Appendix A: Data Sources, Definitions, and Methods

Data Source: The National Hospital Discharge Survey

The National Hospital Discharge Survey (NHDS) is a national probability survey designed to collect and produce national annual estimates of information on inpatients discharged from non-Federal short-stay hospitals in the United States. This survey, which has been conducted annually by the National Center for Health Statistics (NCHS) since 1965, covers the 50 States and the District of Columbia. Since 1988 hospitals with an average length of stay of less than 30 days for all inpatients, general hospitals, or children's general hospitals have been included in the survey. Prior to that time, only hospitals with an average length of stay of less than 30 days were included regardless of specialty. Federal, military, and Department of Veterans Affairs hospitals, as well as hospital units of institutions (such as prison hospitals), and hospitals with fewer than six beds staffed for patient use, are excluded.

Patient-level data collected include age, sex, race, diagnoses and procedures, length of stay, discharge disposition, and expected source of payment. Administrative and hospital data collected include number of hospital beds, type of ownership, and the geographic region of the hospital.

The design of the survey implemented in 1965 continued through 1987; the redesign with a new sample of hospitals, fielded in 1988, is currently in place. The sample for the 1965 NHDS was selected in 1964 from a frame of short-stay hospitals listed in the National Master Facility Inventory. A two-stage stratified sample design was used, with hospitals stratified according to bed size and geographic region. Sample hospitals were selected with probabilities ranging from 1 to 40. Within each participating hospital, a systematic random sample was selected from a daily listing sheet of discharges. Within-hospital sampling rates for discharges varied inversely with the probability of hospital selection, so that the overall probability of selecting a discharge was approximately the same across the sample.

Data collection was conducted by means of manual abstraction of patient information from sampled medical records. Sample selection and transcription of information from inpatient medical records to NHDS survey forms were performed by either the hospital staff or representatives of NCHS or both. In 1985 a new data-collection procedure was introduced. The procedure involved the purchase of computer data tapes from commercial abstracting services that contained the NHDS dataset from some hospitals in the NHDS sample. Discharges on these computer files were subjected to the NHDS sampling specifications, as well as the computer edits and estimation procedures. These two data collection methods, manual and automated, continue to be used in the NHDS.

A redesign of the NHDS was implemented for the 1988 survey (1). Under the redesign hospitals were selected using a modified three-stage sampling design. Units selected at the first stage consisted of primary sampling units (PSUs), which could be either hospitals or geographic areas like counties or townships. PSUs were used for the National Health Interview Survey (NHIS), also conducted by NCHS. Hospitals within PSUs 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.

Appendix A: Data Sources, Definitions, and Methods

The sampling rate was determined by the hospital's sampling stratum and the type of data collection system (manual or automated) used. Discharge records from hospitals submitting data via commercial abstracting services and selected State data systems (approximately 40 percent of sample hospitals) were arrayed by primary diagnoses, patient sex and age group, and date of discharge before sampling.

Although the old and new designs remain quite similar, it is still important to take the redesign into account when conducting trend analyses. Some of the differences between NHDS statistics based on the sample used for the 1965-87 samples, and those based on the sample drawn in 1988, may be due to the survey redesign rather than actual changes in hospital utilization. The injury chartbook tracks and reports on long-term trends in injury hospitalizations over the entire period from 1979-2001 so that the years around the redesign are only analyzed within a larger context. If researchers use NHDS data (including the tables in this chartbook) to compare the years immediately before and after the redesign, they should seek substantiation of the differences they find from other data sources. They should also review the findings from a report comparing estimates from the original and the new design to see the extent to which the redesign made a difference in injury estimates (2).

The NHDS hospital sample is updated every 3 years by continuing the sampling process among hospitals that become eligible for the survey during the intervening years and by deleting hospitals that are no longer NHDS-eligible.

The basic unit of estimation for NHDS is a sampled discharge. The basic estimation procedure involves inflation by the reciprocals of the probabilities of selection. There are adjustments for nonresponding hospitals and discharges, and a post-ratio adjustment to fixed totals is employed.

Since 1979 the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) has been used for classifying diagnoses and procedures in the NHDS (3). Beginning in 1986 the ICD-9-CM has undergone minor annual modifications. These modifications become effective in October of each year and are published in an addendum. Users of the NHDS who wish to conduct trend analyses or other multiple year studies should look up the diagnoses and procedures they are studying in the addenda to be sure they include data on the correct codes for each of the years studied. ICD-9-CM Addenda and a conversion table can be found online at: http://www.cdc.gov/nchs/icd9.htm ; see "Appendix A," ICD-9-CM addenda affecting injuries.

Appendix A: Data Sources, Definitions, and Methods

Definitions and Methods

Alphabetical listing of terms and methodologies used in this report:

Age- Patient's age at the birthday prior to admission to the hospital.

Age adjustment-Age adjustment, using the direct method, is the application of age-specific rates in a population of interest to a standardized age distribution in order to eliminate differences in observed rates that result from age differences in population composition. This adjustment is usually done when comparing two or more populations at one point in time or one population at two or more points in time. Age-adjusted rates are calculated by the direct method as follows:

$$\sum_{i=1}^{n} r_i \times (p_i/P)$$

where r_i = rate in age group i in the population of interest

p_i = standard population in age group i

$$P=\sum_{i=1}^{n} p_i$$

n = total number of age groups over the age range of the age-adjusted rate

Age adjustment by the direct method requires use of a standard age distribution. The standard for age-adjusting estimates from NCHS surveys is the year 2000 projected U.S. resident population. The standardized age distribution used in this report is listed in Table I, page 13.

