shows the probabilities of exceeding the true population value for $\alpha = 0.05$. When $\alpha = 0.05$, the coefficient 1.96 is multiplied by the standard error. One would expect that 5 percent of the time an error will be made. However, this percent is either higher or lower depending upon whether the true variance is larger or smaller than the variance utilized. For example, if $\sqrt{E}$ is 1.3, the number of standard errors being used is $1.96 \times 1.3 = 2.55$. Thus, 0.0120 of the time an error will be made.

Table G indicates that when the correlations are large, the confidence level is zero. This means that using the variance estimate calculated under the assumption of zero correlation leads to no incorrect statements. Correlations that are negative result in confidence levels ranging from 0.0524 to 0.0688. Confidence intervals constructed based on variance estimates that ignore correlations can be distorted. These distortions would also be present in hypothesis testing procedures.

Table G. Effects of independence assumption on confidence intervals using statistics from the National Hospital Discharge Survey and the National Health Interview Survey, by selected characteristics: 1982–84

Characteristic	*Correlation*	E	$\sqrt{E}$	*Probability of incorrect statements ($\alpha = 0.05$)*
Congenital anomaly, under 15 years	0.9570	21.4	4.6	0.0000
15–44 years	0.8813	12.0	3.5	0.0000
Cataract	0.7347	3.1	1.8	0.0006
Arthroplasty and replacement of hip, 65 years and over	0.5272	2.0	1.4	0.0052
Chronic conditions, female	0.3861	1.6	1.3	0.0120
Physician visits, under 6 years	0.1162	1.1	1.0	0.0376
Work-loss days, female	0.0197	1.0	1.0	0.0478
Work-loss days, total	−0.0131	1.0	1.0	0.0524
Short-stay hospital episodes, married, 55–64 years	−0.1233	0.9	0.9	0.0658
Short-stay hospital episodes, divorced	−0.1628	0.9	0.9	0.0688

Components of variance

The correlations for NHIS are not as large as those for NHDS. In both designs the PSU's are the same from year to year; it is the second stage of sampling that varies from one year to the next. In NHDS the same hospitals remain in the survey and, for the specified year, discharges are sampled. The PSU's in NHIS are groups of counties and metropolitan areas; the same PSU's are used from year to year, but different households within the same segments are selected. The large correlations for NHDS probably reflect the fact that the composition of hospitals is not likely to change when measured by discharge and surgical procedures performed from one year to the next. These features are structured by the facilities and staff of the hospital. In contrast, measures of morbidity, such as acute conditions and limitation of activity, are more variable.

If the difference in the values of correlations depends upon the type of primary sampling units, most of the variability in NHDS should be among hospitals, whereas for NHIS the largest component should be the variability among the households within the PSU's. Casady (27) and Bean and Schnack (28) demonstrate how to apply the BRR method to obtain estimates of the components of variance.

To estimate the within component of variance, denoted as V_W, each of the PSU's is considered to be a pseudostratum. Here, the sampled elements are randomly placed in one of two equal-sized groups. Constructing a half-sample thus consists of choosing one of the two groups of elements from each of the PSU's. The data from a half-sample are subject to the same estimation procedure as the data from the total sample, creating another estimate of $\hat{\theta}$. By means of a second orthogonal pattern, B estimates of $\hat{\theta}$ are produced. Then an estimate of the within component of variance is:

$$V_W(\hat{\theta}) = \sum_{i=1}^{B} (\hat{\theta}_i - \hat{\theta})^2 / B$$

The BRR estimator of V_W was applied to both the NHDS and NHIS data. Each of the sampled PSU's in the sample was considered to be a pseudostratum. Then within each of these PSU's, either the hospitals in NHDS or the households in NHIS were randomly allocated into one of two groups.

Instead of displaying the within and between components of variance for each of the NHDS and NHIS statistics, the mean, maximum, and minimum of the percent contribution of the within component of variance are presented by year in tables H–N. Tables 7–12 show the percent contribution for each variable by year. For several statistics in NHIS the within component of variance estimate is larger than the estimate for the total variance. NCHS assumes that the between component of variance is zero in this case.

As observed in table H, the within component of variance for the number of diagnostic discharges contributes, on the average, 13–15 percent of the total variance. Therefore, most of the variability is between rather than within hospitals. For the days-of-care statistics, the mean of the within component of variance increases to 21–23 percent of the total variance (table J). The maximum across the three years is 89 percent. Similar results are found in table K for the within component of variance percent contribution to the total variability for the number of surgical procedures. Here the means are 17 percent for 1982, 18.5 percent for 1983, and 12 percent for 1984. This body of empirical evidence suggests that, regardless of the statistic, the between component of variance is considerably larger than the within component of variance.

The opposite is true for the 60 NHIS statistics. For the acute and chronic conditions, the within component of variance counts for at least half of the variance (table L). The maximum is 97.9 percent for the year 1984. For the other 40 statistics, the average contribution is higher than 50 percent; the mean percent contribution ranges from 71.6 percent to 77.8 percent (tables M and N). These values provide empirical evidence that for NHIS most of the variability is among the households in the second stage of sampling.

Table H. Mean, maximum, and minimum of percent contribution of within component of variance to total variance for number of patients discharged from short-stay hospitals: United States, 1982–84

Summary statistic	*1982*	*1983*	*1984*
		Percent	
Mean	15.5	12.6	13.1
Maximum	64.3	43.1	43.9
Minimum	0.8	0.5	0.6

SOURCE: National Center for Health Statistics, National Hospital Discharge Survey, 1982–84.

