NATIONAL HOSPITAL DISCHARGE SURVEY

1979-2007 Multi-Year Public Use Data File Documentation

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Centers for Disease Control and Prevention
National Center for Health Statistics
Division of Health Care Statistics
Ambulatory and Hospital Care Statistics
3311 Toledo Road
Hyattsville, MD 20782
301.458.4321
NHDS@cdc.gov
http://www.cdc.gov/nchs/nhds.htm

July 2010

NATIONAL HOSPITAL DISCHARGE SURVEY

1979-2007 Multi-Year Public Use Data File Documentation

This document provides information for users of the 1979-2007 National Hospital Discharge Survey (NHDS) multi-year public use data file. For all records in the 1979-2007 NHDS data file, the *International Classification of Diseases, 9th Revision, Clinical Modification* (ICD-9-CM) was used for coding medical diagnoses and procedures. It is important to note, however, that important changes in NHDS methodology, as well as minor modifications to the ICD-9-CM coding system, have occurred during the period from 1979-2007. These changes are discussed in detail in this documentation.

For those familiar with single-year NHDS public-use data files, it is important to note several major differences between the single-year files and the multi-year files.

- Unlike the annual, single-year files, the multi-year data have separate files for newborn infant records and non-newborn records.
- For data years before 1993, the single-year data files included several recodes of basic variables that are not included on the multi-year files.
- On the multi-year files, coding of all variables has been standardized across the data years. Thus, for selected variables, coding may differ between the single-year and multi-year files.
- The record layout has been changed to allow for more efficient data storage.
- Beginning with the 1998 data year, HMO/PPO was added as a value for the two expected source of payment variables. Pre-1998 data years will have missing values for HMO/PPO.
- In 2000, coding for RACE was expanded and the code for "Asian/Pacific Islander" was separated into two categories. Thus, "Pacific Islander" as a separate category is missing before 2000. A code for "Multiple races" was also introduced in 2000.
- Beginning in 2001, two new variables were added to the file: TYPE OF ADMISSION and SOURCE OF ADMISSION, which are located at the end of each record. For the years before 2001, these variables are missing.
- The DRG variable is not available on the multi-year file (but is included on single year files).

Section I describes the survey and includes information on the history and scope of NHDS; the methodology, including data collection and medical coding procedures; population estimates; measurement errors and sampling errors. *Section II* provides technical details about the data file. *Section III* provides a detailed description of each variable in the data file.

Appendix A defines certain terms used in this document. *Appendix B* provides a detailed discussion about the computation of standard errors, and includes a list of the files needed for calculations, which are in a separate directory on this CD-ROM. *Appendix C* describes how to use the ICD-9-CM Addenda and Conversion Table. *Appendix D* gives a list of selected ICD-9-CM procedure codes and the years that they were first used. *Appendix E* shows a list and description of Census population files (also on this CD-ROM) that allow for the calculation of utilization rates. *Appendix F* provides weighted frequencies for selected variables for the purpose of verifying analyses.

TABLE OF CONTENTS

NOTE: The files referred to in Appendices B and E are provided on this CD-ROM in EXCEL format, and located in separate directories. Spreadsheets containing the parameter values for relative standard error curves (for use in calculating errors of NHDS estimates) are in the RSEs folder, and Census population estimates (for use in calculating utilization rates) are in the POPS folder.

I. DESCRIPTION OF THE NATIONAL HOSPITAL DISCHARGE SURVEY

INTRODUCTION. The National Hospital Discharge Survey (NHDS) has been conducted continuously by the National Center for Health Statistics (NCHS) since 1965. NHDS abstracts both demographic and medical information for inpatients discharged from non-federal, short-stay hospitals from a national sample of hospitals in the United States. Based on this information, national and regional estimates of characteristics of patients, lengths of stay, diagnoses, and surgical and non-surgical procedures in hospitals of various bed sizes and types of ownership are produced. The survey design, sampling, and estimation procedures were planned to produce calendar year estimates.

NHDS utilizes a stratified, multi-stage probability design. The original sample, drawn in 1965 and followed through 1987, was based on a two-stage sampling plan. A new sample was drawn in 1988, when a three-stage sampling plan was implemented and several data collection and estimation procedures were revised. The redesign of the survey is important, especially for those conducting trend analyses. Because the new survey differs from the original one in sample design, data collection, and estimation procedures, some of the differences between NHDS statistics based on the 1965-87 sample and statistics based on the sample drawn in 1988 may be due to the survey redesign rather than actual changes in hospital utilization. A report detailing pre- and post-redesign differences has been published (1).

Since 1979, the *International Classification of Diseases, 9th Revision, Clinical Modification*, (ICD-9-CM), has been used for classifying medical diagnoses and procedures in NHDS (2). Beginning in 1986, however, the ICD-9-CM has been modified annually. These modifications become effective in October of each year and are published in an Addendum. Users of NHDS who wish to conduct trend analyses or other multiple year studies must take into account the ICD-9-CM Addenda. The 1986-2006 ICD-9-CM Addenda and Conversion Table are included on this CD in Appendix C.

For a general description of the survey design and data collection procedures, see below. Detailed information on technical aspects of the survey has been published (1, 3, 4). Publications based on the data collected in each survey year can be obtained from the NCHS website: http://www.cdc.gov/nchs/nhds/nhds_products.htm.

HISTORY. In 1962, NCHS began exploring possibilities for conducting a survey to provide information on the utilization of the Nation's hospitals and on the nature and treatment of illness among the hospitalized population. A national advisory group was established, and NCHS undertook planning discussions with other officials of the Public Health Service. Hospitalization material from the Survey Research Center of the University of Michigan, the American Hospital Association, and the Professional Activities Study was examined and evaluated. In 1963, a study by the School of Public Health of the University of Pittsburgh under contract to NCHS demonstrated the feasibility of an NHDS type of program. An additional pilot study using enumerators from the Bureau of the Census was conducted in late 1964 and confirmed the University of Pittsburgh's findings. Finally, with advice and support from the American Hospital Association, the American Medical Association, individual experts, other professional groups, and officials of the U.S. Public Health Service, NCHS initiated the National Hospital Discharge Survey in 1964.

SOURCE OF THE DATA. The National Hospital Discharge Survey (NHDS) covers discharges from noninstitutional hospitals, exclusive of Federal, military, and Veterans Administration hospitals, located in the 50 States and the District of Columbia. Only short-stay hospitals (those with an average length of stay for all patients of less than 30 days) or those whose specialty is general (medical or surgical) or children's general are included in the survey. These hospitals must also have six or more beds staffed for inpatient use. These criteria, used since the survey redesign in 1988, differ slightly from those used under the original design. Prior to 1988, hospitals with an average length of stay of 30 days or more were excluded, regardless of specialty. However, the term "short-stay" continues to be used because 98 percent of the hospitals in NHDS universe fall into this category.

The original universe for the survey consisted of 6,965 short-stay hospitals contained in the 1963 National Master Facility Inventory of Hospitals. This list was updated periodically from lists of hospitals provided by the American Hospital Association. When the survey was redesigned in 1988, the NHDS sampling frame consisted of hospitals that were listed in the April 1987 SMG Hospital Market Data File (5), met the above criteria, and began accepting patients by August 1987. In 2002, the SMG database became the "Healthcare Market Index " and the "Hospital Market Profiling Solution", produced by Verispan LLC, and these products have served as the sampling frame since 2003. The hospital sample was updated in 1991, 1994, 1997, 2000, 2003 and 2006 to allow for hospitals that opened later or changed their eligibility status since the previous sample update. Table 1 (below) shows the number of hospitals in the NHDS universe and sample, as well as the approximate number of sampled abstracts and estimated number of discharges, for each year from 1979 to 2007. From 1984 to 1987, data on the universe of short-stay non-Federal hospitals was obtained from the American Hospital Association, instead of the Master Facility Inventory of Hospitals.

TABLE 1. Number of Hospitals in Universe and Sample, Hospital Participation and Respondent Counts, Approximate Number of Sampled Discharge Abstracts, and Estimated Number of Patients Discharged: National Hospital Discharge Survey, 1979-2007

Data year	Number of Hospitals in NHDS universe	Number of Hospitals in NHDS Sample	Number of Out-of-Scope Hospitals	Number of Refusal Hospitals	Number of Responding Hospitals	Approximate Number of Sampled Patient Abstracts [includes newborn infants]	Estimated Number of Discharges (in 1,000's) [excludes newborn infants]
1979	8,017	544	48	80	416	215,000	36,747
1980	8,017	544	52	72	420	224,000	37,832
1981	8,080	550	51	71	428	227,000	38,543
1982	8,080	550	53	71	426	214,000	38,594
1983	8,130	553	57	78	418	206,000	38,784
1984	6,023	553	60	86	407	192,000	37,162
1985	6,007	558	62	82	414	195,000	35,057
1986	6,007	558	65	75	418	193,000	34,255
1987	6,007	558	66	92	400	181,000	33,387
1988	6,400	542	11	109	422	250,000	31,146
1989	6,400	542	16	118	408	233,000	30,947
1990	6,400	542	23	45	474	266,000	30,788
1991	6,250	528	7	37	484	274,000	31,098
1992	6,250	528	14	20	494	274,000	30,951
1993	6,250	528	15	47	466	235,000	30,825
1994	6,337	525	13	34	478	277,000	30,843
1995	6,337	525	17	42	466	263,000	30,722
1996	6,337	525	18	27	480	282,000	30,545
1997	6,202	513	12	27	474	300,000	30,914
1998	6,202	513	18	17	478	307,000	31,827
1999	6,202	513	26	29	458	300,000	32,132
2000	6,078	519	38	47	434	313,000	31,706
2001	6,078	504	27	29	448	330,000	32,653
2002	6,078	504	30	29	445	327,000	33,727
2003	5,848	501	22	53	426	320,000	34,738
2004	5,848	501	25	37	439	371,000	34,864
2005	5,848	501	28	29	444	375,000	34,667
2006	6,010	501	23	40	438	376,000	34,854
2007	5,983	501	24	55	422	366,000	34,369

SAMPLING DESIGN. The purpose of the NHDS redesign in 1988 was to provide geographic comparability with other surveys conducted by NCHS; to update the sample of hospitals; and to use data already available in automated systems (3). The 1988 redesign added some complexity to the sampling plan of NHDS. Specifically, the redesigned survey employed a modified, three-stage sampling plan, whereas the original design was based on a two-stage sample.

Under the original design, the first-stage sampling units were hospitals, which were selected within 24 bedsize-by-region strata, controlling for type of ownership and Census geographic division. At the second stage, discharges within hospitals were selected by a systematic random sampling technique, using a sampling rate such that the overall probability of selecting a discharge was approximately the same in each hospital-size class.

In the 1988 redesign, units selected at the first stage of sampling consisted of either hospitals or geographic areas, the latter being primary sampling units (PSUs) from the 1985-94 National Health Interview Survey sample. Hospitals within PSU's were then selected at the second stage. Strata at this stage were defined by geographic region, PSU size, abstracting service status, PSU, and hospital specialty-size groups. Within these strata, hospitals were selected with probabilities proportional to their annual number of discharges. At the third stage, a sample of discharges was selected by a systematic random sampling technique. The sampling rate was determined by the hospital's sampling stratum and the type of data collection system (manual or automated) used.

Note again that these changes in the design of the survey may affect trend analyses. Some observed differences between NHDS statistics based on the 1965-87 sample and those based on the redesigned sample may be due to updating the sample and revising data collection and estimation procedures rather than actual changes in hospital utilization. A report comparing selected estimates obtained from the old and the new survey designs has been published (1).

DATA COLLECTION PROCEDURES. Originally, all data collection for NHDS was conducted manually within the hospital, either by hospital personnel or by Bureau of Census staff under contract with NCHS. Currently, approximately 55 percent of the responding hospitals utilize the manual system of sample selection and data abstraction. Of the hospitals using this system in 2007, about 23 percent had the work performed by their own medical records staff. In the remaining hospitals using the manual system, personnel of the U.S. Bureau of the Census did the work on behalf of NCHS. The completed forms, along with sample selection control sheets, were forwarded to NCHS for coding, editing, and weighting.

Beginning in 1985, data from some hospitals were obtained from commercial abstracting services, state computerized data systems, or a hospital's own computer system. Files from these sources contained machine-readable medical record data from which records were systematically sampled by NCHS. In 2007, this method was used for about 45 percent of the respondent hospitals.

Both the medical abstract form used in manual data collection and the automated data contain items relating to the personal characteristics of the patient, including birth date or age, sex, race, and marital status; administrative information, including discharge status, admission date and discharge date; and medical information, including up to seven diagnoses and up to four surgical or non-surgical procedures. Since 1977, patient zip code, expected source of payment, and dates of surgery have also been collected. In 2001, two new variables were added to the abstract form: TYPE OF ADMISSION and SOURCE OF ADMISSION. Certain data elements collected are confidential and are not made available to the public, including date of birth and patient zip code. As noted, the medical abstract form (HDS-1) undergoes periodic updating. Two recent versions used in 2001 and 2007 are included as PDF files on this CD.