For more information on implementing the 2000 population standard for age-adjusting death rates, see Anderson RN, Rosenberg HM. Age Standardization of Death Rates: Implementation of the Year 2000 Standard. National vital statistics reports; vol. 47 no. 3. Hyattsville, Maryland: National Center for Health Statistics.1998 (available online at www.cdc.gov/nchs/data/nvsr/nvsr47/nvs47_03.pdf). For more information on the derivation of age adjustment weights for use with NCHS survey data, see Klein RJ, Schoenborn CA. Age Adjustment Using the 2000 Projected U.S. Population. Healthy People Statistical Notes no. 20. Hyattsville, Maryland: National Center for Health Statistics. 2001 (available online at http://www.cdc.gov/nchs/data/statnt/statnt20.pdf). The year 2000 projected U.S. resident population is available through the Bureau of the Census home page at www.census.gov/.

Appendix A: Data Sources, Definitions, and Methods

Table I. Projected year 2000 U.S. population and proportion distribution by age for age-adjusting rates (adapted for NHDS data)

Age	Population	Proportion distribution (weights)	Standard million
Total	274,634,000	1.000000	1,000,000
Under 15 years	58,964,000	0.214700	214,700
15-24 years	38,077,000	0.138646	138,646
25-44 years	81,892,000	0.298186	298,186
45-64 years	60,991,000	0.222081	222,081
65 years and over	34,710,000	0.126387	126,387

SOURCE: Anderson RN, Rosenberg HM. Age Standardization of Death Rates: Implementation of the Year 2000 Standard. National vital statistics reports; vol 47 no 3. Hyattsville, Maryland: National Center for Health Statistics. 1998.

Age-adjusted percents and means were calculated using rounded values (one decimal place). Age-adjusted rates were calculated using unrounded values. Data that are age adjusted are noted in the figure and table notes. The relative standard errors (RSE) for the unadjusted percents, means, and rates were applied to the age-adjusted percents, means, and rates.

Average length of stay- Mean length of stay for discharges. It is calculated by dividing the total number of days of care, counting the date of admission but not the date of discharge, by the number of discharges. See related Days of care.

Average percent change over time, test of trend, and test of significance between two statistics- In this report, trends in injury statistics (expressed here as rates, means, or percents) are summarized by two measures of average percent change (rather than by absolute change) as a "total" percent change, referred to in the tables as average percent change (e.g., a total decline of 30 percent for the 22 year period 1979-2001) and as an "average annual" percent change (e.g., an average increase of 2 percent per year).

Weighted least squares regression was used to assess trends in annual estimates (4,5) by fitting a linear model to the logarithm of annual estimates (5-7). The inverse of the variance of the logarithm of the annual estimates was used as the weights in a weighted least squares algorithm to obtain the parameter estimate, ß, and corresponding test statistics and standard errors for the linear regression model.

From the fitted regression equation, the average annual percent change (7) in the outcome, $(100\ (e^{ß}\ -1))$, and the total percent change over the 22 year time period, $100\ (e^{ß*22}\ -1)$, were calculated for this report.

The regression model used in this report makes use of all data points for the calculation of the percent change. Another approach that could have been used to calculate change in the rates subtracts the

rate (or percent) for the last year from the value for the first year, dividing by the value at the first year and multiplied by 100 percent.

An estimate of the total percent change based only on the first and last data points differs from an estimate of total percent change based on all data points; this difference can be large if either the first or last data point differs significantly from the trend of the intervening years.

For each model tested, the standard error of the parameter estimate, ß, was used to determine whether the parameter estimate differed from zero; this determination was used to indicate whether or not there was a statistically significant trend in the injury statistic over the time period. [See"Appendix A," Sampling errors] Terms such as "increases" and "decreases" indicate injury trends that were statistically significant at the .05 level. The determination of statistical inference is based on the two tailed t-test. The Bonferroni technique for multiple comparisons was used to establish the critical value for statistically significant differences (0.05 level of significance) for each set of comparisons within each table. Terms relating to differences such as "higher than" indicate that the difference is statistically significant. Terms such as "similar" or "no difference" indicate that the difference is not statistically significant. Terms such as "generally higher" or "in most instances" refer to tests involving multiple comparisons that were significant at least 70% of the time. A lack of comment regarding the difference between any two estimates does not mean that the difference was tested and found to be not significant.

Barell Matrix - The two-dimensional array of ICD-9-CM injury codes (as of 1998). Codes are grouped by body region of the injury and the nature of the injury. This matrix provides a standard format for reporting injury data. This injury diagnosis matrix is a product of the participants in the International Collaborative Effort (ICE) on Injury Statistics. For more information about the Barell Matrix refer to the Web site page: www.cdc.gov/nchs/about/otheract/ice/barellmatrix.htm (8), "See Appendix B," Barell Matrix.