Table J. Mean, maximum, and minimum of percent contribution of within component of variance to total variance for number of days of care for patients discharged from short-stay hospitals: United States, 1982–84

Summary statistic	1982	1983	1984
		Percent	
Mean	22.9	22.6	21.5
Maximum	89.3	68.8	54.3
Minimum	2.7	2.3	3.0

SOURCE: National Center for Health Statistics. National Hospital Discharge Survey, 1982–84.

Table K. Mean, maximum, and minimum of percent contribution of within component of variance to total variance for number of surgical procedures for patients discharged from short-stay hospitals: United States, 1982–84

Summary statistic	1982	1983	1984
		Percent	
Mean	17.0	18.5	12.4
Maximum	69.9	70.8	55.7
Minimum	0.2	0.1	0.1

SOURCE: National Center for Health Statistics. National Hospital Discharge Survey, 1982–84.

Table L. Mean, maximum, and minimum of percent contribution of within component of variance to total variance for number of acute and chronic conditions: United States, 1982–84

Summary statistic	1982	1983	1984
		Percent	
Mean	57.5	57.9	61.7
Maximum	96.9	91.6	97.9
Minimum	1.9	1.6	2.3

SOURCE: National Center for Health Statistics. National Health Interview Survey, 1982–84.

Table M. Mean, maximum, and minimum of percent contribution of within component of variance to total variance for number of physician visits and short-stay hospital episodes: United States, 1982–84

Summary statistic	1982	1983	1984
		Percent	
Mean	75.5	76.7	77.0
Maximum	98.5	97.7	93.2
Minimum	35.9	27.1	56.8

SOURCE: National Center for Health Statistics. National Health Interview Survey, 1982–84.

Table N. Mean, maximum, and minimum of percent contribution of within component of variance to total variance for number of restricted-activity days, bed days, and work-loss days: United States, 1982–84

Summary statistic	1982	1983	1984
		Percent	
Mean	71.6	77.1	77.8
Maximum	97.4	99.9	97.8
Minimum	25.1	41.5	49.7

SOURCE: National Center for Health Statistics. National Health Interview Survey, 1982–84.

Discussion

Both NHDS and NHIS are conducted on a continuing basis for the purpose of producing various estimates. Data from these surveys can be employed to estimate parameters for all the years combined, parameters for the current year, and parameters measuring change from one year to the next. As discussed by Cochran (29), each of these estimators requires a different design to achieve maximum precision. Sampling theory indicates that for estimating change the same sampling units should be retained for both surveys, whereas different sampling units are better for estimates using data from all time periods. An efficient scheme for constructing a current estimate is to retain a fraction λ of the sampling units from the previous survey while selecting $1 - \lambda$ fraction of new units. The optimum percentage to retain depends upon the correlation. Making specific assumptions about the design, Cochran (29) shows that for sampling on two to four occasions, as the correlation increases, the optimum to retain decreases. For example, when sampling at two time periods, if the correlation is 0.7, the optimum to retain is 42 percent, but if the correlation increases to 0.8, the percentage should be only 38 percent. For this design, the best type of estimator is a double-sampling regression estimate.

The findings for number of discharges, days of care, and surgical procedures clearly indicate that variance estimates based on the assumption that statistics from year to year are independent are in error. The overestimation of variance becomes larger as the correlation nears one. These empirical results demonstrate that NHDS can be used to detect very small differences from year to year. However, with large correlations the design of 100-percent overlap is not the most efficient for maximizing the precision of yearly estimates. As Cochran (29) demonstrated for simple random sampling and a regression estimator, if the correlation is 0.9 and the number of occasions is four, the percentage gain in efficiency is 59 percent when the overlap is 50 percent; the gain in efficiency drops to 40 percent when the overlap is 75 percent. Research is required to determine if this holds for NHDS. Because the hospitals are the same for each survey, the NHDS design is not very efficient at producing data for combining across the years.

However, for reasons of economy, selecting a different sample each year may not be feasible. A possible compromise that should be considered by NCHS is a design whereby part of the sample is changed from year to year. Another consideration is using a regression-type estimator that would take into account the estimates produced from the previous year.

The formula for the variances of differences given in the appendixes to NHIS reports is based on the assumption that the sample for a given year, say 1982, is independent of other sample years even though the primary sampling units are the same across the years. By assuming independence, data are compared using the formula:

$$V(Y_1 - Y_2) = V(Y_1) + V(Y_2)$$

The investigation of this assumption for 60 NHIS statistics indicates that the correlations range from a negative value of –0.20 to a high positive value of 0.39. The evidence here suggests that NHIS is great for adding yearly data together, but the design is not as efficient for examining differences from one year to the next. The results do not provide any clear findings on how an annual point estimate would compare with a yearly regression estimate.

A caveat to these conclusions is that the empirical results are for totals only. Correlations for means and proportions were not calculated. Parsons (30) estimated correlations for selected statistics from the 1985 and 1986 NHIS. For eight estimates of totals, the correlations ranged from 0.03 to 0.36, which are similar to the values observed in this study. The range of correlations for 26 means and proportions was from a low of –0.03 to a high of 0.38, which is similar to values observed for totals. No results are available for means and proportions for NHDS. A further evaluation of means and proportions for both surveys is warranted.