THE UNIFORM HOSPITAL DISCHARGE DATA SET (UHDDS). Starting with 1979 data, NHDS has followed guidelines of the Uniform Hospital Discharge Data Set (UHDDS) within the confines of its contractual agreement with participating hospitals. The UHDDS is a uniformly defined, minimum data set (6). Items for the data set were selected on the basis of their usefulness to a broad range of organizations and agencies requiring hospital information, uniformity of definition, and general availability from medical records and abstract services.

MEDICAL CODING AND EDIT. The medical information manually recorded on the sampled patients' abstracts was coded centrally by NCHS staff using the ICD-9-CM. A maximum of seven diagnostic codes was assigned for each sampled abstract; in addition, if the medical information included surgical or non-surgical procedures, a maximum of four codes for these procedures was assigned. Since 1991, all of the diagnostic and procedure codes in the ICD-9-CM have been utilized. However, for the years 1979 to 1990, some procedure codes were not utilized and so it is not possible to produce estimates for those codes. Appendix C contains a listing of the procedures not coded and the year the code was first used. It is important to note that the ICD-9-CM serves as a basis for classifying morbidity information on medical records, and as a tool for generating basic health statistics. As it is used in NHDS, it is not intended to provide a complete clinical picture of a patient. More information about the ICD-9-CM can be found at: http://www.cdc.gov/nchs/icd.htm.

NHDS usually presents diagnoses and procedures in the order they were listed on the abstract form or obtained from abstract services, however, there were exceptions. For women discharged after childbirth, a code of V27 from the supplemental classifications was entered as the first-listed code, with a code designating either normal or complicated delivery in the second-listed position. In another exception, a decision was made to reorder some acute myocardial infarction (AMI) diagnoses. If an acute myocardial infarction was listed with other circulatory diagnoses and was other than the first entry, it was reordered to the first position. If a code from Chapter 16 (Symptoms, Signs, and Ill-Defined Conditions), appeared as a first-listed code and a diagnosis appeared as a secondary code, the diagnosis code was moved to the first position and the symptom code was moved back.

In 2002, the ICD-9-CM Coordination and Maintenance Committee created procedure Chapter 00 – Procedures and Interventions, Not Elsewhere Classified – as a way of handling space limitations in the existing hierarchical structure and alleviating inappropriate categorization of new procedures. Since October addendum changes are not implemented in NHDS until the following data collection year, 2003 was the first year these codes were used.

Following conversion of the information on the medical abstracts to a computer file and combining it with the automated data files, a final medical edit was accomplished by computer inspection and by a manual review of rejected records. Priority was given to medical information in the editing.

MEASUREMENT ERRORS. As in any survey, results were subject to nonsampling or measurement errors, which included errors due to hospital nonresponse, missing abstracts, information incompletely or inaccurately recorded on abstract forms, and processing errors. In general, less than one percent of the discharge records failed to include the age or sex of the patient. If the hospital record did not state the age or sex of the patient, it was imputed by assigning the patient an age or sex consistent with the age or sex of other sampled patients with the same first-listed diagnosis code.

Data on race are missing for about 15 percent of all discharges from all years, but this varies by year. Except for one year, no attempt was made to impute for these missing values. In 1981, "race not stated" values were imputed so there are no "not stated" cases for that year. Details about the underreporting of race in NHDS can be found at: http://www.cdc.gov/nchs/data/ad/ad265.pdf.

For data years before 1996, if dates of admission or discharge were not given, and if they could not be obtained from the monthly sample listing sheet transmitted by the sample hospital, a length of stay was imputed by assigning the patient a stay characteristic of the stays of other patients of the same age. For records where the length of stay and the discharge month were known, a discharge day of the 20th of the month was assigned to the record, and the admission date was computed based on the given length of stay.

In 1996, an updated edit program was developed and implemented. The updated program followed the same general specifications as the previous edit program and was designed to make as few changes as possible in the data. However, there may be some minor anomalies that would be apparent when examining data over time, performing trend analyses, or examining combinations of variables. Particular features of the edit program implemented in 1996 that may affect certain variables are:

- An improved imputation procedure for missing age and sex data was developed, which maintains the known distribution of these variables, according to categories of the first-listed diagnosis.
- There is no longer a re-ordering of the procedure codes.
- Principal and additional expected sources of payment are no longer re-ordered, with one exception: "Self-Pay" is listed as the principal source only if there are no other sources, or the only other source is "Not Stated"; otherwise it must be listed after every other source (except "Not Stated").
- An arbitrary month of admission is no longer assigned to records received from abstract services that do not provide the exact date of admission and discharge.
- Beginning in the 2004 data collection year, if a hospital failed to provide month of discharge but did provide the quarter of discharge, a discharge month within the quarter was sequentially assigned to each record. For example, for discharges within the first quarter, a discharge month of January, February, or March was assigned.

Other edit and imputation procedures may have been applied to data received in automated form.

SAMPLING ERRORS AND ROUNDING OF NUMBERS. The standard error is primarily a measure of sampling variability that occurs by chance because only a sample rather than the entire universe is surveyed. The relative standard error (RSE) of an estimate is obtained by dividing the standard error by the estimate itself. When the resulting value is multiplied by 100, the relative standard error is expressed as a percent of the estimate. The RSE is used as a guide to the reliability of the estimate (see ***Presentation of Estimates*** below).

Since 1988, estimates of sampling variability have been calculated with SUDAAN software, which computes standard errors by using a first-order Taylor series approximation of the deviation of estimates from their expected values. A description of the software and its approach was published by Bieler and Williams (7). Before 1988, standard error estimates were produced using a computerized routine based on a rigorously unbiased algebraic estimator of the variance.

To obtain standard errors that would be applicable for a wide variety of statistics and that could be prepared at a moderate cost, numerous variances were calculated and a best fit formula was derived. This formula, which is based on an empirically determined relationship between the size of an estimate, X, and its relative variance, was used to produce generalized variance curves. These curves provide approximations to the relative standard errors that are applicable to estimates of discharges, first- or all-listed diagnoses, all-listed procedures, and days of care, either aggregated or disaggregated by selected patient or hospital characteristics.

For the years 1979 through 1987, curves are represented in tables containing estimates of different sizes and their approximate relative standard errors. Linear interpolation is then used to obtain the RSE for a specific estimate. For the years 1988 through 2007, tables contain parameter values that can be substituted in a mathematical formula to produce approximate relative standard errors. Instructions on how to use the tables and/or the parameter values are given in Appendix B.

PRESENTATION OF NHDS ESTIMATES. Based on consideration of the complex sample design of NHDS, the following guidelines are recommended for using and reporting NHDS estimates:

- If the sample size is less than 30, the value of the estimate is not reported.
- If the sample size is 30-59, the value of the estimate is reported but should not be assumed reliable.
- If the sample size is 60 or more and the relative standard error is less than 30 percent, the estimate may be reported.
- If the relative standard error of any estimate is over 30 percent, the estimate is considered to be unreliable. It is left to the author to decide whether or not to report it. However, if the author chooses to present the unreliable estimate, the consumer of the statistic must be informed that the statistic is not reliable.

POPULATION ESTIMATES. Hospital utilization rates are computed using U.S. Census Bureau population estimates as denominators. Before 1981, rates of discharges and days of care that appeared in published reports from NCHS were calculated using estimates of the civilian non-institutional population (CNP). However, beginning in 1981, estimates of the civilian resident population (CRP) were used to calculate hospital utilization rates. The CRP was determined to be more appropriate because persons in institutions, for example nursing home patients, are hospitalized when necessary. A report has been published which discusses differences in discharge rates based on the different denominators (8).

Files containing estimates of the civilian resident population as of July 1 of each year from 1979 to 2007 are provided with this documentation. Population estimates for 1979 were adjusted based on the 1980 census. The estimates for 1980-1989 have been adjusted based on the 1990 decennial census. Population estimates for 1990-1999 have been adjusted for net underenumeration using the 1990 National Population Adjustment Matrix. Population estimates for 2000-2007 are based on results of the 2000 Census.

Note that rates calculated with the 2000-based population estimates may differ slightly from those appearing in published NCHS reports for 2000 or those calculated from population estimates disseminated with the 2000 NHDS data file and documentation. At the time of release of those reports and data files, only 1990-based Census estimates for 2000 were available.

Because of new federal guidelines implemented in the 2000 Census which regulate the reporting of race data, population estimates by race based on the 2000 Census are not directly comparable with estimates from earlier censuses. See Appendix E for further explanation.

MONTHLY AND SEASONAL ESTIMATES CAUTION. An important difference between the old and new designs is the method used to adjust for nonresponse. The result of this difference is that monthly and seasonal estimates under the new design may be skewed. While the effect is believed to be small, it is recommended that partial year estimates not be produced for 1988 and later years. The reasons for this are explained below.

In the old design, weights for responding hospitals were adjusted each month to account for hospitals that did not respond for that month. In the new design, the type of nonresponse adjustment applied depended on whether the hospital was considered a nonrespondent or a partial respondent. A nonresponding hospital was one that failed to provide at least half of the expected number of discharges for at least half of the months for which it was inscope. In this case, weights of discharges from hospitals similar to the nonresponding hospital were inflated to account for discharges of the nonrespondent hospital. However, this adjustment was performed just once, after the close out of the survey for the year, instead of monthly as before.

For partially responding hospitals, one or both of two adjustments were made. If the hospital provided at least half, but not all, of the expected number of abstracts for a given month, the weights of the abstracts actually collected for that month were inflated to account for the missing abstracts. If fewer than half of the expected number of abstracts were provided, the weights of the abstracts provided were inflated by a factor of two, then another adjustment was made to account for the excess nonresponse. In the second adjustment, the weights of the discharges in the hospital's respondent months were inflated by ratios that varied by category of first-listed ICD-9-CM diagnostic code. This adjustment ratio was based on the hospital's month(s) of nonresponse and the month-by-month distributions of first-listed diagnostic groups among discharges from hospitals that responded for all twelve months. The ratio accounts for the seasonality in the occurrence of the first-listed diagnostic groups for annual statistics, but not for partial year estimates. In the 2007 NHDS, 93 percent of the 422 responding hospitals provided data for all twelve months, and 98 percent provided at least nine months of data.

CONFIDENTIALITY. Persons using the public use file agree to abide by the confidentiality restrictions that accompany use of the data. Specifically, they agree that, in the event of inadvertent discovery of the identity of any individual or establishment, then: (a) no use will be made of this knowledge; (b) the director of NCHS will be advised of the incident; (c) the information that would identify the individual or establishment will be safe-guarded or destroyed, as requested by NCHS; and (d) no one else will be informed of the discovered identity.

Maintaining the confidentiality of survey respondents, whether individuals or establishments, is a responsibility of NCHS as described in section 308(d) of the Public Health Service Act. As such it may be necessary for NCHS to block the release of data or modify variables that may, because of their unique nature, lead to inadvertent disclosure of the identity of a participating facility or respondent.

HOW TO USE THE DATA FILE. NHDS records contain weights to allow inflation to national or regional estimates. The weight for each record is found in location 21-25. To produce an estimate of the number of discharges, the weights for the desired records must be summed. To produce an estimate for number of days of care, the weight must be multiplied by the days of care (location 13-16) and these products summed. Estimates apply to the calendar year (January-December). Appendix F contains weighted frequencies for selected variables. These may be used as a cross-check when analyzing NHDS data.

QUESTIONS. Questions concerning NHDS data should be directed to:

Centers for Disease Control and Prevention
National Center for Health Statistics
Division of Health Care Statistics
Ambulatory and Hospital Care
3311 Toledo Road
Hyattsville, Maryland 20782
Phone: 301.458.4321
Fax: 301.458.4032
Email: NHDS@cdc.gov

For more information about NHDS, including links to publications and public-use data files, visit the NCHS website: http://www.cdc.gov/nchs/nhds.htm

For email discussions and dissemination of NHDS data, join the Hospital Discharge and Ambulatory Surgery Data listserv (HDAS-DATA). To join, in the body of an email message (leaving the subject line blank), type: subscribe hdas-data Your Name
Then send this message to: listserv@cdc.gov

REFERENCES

1. Haupt B, Kozak LJ. Estimates from Two Survey Designs: National Hospital Discharge Survey. National Center for Health Statistics. Vital and Health Stat 13 (111). 1992. http://www.cdc.gov/nchs/data/series/sr_13/sr13_111.pdf

2. U.S. Department of Health and Human Services. Centers for Disease Control and Prevention, Centers for Medicare and Medicaid Services. Official version International Classification of Diseases, Ninth Revision, Clinical Modification, Sixth Edition. DHHS Pub No. (PHS) 06-1260

3. Dennison CF, Pokras R. Design and Operation of the National Hospital Discharge Survey: 1988 Redesign. National Center for Health Statistics. Vital and Health Stat 1 (39). 2000. http://www.cdc.gov/nchs/data/series/sr_01/sr01_039.pdf

4. 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. www.cdc.gov/nchs/data/series/sr_02/sr02_039.pdf

5. SMG Marketing Group, Inc. Hospital Market Database. Chicago: Healthcare Information Specialists. April 1987, April 1991, April 1994, April 1997, April 2000; Verispan, L.L.C. Healthcare Market Index and Hospital Market Profiling Solution, 2003 and 2006.