Days of care - The total number of days a patient spent in the hospital. A stay of less than 1 day (patient admission and discharge on the same day) is counted as 1 day. For patients admitted and discharged on different days, the number of days of care is computed by counting all the days from (and including) the day of admission to (but not including) the day of discharge. See related: Average length of stay, Discharge, Hospital, Patient.

Discharge - A completed inpatient hospitalization. A hospitalization may be completed by death or by releasing the patient to the customary place of residence, a nursing home, another hospital, or other location, or if the patient leaves against medical advice. Persons with multiple discharges during a calendar year may be sampled more than once, thus estimates are for discharges, not individual persons. See related Average length of stay, Days of care, Patient.

Discharge Diagnosis - See First-listed diagnosis.

Appendix A: Data Sources, Definitions, and Methods

Discharge Disposition - The disposition of a patient on termination of hospitalization into one of the following categories:

- Home or routine discharge - patient returned to previous place of residence after discharge from the hospital.
- Long-term care institution - patient entered a nursing home including skilled nursing facilities, extended care facilities, custodial care facilities, or other long term care placement upon discharge from the hospital.
- Short-term facility - patient transferred to another short-term hospital at discharge, including short-term maternity hospitals.
- Dead - patient who died during the inpatient stay.
- Other or not stated - patient who has no discharge disposition listed or other disposition which does not fit into the above categories.

Expected source of payment - The principal expected source of payment for the hospitalization. Terms used in figure 31, data table 31, and in appendix table 31 are defined below.

- Medicare - The health insurance program for the aged and disabled administered by the Centers for Medicare and Medicaid Services (formerly the Health Care Financing Administration).
- Medicaid - A jointly funded Federal-State health insurance program providing medical care to those unable to afford it.
- Worker's compensation - A State or municipal disability insurance or industrial accident insurance.
- Private health insurance- Includes HMO/PPO, Blue Cross/Blue Shield, and other private.
  - -HMO/PPO - Any health maintenance organization (HMO) or preferred provider organization (PPO) sponsored by consumers, communities, physicians, or hospitals.
  - -Blue Cross/Blue Shield and other private - A private insurance plan not specified as an HMO/PPO. This includes Blue Cross/Blue Shield plans, medical coverage provided by life insurance companies, casualty insurance companies, health insurance companies, and independent plans such as employer/union-sponsored plans and/or self-funded plans (partial or total).
- Self-pay - The majority of the costs for the hospitalization were expected to be paid by the patient, spouse, family, or next-of-kin.
- Other government- Other Federal, State, or local government other than worker's compensation, Medicare, and Medicaid not listed separately including casualty insurance paid by the State, Federal or State medical research grant.
- No charge- Patients admitted with the understanding that payment would not be expected because the medical services are free, e.g., charity patients or research or teaching patients.
- Other and not stated- Other nonprofit source of payment, e.g., church, Shriner's, etc., and no source of payment indicated.

External cause of injury code - External cause of injury and poisoning codes are a supplemental component of the ICD-9-CM. Each code begins with an "E" and is followed by a three- or four-digit number (e.g., E800.1). External cause codes are intended to be used in addition to a code from the main chapters of the ICD-9-CM (3). External cause codes provide information on the circumstances and causes of injuries.

ICD-9-CM codes E800-E999 excluding E849 (place of occurrence code), E869.4 (Accidental poisoning by second-hand tobacco smoke), E870-E879 (Misadventures to patients during surgical and medical care), E930-E949 (Drugs, medicinal, and biological substances causing adverse effects in therapeutic use),

E967 (Perpetrator of child and adult abuse) were used in this report (with the exception of figure 23 which includes all codes from E800-E999). If more than one external cause code is listed for a given record, this report used only the first external cause code listed in the diagnostic code fields which met the criteria for this report (9,10). See "Appendix C," ICD-9-CM External Cause of Injury Matrix.

First-listed diagnosis - The diagnosis established after study to be chiefly responsible for the admission of the patient to the hospital is the principal diagnosis. The first-listed diagnosis is the one specified as the principal diagnosis on the face sheet or discharge summary of the medical record. However, if the principal diagnosis is not specified, the first one listed on the face sheet or discharge summary of the medical record is used. The number of first-listed diagnoses is equivalent to the number of discharges.

Geographic region - Hospitals are classified by location in one of the four geographic regions of the United States that correspond to those used by the U.S. Bureau of the Census. The four geographic regions are as follows:

- Northeast - Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New York, New Jersey, and Pennsylvania

- Midwest - Ohio, Indiana, Illinois, Michigan, Wisconsin, Minnesota, Iowa, Missouri, North Dakota, South Dakota, Nebraska, and Kansas

- South - Delaware, Maryland, District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida, Kentucky, Tennessee, Alabama, Mississippi, Arkansas, Louisiana, Oklahoma, and Texas

- West - Montana, Idaho, Wyoming, Colorado, New Mexico, Arizona, Utah, Nevada, Washington, Oregon, California, Alaska, and Hawaii

Hospital - The NHDS includes hospitals with an average length of stay for all inpatients of less than 30 days (short stay) or those whose specialty is general (medical or surgical) or children's general. Federal hospitals, hospital units of institutions, and hospitals with less than six beds staffed for patient use are excluded.