In summary, making the assumption of independence in NHDS may result in extreme overestimates of variance, at least for totals. The results clearly indicate that NHDS is ideally suited for measuring change over time. This feature of the survey should be utilized more than it currently is. In NHIS when the assumption of independence does not hold, the findings show that the variance estimate is not an excessive overestimation. The variance estimated was never more than 1.6 times as large as it should be. One possible explanation is the difference in primary sampling units. Empirical evidence suggests that for NHDS the within component of variance is less than 26 percent of the total variance, whereas for NHIS the within component of variance accounts for at least half the variability.

References

1. Simmons WR, Schnack GA. Development of the design of the NCHS Hospital Discharge Survey. National Center for Health Statistics. Vital Health Stat 2(39). 1977.
2. Shimizu IM. The new statistical design of the National Hospital Discharge Survey. In: Proceedings of the Survey Research Methods Section of the American Statistical Association. Anaheim, California: American Statistical Association. 702–6. 1990.
3. National Center for Health Statistics. The National Health Interview Survey design, 1973–84, and procedures, 1975–83. National Center for Health Statistics. Vital Health Stat 1(18). 1985.
4. Massey JT, Moore TF, Parsons VL, Tadros W. Design and estimation for the National Health Interview Survey, 1985–94. National Center for Health Statistics. Vital Health Stat 2(110). 1989.
5. McCarthy PJ. Replication: an approach to the analysis of data from complex surveys. National Center for Health Statistics. Vital Health Stat 2(14). 1966.
6. McCarthy PJ. Pseudoreplication: further evaluation and application of the balanced half-sample technique. National Center for Health Statistics. Vital Health Stat 2(31). 1969.
7. Kish L, Frankel MR. Balanced repeated replication for analytical statistics. In: Goldfield ED, ed. Proceedings of the Social Statistics Section of the American Statistical Association. Pittsburgh: American Statistical Association. 2–10. 1968.
8. Kish L, Frankel MR. Balanced repeated replication for standard errors. J Am Stat Assoc 65:1071–94. 1970.
9. Frankel MR. Inference from survey samples: an empirical investigation. Institute for Social Research. Ann Arbor, Michigan: University of Michigan. 1971.
10. Koch GG, Lemeshow S. An application of multivariate analysis to complex sample survey data. J Am Stat Assoc 67:780–2. 1972.
11. Bean JA. Distribution and properties of variance estimators for complex multistage probability samples: an empirical distribution. National Center for Health Statistics. Vital Health Stat 2(65). 1975.
12. Krewski D, Rao JNK. Inference from stratified samples: properties of linearization, jackknife, and balanced repeated replication methods. Ann Stat 9:1010–9. 1981.
13. Gurney M, Jewett RS. Constructing orthogonal replications for variance estimation. J Am Stat Assoc 70:819–21. 1975.
14. Krewski D. On the stability of some replication variance estimators in the linear case. J Stat Plan Infer 2:45–51. 1978.
15. Koch GG, Freeman DH Jr, Freeman JL. Strategies in the multivariate analysis of data from complex surveys. Int Stat Rev 43:59–78. 1975.
16. Freeman DH Jr. The regression analysis of data from complex surveys: an empirical investigation of covariance matrix estimation. Institute of Statistics Mimemo Series no. 1020. Chapel Hill, North Carolina: University of North Carolina. 1975.
17. Freeman DH Jr, Freeman JL, Brock DB, Koch GG. Strategies in multivariate analysis of data from complex surveys II: an application to the United States National Health Interview Survey. Int Stat Rev 44:317–30. 1976.
18. Freeman DH Jr, Freeman JL, Koch GG, and Brock DB. An analysis of physician visit data from a complex sample survey. Am J Public Health 66:979–83. 1976.
19. Freeman DH Jr, Brock DB. The role of covariance matrix estimation in the analysis of complex sample survey data. In: Namboodiri NK, ed. Survey sampling and measurement. New York, San Francisco, and London: Academic Press, Inc. 121–40. 1978.
20. SAS Institute, Inc. SAS's user guide: basics. 1985 ed. Cary, North Carolina: SAS Institute, Inc. 1985.
21. Stanek EJ. The properties of balanced half-sample variance estimates in complex surveys when strata are paired to form pseudo-strata. Biostatistic-Epidemiology Program Series no 77–8. Amherst, Massachusetts: University of Massachusetts. 1977.
22. Simmons WR, Baird JT. Pseudoreplication in the NCHS Health Examination Survey. In: Goldfield ED, ed. Proceedings of the Social Statistics Section of the American Statistical Association. Pittsburgh: American Statistical Association. 19–30. 1968.
23. Lemeshow S. The use of unique statistical weights for estimating variances with the balanced half-sample technique. In: Goldfield ED, ed. Proceedings of the Social Statistics Section of the American Statistical Association, part II. Boston: American Statistical Association. 507–512. 1976.
24. Graves EJ. Utilization of short-stay hospitals, United States: 1984 annual summary. National Center for Health Statistics. Vital Health Stat 13(84). 1986.
25. Kish L. Confidence intervals for clustered samples. Am Sociol Rev 22:154–65. 1957.
26. Krewski D. Jackknifing u-statistics in finite populations. Commun Statisti-Theory Meth A(7):1–12. 1978.
27. Casady RJ. The estimation of variance components using balanced repeated replication. In: Goldfield ED, ed. Proceedings of the Social Statistics Section of the American Statistical Association. Atlanta: American Statistical Association. 352–56. 1975.
28. Bean JA, Schnack GA. An application of balanced repeated replication to the estimation of variance components. In: Goldfield ED, ed. Proceedings of the Social Statistics

Section of the American Statistical Association, part II. Chicago: American Statistical Association. 938–42. 1978.