6. Office of the Secretary, Department of Health and Human Services: Health Information Policy Council: 1984 Revision of the Uniform Hospital Discharge Data Set. Federal Register, Volume 50, No. 147. July 31, 1985.

7. Bieler GS, Williams RL. Analyzing Survey Data Using SUDAAN Release 7.5. Research Triangle Institute: Research Triangle Park, N.C. 1997.

8. Pokras R, Kozak LJ. Adjustment of Hospital Utilization Rates--United States, 1965-80. National Center for Health Statistics. Vital and Health Stat 13(81). 1984. http://www.cdc.gov/nchs/data/series/sr_13/sr13_081.pdf

II. TECHNICAL DESCRIPTION OF DATA FILE

Data Set Name (Non-newborns, 1979-1989)	NHDS7989.NOTNB.TXT
Data Set Name (Non-newborns, 1990-1999)	NHDS9099.NOTNB.TXT
Data Set Name (Non-newborns, 2000-2007)	NHDS0007.NOTNB.TXT
Data Set Name (Newborns, 1979-2006)	NHDS7907.NEWBORN.TXT
Record Length	87
Number of Records (Non-newborns, 1979-1989)	2,112,113
Number of Records (Non-newborns, 1990-1999)	2,473,112
Number of Records (Non-newborns, 2000-2007	2,477,365
Number of Records (Newborns, 1979-2007)	825,075

III. RECORD LAYOUT: LOCATION AND CODING OF DATA ELEMENTS

This section provides detailed information for each sampled record on the file, with a description of the coding of each item included on the record. Data elements are arranged sequentially according to their physical location on the record. Unless otherwise stated in the Item Description, the data are derived from the abstract form or from automated sources. The SMG Hospital Market Data File and the hospital interview are alternate sources of data; some other items are computer generated.

NATIONAL HOSPITAL DISCHARGE SURVEY
1979-2007 MULTI-YEAR DATA FILE: LAYOUT AND CODING OF DATA ITEMS

Item Number	Location	Number of Positions	Item description	Code description
1	1-2	2	Last 2 digits of survey year	79-07=1979 to 2007
2	3	1	Newborn status	1=Newborn 2=Not newborn
3	4	1	Units for age	1=Years 2=Months 3=Days
4	5-6	2	Age in years, months, or days	If units=years: 00-99* If units=months: 01-11 If units=days: 00-31 *Ages 100 and over were recoded to 99
5	7	1	Sex	1=Male 2=Female
6	8	1	Race	1=White 2=Black 3=American Indian/Alaskan Native 4=Asian 5=Native Hawaiian/Other Pacific Islander 6=Other 8=Multiple race indicated (begun in 2000) 9=Not stated ***NOTES**: In 1979, only 1, 2, and 9 are available. In 1981, "not stated" values were imputed so the value "9" does not occur for that year. Before 2000, "Asian" and "NH/OPI" were grouped together, so the value "5" for "NH/OPI" is missing prior to 2000.
7	9	1	Marital status	1=Married 2=Single 3=Widowed 4=Divorced 5=Separated 6=Unknown (only from 1979-1995) 9=Not stated ***NOTE**: From 1979-1995, "unknown" cases were coded as value 6 and "not stated" cases were coded as value 9. Starting in 1996, no distinction was made between "unknown" and "not stated", so all cases of unknown or unstated marital status were coded as value 9.

Item Number	Location	Number of Positions	Item description	Code description
8	10-11	2	Admission month	01-12=January to December 99=Missing (beginning in 1996)
9	12	1	Discharge status	1=Routine/discharged home 2=Left against medical advice 3=Discharged/transferred to short-term facility 4=Discharged/transferred to long-term care institution 5=Alive, disposition not stated (not coded in 1979 and 1980) 6=Dead 9=Not stated or not reported ***NOTE:** In 1979 and 1980, the value “3” indicated “Discharged/transferred to unspecified facility” and the value “4” indicated “Discharged/transferred to organized home care”.
10	13-16	4	Days of care	Use to calculate number of days of care. Values of zero generated by the computer from admission and discharge dates were changed to one. (Discharges for which dates of admission and discharge are the same are identified in item 11 below)
11	17	1	Length of stay flag	0=Less than 1 day 1=One day or more
12	18	1	Geographic region	1=Northeast 2=Midwest 3=South 4=West
13	19	1	Number of beds, recode	1=6-99 2=100-199 3=200-299 4=300-499 5=500 and over
14	20	1	Hospital ownership	1=Proprietary 2=Government 3=Nonprofit, including church
15	21-25	5	Analysis weight	Use to obtain weighted estimates
16	26-27	2	First two digits of survey year	Either 19 or 20

Item Number	Location	Number of Positions	Item description	Code description
17	28-32	5	Diagnosis code #1	*
18	33-37	5	Diagnosis code #2	*
19	38-42	5	Diagnosis code #3	*
20	43-47	5	Diagnosis code #4	*
21	48-52	5	Diagnosis code #5	*
22	53-57	5	Diagnosis code #6	*
23	58-62	5	Diagnosis code #7	*
24	63-66	4	Procedure code #1	*
25	67-70	4	Procedure code #2	*
26	71-74	4	Procedure code #3	*
27	75-78	4	Procedure code #4	*
28	79-80	2	Discharge month	01-12=January to December
29	81-82	2	Principal expected source of payment	01=Worker's comp 02=Medicare 03=Medicaid 04=Other government 05=Blue Cross/Blue Shield 06=HMO/PPO (added in 1998) 07=Other private 08=Self-pay 09=No charge 10=Other 99=Not stated
30	83-84	2	Secondary expected source of payment	Same coding as item 29 above
31	85	1	Type of Admission	1 = Emergency 2 = Urgent 3 = Elective 4 = Newborn 9 = Not available

Item Number	Location	Number of Positions	Item description	Code description
32	86-87	2	Source of Admission	01 = Physician referral 02 = Clinical referral 03 = HMO referral 04 = Transfer from a hospital 05 = Transfer from skilled nursing facility 06 = Transfer from other health facility 07 = Emergency room 08 = Court/law enforcement 09 = Other 99 = Not available

* Diagnosis and procedure codes are in compliance with the ***International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM)*** *(2)*. For diagnosis codes, there is an implied decimal between the 3rd and 4th digits. For E-codes, the implied decimal is between the 4th and 5th digits. For inapplicable 4th or 5th digits of diagnosis codes, a dash (-) is inserted. For procedure codes, there is an implied decimal between the 2nd and 3rd digits. For inapplicable 3rd or 4th digits of procedure codes, a dash (-) is inserted.

APPENDIX A:

DEFINITION OF TERMS

Terms relating to hospitals and hospitalization

Hospitals: Short stay hospitals or hospitals whose specialty is general (medical or surgical), or children's general. Hospitals must have 6 beds or more staffed for patients use. Federal hospitals and hospital units of institutions are not included.

Type of ownership of hospital: The type of organization that controls and operates the hospital. Hospitals are grouped as follows:

> **Not for Profit**: Hospitals operated by a church or another not for profit organization.
>
> **Government**: Hospitals operated by State and local government.
>
> **Proprietary**: Hospitals operated by individuals, partnerships, or corporations for profit.

Bed size of hospital: Size is measured by the number of beds, cribs, and pediatric bassinets regularly maintained (set up and staffed for use) for patients, not including bassinets for newborn infants. The classification of hospitals by bed size is based on the number of beds at or near midyear as reported by the hospital.

Patient: A person who is formally admitted to the inpatient service of a short-stay hospital for observation, care, diagnosis, or treatment, or by birth.

Discharge: The formal release of a patient by a hospital; that is, the termination of a period of hospitalization by death or by disposition to place of residence, nursing home, or another hospital. The terms "discharges" and "patients discharged" are used synonymously.

Discharge rate: The ratio of the number of hospital discharges during the year to the number of persons in the civilian population on July 1 of that year.

Days of care: The total number of patient days accumulated at time of discharge by patients discharged from short- stay hospitals during a year. A stay of less than 1 day (patient admission and discharge on the same day) is counted as 1 day in the summation of total days of care. For patients admitted and discharged on different days, the number of days of care is computed by counting all days from (and including) the date of admission to (but not including) the date of discharge.

Rate of days of care: The ratio of the number of patient days accumulated at time of discharge to the number of persons in the civilian population on July 1 of that year.

Average length of stay: The total number of days of care accumulated at time of discharge by patients discharged during the year, divided by the number of patients discharged.

Terms relating to diagnoses and procedures

Discharge diagnoses: One or more diseases or injuries (or some factor that influences health status and contact with health services that is not itself a current illness or injury) listed by the attending physician on the medical record of a patient. In NHDS, discharge (or final) diagnoses listed on the face sheet (summary sheet) of the medical record are transcribed in the order listed. Each sample discharge is assigned a maximum of seven five-digit codes according to ICD-9-CM (2).

Principal diagnosis: The condition established after study to be chiefly responsible for occasioning the admission of the patient to the hospital for care.

First-listed diagnosis: The coded diagnosis identified as the principal diagnosis or listed first on the face sheet of the medical record if the principal diagnosis cannot be identified. The number of first-listed diagnoses is equivalent to the number of discharges.

Procedure: One or more surgical or non-surgical operations, procedures, or special treatments listed by the physician on the medical record. In NHDS, all terms listed on the face sheet (summary sheet) of the medical record under the caption "operation," "operative procedures," "operations and/or special treatment," and the like are transcribed in the order listed. A maximum of four procedures are coded.

Rate of procedures: The ratio of the number of all-listed procedures during a year to the number of persons in the civilian population on July 1 of that year determines the rate of procedures.

Demographic terms

Age: Refers to the age of the patient on the birthday prior to admission to the hospital inpatient service.

Population: Civilian population is the resident population excluding members of the Armed Forces.

Geographic regions: Hospitals are classified by location in one of the four geographic regions of the United States corresponding to those used by the U.S. Bureau of the Census:

U.S. CENSUS REGIONS

NORTHEAST	MIDWEST	SOUTH	WEST
Maine	Michigan	Delaware	Montana
New Hampshire	Ohio	Maryland	Idaho
Vermont	Illinois	District of Columbia	Wyoming
Massachusetts	Indiana	Virginia	Colorado
Connecticut	Wisconsin	West Virginia	New Mexico
Rhode Island	Minnesota	North Carolina	Arizona
New York	Iowa	South Carolina	Utah
New Jersey	Missouri	Georgia	Nevada
Pennsylvania	North Dakota	Florida	Washington
	South Dakota	Kentucky	Oregon
	Nebraska	Tennessee	California
	Kansas	Alabama	Hawaii
		Mississippi	Alaska
		Arkansas	
		Louisiana	
		Oklahoma	
		Texas	

APPENDIX B:

COMPUTATION OF STANDARD ERRORS

The standard error is primarily a measure of sampling variability that occurs by chance because only a sample rather than the entire universe is surveyed. The relative standard error of an estimate is obtained by dividing the standard error by the estimate itself. When the resulting value is multiplied by 100, the relative standard error is expressed as a percent of the estimate.

Before 1988, standard error estimates were produced using a computerized routine based on a rigorously unbiased algebraic estimator of the variance. Since 1988, estimates of sampling variability have been calculated with SUDAAN software, which computes standard errors by using a first-order Taylor series approximation of the deviation of estimates from their expected values. A description of the software and its approach was published by Bieler and Williams (7). Use of SUDAAN with NHDS data for standard error calculation may be available to researchers outside of the CDC/NCHS community through the NCHS Research Data Center (RDC). More information can be found on the NCHS RDC website: http://www.cdc.gov/rdc/.

In order to obtain standard errors that would be applicable for a wide variety of statistics and that could be prepared at a moderate cost, a method using generalized variance curves was developed. Numerous variances were calculated and a best fit formula was derived which was based on an empirically determined relationship between the size of an estimate, X, and its relative variance. The relative standard error was then obtained by taking the square root of the relative variance. These generalized variance curves provided approximations to the relative standard errors that were applicable to estimates of discharges, first- or all-listed diagnoses, all-listed procedures, and days of care, either aggregated or disaggregated by selected patient or hospital characteristics.