ICD-9-CM addenda affecting injuries - Assignment of new diagnostic and procedure codes, fourth and fifth-digit expansion of codes, as well as code deletions, are contained in addenda 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 Health Care Financing Administration).

Since 1986 the ICD-9-CM has been updated annually with the exception of 1999. (No addendum was released in 1999 because of concerns about instituting coding changes before the millennium crossover.) Addenda to the ICD-9-CM become effective on October 1, but the new codes included in NHDS data for October 1 to December 31 of each year are converted back to their previous code assignments so that annual estimates are based on consistent coding.

Table II shows the ICD-9-CM addenda for injuries, the date they were introduced, and the code to which the diagnosis or procedure had been assigned. Specific titles and more detailed information about the coding system can be found in appropriate volumes of the ICD-9-CM (3).

Appendix A: Data Sources, Definitions, and Methods

Table II. Changes in injury ICD-9-CM diagnoses codes from the 1986-2000 addenda

Current code(s) assignment	Effective October 1	Previous code(s) assignment
864.05	1992	864.09
864.15	1992	864.19
909.5	1994	909.9
922.31-922.33	1996	922.3
925.1-925.2	1993	925
959.0 (code title restated)	1997	959.0
959.01	1997	854.00
959.09	1997	959.0
965.61	1998	965.6
965.69	1998	965.6
989.81-989.84	1995	989.8
989.89	1995	989.8
995.50-995.55	1996	995.5
995.59	1996	995.5
995.60-995.69	1993	995.0
995.7	2000	None
995.80	1996	995.81
995.81 (code title restated)	1996	995.81
995.82-995.85	1996	995.81
E854.8	1995	E858.8
E869.4	1994	E869.8
E880.1	1995	E884.9
E884.3-E884.4	1995	E884.2
E884.5-E884.6	1995	E884.9
E885.1-E885.4, E885.9	2000	E885
E906.5	1995	E906.3
E908.0-E908.4	1995	E908
E908.8-E908.9	1995	E908
E909.0-E909.4	1995	E909
E909.8-E909.9	1995	E909
E920.5	1995	E920.4
E922.4	1997	E917.9
E924.2	1995	E924.0
E928.3	2000	E928.8
E955.6	1997	E955.9
E967.2	1996	E967.0
E967.3	1996	None
E967.4-E967.8	1996	E967.1
E968.7	2000	E968.8
E968.5	1995	E968.8
E968.6	1997	E968.8
E985.6	1997	E985.4

An online version of Table II can be found at: www.cdc.gov/nchs/data/icd9/icdcnv05.pdf

Appendix A: Data Sources, Definitions, and Methods

ICD-9-CM External Cause of Injury Matrix - The two-dimensional array of ICD-9-CM external cause of injury codes is grouped by mechanism and intent of injury e.g., fall (mechanism) and unintentional (intent). The categories within the matrix are mutually exclusive. This framework was developed by CDC in collaboration with members of the American Public Health Association's Injury Control and Emergency Health Services Section (ICEHS) (10). The International Collaborative Effort (ICE) on Injury Statistics endorsed the matrix for international comparisons. More information can be found online at http://www.cdc.gov/ncipc/whatsnew/matrix2.htm. See related External cause of injury code and "Appendix C", ICD-9-CM External Cause of Injury Matrix.

Injury discharge - Hospitalizations where the principal diagnosis (first-listed) is an ICD-9-CM diagnosis including 800-909.2, 909.4, 909.9, 910-994.9, 995.5-995.59, 995.80-995.85 (9).

International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) - The United States currently uses ICD-9-CM to code morbidity diagnoses and inpatient procedures. The ICD-9-CM is based on and is compatible with the World Health Organization's International Classification of Diseases, Ninth Revision. ICD-9-CM is divided into 17 chapters and 2 supplemental classifications. The chapters are arranged primarily by body system. In addition, there are chapters for infectious and parasitic diseases; neoplasms; endocrine, metabolic, and nutritional diseases; mental disorders; complications of pregnancy, childbirth and puerperium; certain conditions originating in the perinatal period; congenital anomalies; and symptoms, signs and ill-defined conditions. The two supplemental classifications are for factors influencing health status and contact with health services and classification of external causes of injury and poisoning. More information can be found online at www.cdc.gov/nchs/icd9.htm

International Collaborative Effort on Injury Statistics (ICE on Injury Statistics) - An international activity sponsored by the Centers for Disease Control and Prevention's National Center for Health Statistics. The ICE on Injury Statistics also receives funding from the National Institutes of Health's National Institute of Child Health and Human Development. The purpose of the ICE on Injury Statistics is to improve international comparability and quality of injury data. The ultimate goal is to provide the data needed to better understand the causes of injury and the most effective means of prevention. More information can be found online at www.cdc.gov/nchs/advice.htm.