29. Cochran WG. Double sampling. In: Sampling techniques. 3d ed. New York, Santa Barbara, London, Sydney, and Toronto: John Wiley & Sons, Inc. 327–58. 1977.

30. Parsons VL. Estimation of year-to-year covariances for a national health survey. In: Proceedings of the Survey Research Methods Section of the American Statistical Association. New Orleans: American Statistical Association. 210–5. 1988.

List of detailed tables

Table 1. Number of patients discharged from short-stay hospitals, standard errors, and standard errors of difference, by selected characteristics: United States, 1982–84

Characteristic	1982		1983		1984		Standard error of difference	
	Number of patients discharged	Standard error	Number of patients discharged	Standard error	Number of patients discharged	Standard error	1982–83	1983–84
	Number in thousands							
Female	21,554	635	22,027	687	21,380	695	284	184
15–44 years	14,515	520	14,537	547	13,973	530	216	161
Male	14,428	436	14,805	446	14,296	439	186	140
65 years and over	9,982	316	10,752	362	10,801	375	150	121
45–64 years	8,094	245	8,100	249	7,808	255	119	82
Females with deliveries	3,695	216	3,795	229	3,734	224	73	65
Under 15 years	3,392	214	3,443	249	3,094	233	75	54
Mental disorders	1,614	103	1,627	108	1,630	105	34	37
Mental disorders, 15–44 years	892	71	912	77	934	74	22	28
Acute myocardial infarction	631	29	642	30	670	29	22	21
Cataract	520	38	562	44	467	31	20	28
Atherosclerotic heart disease	465	25	438	28	325	21	17	16
Inguinal hernia, male	417	15	403	17	378	18	15	16
Cataract, 65 years and over	399	30	451	36	384	26	17	23
Malignant neoplasm of lung	301	14	320	17	325	21	16	17
Malignant neoplasm of breast, female	210	11	230	12	225	13	11	13
Malignant neoplasm of lung, male	188	11	203	11	205	14	12	14
Fracture of neck of femur, 65 years and over	177	10	184	11	198	11	9	10
Congenital anomaly, under 15 years	161	28	178	34	164	31	11	10

SOURCE: National Center for Health Statistics. National Hospital Discharge Survey, 1982–84.

Table 2. Number of days of care for patients discharged from short-stay hospitals, standard errors, and standard errors of difference, by selected characteristics: United States, 1982–84

Characteristic	1982		1983		1984		Standard error of difference	
	Number of days of care	Standard error	Number of days of care	Standard error	Number of days of care	Standard error	1982–83	1983–84
	Number in thousands							
Female	146,117	3,856	145,768	4,083	134,533	3,772	2,189	1,541
Male	108,167	3,042	109,100	3,046	100,161	2,981	1,493	1,285
65 years and over	100,807	3,148	104,284	3,346	96,016	3,219	1,654	1,394
15–44 years	74,416	2,468	73,328	2,674	68,586	2,381	1,268	1,035
45–64 years	63,517	1,860	61,514	1,717	56,172	1,724	1,011	822
Mental disorders	19,541	1,610	20,259	1,763	19,420	1,676	686	641
Under 15 years	15,544	1,124	15,740	1,375	13,919	1,349	505	376
Females with deliveries	13,172	695	13,341	756	12,810	720	329	242
Mental disorders, 15–44 years	10,534	1,131	11,102	1,271	11,037	1,239	506	466
Acute myocardial infarction	7,074	357	6,980	343	6,655	310	338	282
Atherosclerotic heart disease	4,130	278	3,681	245	2,386	176	209	150
Fracture of neck of femur, 65 years and over	3,318	193	3,267	219	3,131	202	202	230
Malignant neoplasm of lung	3,215	173	3,375	174	3,089	169	187	177
Malignant neoplasm of breast, female	2,082	140	2,167	118	1,881	149	135	158
Malignant neoplasm of lung, male	2,062	121	2,094	116	1,916	129	137	147
Inguinal hernia, male	1,872	88	1,620	67	1,416	77	87	85
Cataract	1,519	104	1,431	107	1,105	72	58	73
Cataract, 65 years and over	1,184	86	1,167	91	897	59	53	65
Congenital anomaly, under 15 years	891	187	985	206	932	213	77	98

SOURCE: National Center for Health Statistics. National Hospital Discharge Survey, 1982–84.