This appendix contains information needed to produce generalized errors for NHDS statistics. For the years 1979 through 1987, curves are represented in tables containing selected estimates of different sizes and their approximate relative standard errors. For the years 1988 through 2006, tables containing parameter values for relative standard error curves are given. The RSE tables and curves for each data year are contained in the folder named RSES on this CD. Filenames are of the form RSEXXXX.XLS, where XXXX is replaced by the data year desired. Instructions on how to use this information follow.

COMPUTATION OF RELATIVE STANDARD ERRORS FOR AGGREGATE ESTIMATES, 1979 THROUGH 1987 DATA YEARS

For each of the years from 1979 through 1987, this CD contains spreadsheet files of approximate relative standard errors (RSEs) for estimates of various sizes for discharges, first- or all-listed diagnoses, all-listed procedures, and days of care. Because RSEs may differ depending on the characteristic being estimated, more than one RSE curve is usually presented. Please be aware there are different RSE curves for diagnoses, days of care and procedures. Also, since it is not possible to provide exact standard errors for every size estimate, it is necessary to use arithmetic interpolation to obtain the RSE for an estimate not included in the table. Linear interpolation is used for simplicity and without loss of accuracy, even though the curves are not strictly speaking linear functions.

For example, in 1979 the estimated number of appendectomies (ICD-9-CM code 47.0) performed on patients 15 years and older discharged from short-stay hospitals was 232,000. Referring to the file, RSE1979.xls, there is no curve by age, so the one entitled "All Other Characteristics" is used. Also, the estimate 232,000 is not listed, so in order to obtain an approximate RSE, arithmetic interpolation is performed as follows.

Step 1. In the column headed "Size of Estimate", locate the two adjacent values between which the estimate of interest is located. In this example, they would be 100,000 and 250,000.

Step 2. For these estimates, compute estimated standard errors, using the corresponding RSEs from the column headed "All Other Variables".

SE (100,000) = 9.9% * 100,000 = 9,900
SE (250,000) = 8.6% * 250,000 = 21,500

Step 3. Calculate the proportional part of the interval between 100,000 and 250,000 which falls between 100,000 and 232,000.

P = (232,000 - 100,000) / (250,000 - 100,000) = .88

Step 4. Calculate the estimated standard error of 232,000 by subtracting the proportional part of the interval between the two standard errors from the standard error of 100,000.

SE (232,000) = SE (100,000) - P * (SE(100,000) - SE(250,000))
SE (232,000) = 9,900 - .88 * (9,900 - 21,500) = 20,108

The relative standard error can be obtained by dividing the standard error by the estimate:

RSE (232,000) = 20,108 / 232,000 = .087

When multiplied by 100, the RSE is expressed as a percent of the estimate (i.e. 8.7%).

The standard error can be employed to generate confidence intervals for statistical testing. In this example, the two-tailed, 95% confidence interval for the estimate of appendectomies for inpatients aged 15 and older in 1979 is:

LOWER LIMIT: 232,000 - 1.96 * 20,108 = 193,000
UPPER LIMIT: 232,000 + 1.96 * 20,108 = 271,000

COMPUTATION OF RELATIVE STANDARD ERRORS FOR PERCENTS, 1979 THROUGH 1987 DATA YEARS

The relative standard error of a percent in which both the numerator (X) and denominator (Y) are from NHDS is estimated by:

$$\mathbf{RSE(X/Y) = SQRT\ \{[SE(X)^2 / X^2] - [SE(Y)^2 / Y^2]\}}$$

To verbally clarify this formula, the RSE(X/Y) is obtained by taking the square root of the difference between two quantities. The first of the two quantities is obtained by dividing the squared standard error of X by X-squared; the second of the two quantities is obtained by dividing the squared standard error of Y by Y-squared. When RSE(X/Y) is multiplied by 100, then RSE(X/Y) is expressed as a percent of the estimate.

For example, the estimated 232,000 appendectomies performed on patients aged 15 years and older represent 74.6% of the estimated 311,000 appendectomies in 1979. To compute the relative standard error of this percent, the standard errors of both the numerator and the denominator are needed. The standard error of the numerator is given above. The standard error of the denominator can be calculated using the procedure described in the preceding section and is found to be 25,770. Using these figures in the formula gives:

RSE (.746) = SQRT [(20,108^2 / 232,000^2) - (25,770^2 / 311,000^2)] = .025

Expressed as a percent, RSE (.746) = 2.5%.

The standard error of the percent can be obtained by multiplying the percent by its RSE:
SE (.746) = RSE (.746) * .746 = .025 * .746 = .019

The standard error can be employed to generate confidence intervals around the estimate, as shown above.

COMPUTATION OF RELATIVE STANDARD ERRORS FOR AGGREGATE ESTIMATES, 1988 THROUGH 2007 DATA YEARS

The relative standard error of an estimate, RSE(X), may be calculated from the formula:

RSE(X) = SQRT (a + b/X)

with a and b provided in the accompanying files. When multiplied by 100, RSE(X) is expressed as a percent of X.

For example, in 1992 the estimated number of discharges from short-stay hospitals for females with a first-listed diagnosis of atherosclerotic heart disease (ICD-9-CM code 414.0) was 130,000. Using the file, RSE1992.xls, in Appendix D for estimates by sex, the value of a is .00089 and the value for b is 404.530. Thus,
RSE (130,000) = SQRT [.00089 + (404.530 / 130,000)] = .0633

Expressed as a percent, RSE (130,000) = 6.33%.

The standard error of the estimate is obtained by multiplying the relative standard error by the estimate itself:
SE (130,000) = 130,000 * .0633 = 8,229

The standard error can be employed to generate confidence intervals for statistical testing. In this example, the two-tailed, 95% confidence interval for the estimate of female inpatients with a first-listed diagnosis of atherosclerotic heart disease in 1992 is:
LOWER LIMIT: 130,000 - 1.96 * 8,229 = 114,000
UPPER LIMIT: 130,000 + 1.96 * 8,229 = 146,000

COMPUTATION OF RELATIVE STANDARD ERRORS FOR ESTIMATES OF PERCENTS, 1988 THROUGH 2007 DATA YEARS

Approximate relative standard errors for estimates of percents may be calculated from the tables in Appendix D also. The relative standard error for a percent, 100 * p ($0<p<1$), may be calculated using the formula:

RSE (p) = SQRT [b * (1 - p)/(p * X)]

where 100 * p is the percent of interest, X is the base of the percent, and b is the parameter b in the formula for approximating the RSE(X). Values for b are given in the accompanying files.

For example, in 1992 the estimated number of discharges from short-stay hospitals which were female was 18,545,000. This is 59.9 percent of the estimated 30,951,000 discharges for that year. Using the file, RSE1992.xls, in Appendix D for estimates by sex, the value of b is found to be 404.530. Thus,

RSE (.599) = SQRT [404.530 * (1 - .5899) / (.599 * 30,951,000)] = 0.00296

The relative standard error for the estimate of interest is 0.00296. Expressed as a percent, RSE (.599) = .296%. From this the standard error is obtained by multiplying the relative standard error by the estimate:

SE (.599) = .599 * 0.00296 = 0.00177.

The standard error can be employed to generate confidence intervals for statistical testing, as shown above.

COMPUTATION OF RELATIVE STANDARD ERRORS OF RATES IN WHICH THE DENOMINATOR HAS NO SAMPLING ERROR

It is generally assumed that population estimates which are obtained from the Bureau of the Census for certain overall totals, such as the U.S. population and subgroups disaggregated by age, sex, race, and region, are not subject to sampling error or that the error may be small enough to be considered negligible. The relative standard error of rates formed with these populations as the denominator is the relative standard error of the numerator. Thus, to obtain the standard error of the rate, simply multiply the rate itself by the RSE of the numerator.

COMPUTATION OF RELATIVE STANDARD ERRORS FOR MULTIPLE YEAR ESTIMATES, 1979 THROUGH 2007

This section presents procedures which may be used to approximate sampling errors of estimates based on multiple years of data collected under either or both of the 1965 and 1988 NHDS sample designs. These procedures will permit approximation of variances for multi-year estimates based on the 1988 sample design. It is believed the approximations are conservative and, hence, should not indicate significance when the contrary is actually true.

VARIANCE OF ESTIMATED TOTALS FOR MULTIPLE YEARS

For discussion purposes, let X' equal the multi-year aggregate estimate for the characteristic of interest. That is, X' = sum of annual aggregate estimates for any characteristic of interest for any number of years (say, Y years).

If the annual estimates are all from years between 1965 and 1987, then X'65 = sum of annual aggregate estimates for any characteristic of interest for any number of years between 1965 and 1987.

If the annual estimates are all from the years 1988 to the present, then X'88 = sum of annual aggregate estimates for any characteristic of interest for any number of years between 1988 and 2007.

If the annual estimates are from years in both the 1965 and 1988 samples, then X' = X'65 + X'88.

The procedure for approximating the variance of X' differs depending on the set Y of years in the database for X'.

SITUATION A:

When the database includes only years from the period 1965-1987, the following formula is used:

VAR(X'65) = (X'65)^2 * [RSE(X'65|Y years)]^2 (1)

where

RSE(X'65|Y years) = [SQRT (R|Y years)] * max [RSE (X'65|1 year)] (2)

and

max [RSE (X'65|1 year)] = maximum relative standard error over Y years calculated as though X'65 were based only on data from the single year, Y.

(Instructions for determining the RSE for a single year estimate are given above.)

The value R in (2) above is calculated as follows:
(Exhibit B1 in this Appendix presents a derivation of the R factor).

R|Y years = [(alpha * average-n) + 500] / [alpha * (average-n + 500)] (3)

Where
alpha = the number of years from the 1965-1987 period which are included in the data base for X'65
average-n = the average number of sampled discharges per hospital per year in the data base for X'65 (Exhibit B1 explains the calculation of average-n).

For example, assume an estimated 10,000 discharges during the data years 1986-1987 were to patients receiving surgery "S" during their stays. Here,

X'65 = 10,000 discharges, Y = {1986, 1987}, and alpha = 2.

RSE (X'65|1 year, 1986) = .102 (From published RSE curves)

RSE (X'65|1 year, 1987) = .105 (From published RSE curves)

Hence, max [RSE (X'65|1 year)] = max [.102, .105] = .105

Using data from Exhibit B2 in this Appendix, the average number of discharges per respondent hospital per year is calculated as:

average-n = (193,000 + 181,000) / (418 + 400) = 457.2

The use of alpha and the above in equation (3) gives:

R|Y years = [(alpha * average-n) + 500] / [alpha * (average-n + 500)]
= [2 (457.2) + 500] / [2 (457.2 + 500)]
= 1,414.4 / 1914.4
= .739

Substituting in equation (2),

RSE(X'65|2 years) = [SQRT (.739)] * .105 = .09

And, by equation (1), the variance of the multi-year aggregate X'65 is

Var (X'65|years 1986 & 1987) = $\{[(10{,}000)^2] * [(.09)^2)]\}$ = 810,000.

SITUATION B:

When the database includes only years from the period 1988 through the present, then equation (1) above is used, but the calculation of the single year RSE is different. The following formula is used to calculate the RSE of any single year estimate from 1988 through 2007.

RSE(X'88 |1 year) = SQRT (a + b/X) (4)

The parameters a and b are given in the files accompanying this document. (This approximation for the variance of X'88 ignores any reductions in the variance which may result from relationships between the first, second, and third stage components in the variance when estimates are based on the 1988 NHDS sample design. That relationship and the magnitude of its effects on the variances is unknown at this writing.)

For example, assume an estimated 10,000 discharges during the data years 1988-1989 were to patients receiving surgery "S" during their stays. Here,

X'88 = 10,000 discharges, Y = {1988, 1989}, and alpha = 2.

Using equation (4) above, the following is obtained:

Year	**A**	**b**	**RSE (10,000)**
1988	**.001591421**	**403.123981**	**0.2047 (based on 1988 Data File Documentation)**
1989	**.0029026252**	**478.865192**	**0.2254 (based on 1989 Data File Documentation)**

Hence, max [RSE (X'88|1 year)] = max [.2047, .2254] = .2254

And, by equation (1), the variance of the multi-year aggregate X'88 is:

Var (X'88|years 1988 & 1989) = $\{[(10{,}000)^2] * [(0.2254)^2]\}$ = 5,081,000

SITUATION C:

When the data base Y includes years from both the 1965-1987 and the 1988-2007 periods, the estimate X' and the data base set of years Y may be written as:

X' = X'65 + X'88 and Y = Y65 + Y88

where the X' and Y denote, respectively, the multi-year estimate and the set of years when data collection is restricted to either the 1965 sample design or the 1988 sample design. Then, because the two samples were selected independently of each other, the variance becomes

Var (X') = Var (X'65) + Var (X'88) , (5)

Use the procedures given above to approximate Var(X'65) and Var (X'88).