Appendix A: Data Sources, Definitions, and Methods

Figure 1. Number of diagnoses A and B plotted on a linear scale, 1970-2000

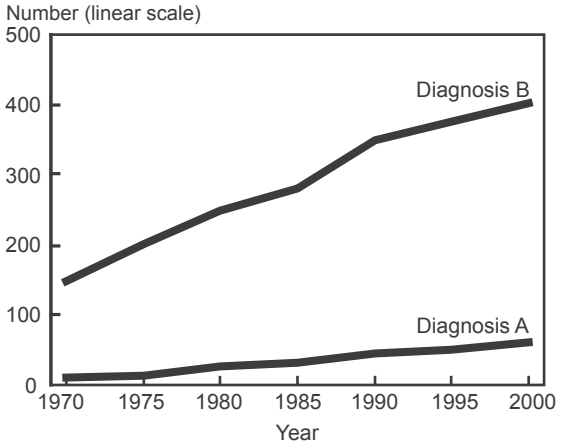
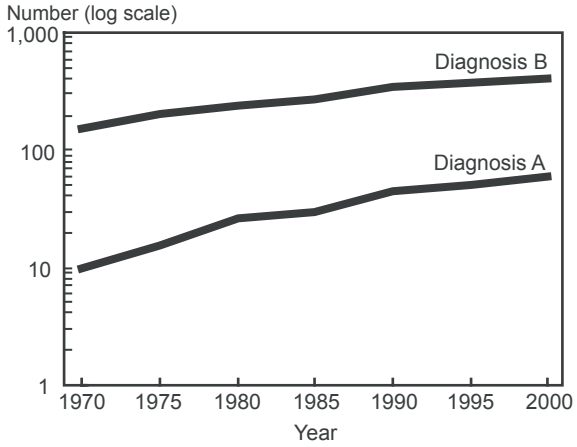


Figure 2. Number of diagnoses A and B plotted on a log scale, 1970-2000



Length of stay - See Average length of stay.

Logarithmic scale - A scale used to emphasize relative changes in numbers. The choice of a linear or logarithmic (log) scale depends on what the analyst/author wants to emphasize about the graph for the audience-the absolute or the relative changes in numbers. The absolute change is the arithmetic difference between two values. The relative change is the percent difference between two values.

The linear scale is the scale most frequently used and recognized, and it emphasizes the absolute changes between data points over time (11). Figure 1 is an example of data plotted on a linear (both x and y axes) scale. The absolute change from 1970 to 2000 for diagnosis A was 50 (from 10 to 60) and for diagnosis B it was 254 (from 150 to 404).

Logarithmic scales, on the other hand, emphasize the relative or percentage change between data points. Figure 2 is a semi-log scale graph (x-axis is linear and y-axis is logarithmic) of the same information as in figure 1. Equal distances on a log scale represent an equal percentage change. This feature makes a log graph particularly useful for showing rates of change in data. Thus, diagnosis A increased 500 percent from 1970 to 2000 while diagnosis B increased 169 percent.

If the important piece of information to be gleaned from these data is the greater percentage change in diagnosis A compared with diagnosis B, then the presentation on the log scale (figure 2) makes that very clear. The linear scale (figure 1) more clearly emphasizes the greater absolute magnitude of diagnosis B compared with A. In addition, trends can be shown on a log scale to enable measures with large differences in magnitude to be shown on the same chart. One potential disadvantage to using the log scale is that the absolute magnitude of changes may appear less dramatic (12).

Appendix A: Data Sources, Definitions, and Methods

To properly interpret data on a log scale, the following points should be kept in mind:

1. A sloping straight line indicates a constant rate (not amount) of increase or decrease in the values.
2. A horizontal line indicates no change.
3. The slope of the line indicates the rate of increase or decrease.
4. Parallel lines, regardless of their magnitude, depict similar rates of change (11).

Because this report is designed to emphasize relative rates of change in injury hospitalizations from 1979 to 2001, we have chosen to present the data using the logarithmic scale. All of the actual data points are given in the detailed data and appendix tables.

Measurement errors - As in any survey, results are subject to nonsampling or measurement errors, which include errors due to hospital nonresponse, missing abstracts, information incompletely or inaccurately recorded on abstract forms, and processing errors. In general, less than one-half of 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 were missing on average for about 15 percent of all discharges for all years, but this varies by year. The percent of race reporting for injury discharges was similar to the percent of race reporting for noninjury discharges. Except for one year, no attempt was made to impute for these missing values. In 1981 "race not stated" values were imputed for approximately 11 percent of the records so there are no "not stated" cases for that year. Data by race are provided by the hospital and are not based on patient self-report. Details about the underreporting of race in the NHDS can be found online at: www.cdc.gov/nchs/data/ad/ad265.pdf.

For data years before 1996, if dates of admission or discharge were missing or invalid, a length of stay was imputed by assigning the patient a stay characteristic of the stays of other patients of the same age.

A new edit program was developed and implemented for the NHDS beginning in 1996. The updated edit 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 new edit program 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 any 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.

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

Appendix A: Data Sources, Definitions, and Methods

Patient - A patient is a person who is formally admitted to the inpatient service of a hospital for observation, care, diagnosis, or treatment. Persons with multiple discharges during the year may be sampled more than once, thus estimates are for discharges, not individual persons. See related Average length of stay, Days of care, Discharge, Hospital.

Percent change - See Average percent change over time.

Population estimates - Hospital utilization rates are computed using U.S. Census Bureau population estimates as denominators.