Table 3. Number of surgical procedures for patients discharged from short-stay hospitals, standard errors, and standard errors of difference, by selected characteristics: United States, 1982–84

Characteristic	1982		1983		1984		Standard error of difference	
	Number of surgical procedures	Standard error	Number of surgical procedures	Standard error	Number of surgical procedures	Standard error	1982–83	1983–84
	Number in thousands							
All procedures:								
Total	33,519	1,186	35,223	1,340	35,811	1,375	619	423
Female	19,889	793	20,737	881	20,897	867	391	256
15–44 years	14,281	647	14,373	667	14,094	643	273	219
Male	13,631	465	14,486	517	14,914	559	261	214
65 years and over	8,132	356	9,295	487	10,108	531	253	189
45–64 years	7,747	309	8,077	338	8,243	355	170	133
Under 15 years	3,360	189	3,478	214	3,366	217	81	76
Circumcision, under 15 years	1,223	68	1,226	77	1,251	80	41	34
Cesarean section:								
All ages	686	48	766	60	785	63	26	22
15–44 years	683	48	764	60	783	63	26	22
Hysterectomy	615	48	635	46	638	47	23	18
Bilateral occlusion of fallopian tube, 15–44 years	558	54	532	51	466	45	20	19
Extraction of lens, 65 years and over	425	34	473	36	398	31	19	23
Cardiac catheterization	420	36	478	43	516	49	19	20
Prostatectomy	328	18	332	19	347	22	15	15
Tonsillectomy, under 15 years	249	18	255	17	209	18	13	12
Prostatectomy, 65 years and over	241	14	256	15	264	17	12	13
Hysterectomy, 45–64 years	171	13	169	10	180	13	9	10
Direct heart revascularization	154	18	173	18	171	19	10	10
Myringotomy, under 15 years	132	12	152	17	114	14	10	9
Arthroplasty and replacement of hip:								
65 years and over	93	6	112	8	135	11	7	10
Female	91	6	103	8	124	10	7	10
Hysterectomy, 65 years and over	55	4	49	5	58	6	5	7

SOURCE: National Center for Health Statistics. National Hospital Discharge Survey, 1982–84.

Table 4. Number of acute and chronic conditions, standard errors, and standard errors of difference, by selected characteristics: United States, 1982–84

Characteristic	1982		1983		1984		Standard error of difference	
	Number of conditions	Standard error	Number of conditions	Standard error	Number of conditions	Standard error	1982–83	1983–84
Acute conditions	Number in thousands							
Total	379,598	7,665	400,707	6,635	410,812	6,387	9,730	8,634
White, 25–34 years	58,686	2,435	60,095	2,076	63,126	2,275	3,369	2,969
Female, 6–16 years	45,124	1,826	45,470	1,841	45,810	1,770	2,744	2,441
Male, 6–16 years	41,766	1,855	45,766	1,903	47,082	1,717	2,603	2,744
White, 55–64 years	20,578	1,116	20,338	1,022	21,522	1,181	1,509	1,687
Married, 45–54 years	19,526	1,165	18,625	1,192	17,403	990	1,645	1,610
Female, 55–64 years	14,648	959	14,008	1,024	14,012	919	1,404	1,344
75 years and over	7,969	751	9,581	808	10,245	844	1,058	1,092
Separated, 17–24 years	908	240	898	220	797	262	330	342
Black, 65–74 years	678	151	1,482	376	1,564	323	391	499
Chronic conditions								
Total	227,114	2,970	229,322	3,271	231,606	3,179	4,382	3,686
Female	117,579	1,703	118,625	1,764	119,771	1,719	2,559	1,930
25–34 years	38,499	834	39,149	815	40,003	934	1,221	1,112
Married, 25–34 years	27,520	778	27,272	750	28,199	818	1,038	986
White, 45–54 years	19,400	605	19,470	573	19,374	580	823	869
Male, 6–16 years	19,225	731	19,251	669	19,153	618	928	913
Female, 45–54 years	11,540	419	11,478	315	11,502	398	515	526
Income less than $5,000, 17–24 years	3,892	319	3,936	358	3,386	301	487	465
Black, 35–44 years	2,889	233	2,999	289	3,168	266	334	370
Income less than $5,000, 65–74 years	1,565	134	1,407	145	1,124	143	192	206

SOURCE: National Center for Health Statistics. National Health Interview Survey, 1982–84.

Table 5. Number of physician visits and short-stay hospital episodes, standard errors, and standard errors of difference, by selected characteristics: United States, 1982–84

Characteristic	1982		1983		1984		Standard error of difference	
	Number of visits or episodes	Standard error	Number of visits or episodes	Standard error	Number of visits or episodes	Standard error	1982–83	1983–84
Physician visits	Number in thousands							
Total	799,980	8,325	808,320	10,064	841,760	10,109	10,992	12,108
Male	322,307	4,630	331,788	4,944	345,555	6,255	6,416	7,722
Black	93,158	3,576	92,198	3,682	103,190	6,182	4,765	7,225
White, 17–24 years	84,617	1,990	82,592	2,427	79,350	1,911	2,933	2,531
White, 55–64 years	83,482	2,848	80,999	2,866	83,550	2,297	3,654	3,108
Under 6 years	77,967	1,617	83,171	1,825	87,455	2,123	2,293	2,673
Married, 35–44 years	69,487	2,113	71,771	1,866	75,432	2,907	2,656	3,571
Divorced	45,325	1,716	49,610	2,255	52,605	3,120	2,799	3,867
Female, 65–74 years	48,856	2,115	46,099	1,828	49,763	1,777	2,722	2,658
Female, 75 years and over	33,828	1,212	36,185	1,507	41,131	2,428	1,547	2,710
Short-stay hospital episodes								
Total	30,041	455	30,241	402	29,294	435	520	501
White	25,854	421	26,161	390	25,534	404	448	450
Female	17,806	307	17,949	291	17,484	310	347	403
Male	12,235	250	21,292	233	11,809	247	335	313
75 years and over	3,128	121	3,330	136	3,522	126	162	162
Married, 55–64 years	3,013	117	3,312	141	2,826	110	166	189
6–16 years	1,697	87	1,674	71	1,511	70	113	95
Divorced	1,682	95	1,725	94	1,678	81	144	129
Male, 17–24 years	1,048	76	993	61	882	66	94	78
Black, 45–54 years	393	47	335	36	325	41	56	63

SOURCE: National Center for Health Statistics. National Health Interview Survey, 1982–84.