For example, assume an estimated 10,000 discharges had surgery "S" during the data years 1986-87 and another 10,000 discharges had surgery "S" during the data years 1988-89.

Here,

X' = X'65 + X'88 = 10,000 + 10,000 = 20,000 discharges
Y = Y65 + Y88 = {1986, 1987} + {1988, 1989} and
alpha = 4.

From the examples in situations A and B,
VAR (10,000|years 1986 and 1987) = 810,000
VAR (10,000|years 1988 and 1989) = 5,081,000

Hence, by equation (5), the variance for the multi-year aggregate X' is:

VAR (20,000|years 1986-1989) = 810,000 + 5,081,000 = 5,891,000

VARIANCE OF ESTIMATED AVERAGE ANNUAL TOTALS

The variance for an average annual total is derived by first approximating the variance for X', according to the instructions for the multi-year aggregate estimate.
Then, let X' / alpha = the average annual aggregate for the characteristic of interest for alpha years of data, where
X' = the multi-year aggregate estimate for the characteristic of interest, and
alpha = the total number of years included in the data base for X'
(that is, the number of years in the set of Y years).

Then, the variance for the estimated average annual aggregate is:

VAR (X' / alpha) = VAR(X') / (alpha)^2 (6)

SITUATION A:

For the average annual aggregate per year, when the data years are all between 1965 and 1987, using the example cited above,

X'65 / 2 = 10,000 / 2 = 5,000.

Substituting in equation (6) gives the variance of X'65 / 2:

$$\begin{aligned} VAR(X'65 / 2 \mid \text{years } 1986, 1987) &= VAR\,(X'65) / 2^2 \\ &= 810{,}000 / 2^2 \\ &= 203{,}000 \end{aligned}$$

SITUATION B:

For the average annual aggregate per year, when the data years are all between 1988 and 2007, using the example cited above, X'88 / 2 = 10,000 / 2 = 5,000

Substituting in equation (6) gives the variance of X'88 / 2:

$$\begin{aligned} VAR(X'88 / 2 \mid \text{years } 1988, 1989) &= VAR\,(X'88) / 2^2 \\ &= 5{,}081{,}000 / 2^2 \\ &= 1{,}270{,}000 \end{aligned}$$

SITUATION C:

For the average annual aggregate per year, when the data years included are from both the 1965 design and the 1988 redesign, using the example cited above,

$$X' / 4 = 20{,}000/4 = 5{,}000$$

Substituting in equation (6) gives the variance of X' / 4:

$$\begin{aligned} VAR(X' / 4 \mid \text{years } 1986, 1987, 1988, 1989) &= VAR\,(X') / 4^2 \\ &= 5{,}891{,}000 / 4^2 \\ &= 368{,}000 \end{aligned}$$

Exhibit B1:

FACTOR R FOR APPROXIMATING REL-VARIANCES FOR MULTI-YEAR ESTIMATES FROM THE 1965 NHDS SAMPLE DESIGN

Here, it is assumed the multi-aggregate estimate X' is based on data collected in NHDS during two or more of the years 1965-1987. Then the rel-variance of X'65 may be approximated by

$RSE^2(X'65|Y \text{ years}) = R * \max [RSE^2(X'65|1 \text{ year})]$ where

X'65 = the multi-year aggregate estimate of interest.

Y = the set of years in the data base for X'65.

$RSE^2(X'65)$ = the rel-variance of X'65.

For simplicity, assume that the number of sampled hospitals and sampled discharge abstracts in NHDS remain constant over the data years that serve as the reference period for the estimate X'65. Also assume that the total number of abstracts submitted to NHDS each of those years is constant. Let:

m = the number of hospitals participating in NHDS

n = the number of sample abstracts in NHDS each year

B = the between hospital rel-variance for the population of discharges

W = the within hospital rel-variance for the population of discharges

alpha = the number of years in the data base Y

A streamlined approximation to the rel-variance of X'65 based on a single year of data is then:

$$RSE^2(X'65|1 \text{ year}) = (B^2 / m) + (W^2 / n) = B^2 \{(1 / m) + [(W / B) / n]\}$$

During the original work on NHDS design, values of the ratio W/B were determined and found to lie in the range of 300 - 700 with some concentration near 500 (4). Hence, for the general statistic, the effect of using alpha years of data can be roughly approximated by setting W/B equal to 500. Then the rel-variance for one year of data is approximately:

$$RSE^2(X'65|1 \text{ year}) = B^2 [(1 / m) + (500 / n)]$$

and the rel-variance for estimates based on alpha years of data is approximately

$$RSE^2(X'65|\text{alpha years}) = B^2 \{(1 / m) + [500 / (\text{alpha}*n)]\}$$

The ratio of the two rel-variances can thus be formulated as:

$$[RSE^2(X'65|\text{alpha years})] / [RSE^2(X'65|1 \text{ year})] =$$

$$\{[\text{alpha} * (n / m)] + 500\} / \{\text{alpha} * [(n / m) + 500]\} = R$$

Because in reality, sample sizes can vary across the Y years, the average number of sampled abstracts per hospital per year over the Y years in the data base is used in place of n/m in R. That is, in place of n/m, use average-n which is defined as:
average-n = the total number of sampled discharges over the years included in the data base (n), divided by the sum of the number of respondent hospitals over the same time period (m).

NOTE: the number of sampled discharges per year and the number of respondent hospitals per year appear in Exhibit B2 (below).

Hence, R = [(alpha * average-n) + 500] / [alpha * (average-n + 500)].

ExhibitB2:

NUMBERS OF HOSPITALS PARTICIPATING IN NHDS AND APPROXIMATE TOTAL NUMBER OF SAMPLED DISCHARGE ABSTRACTS, 1979-2007

DATA YEAR	NUMBER OF HOSPITALS	APPROXIMATE NUMBER OF SAMPLED DISCHARGE ABSTRACTS
1965	296	100,000
1966	300	137,000
1967	289	145,000
1968	413	210,000
1969	402	208,000
1970	395	205,000
1971	179	200,000
1972	424	225,000
1973	424	225,000
1974	426	227,000
1975	432	232,000
1976	419	223,000
1977	423	224,000
1978	413	219,000
1979	416	215,000
1980	420	224,000
1981	428	227,000
1982	426	214,000
1983	418	206,000
1984	407	192,000
1985	414	195,000
1986	418	193,000
1987	400	181,000
1988	422	250,000
1989	408	233,000
1990	474	266,000
1991	484	274,000
1992	494	274,000
1993	466	235,000
1994	478	277,000
1995	466	263,000
1996	480	282,000
1997	474	300,000
1998	478	307,000
1999	458	300,000
2000	434	313,000
2001	448	330,000
2002	445	327,000
2003	426	320,000
2004	439	371,000
2005	444	375,000
2006	438	376,000
2007	422	366,000

NAMES OF FILES CONTAINING RELATIVE STANDARD ERROR (RSE) CURVES (THESE ARE EXCEL SPREADSHEET FILES CONTAINED ON THIS CD):

RSE1979.XLS	Approximate relative standard errors of estimated statistics, 1979
RSE1980.XLS	Approximate relative standard errors of estimated statistics, 1980
RSE1981.XLS	Approximate relative standard errors of estimated statistics, 1981
RSE1982.XLS	Approximate relative standard errors of estimated statistics, 1982
RSE1983.XLS	Approximate relative standard errors of estimated statistics, 1983
RSE1984.XLS	Approximate relative standard errors of estimated statistics, 1984
RSE1985.XLS	Approximate relative standard errors of estimated statistics, 1985
RSE1986.XLS	Approximate relative standard errors of estimated statistics, 1986
RSE1987.XLS	Approximate relative standard errors of estimated statistics, 1987
RSE1988.XLS	Estimated parameters for relative standard error equations by selected patient and hospital characteristics, 1988
RSE1989.XLS	Estimated parameters for relative standard error equations by selected patient and hospital characteristics, 1989
RSE1990.XLS	Estimated parameters for relative standard error equations by selected patient and hospital characteristics, 1990
RSE1991.XLS	Estimated parameters for relative standard error equations by selected patient and hospital characteristics, 1991
RSE1992.XLS	Estimated parameters for relative standard error equations by selected patient and hospital characteristics, 1992
RSE1993.XLS	Estimated parameters for relative standard error equations by selected patient and hospital characteristics, 1993
RSE1994.XLS	Estimated parameters for relative standard error equations by selected patient and hospital characteristics, 1994
RSE1995.XLS	Estimated parameters for relative standard error equations by selected patient and hospital characteristics, 1995
RSE1996.XLS	Estimated parameters for relative standard error equations by selected patient and hospital characteristics, 1996
RSE1997.XLS	Estimated parameters for relative standard error equations by selected patient and hospital characteristics, 1997
RSE1998.XLS	Estimated parameters for relative standard error equations by selected patient and hospital characteristics, 1998
RSE1999.XLS	Estimated parameters for relative standard error equations by selected patient and hospital characteristics, 1999
RSE2000.XLS	Estimated parameters for relative standard error equations by selected patient and hospital characteristics, 2000
RSE2001.XLS	Estimated parameters for relative standard error equations by selected patient and hospital characteristics, 2001
RSE2002.XLS	Estimated parameters for relative standard error equations by selected patient and hospital characteristics, 2002
RSE2003.XLS	Estimated parameters for relative standard error equations by selected patient and hospital characteristics, 2003
RSE2004.XLS	Estimated parameters for relative standard error equations by selected patient and hospital characteristics, 2004
RSE2005.XLS	Estimated parameters for relative standard error equations by selected patient and hospital characteristics, 2005
RSE2006.XLS	Estimated parameters for relative standard error equations by selected patient and hospital characteristics, 2006
RSE2007.XLS	Estimated parameters for relative standard error equations by selected patient and hospital characteristics, 2007

APPENDIX C:

INTERNATIONAL CLASSIFICATION OF DISEASES, 9TH REVISION, CLINICAL MODIFICATION

ADDENDUM AND CONVERSION TABLE

The In*ternational Classification of Diseases, 9th Revision, Clinical Modification* has been used for coding medical data in NHDS since 1979. However, since 1986, the classification system has undergone annual updating, which includes assignment of new diagnostic and procedure codes, fourth and/or fifth digit expansion of codes, and code deletions. These changes have been released for the years 1986 through 2007, except for 1999 when no Addendum was released because of concern about possible complications for instituting coding changes prior to the millennium crossover. These changes are published in an Addendum and become effective October 1 of the calendar year. Addenda are developed by the ICD-9-CM Coordination and Maintenance Committee and approved by the Director of NCHS and the Administrator of the Centers for Medicare and Medicaid Services (formerly the Health Care Financing Administration).

In 2002, the ICD-9-CM Coordination and Maintenance Committee created procedure Chapter 00 – Procedures and Interventions, Not Elsewhere Classified – as a way of handling space limitations in the existing hierarchical structure and alleviating inappropriate categorization of new procedures. Since October addendum changes are not implemented in NHDS until the following data collection year, 2003 was the first year these codes were used.

Users who wish to produce trend statistics must be aware of the dynamic nature of the ICD-9-CM coding system. Because of the Addendum changes, estimates for some codes may show discontinuity across years. Also, the meaning of the original and revised code may be different, thus causing problems with interpretation across years. Further, to address the potential problem that October coding changes would cause for calendar year estimation, all coding which was consistent with the October Addendum of the survey year was changed back to the coding previous to the Addendum changes. This "backward coding" was performed for the 1992 through 2007 data years, and helps to avoid the problem of partial year estimates being mistaken for full year estimates.

In order to assist users in data analysis, a **conversion table** is provided that shows for each new code, its date of introduction and the previously assigned code equivalent, which had been used for reporting the selected diagnosis or procedure prior to issuance of the new code. This table is available on this CD as a separate PDF file.

APPENDIX D:

PROCEDURES NOT CODED, 1979-1990

Historically, operative and surgical procedures were categorized into four classes, according to UHDDS guidelines. Classes 1-3 consisted of significant procedures, i.e. those that carried an operative or anesthetic risk or required highly trained personnel, special facilities, or special equipment. Class 4 procedures were not considered significant; therefore, reporting was optional. From 1979 through the middle of 1983, only three Class 4 procedures were coded for NHDS: circumcision, episiotomy, and removal of intrauterine contraceptive device. Since the last half of 1983, additional codes were added, until 1991, when all procedures were coded.

It is important to remember that, because many procedure codes were not utilized between 1979 and 1990, it is not possible to produce estimates for those codes. The table contained in this appendix lists the procedures not coded and the year the code was first used. Caution should be exercised when producing estimates for these codes.