Estimates of the civilian resident population as of July 1 of each year from 1979 to 2001 are used to calculate rates in this report. These are available in the multi-year public use data file for 1979 to 2000, and 2001 population estimates are provided in the 2001 public use data file. Population estimates for 1979 were adjusted based on the 1980 census. The estimates for 1980-89 have been adjusted based on the 1990 decennial census. Population estimates for 1990-2000 have been adjusted for underenumeration using the 1990 National Population Adjustment Matrix. Population estimates for 2000 are based on the 1990 census because population estimates from the 2000 census were not available when this report was prepared. Population estimates for 2001 are based on the 2000 census. Due to these updates and adjustments, it should be noted that rates calculated with these estimates may differ slightly from those appearing in published NCHS reports or those calculated from population estimates disseminated with the NHDS single-year file documentation.

Rate - A rate is a measure of an event, disease, or condition in relation to a unit of population for a specified time. For example:

Discharge rate per 10,000 population=(Number of annual discharges / annual population) *10,000

Relative standard error - The relative standard error (RSE) is a measure of an estimate's reliability. The RSE of an estimate is obtained by dividing the standard error of the estimate (SE(r)) by the estimate itself (r). This quantity is expressed as a percent of the estimate and is calculated as follows: RSE = 100 x (SE(r)/r). The RSE is used as a guide to the reliability of the estimate (see Standards of reliability). See related Sampling errors.

Sampling errors - Error introduced by chance because only a sample rather than the entire universe is surveyed.

Before 1988 standard error estimates for NHDS were produced using a computerized routine based on a rigorously unbiased 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 this report, the standard error estimates for 1979-87 were generated using this method.

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. Bieler and Williams published a description of the software and its approach (13). For this report the standard errors for 1988-2001 were generated using this software. See related Relative standard error.

Standards of reliability - Based on consideration of the complex sample design of the NHDS, the following guidelines are recommended for using NHDS estimates and are used throughout this chartbook:

Estimates with relative standard errors (see Relative standard error definition) of more than 30 percent or that are based on fewer than 30 records are not presented due to low reliability; only an asterisk (*) appears in the tables. Estimates based on 30-59 records are preceded by an asterisk to indicate that they also have low reliability. Only an asterisk is shown in the tables if the estimate is less than 5,000 because these estimates generally have a relative standard error of more than 30 percent or a sample size of less than 30. Estimates of discharges of less than 9,000 are preceded by an asterisk because they are generally based on fewer than 60 records. Days of care estimates derived from the smaller estimates described in this paragraph are also replaced by or preceded by asterisks.

STIPDA or State and Territorial Injury Prevention Directors Association- The national nonprofit organization with membership of public health injury professionals that represents all States and territories. Its mission is to promote, sustain, and enhance the ability of State and territorial public health departments to reduce death and disability associated with injuries. This is accomplished by disseminating information on state-of-the-art injury prevention and control policies and strategies (9). More information about STIPDA can be found online at: www.stipda.org.

Test of trend - See Average percent change over time, test of trend, and test of significance between two statistics.

Appendix A: Data Sources, Definitions, and Methods

References

1. Dennison CF, Pokras R. Design and Operation of the National Hospital Discharge Survey: 1988 redesign. National Center for Health Statistics. Vital Health Stat 1 (39). 2000.
2. Haupt B, Kozak LJ. Estimates from two survey designs: National Hospital Discharge Survey. National Center for Health Statistics. Vital Health Stat 13 (111). 1992.
3. International Classification of Diseases, 9th Revision, Clinical Modification, 6th Edition. U.S. Department of Health and Human Services, National Center for Health Statistics, Health Care Financing Administration. 1998. http://www.cdc.gov/nchs/icd9.htm
4. Gillum BS, Graves EJ, Kozak LJ. Trends in hospital utilization: United States 1988- 92. National Center for Health Statistics. Vital Health Stat 13(124). 1996.
5. Gardocki GJ, Pokras R. Utilization of short-stay hospitals by persons with heart disease and malignant neoplasms: National Hospital Discharge Survey, United States, 1977. National Center for Health Statistics. Vital Health Stat 13(52). 1981.
6. Sirken MG, Shimizu BI, French DK, Brock, DB. Manual on standards and procedures for reviewing statistical reports. Rockville, Maryland: National Center for Health Statistics. 1974.
7. Kleinman JC. "Methodological Issues in the Analysis of Vital Statistics" in Kiely, M: Reproductive and perinatal epidemiology, pp 448-68. Boca Raton, Florida: CRC Press. 1991.
8. Barell V, Aharonson-Daniel L, Fingerhut LA, MacKenzie EJ, et al. An introduction to the Barell body region by nature of injury diagnosis matrix. Inj Prev 8:91-6. 2002.
9. Injury Surveillance Workgroup. Consensus Recommendations for using hospital discharge data for injury surveillance. Marietta, Georgia: State and Territorial Injury Prevention Directors Association. 2003.
10. Centers for Disease Control and Prevention. Recommended framework for presenting injury mortality data. MMWR 46(RR14):1-30. 1997.
11. Page RM, Cole GE, Timmreck TC: Basic epidemiological methods and biostatistics: A practical guidebook. Sudbury, Massachusetts: Jones and Bartlett Publishers. 1995.
12. Jekel JF, Elmore JG, Katz DL: Epidemiology biostatistics and preventive medicine. Philadelphia, Pennsylvania: W.B. Saunders Company. 1996.
13. Bieler GS, Williams RL. Analyzing Survey Data Using SUDAAN Release 7.5. Research Triangle Institute: Research Triangle Park, North Carolina. 1997.