Table 6. Number of restricted-activity days, bed days, and work-loss days, standard errors, and standard errors of difference, by selected characteristics: United States, 1982–84

Characteristic	1982		1983		1984		Standard error of difference	
	Number of days	Standard error	Number of days	Standard error	Number of days	Standard error	1982–83	1983–84
Restricted-activity days	Number in thousands							
Total	3,253,254	59,036	3,318,069	47,805	3,427,193	50,960	69,712	58,830
White	2,746,275	48,420	2,813,111	46,009	2,904,705	51,284	52,172	58,632
Male	1,372,863	34,439	1,364,731	27,972	1,454,136	31,144	44,203	37,368
55–64 years	514,393	20,568	501,147	19,684	511,634	20,659	27,929	29,529
Black	438,323	29,174	451,903	22,874	460,593	19,943	38,133	28,573
Male, 75 years and over	92,191	8,587	119,477	9,041	125,864	10,213	12,509	12,472
Black, 45–54 years	66,922	6,939	51,848	5,468	54,137	6,186	9,432	8,104
Bed days								
Total	1,444,556	34,768	1,529,698	31,078	1,508,203	29,954	43,529	42,235
White	1,182,455	27,857	1,264,351	28,185	1,251,628	27,523	32,083	38,494
Female	852,706	22,910	897,783	25,100	884,031	21,982	30,397	32,341
Male	591,850	20,636	631,915	15,848	624,172	16,446	25,967	21,849
Male, married	326,080	16,128	339,074	13,514	349,820	12,702	19,896	19,127
6–16 years	145,403	6,858	157,084	6,016	155,876	6,615	8,303	9,779
Female, black	144,512	11,742	143,765	11,225	135,590	8,141	16,909	12,869
Work-loss days								
Total	452,615	12,151	419,249	11,784	513,896	13,757	16,538	18,231
White, married	269,623	10,207	245,046	9,405	295,453	11,193	11,486	14,963
Female	227,451	8,684	212,161	8,949	254,601	10,434	11,825	13,611
25–34 years	125,901	6,821	116,192	5,484	142,572	6,752	8,914	8,611
Black	53,450	4,834	49,996	4,978	70,026	5,304	6,980	7,012
Separated	12,983	1,659	16,876	2,605	16,756	2,944	2,999	3,658

SOURCE: National Center for Health Statistics. National Health Interview Survey, 1982–84.

Table 7. Percent contribution of within component of variance to total variance for number of patients discharged from short-stay hospitals, by selected characteristics: United States, 1982–84

Characteristic	*1982*	*1983*	*1984*
		Percent	
Female	0.8	0.5	0.6
Male	0.9	0.9	1.0
15–44 years	1.3	1.0	1.7
65 years and over	1.2	1.5	1.2
45–64 years	2.5	2.9	3.0
Females with deliveries	2.1	1.8	2.3
Under 15 years	1.5	1.4	1.6
Mental disorders	3.7	3.6	3.8
Mental disorders, 15–44 years	4.0	3.9	4.1
Acute myocardial infarction	14.9	17.1	20.6
Cataract	9.3	5.7	12.3
Atherosclerotic heart disease	14.2	17.2	14.3
Inguinal hernia, male	64.3	29.4	36.9
Cataract, 65 years and over	9.6	5.7	14.0
Malignant neoplasm of lung	41.4	30.2	19.9
Malignant neoplasm of breast, female	35.4	43.1	27.1
Malignant neoplasm of lung, male	38.4	37.8	34.2
Fracture of neck of femur, 65 years and over	45.3	32.7	43.9
Congenital anomaly, under 15 years	3.5	3.0	6.4

SOURCE: National Center for Health Statistics, National Hospital Discharge Survey, 1982–84.

Table 8. Percent contribution of within component of variance to total variance for number of days of care for patients discharged from short-stay hospitals, by selected characteristics: United States, 1982–84

Characteristic	*1982*	*1983*	*1984*
		Percent	
Female	2.7	3.5	3.6
Male	5.5	5.5	5.6
15–44 years	3.4	4.2	3.8
65 years and over	3.3	3.6	5.1
45–64 years	6.5	9.4	7.2
Mental disorders	6.8	4.9	3.0
Under 15 years	4.1	2.3	5.0
Females with deliveries	3.1	3.7	4.2
Mental disorders, 15–44 years	6.4	5.9	3.5
Acute myocardial infarction	25.2	27.6	30.4
Atherosclerotic heart disease	18.6	30.6	22.0
Cataract, 65 years and over	53.3	51.2	47.1
Malignant neoplasm of lung	49.5	58.2	53.8
Malignant neoplasm of breast, female	55.9	66.1	42.8
Malignant neoplasm of lung, male	64.3	68.8	54.3
Inguinal hernia, male	89.3	43.2	52.0
Cataract	14.3	12.3	24.7
Cataract, 65 years and over	15.7	14.2	27.1
Congenital anomaly, under 15 years	7.4	13.6	13.4

SOURCE: National Center for Health Statistics, National Hospital Discharge Survey, 1982–84.