CLASS 4 PROCEDURES AND YEAR CODE FIRST USED IN NHDS

PROCEDURE CODE	FIRST USED
01.18-01.19	1983
03.39	1983
04.19	1983
05.19	1983
06.19	1983
07.19	1983
08.19	1989
08.91-08.93	1983
09.19	1983
09.41-09.49	1983
10.29	1983
11.29	1983
12.29	1983
14.19	1983
15.09	1983
16.21	1989
16.29	1983
18.01	1989
18.11	1989
18.19	1989
20.39	1983
21.21	1989
21.29	1989
22.19	1989
24.19	1989
25.09	1989
25.91	1989
26.19	1989
27.29	1989
27.91	1989
28.19	1983
29.19	1989
31.48-31.49	1989
33.28-33.29	1983
34.28-34.29	1983
37.29	1989
38.29	1983
40.19	1983
41.38-41.39	1989
42.29	1989
44.91	1989
45.19	1989
45.28-45.29	1989
48.23	1989
48.29	1989
49.21	1989
49.29	1989
49.41	1989

PROCEDURE CODE	FIRST USED
50.19	1983
51.19	1983
52.19	1983
54.29	1983
55.29	1983
56.39	1983
57.39	1983
58.29	1989
59.29	1983
60.18	1983
61.19	1989
62.19	1983
63.09	1983
64.0-	1979
64.19	1989
64.91	1989
64.94	1989
65.19	1983
66.19	1983
67.19	1983
68.19	1983
69.92	1989
70.21	1989
70.29	1983
71.19	1983
73.6-	1979
73.91-73.92	1989
75.35	1989
76.19	1983
78.80-78.89	1983
81.98	1983
83.29	1983
85.19	1989
86.19	1989
86.92	1989
87.09-87.12	1991
87.16-87.17	1991
87.22-87.29	1991
87.36-87.37	1989
87.39	1991
87.43-87.49	1991
87.69	1989
87.79	1989
87.85-87.89	1991
87.92	1991
87.95-87.99	1991
88.09	1991
88.16-88.31	1991
88.33	1991
88.35	1991

PROCEDURE CODE	FIRST USED
88.37	1991
88.39	1991
89.01-89.09	1991
89.11-89.13	1991
89.15-89.16	1991
89.26-89.31	1991
89.33-89.39	1991
89.45-89.49	1989
89.51-89.53	1989
89.55-89.59	1989
89.66	1989
89.7-	1991
90.01-91.99	1991
93.01-93.25	1991
93.27-93.28	1991
93.31-93.39	1991
93.42-93.44	1989
93.61-93.89	1991
93.91	1991
93.94	1989
93.96	1989
93.99	1989
94.01-94.23	1991
94.25	1991
94.29-94.59	1991
95.01-95.03	1991
95.05-95.11	1991
95.14-95.15	1991
95.31-95.48	1991
95.49	1989
96.09	1989
96.11-96.19	1991
96.26-96.28	1991
96.34-96.59	1991
97.01-97.04	1991
97.14-97.69	1991
97.71	1979
97.72-97.89	1991
99.02-99.14	1991
99.16-99.24	1991
99.26-99.59	1991
99.71-99.79	1989
99.82-99.84	1990
99.91-99.99	1990

APPENDIX E:

CENSUS POPULATION ESTIMATES, 1979-2007

This appendix describes files included on this CD which contain Census Bureau estimates of the U.S. civilian resident population for the years 1979 through 2007. These files are EXCEL spreadsheets with a XLS extension and are located in the POPS folder. There are two types of files, named either **XXXXRACE** or **XXXXREG**. To select a given year, the "**XXXX**" can be replaced by the year desired between 1979 and 2007.

For the years 1980 through 2007, the **XXXXRACE** tables provide estimates disaggregated by sex, race, and single-year age groupings and the **XXXXREG** tables provide estimates disaggregated by sex, region, and single-year age groupings. There is only one spreadsheet for 1979 because estimates disaggregated by region were not available for that year. And the figures in the 1979 spreadsheet have been rounded to the nearest thousand, whereas those for all the other years are unrounded.

The 1979 population estimates were adjusted based on the 1980 census. Population estimates for 1980-1989 were adjusted based on the 1990 decennial census. Estimates for 1990-1999 have been adjusted for net underenumeration using the 1990 National Population Adjustment Matrix. The 2000-2007 estimates are based on the results of the 2000 Census.

It should be noted that rates calculated with these estimates may differ slightly from those in published NCHS reports or those calculated from population estimates disseminated with NHDS annual data file documentation. Before 1981, estimates of rates of discharges and days of care that appeared in published reports from NCHS were calculated using the civilian noninstitutional population. However, beginning in 1981, the civilian resident population estimates were used. The civilian resident population was determined to be more appropriate because persons in institutions, for example nursing home patients, are hospitalized when necessary. A report has been published which discusses differences in discharge rates based on the different denominators (8).

Because population estimates for 2000-2007 are based on the 2000 Census, researchers examining trends across time may observe discontinuities which may be due to the denominators and not necessarily attributable to real temporal changes. However, a comparison of both the population estimates and hospitalization rates for 2000, based on the 1990-based and the 2000 Census estimates for selected sex by age groups, suggests that differences are not large, averaging under 3 percentage points. Researchers interested in further information about this issue can contact the Ambulatory and Hospital Care Statistics Branch at 301.458.4321.

NOTE ON RACE DATA IN NHDS AND CENSUS: In 1997, the Federal Office of Management and Budget (OMB) revised standards that regulated how the Federal government would collect and report data on race and ethnicity in the 2000 Census. In addition to changes in some of the racial categories previously reported, it also permitted respondents to self-identify with more than one racial group. The goal was to improve the accuracy of information on racial diversity in the United States.

The major implication of the new Federal guidelines is that Census 2000 race data are not directly comparable with race data from the 1990 or earlier censuses. A number of new tabulations of racial categories are now available, but NHDS utilizes tabulations based on six race-alone and one multiple race categorization. The six single race-alone groups are White, African-American, American Indian and Alaskan Native, Asian, Native Hawaiian and Other Pacific Islander, and Some Other Race; and the multiple-race category groups together all respondents who identified with two or more races. These categories are mutually exclusive and when summed together add to 100 percent of the US population.

It is not known to what extent these groupings differ from earlier ones where no attempt was made to identify respondents with multi-racial backgrounds. Census cautions that direct comparisons of racial categories from the 1990's to 2000 cannot be made, and recommends that the data user decide whether the single race-alone estimate is appropriate for their analysis.

The Census population tables provided in this documentation contain groupings for three primary racial groups: White, Black/African American, and All Other Races. The reason for this is simply that NHDS statistics based on the smaller racial groups (e.g. Asian, American Indian/Alaskan Native, and Native Hawaiian/Other Pacific Islander) often do not meet NCHS standards for reliability of published estimates. Calculating rates with NHDS data by race is complicated by the fact that there is substantial underreporting of race in the survey (30% nonresponse in 2007). Extreme caution should be exercised when using NHDS race data, especially when reporting population-based utilization rates.

The OMB standards discussed above do not apply to how hospitals record patient information in the medical record, which is the source document for NHDS. As a result, reporting of multiple races in NHDS is almost non-existent. For the 2007 NHDS, 152 of the 366,000 sample records had more than one race marked on the patient abstract form.

NAMES OF FILES CONTAINING CENSUS POPULATION TABLES: (THESE ARE EXCEL SPREADSHEET FILES INCLUDED ON THIS CD):

1979RACE.XLS	Civilian Population of the US, Estimates by Age, Sex, and Race, July 1, 1979
1980RACE.XLS	Civilian Population of the US, Estimates by Age, Sex, and Race, July 1, 1980
1981RACE.XLS	Civilian Population of the US, Estimates by Age, Sex, and Race, July 1, 1981
1982RACE.XLS	Civilian Population of the US, Estimates by Age, Sex, and Race, July 1, 1982
1983RACE.XLS	Civilian Population of the US, Estimates by Age, Sex, and Race, July 1, 1983
1984RACE.XLS	Civilian Population of the US, Estimates by Age, Sex, and Race, July 1, 1984
1985RACE.XLS	Civilian Population of the US, Estimates by Age, Sex, and Race, July 1, 1985
1986RACE.XLS	Civilian Population of the US, Estimates by Age, Sex, and Race, July 1, 1986
1987RACE.XLS	Civilian Population of the US, Estimates by Age, Sex, and Race, July 1, 1987
1988RACE.XLS	Civilian Population of the US, Estimates by Age, Sex, and Race, July 1, 1988
1989RACE.XLS	Civilian Population of the US, Estimates by Age, Sex, and Race, July 1, 1989
1990RACE.XLS	Civilian Population of the US, Estimates by Age, Sex, and Race, July 1, 1990, Adjusted for net underenumeration in 1990 Census
1991RACE.XLS	Civilian Population of the US, Estimates by Age, Sex, and Race, July 1, 1991, Adjusted for net underenumeration in 1990 Census
1992RACE.XLS	Civilian Population of the US, Estimates by Age, Sex, and Race, July 1, 1992, Adjusted for net underenumeration in 1990 Census
1993RACE.XLS	Civilian Population of the US, Estimates by Age, Sex, and Race, July 1, 1993, Adjusted for net underenumeration in 1990 Census
1994RACE.XLS	Civilian Population of the US, Estimates by Age, Sex, and Race, July 1, 1994, Adjusted for net underenumeration in 1990 Census
1995RACE.XLS	Civilian Population of the US, Estimates by Age, Sex, and Race, July 1, 1995, Adjusted for net underenumeration in 1990 Census
1996RACE.XLS	Civilian Population of the US, Estimates by Age, Sex, and Race, July 1, 1996, Adjusted for net underenumeration in 1990 Census
1997RACE.XLS	Civilian Population of the US, Estimates by Age, Sex, and Race, July 1, 1997, Adjusted for net underenumeration in 1990 Census
1998RACE.XLS	Civilian Population of the US, Estimates by Age, Sex, and Race, July 1, 1998, Adjusted for net underenumeration in 1990 Census
1999RACE.XLS	Civilian Population of the US, Estimates by Age, Sex, and Race, July 1, 1999, Adjusted for net underenumeration in 1990 Census
2000RACE.XLS	Civilian Population of the US, Estimates by Age, Sex, and Race, July 1, 2000, based on 2000 Census
2001RACE.XLS	Civilian Population of the US, Estimates by Age, Sex, and Race, July 1, 2001, based on 2000 Census
2002RACE.XLS	Civilian Population of the US, Estimates by Age, Sex, and Race, July 1, 2002, based on 2000 Census
2003RACE.XLS	Civilian Population of the US, Estimates by Age, Sex, and Race, July 1, 2003, based on 2000 Census
2004RACE.XLS	Civilian Population of the US, Estimates by Age, Sex, and Race, July 1, 2004, based on 2000 Census
2005RACE.XLS	Civilian Population of the US, Estimates by Age, Sex, and Race, July 1, 2005, based on 2000 Census
2006RACE.XLS	Civilian Population of the US, Estimates by Age, Sex, and Race, July 1, 2006, based on 2000 Census
2007RACE.XLS	Civilian Population of the US, Estimates by Age, Sex, and Race, July 1, 2007, based on 2000 Census