Appendix B: The Barell injury diagnosis matrix; classification by body region and nature of injury (based on five digit ICD-9-CM codes)

Body region				Nature of injury											
				Fracture 800-829	Dislocation 830-839	Sprains & strains 840-848	Internal 850-854,860-869 952, 995.55	Open wound 870-884, 890-894	Amputations 885-887, 895-897	Blood vessels 900-904	Contusion/ superficial 910-924	Crush 925-929	Burns 940-949	Nerves 950-951 953-957	Unspecified 959
Head and neck	Traumatic brain injury	1	Type 1 TBI	800, 801, 803, 804(.1-.4,.6-.9) 800, 801, 803, 804(.03-.05,.53-.55)			850(.2-.4) 851-854*, 995.55							950.1-.3	
		2	Type 2 TBI	800, 801, 803, 804(.00,.02,.06,.09), 800, 801, 803, 804(.50,.52,.56,.59)			850(.0,.1,.5,.9)								
		3	Type 3 TBI	800, 801, 803, 804(.01,.51)											
	Other head, face and neck	4	Other head					873.0-.1,.8-.9					941.x6	951	959.01*
		5	Face	802	830	848.0-.1		872, 873.2-.7					941.x1,.x3-.x5,.x7	/	
		6	Eye					870-871			918, 921		940, 941.x2	950(.0,.9)	
		7	Neck	807.5-.6		848.2		874				925.2	941.x8	953.0, 954.0	
		8	Head, face and neck unspecified							900	910, 920	925.1	941.x0,.x9, 947.0	957.0	959.09
Spine and back	Spinal cord (SCI)	9	Cervical SCI	806.0-.1			952.0								
		10	Thoracic/dorsal SCI	806.2-.3			952.1								
		11	Lumbar SCI	806.4-.5			952.2								
		12	Sacrum coccyx SCI	806.6-.7			952.3-.4								
		13	Spine + back unspecified SCI	806.8-.9			952.8-.9								
	Vertebral column (VCI)	14	Cervical VCI	805.0-.1	839.0-.1	847.0									
		15	Thoracic/dorsal VCI	805.2-.3	839.21,.31	847.1									
		16	Lumbar VCI	805.4-.5	839.20,.30	847.2									
		17	Sacrum coccyx VCI	805.6-.7	839(.41-.42, .51-.52)	847.3-.4									
		18	Spine + back unspecified VCI	805.8-.9	839(.40,.49,.50,.59)										
Torso	Torso	19	Chest (thorax)	807.0-.4	839.61,.71	848.3-.4	860-862	875, 879.0-.1		901	922(.0,.1,.33)	926.19	942.x1-x2	953.1	
		20	Abdomen				863-866, 868	879.2-.5		902.0-.4	922.2	/	942.x3, 947.3	953.2, 953.5	
		21	Pelvis & urogenital	808	839.69,.79	846, 848.5	867	877-878		902(.5, .81-.82)	922.4	926(.0, .12)	942.x5, 947.4	953.3	
		22	Trunk	809				879.6-.7			911, 922.8-.9	926.8-.9	942.x0, 942.x9	954.1, .8-.9	959.1
		23	Back and buttock			847.9		876			922.31-.32	926.11	942.x4		
Extremities	Upper	24	Shoulder & upper arm	810-812	831	840		880	887.2-.3		912, 923.0	927.0	943.x3-.x6		959.2
		25	Forearm & elbow	813	832	841		881.x0-x1	887.0-.1		923.1	927.1	943.x1-x2		
		26	Wrist, hand & fingers	814-817	833, 834	842		881.x2,882, 883	885-886		914-915, 923.2-.3	927.2-.3	944		959.4-.5
		27	Oher & unspecified	818		/		884	887.4-.7	903	913,923.8,.9	927.8-.9	943.x0,.x9	953.4, 955	959.3
	Lower	28	Hip	820	835	843					924.01	928.01			
		29	Upper leg & thigh	821					897.2-.3		924.00	928.00	945.x6		
		30	Knee	822	836	844.0-.3					924.11	928.11	945.x5		
		31	Lower leg & ankle	823-824	837	845.0			897.0-.1		924.10,.21	928.10,.21	945.x3-.x4		
		32	Foot & toes	825-826	838	845.1		892-893	895-896		917, 924.3,.20	928.3,.20	945.x1-.x2		
		33	Other & unspecified	827		844.8,.9		890-891,894	897.4-.7	904.0-.8	916, 924.4-.5	928.8,.9	945.x0-.x9		959.6-.7
Unclassifiable by site	Other & unspecified	34	Other/multiple	819, 828						902.87,.89			947.1-.2	953.8, 956	
		35	Unspecified site	829	839.8-.9	848.8-.9	869	879(.8-.9)		902.9, 904.9	919, 924.8,.9	929	946, 947.8,.9 948, 949	953.9, 957.1,.8,.9	959.8,.9
	System wide	36	System-wide & late effects	Foreign body (930-939), Early complications of trauma (958), Poisoning (960-979), Toxic Effects (980-989), Other and unspecified effects of external cause (990-994), Child and adult maltreatment (995.50-.54,.59, 995.80-.85), Late effects of injuries, poisonings, toxic effects and other external causes (905-909) excluding 909(.3, .5)											

For purposes of classification, head injuries are labeled as Type 1 TBI if there is recorded evidence of an intracranial injury or a moderate or a prolonged loss of consciousness (LOC), Shaken Infant Syndrome (SIS), or injuries to the optic nerve pathways. Type 2 TBI includes injuries with no recorded evidence of intracranial injury, and LOC of less than one hour, or LOC of unknown duration, or unspecified level of consciousness. Type 3 TBI includes patients with no evidence of intracranial injury and no LOC.