Table 9. Percent contribution of within component of variance to total variance for number of surgical procedures for patients discharged from short-stay hospitals, by selected characteristics: United States, 1982–84

Characteristic	1982	1983	1984
		Percent	
All procedures:			
Total	0.1	0.1	0.1
Female	0.4	0.4	0.4
15–44 years	0.7	0.6	0.8
Male	1.0	0.9	1.1
65 years and over	1.3	0.8	0.7
45–64 years	2.3	1.2	1.6
Under 15 years	2.0	2.2	2.3
Circumcision, under 15 years	7.0	5.3	7.7
Cesarean section:			
All ages	6.5	9.4	5.2
15–44 years	6.5	9.4	5.1
Hysterectomy	7.8	11.1	9.2
Bilateral occlusion of fallopian tube, 15–44 years	6.2	7.5	6.9
Extraction of lens, 65 years and over	7.6	5.6	8.2
Cardiac catheterization	5.7	7.4	5.0
Prostatectomy	19.0	23.6	21.5
Tonsillectomy, under 15 years	19.2	35.9	19.2
Prostatectomy, 65 years and over	33.2	32.5	24.0
Hysterectomy, 45–64 years	27.3	37.8	36.8
Direct heart revascularization	17.0	16.7	13.1
Myringotomy, under 15 years	28.1	18.5	11.2
Arthroplasty and replacement of hip:			
65 years and over	64.9	60.4	22.8
Female	58.5	67.2	27.3
Hysterectomy, 65 years and over	69.9	70.8	55.7

SOURCE: National Center for Health Statistics. National Hospital Discharge Survey, 1982–84.

Table 10. Percent contribution of within component of variance to total variance for number of acute and chronic conditions, by selected characteristics: United States, 1982–84

Characteristic	1982	1983	1984
Acute conditions		Percent	
Total	4.3	3.8	4.4
White, 25–34 years	39.3	72.1	60.2
Female, 6–16 years	68.5	78.9	53.5
Male, 6–16 years	55.7	64.9	97.9
White, 55–64 years	96.9	91.6	85.1
Married, 45–54 years	92.8	56.9	([1])
Female, 55–64 years	([1])	65.2	94.3
75 years and over	79.4	91.4	62.0
Separated, 17–24 years	([1])	([1])	70.1
Black, 65–74 years	71.0	80.0	94.2
Chronic conditions			
Total	1.9	1.6	2.3
Female	27.1	24.1	31.5
25–34 years	76.5	61.5	49.9
Married, 25–34 years	60.0	42.0	48.4
White, 45–54 years	74.3	72.5	92.2
Male, 6–16 years	40.0	45.6	59.8
Female, 45–54 years	([1])	([1])	([1])
Income less than $5,000, 17–24 years	73.4	37.3	60.0
Black, 35–44 years	72.7	66.1	65.6
Income less than $5,000, 65–74 years	([1])	85.9	88.2

[1] The within component of variance estimate was larger than the total variance estimate.

SOURCE: National Center for Health Statistics. National Health Interview Survey, 1982–84.

Table 11. Percent contribution of within component of variance to total variance for number of physician visits and short-stay hospital episodes, by selected characteristics: United States, 1982–84

Characteristic	1982	1983	1984
Physician visits		Percent	
Total	53.3	35.7	72.7
Male	59.6	71.8	(1)
Black	35.9	27.1	67.9
White, 17–24 years	87.1	82.7	80.3
White, 55–64 years	71.5	78.3	83.0
Under 6 years	76.3	63.7	56.8
Married, 35–44 years	68.9	91.4	64.7
Divorced	98.5	88.6	93.2
Female, 65–74 years	(1)	78.0	(1)
Female, 75 years and over	98.1	89.3	74.9
Short-stay hospital episodes			
Total	62.6	87.2	75.8
White	64.4	83.6	68.4
Female	73.3	87.7	73.9
Male	98.4	(1)	92.3
75 years and over	82.5	89.9	(1)
Married, 55–64 years	88.7	(1)	(1)
6–16 years	76.4	(1)	83.2
Divorced	83.3	97.7	90.7
Male, 17–24 years	92.7	(1)	76.3
Black, 45–54 years	62.9	74.4	77.8

[1]The within component of variance estimate was larger than the total variance estimate.

SOURCE: National Center for Health Statistics. National Health Interview Survey, 1982–84.

Table 12. Percent contribution of within component of variance to total variance for number of restricted-activity days, bed days, and work-loss days, by selected characteristics: United States, 1982–84

Characteristic	1982	1983	1984
Restricted-activity days		Percent	
Total	45.0	82.1	62.1
White	59.4	77.7	49.7
Male	56.0	89.2	80.2
55–64 years	83.5	62.0	75.8
Black	25.2	41.5	88.8
Male, 75 years and over	84.2	(1)	97.8
Black, 45–54 years	84.8	(1)	84.8
Bed days			
Total	58.9	77.4	72.7
White	76.6	79.1	65.8
Female	89.2	64.1	77.2
Male	64.9	(1)	(1)
Male, married	69.0	(1)	(1)
6–16 years	82.3	99.9	79.4
Female, black	59.3	76.9	(1)
Work-loss days			
Total	(1)	(1)	(1)
White, married	(1)	77.6	89.9
Female	97.4	(1)	87.5
25–34 years	84.4	(1)	(1)
Black	96.1	91.3	(1)
Separated	(1)	84.4	77.1

[1]The within component of variance estimate was larger than the total variance estimate.