1980REG.XLS	Civilian Population of the US, Estimates by Age, Sex, and Region, July 1, 1980
1981REG.XLS	Civilian Population of the US, Estimates by Age, Sex, and Region, July 1, 1981
1982REG.XLS	Civilian Population of the US, Estimates by Age, Sex, and Region, July 1, 1982
1983REG.XLS	Civilian Population of the US, Estimates by Age, Sex, and Region, July 1, 1983
1984REG.XLS	Civilian Population of the US, Estimates by Age, Sex, and Region, July 1, 1984
1985REG.XLS	Civilian Population of the US, Estimates by Age, Sex, and Region, July 1, 1985
1986REG.XLS	Civilian Population of the US, Estimates by Age, Sex, and Region, July 1, 1986
1987REG.XLS	Civilian Population of the US, Estimates by Age, Sex, and Region, July 1, 1987
1988REG.XLS	Civilian Population of the US, Estimates by Age, Sex, and Region, July 1, 1988
1989REG.XLS	Civilian Population of the US, Estimates by Age, Sex, and Region, July 1, 1989
1990REG.XLS	Civilian Population of the US, Estimates by Age, Sex, and Region, July 1, 1990, Adjusted for net underenumeration in 1990 Census
1991REG.XLS	Civilian Population of the US, Estimates by Age, Sex, and Region, July 1, 1991, Adjusted for net underenumeration in 1990 Census
1992REG.XLS	Civilian Population of the US, Estimates by Age, Sex, and Region, July 1, 1992, Adjusted for net underenumeration in 1990 Census
1993REG.XLS	Civilian Population of the US, Estimates by Age, Sex, and Region, July 1, 1993, Adjusted for net underenumeration in 1990 Census
1994REG.XLS	Civilian Population of the US, Estimates by Age, Sex, and Region, July 1, 1994, Adjusted for net underenumeration in 1990 Census
1995REG.XLS	Civilian Population of the US, Estimates by Age, Sex, and Region, July 1, 1995, Adjusted for net underenumeration in 1990 Census
1996REG.XLS	Civilian Population of the US, Estimates by Age, Sex, and Region, July 1, 1996, Adjusted for net underenumeration in 1990 Census
1997REG.XLS	Civilian Population of the US, Estimates by Age, Sex, and Region, July 1, 1997, Adjusted for net underenumeration in 1990 Census
1998REG.XLS	Civilian Population of the US, Estimates by Age, Sex, and Region, July 1, 1998, Adjusted for net underenumeration in 1990 Census
1999REG.XLS	Civilian Population of the US, Estimates by Age, Sex, and Region, July 1, 1999, Adjusted for net underenumeration in 1990 Census
2000REG.XLS	Civilian Population of the US, Estimates by Age, Sex, and Region, July 1, 2000, based on 2000 Census
2001REG.XLS	Civilian Population of the US, Estimates by Age, Sex, and Region, July 1, 2001, based on 2000 Census
2002REG.XLS	Civilian Population of the US, Estimates by Age, Sex, and Region, July 1, 2002, based on 2000 Census
2003REG.XLS	Civilian Population of the US, Estimates by Age, Sex, and Region, July 1, 2003, based on 2000 Census
2004REG.XLS	Civilian Population of the US, Estimates by Age, Sex, and Region, July 1, 2004, based on 2000 Census
2005REG.XLS	Civilian Population of the US, Estimates by Age, Sex, and Region, July 1, 2005, based on 2000 Census
2006REG.XLS	Civilian Population of the US, Estimates by Age, Sex, and Region, July 1, 2006, based on 2000 Census
2007REG.XLS	Civilian Population of the US, Estimates by Age, Sex, and Region, July 1, 2007, based on 2000 Census

APPENDIX F:

FREQUENCIES FOR SELECTED VARIABLES, 1979-2007

The following tabulations are provided for the purpose of verifying that data processing of the multi-year file is accurate. Included are tables of estimated discharges, disaggregated by sex, age groups, race, discharge status, and region; estimated days of care; estimated numbers of selected first-listed and all-listed diagnoses; estimated numbers of selected all-listed procedures; and estimated newborn infant discharges, disaggregated by sex, region, and health status. Note that these are weighted frequencies, which were produced by applying the weight variable (columns 21-25) to the analysis.

The following ICD-9-CM codes are used for the diagnoses and procedures shown in these tables:

DIAGNOSIS	**ICD-9-CM code**
Delivery	V27
Septicemia	038
Malignant neoplasms	140-208, 230-234
Diabetes	250
Psychoses	290-299
Acute myocardial infarction	410
Asthma	493
Injury & poisoning	800-999

PROCEDURE	**ICD-9-CM code**
Cesarean section	74.0-, 74.1-, 74.2-, 74.4-, 74.99
Hysterectomy	68.3-68.7, 68.9 (after 1992)
CABG	36.1
Prostatectomy	60.2-60.6
Appendectomy	47.0

SICK BABY (defined for Newborn Infants only, i.e. those who have a V30-V39 as a first-listed diagnosis):

1979-1994: record had at least one additional diagnosis code between 001-999 or at least one risk-related diagnosis, V01-V82, except V30-V39.

1995-2007: record had at least one additional diagnosis code between 001-999 only.

FREQUENCIES FOR NON-NEWBORNS:

	DISCHARGES - ALL	DISCHARGES – MALE	DISCHARGES - FEMALE	DAYS OF CARE – ALL
1979	36,746,563	14,704,807	22,041,756	264,173,219
1980	37,831,559	15,145,249	22,686,310	274,508,358
1981	38,543,980	15,379,076	23,164,904	277,229,888
1982	38,593,355	15,469,954	23,123,401	272,626,923
1983	38,782,663	15,573,018	23,209,645	268,337,331
1984	37,162,124	14,899,095	22,263,029	244,651,694
1985	35,056,134	14,160,466	20,895,668	226,217,043
1986	34,255,915	13,949,308	20,306,607	218,496,028
1987	33,386,694	13,568,200	19,818,494	214,942,242
1988	31,146,360	12,641,965	18,504,395	203,677,561
1989	30,946,718	12,583,154	18,363,564	200,826,678
1990	30,787,939	12,279,604	18,508,335	197,421,786
1991	31,098,002	12,478,352	18,619,650	199,098,663
1992	30,950,912	12,406,079	18,544,833	190,385,530
1993	30,825,148	12,262,445	18,562,703	184,601,125
1994	30,843,087	12,293,339	18,549,748	177,178,999
1995	30,722,351	12,197,504	18,524,847	164,626,802
1996	30,544,614	12,109,749	18,434,865	159,882,807
1997	30,914,167	12,267,559	18,646,608	157,458,023
1998	31,826,545	12,468,847	19,357,698	160,913,782
1999	32,131,877	12,748,152	19,383,725	160,128,236
2000	31,705,672	12,513,977	19,191,695	155,856,547
2001	32,652,589	12,851,779	19,800,810	159,364,656
2002	33,726,611	13,388,628	20,337,983	164,152,068
2003	34,738,411	13,874,228	20,864,183	167,301,012
2004	34,864,166	13,844,420	21,019,746	167,880,848
2005	34,667,315	13,901,651	20,765,664	165,924,644
2006	34,853,896	13,990,227	20,863,669	166,341,655
2007	34,368,888	13,833,726	20,535,162	166,440,669

DISCHARGES BY AGE GROUP

	UNDER 15 YEARS	15 TO 44 YEARS	45 TO 64 YEARS	65 YEARS AND OVER
1979	3,641,112	15,487,593	8,532,087	9,085,771
1980	3,672,493	15,635,443	8,659,807	9,863,816
1981	3,733,060	15,725,148	8,677,432	10,408,340
1982	3,654,396	15,553,926	8,688,003	10,697,030
1983	3,653,950	15,268,547	8,558,436	11,301,730
1984	3,208,176	14,532,906	8,194,831	11,226,211
1985	2,971,825	13,966,558	7,609,571	10,508,180
1986	2,782,612	13,457,746	7,299,897	10,715,660
1987	2,687,520	13,141,509	7,099,013	10,458,652
1988	2,610,430	11,933,540	6,456,122	10,146,268
1989	2,597,343	11,848,274	6,271,496	10,229,605
1990	2,411,674	11,798,790	6,244,267	10,333,208
1991	2,498,479	11,620,374	6,173,130	10,806,019
1992	2,531,414	11,226,701	6,328,830	10,863,967
1993	2,141,152	11,200,025	6,283,434	11,200,537
1994	2,249,317	10,956,147	6,311,114	11,326,509
1995	2,405,422	10,593,181	6,167,656	11,556,092
1996	2,206,856	10,325,208	6,294,238	11,718,312
1997	2,311,503	10,029,526	6,377,224	12,195,914
1998	2,298,531	10,376,428	6,695,652	12,455,934
1999	2,458,302	10,092,435	6,898,581	12,682,559
2000	2,382,813	9,968,908	6,958,001	12,395,950
2001	2,559,809	10,343,100	7,224,268	12,525,412
2002	2,539,536	10,736,431	7,723,290	12,727,354
2003	2,571,263	10,831,323	8,119,593	13,216,232
2004	2,570,979	10,799,996	8,322,634	13,170,557
2005	2,430,926	10,658,706	8,349,380	13,228,303
2006	2,297,691	10,799,738	8,685,968	13,070,499
2007	2,180,985	10,571,285	8,753,338	12,863,280

DISCHARGES BY REGION

	NORTHEAST	MIDWEST	SOUTH	WEST
1979	7,785,635	10,647,226	12,424,987	5,888,715
1980	7,868,414	10,877,714	12,982,849	6,102,582
1981	7,821,859	11,131,851	13,202,327	6,387,943
1982	7,847,372	10,937,996	13,435,287	6,372,700
1983	7,792,914	10,492,289	13,883,629	6,613,831
1984	7,407,768	9,899,132	13,450,674	6,404,550
1985	7,168,298	9,111,370	12,274,465	6,502,001
1986	6,955,118	8,930,808	11,892,160	6,477,829
1987	6,698,861	8,718,351	11,291,607	6,677,875
1988	7,078,483	7,832,097	10,844,990	5,390,790
1989	7,044,102	7,675,518	10,959,504	5,267,594
1990	6,895,148	7,620,112	11,173,053	5,099,626
1991	7,153,072	7,315,173	11,289,656	5,340,101
1992	7,140,865	7,121,396	11,255,745	5,432,906
1993	6,964,731	7,097,084	11,580,414	5,182,919
1994	7,128,436	7,133,439	11,310,254	5,270,958
1995	7,051,480	6,994,250	11,373,398	5,303,223
1996	6,665,339	7,106,564	11,085,190	5,687,521
1997	6,679,107	7,234,030	11,445,029	5,556,001
1998	6,818,497	7,365,774	12,021,611	5,620,663
1999	6,952,345	7,368,289	12,006,628	5,804,615
2000	7,102,516	7,207,208	12,015,500	5,380,448
2001	7,026,430	7,428,559	12,744,295	5,453,305
2002	6,989,562	7,503,133	12,994,445	6,239,471
2003	7,266,500	7,785,932	13,055,408	6,630,571
2004	7,387,692	7,680,065	13,238,977	6,557,432
2005	7,191,805	7,947,947	12,906,556	6,621,007
2006	7,276,715	7,951,159	13,139,572	6,486,450
2007	7,361,332	7,706,405	12,689,036	6,612,115

DISCHARGES BY RACE

	WHITE	BLACK	OTHER	RACE NOT STATED
1979	27,450,925	3,957,844	-	5,337,794
1980	28,484,009	4,176,172	702,456	4,468,922
1981	32,242,003	5,003,852	1,298,125	-
1982	29,879,874	4,216,328	926,139	3,571,014
1983	30,106,094	4,254,555	980,708	3,441,306
1984	28,448,523	4,265,667	1,035,836	3,412,098
1985	26,378,763	4,109,162	1,071,573	3,496,636
1986	25,363,328	4,015,423	1,100,707	3,776,457
1987	24,375,931	3,866,448	1,308,034	3,836,281
1988	23,321,644	3,828,797	1,179,155	2,816,764
1989	22,677,616	3,891,025	976,405	3,401,672
1990	21,375,856	3,611,203	958,238	4,842,642
1991	20,816,245	3,717,049	1,036,368	5,528,340
1992	20,017,886	3,692,237	1,127,549	6,113,240
1993	20,101,017	3,659,646	1,252,661	5,811,824
1994	20,003,156	3,711,210	1,351,413	5,777,308
1995	19,950,723	3,887,498	1,097,619	5,786,511
1996	19,738,155	3,779,331	1,351,228	5,675,900
1997	19,971,406	3,689,251	1,512,058	5,741,452
1998	20,620,239	3,750,056	1,672,637	5,783,613
1999	20,757,130	3,788,248	1,730,946	5,855,553
2000	19,165,231	3,571,561	1,071,400	7,897,480
2001	20,332,367	3,900,106	1,210,672	7,209,444
2002	20,805,709	3,994,884	1,296,499	7,629,519
2003	21,291,963	4,101,921	1,390,922	7,953,605
2004	20,979,123	4,118,156	1,525,649	8,241,238
2005	20,897,155	4,108,786	1,429,735	8,231,639
2006	20,906,742	4,240,877	1,343,638	8,362,639
2007	20,482,940	4,320,948	1,249,025	8,315,975

FIRST-LISTED DIAGNOSIS

	DELIVERY	SEPTICEMIA	MALIGNANT NEOPLASMS	DIABETES
1979	3,645,988	54,855	1,793,157	600,193
1980	3,762,253	59,492	1,872,506	645,187
1981	3,912,917	58,634	1,993,092	655,118
1982	3,944,532	65,660	2,016,166	661,156
1983	3,975,611	92,024	2,064,744	674,709
1984	3,853,144	127,292	2,059,161	592,507
1985	3,854,030	148,731	1,910,954	479,789
1986	3,762,138	175,713	1,859,641	491,150
1987	3,910,926	172,127	1,879,018	473,863
1988	3,780,654	193,359	1,670,159	453,671
1989	3,936,703	196,001	1,608,186	437,965
1990	4,025,456	215,602	1,571,191	419,998
1991	3,973,087	240,434	1,593,890	429,455
1992	3,909,719	279,055	1,577,277	475,763
1993	4,014,893	270,271	1,481,949	464,212
1994	3,901,426	301,756	1,443,118	502,160
1995	3,766,006	307,860	1,414,338	493,121
1996	3,829,763	355,134	1,374,138	503,064
1997	3,807,787	351,967	1,335,679	507,165
1998	4,000,061	347,172	1,266,308	512,598
1999	3,810,079	340,883	1,274,268	544,875
2000	3,737,621	326,153	1,156,329	557,080
2001	3,839,229	314,712	1,211,930	562,056
2002	3,951,450	341,232	1,207,612	577,434
2003	4,022,483	365,827	1,267,101	596,548
2004	4,135,702	409,607	1,205,321	598,890
2005	4,038,350	489,527	1,195,829	584,884
2006	4,127,491	530,434	1,207,753	584,133
2007	4,136,028	592,135	1,224,585	626,449