*NOTES: 959.01 (added to ICD-9-CM in 1997) is not intended to be assigned to TBI cases; however, in the USA it has been assigned incorrectly to a substantial proportion of cases previously coded 854.
The Matrix is available on the internet at www.cdc.gov/nchs/about/otheract/ice/barellmatrix.htm. Empty cells = No diagnosis codes.

Appendix C: Recommended framework of E-code groupings for presenting injury morbidity data (August 16, 2004)

Appendix C was corrected July 2005. Codes outlined in yellow box have been shifted down one row from the original print and online versions.

Mechanism/cause	Manner/Intent				
	Unintentional	Self-inflicted	Assault	Undetermined	Other[1]
Cut/pierce	E920.0-.9	E956	E966	E986	E974
Drowning/submersion	E830.0-.9, E832.0-.9, E910.0-.9	E954	E964	E984	
Fall	E880.0-E886.9, E888	E957.0-.9	E968.1	E987.0-.9	
Fire/burn	E890.0-E899, E924.0-.9	E958.1,.2,.7	E961, E968.0,.3, E979.3	E988.1,.2,.7	
Fire/flame	E890.0-E899	E958.1	E968.0, E979.3	E988.1	
Hot object/substance	E924.0-.9	E958.2,.7	E961, E968.3	E988.2,.7	
Firearm	E922.0-.3,.8, .9	E955.0-.4	E965.0-.4, E979.4	E985.0-.4	E970
Machinery	E919 (.0-.9)				
Motor vehicle traffic[2]	E810-E819 (.0-.9)	E958.5	E968.5	E988.5	
Occupant	E810-E819 (.0,.1)				
Motorcyclist	E810-E819 (.2,.3)				
Pedal cyclist	E810-E819 (.6)				
Pedestrian	E810-E819 (.7)				
Unspecified	E810-E819 (.9)				
Pedal cyclist, other	E800-E807 (.3), E820-E825 (.6), E826.1,.9, E827-E829(.1)				
Pedestrian, other	E800-807(.2), E820-E825(.7), E826-E829(.0)				
Transport, other	E800-E807 (.0,.1,.8,.9), E820-E825 (.0-.5,.8,.9), E826.2-.8, E827-E829 (.2-.9), E831.0-.9, E833.0-E845.9	E958.6		E988.6	
Natural/environmental	E900.0-E909, E928.0-.2	E958.3		E988.3	
Bites and stings	E905.0-.6,.9, E906.0-.4,.5,.9				
Overexertion	E927				
Poisoning	E850.0-E869.9	E950.0-E952.9	E962.0-.9	E980.0-E982.9	E972
Struck by, against	E916-E917.9		E960.0; E968.2		E973, E975
Suffocation	E911-E913.9	E953.0-.9	E963	E983.0-.9	
Other specified and classifiable[3]	E846-E848, E914-E915, E918, E921.0-.9, E922.4,.5 E923.0-.9, E925.0-E926.9, E928.3, E929.0-.5	E955.5,.6,.7,.9, E958.0,.4	E960.1, E965.5-.9, E967.0-.9, E968.4,.6, .7, E979.0-.2, E979.5-.9	E985.5,.6,.7, E988.0,.4	E971, E978, E990-E994, E996, E997.0-.2
Other specified, not elsewhere classifiable	E928.8, E929.8	E958.8, E959	E968.8, E969	E988.8, E989	E977, E995, E997.8, E998, E999
Unspecified	E887, E928.9, E929.9	E958.9	E968.9	E988.9	E976, E997.9
All injury	E800-E869, E880-E929	E950-E959	E960-E969, E979	E980-E989	E970-E978, E990-E999
Adverse effects					E870-E879, E930.0-E949.9
Medical care					E870-E879
Drugs					E930.0-E949.9
All external causes					E800-E999

[1]Includes legal intervention (E970-E978) and operations of war (E990-E999).
[2]Three 4th-digit codes (.4 [occupant of streetcar], .5 [rider of animal], .8 [other specified person]) are not presented separately because of small numbers. However, because they are included in the overall motor vehicle traffic category, the sum of these categories can be derived by subtraction.
[3]ICD-9-CM E849 code should never be first-listed E code and should only appear as an additional code to specify the place of occurrence of the injury incident and has been excluded from the matrix.

NOTES: ICD-9-CM codes were updated as of August 2004. The matrix is avilable onthe internet at www.cdc.gov/ncipc/whatsnew/matrix2.htm. Empty cells = No diagnosis codes.