SOURCE: National Center for Health Statistics. National Health Interview Survey, 1982–84.

☆ U.S. GOVERNMENT PRINTING OFFICE: 1992— 3 1 2 - 0 8 2/ 6 0 0 0 3

Vital and Health Statistics series descriptions

SERIES 1. **Programs and Collection Procedures**—Reports describing the general programs of the National Center for Health Statistics and its offices and divisions and the data collection methods used. They also include definitions and other material necessary for understanding the data.

SERIES 2. **Data Evaluation and Methods Research**—Studies of new statistical methodology including experimental tests of new survey methods, studies of vital statistics collection methods, new analytical techniques, objective evaluations of reliability of collected data, and contributions to statistical theory. Studies also include comparison of U.S. methodology with those of other countries.

SERIES 3. **Analytical and Epidemiological Studies**—Reports presenting analytical or interpretive studies based on vital and health statistics, carrying the analysis further than the expository types of reports in the other series.

SERIES 4. **Documents and Committee Reports**—Final reports of major committees concerned with vital and health statistics and documents such as recommended model vital registration laws and revised birth and death certificates.

SERIES 5. **Comparative International Vital and Health Statistics Reports**—Analytical and descriptive reports comparing U.S. vital and health statistics with those of other countries.

SERIES 6. **Cognition and Survey Measurement**—Reports from the National Laboratory for Collaborative Research in Cognition and Survey Measurement using methods of cognitive science to design, evaluate, and test survey instruments.

SERIES 10. **Data From the National Health Interview Survey**—Statistics on illness, accidental injuries, disability, use of hospital, medical, dental, and other services, and other health-related topics, all based on data collected in the continuing national household interview survey.

SERIES 11. **Data From the National Health Examination Survey and the National Health and Nutrition Examination Survey**—Data from direct examination, testing, and measurement of national samples of the civilian noninstitutionalized population provide the basis for (1) estimates of the medically defined prevalence of specific diseases in the United States and the distributions of the population with respect to physical, physiological, and psychological characteristics and (2) analysis of relationships among the various measurements without reference to an explicit finite universe of persons.

SERIES 12. **Data From the Institutionalized Population Surveys**—Discontinued in 1975. Reports from these surveys are included in Series 13.

SERIES 13. **Data on Health Resources Utilization**—Statistics on the utilization of health manpower and facilities providing long-term care, ambulatory care, hospital care, and family planning services.

SERIES 14. **Data on Health Resources: Manpower and Facilities**—Statistics on the numbers, geographic distribution, and characteristics of health resources including physicians, dentists, nurses, other health occupations, hospitals, nursing homes, and outpatient facilities.

SERIES 15. **Data From Special Surveys**—Statistics on health and health-related topics collected in special surveys that are not a part of the continuing data systems of the National Center for Health Statistics.

SERIES 16. **Compilations of Advance Data From Vital and Health Statistics**—These reports provide early release of data from the National Center for Health Statistics' health and demographic surveys. Many of these releases are followed by detailed reports in the Vital and Heath Statistics Series.

SERIES 20. **Data on Mortality**—Various statistics on mortality other than as included in regular annual or monthly reports. Special analyses by cause of death, age, and other demographic variables; geographic and time series analyses; and statistics on characteristics of death not available from the vital records based on sample surveys of those records.

SERIES 21. **Data on Natality, Marriage, and Divorce**—Various statistics on natality, marriage, and divorce other than as included in regular annual or monthly reports. Special analyses by demographic variables; geographic and time series analyses; studies of fertility; and statistics on characteristics of births not available from the vital records based on sample surveys of those records.

SERIES 22. **Data From the National Mortality and Natality Surveys**—Discontinued in 1975. Reports from these sample surveys based on vital records are included in Series 20 and 21, respectively.

SERIES 23. **Data From the National Survey of Family Growth**—Statistics on fertility, family formation and dissolution, family planning, and related maternal and infant health topics derived from a periodic survey of a nationwide probability sample of women 15–44 years of age.

SERIES 24. **Compilations of Data on Natality, Mortality, Marriage, Divorce, and Induced Terminations of Pregnancy**—Advance reports of births, deaths, marriages, and divorces are based on final data from the National Vital Statistics System and are published annually as supplements to the Monthly Vital Statistics Report (MVSR). These reports are followed by the publication of detailed data in Vital Statistics of the United States annual volumes. Other reports including induced terminations of pregnancy issued periodically as supplements to the MVSR provide selected findings based on data from the National Vital Statistics System and may be followed by detailed reports in Vital and Health Statistics Series.

For answers to questions about this report or for a list of titles of reports published in these series, contact:

Scientific and Technical Information Branch
National Center for Health Statistics
Centers for Disease Control
Public Health Service
6525 Belcrest Road, Room 1064
Hyattsville, Md. 20782

301-436-8500

DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service
Centers for Disease Control
National Center for Health Statistics
6525 Belcrest Road
Hyattsville, Maryland 20782

DHHS Publication No. (PHS) 92-1388, Series 2, No. 114