	PSYCHOSES	ACUTE MYOCARDIAL INFARCTION (AMI)	ASTHMA	INJURY & POISONING
1979	511,584	432,997	339,260	3,634,927
1980	506,656	431,494	408,308	3,592,833
1981	568,671	465,038	417,594	3,583,875
1982	573,969	680,667	434,157	3,568,431
1983	576,433	676,458	459,141	3,450,082
1984	624,746	700,234	464,754	3,472,232
1985	701,124	754,511	461,655	3,303,164
1986	766,190	757,766	476,977	3,224,625
1987	813,842	759,823	454,110	3,026,710
1988	781,216	716,394	479,462	2,816,692
1989	772,830	694,880	475,283	2,806,379
1990	811,663	674,716	476,060	2,773,767
1991	902,449	697,091	490,244	2,767,962
1992	908,006	747,055	463,386	2,701,184
1993	1,053,770	744,544	468,453	2,718,252
1994	1,243,157	759,170	451,211	2,605,386
1995	1,205,988	770,802	510,629	2,591,349
1996	1,211,645	825,239	473,523	2,549,628
1997	1,220,554	756,497	483,580	2,520,417
1998	1,253,397	782,723	423,139	2,540,057
1999	1,308,629	828,600	478,087	2,565,430
2000	1,445,316	780,588	464,968	2,466,407
2001	1,633,053	794,789	454,038	2,612,727
2002	1,704,304	817,875	483,755	2,697,138
2003	1,585,587	766,954	574,078	2,832,724
2004	1,601,312	732,263	497,085	2,844,647
2005	1,747,396	682,699	488,594	2,825,947
2006	1,746,814	646,650	443,569	2,967,647
2007	1,690,339	577,292	455,865	2,855,413

ALL-LISTED DIAGNOSES

	TOTAL DIAGNOSES (INCLUDING ECODES)	SEPTICEMIA	MALIGNANT NEOPLASMS	DIABETES
1979	87,827,421	157,201	2,887,931	2,136,178
1980	92,186,265	167,298	3,011,597	2,279,410
1981	96,379,456	179,721	3,197,975	2,395,492
1982	99,889,516	209,174	3,261,378	2,562,028
1983	104,282,026	256,814	3,284,594	2,774,733
1984	106,727,848	325,239	3,815,918	2,940,819
1985	103,063,247	376,900	3,739,439	2,756,904
1986	105,310,267	433,149	3,744,660	2,863,645
1987	106,019,667	438,527	3,768,126	2,901,699
1988	102,898,927	473,053	3,547,167	2,917,796
1989	104,219,232	482,964	3,588,880	2,929,710
1990	105,167,895	479,603	3,457,063	3,007,518
1991	109,432,651	516,821	3,604,767	3,304,513
1992	111,584,770	564,740	3,598,145	3,539,249
1993	115,286,773	551,449	3,486,860	3,669,661
1994	118,313,274	553,182	3,403,488	3,740,073
1995	121,167,653	587,651	3,472,180	3,952,279
1996	125,078,947	661,377	3,502,145	4,157,040
1997	129,657,636	679,443	3,381,752	4,211,109
1998	135,675,911	668,639	3,452,654	4,456,244
1999	139,915,090	691,868	3,454,977	4,571,322
2000	142,118,933	636,877	3,236,023	4,658,343
2001	148,632,364	646,654	3,252,065	4,871,331
2002	157,933,984	706,523	3,275,816	5,116,921
2003	165,907,405	761,974	3,602,481	5,397,647
2004	169,989,068	803,299	3,486,277	5,435,164
2005	173,682,720	861,257	3,439,211	5,480,385
2006	178,753,357	910,127	3,448,775	5,501,458
2007	176,159,265	956,866	3,480,301	5,383,577

	PSYCHOSES	ACUTE MYOCARDIAL INFARCTION (AMI)	ASTHMA	INJURY & POISONING
1979	844,025	688,256	512,964	6,390,512
1980	860,709	683,416	603,863	6,435,257
1981	965,847	772,357	627,926	6,476,685
1982	1,020,657	785,267	664,081	6,447,751
1983	1,073,172	784,672	708,607	6,340,400
1984	1,216,067	811,099	738,631	6,555,550
1985	1,336,485	870,981	742,906	6,426,072
1986	1,472,944	869,982	780,493	6,369,803
1987	1,562,034	872,189	784,536	6,353,854
1988	1,527,293	818,819	827,683	6,066,482
1989	1,568,434	797,551	835,203	5,939,040
1990	1,635,736	785,806	861,036	6,039,311
1991	1,846,023	809,233	923,066	6,059,732
1992	1,974,622	830,183	955,659	5,927,373
1993	2,218,577	839,502	999,768	6,013,845
1994	2,549,059	865,039	1,053,348	5,972,407
1995	2,544,787	857,664	1,176,094	5,965,231
1996	2,611,842	919,247	1,183,413	5,918,968
1997	2,747,776	870,820	1,242,049	5,890,989
1998	2,868,139	917,746	1,255,838	5,960,744
1999	3,027,574	980,652	1,347,468	6,037,891
2000	3,400,302	941,335	1,391,110	5,770,078
2001	3,637,158	970,056	1,486,111	6,008,225
2002	3,949,125	982,110	1,624,474	6,311,191
2003	3,828,118	956,292	1,842,260	6,506,529
2004	3,949,188	901,997	1,869,623	6,914,310
2005	4,153,897	864,948	1,927,629	7,121,636
2006	4,148,405	837,596	1,910,879	7,416,226
2007	4,165,231	747,859	1,926,090	7,153,496

ALL-LISTED PROCEDURES

	TOTAL PROCEDURES	CESAREAN SECTION	HYSTERECTOMY
1979	29,602,619	599,140	638,898
1980	31,411,526	619,499	649,446
1981	33,635,312	701,616	672,706
1982	34,632,285	730,339	649,998
1983	35,938,870	807,513	672,254
1984	36,155,105	812,661	664,487
1985	36,759,878	876,753	670,332
1986	38,000,321	905,616	644,188
1987	39,118,398	952,615	654,595
1988	39,191,713	933,397	578,346
1989	40,042,644	937,899	540,665
1990	40,505,676	945,330	591,066
1991	43,922,460	933,325	546,341
1992	42,627,290	920,872	579,966
1993	41,608,334	917,313	561,693
1994	40,710,173	858,417	555,987
1995	39,807,237	784,856	582,778
1996	40,397,052	835,136	590,740
1997	40,509,419	819,832	603,132
1998	41,499,972	899,914	644,910
1999	41,314,946	841,374	616,482
2000	39,981,139	855,296	633,204
2001	41,039,170	972,921	648,960
2002	42,533,221	1,058,565	669,252
2003	43,888,936	1,130,096	615,396
2004	45,022,620	1,236,412	617,228
2005	44,950,086	1,262,182	574,988
2006	45,963,407	1,294,711	569,303
2007	44,993,243	1,338,776	516,568

	CABG	PROSTATECTOMY	APPENDECTOMY
1979	114,002	292,958	311,043
1980	137,145	334,946	290,594
1981	159,471	347,996	311,821
1982	170,117	358,306	276,638
1983	191,308	356,830	282,080
1984	202,212	361,119	294,242
1985	230,353	366,804	283,156
1986	284,464	366,713	275,192
1987	331,793	410,132	302,811
1988	352,942	357,537	272,980
1989	368,210	376,315	252,749
1990	392,111	363,789	273,725
1991	407,353	363,472	254,992
1992	468,354	352,706	260,804
1993	485,399	317,184	250,084
1994	501,210	262,551	279,774
1995	573,133	239,255	237,126
1996	598,288	202,547	262,503
1997	607,161	223,117	262,570
1998	553,384	202,747	277,832
1999	571,416	192,075	290,023
2000	518,730	184,126	303,195
2001	516,127	185,247	312,496
2002	515,126	194,987	328,949
2003	467,335	166,630	330,762
2004	427,055	160,777	331,741
2005	465,658	155,034	341,332
2006	443,507	168,270	341,488
2007	405,115	155,674	326,422

FREQUENCIES FOR NEWBORN INFANTS:

	TOTAL	MALE	FEMALE	WELL BABY	SICK BABY
1979	3,680,008	1,906,207	1,773,801	2,751,669	928,339
1980	3,823,935	1,949,729	1,874,206	2,815,819	1,008,116
1981	3,841,360	1,987,109	1,854,251	2,797,195	1,044,165
1982	3,926,071	2,019,681	1,906,390	2,833,251	1,092,820
1983	3,828,449	1,971,712	1,856,737	2,729,574	1,098,875
1984	3,857,445	2,002,965	1,854,480	2,690,509	1,166,936
1985	3,793,931	1,953,429	1,840,502	2,475,143	1,318,788
1986	3,751,460	1,953,620	1,797,840	2,376,955	1,374,505
1987	3,970,798	2,032,577	1,938,221	2,434,603	1,536,195
1988	3,732,535	1,881,926	1,850,609	2,221,710	1,510,825
1989	3,884,481	1,989,462	1,895,019	2,256,737	1,627,744
1990	3,869,089	1,981,939	1,887,150	2,329,930	1,539,159
1991	3,879,666	2,002,719	1,876,947	2,293,271	1,586,395
1992	3,688,513	1,926,108	1,762,405	2,178,478	1,510,035
1993	3,578,666	1,823,043	1,755,623	2,074,364	1,504,302
1994	3,749,285	1,934,737	1,814,548	2,010,957	1,738,328
1995	3,630,558	1,872,083	1,758,475	2,168,523	1,462,035
1996	3,925,871	1,998,799	1,927,072	2,436,231	1,489,640
1997	3,789,839	1,931,256	1,858,583	2,325,682	1,464,157
1998	3,837,724	1,976,304	1,861,420	2,315,945	1,521,779
1999	3,726,258	1,893,106	1,833,152	2,293,557	1,432,701
2000	3,642,514	1,869,453	1,773,061	2,256,110	1,386,404
2001	3,658,669	1,865,637	1,793,032	2,218,659	1,440,010
2002	3,789,310	1,940,256	1,849,054	2,290,208	1,499,102
2003	3,875,317	2,018,078	1,857,239	2,222,777	1,652,540
2004	3,907,910	2,019,522	1,888,388	2,241,371	1,666,539
2005	3,999,112	2,044,389	1,954,723	2,200,973	1,798,139
2006	4,019,881	2,040,679	1,979,202	2,170,604	1,849,277
2007	4,047,423	2,103,842	1,943,581	2,266,572	1,780,851

	NORTHEAST	MIDWEST	SOUTH	WEST
1979	718,022	1,028,740	1,246,884	686,362
1980	735,271	1,043,176	1,315,209	730,279
1981	731,516	1,031,989	1,302,700	775,155
1982	733,230	1,002,579	1,404,114	786,148
1983	716,646	953,156	1,387,511	771,136
1984	688,886	938,452	1,447,418	782,689
1985	665,177	908,868	1,363,567	856,319
1986	685,520	910,701	1,250,755	904,484
1987	699,899	937,561	1,340,844	992,494
1988	736,577	887,419	1,226,961	881,578
1989	751,647	952,751	1,269,061	911,022
1990	732,771	905,076	1,323,844	907,398
1991	714,390	876,902	1,333,637	954,737
1992	684,583	806,457	1,264,024	933,449
1993	689,268	793,118	1,289,740	806,540
1994	751,139	816,617	1,315,455	866,074
1995	699,331	785,845	1,321,036	824,346
1996	645,439	825,932	1,403,164	1,051,336
1997	604,959	850,556	1,423,281	911,043
1998	731,478	865,737	1,328,003	912,506
1999	696,833	800,581	1,340,857	887,987
2000	754,967	792,441	1,343,998	751,108
2001	762,384	777,929	1,393,810	724,548
2002	724,777	764,279	1,354,445	945,809
2003	712,347	748,804	1,408,710	1,005,456
2004	702,190	802,719	1,403,084	999,917
2005	739,494	805,593	1,428,531	1,025,494
2006	722,066	833,077	1,464,410	1,000,328
2007	701,994	858,696	1,498,366	988,367