2000 NHDS DATA FILE DOCUMENTATION

I. DESCRIPTION OF THE NATIONAL HOSPITAL DISCHARGE SURVEY

Introduction. This document and its appendices contain information for users of the 2000 National Hospital Discharge Survey (NHDS) public use data file. Conducted annually by the National Center for Health Statistics, NHDS collects medical and demographic information from a sample of discharge records selected from a national sample of non-Federal, short-stay hospitals. The data serve as a basis for calculating statistics on hospital inpatient utilization in the United States. For a brief description of the survey design and data collection procedures, see below. For a more detailed description of the survey design, data collection procedures, and the estimation process, see Reference 1. Publications based on the data for each survey year can be obtained from the NCHS website at: http://www.cdc.gov/nchs/about/major/hdasd/listpubs.htm .

History. To provide more complete and precise information on the utilization of the Nation's hospitals and on the nature and treatment of illness among the hospitalized population, in 1962 the NCHS began exploring possibilities for surveying morbidity in hospitals. A national advisory group was established. The NCHS conducted 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 the 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, the NCHS initiated the National Hospital Discharge Survey in 1964.

SURVEY METHODOLOGY

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 (hospitals 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 patient use. These criteria, used from 1988 through the current survey year, differ slightly from those used prior to 1988.

Beginning in 1988, the NHDS sampling frame consisted of hospitals that were listed in the April 1987 SMG Hospital Market Data File (2), met the above criteria, and began accepting patients by August 1987. The hospital sample was updated in 1991, 1994, 1997, and 2000 to allow for hospitals that opened later or changed their eligibility status since the previous sample update. In 2000, the sample consisted of 509 hospitals. Of the 509 hospitals, 28 were found to be out-of-scope (ineligible) because they went out of business or otherwise failed to meet the criteria for the NHDS universe. Of the 481 in-scope (eligible) hospitals, 434 hospitals responded to the survey.

Sample design and data collection. The NCHS has conducted the NHDS continuously since 1965. The original sample was selected in 1964 from a frame of short-stay hospitals listed in the National Master Facility Inventory (NMFI). That sample was updated periodically with samples of hospitals that opened later. Sample hospitals were selected with probabilities ranging from certainty for the largest hospitals to 1 in 40 for the smallest hospitals. Within each sampled hospital, a systematic random sample of discharges was selected. A report on the design and development of the original NHDS has been published (3).

In 1988, the NHDS was redesigned to provide geographic sampling comparability with other surveys conducted by the NCHS; to update the sample of hospitals selected into the survey; and to maximize

the use of data collected through automated systems. As did the original design, the redesigned NHDS sample included with certainty the largest hospitals. The remaining sample of hospitals was based on a stratified, three-stage design. The first stage consisted of selecting 112 primary sampling units (PSU's) that comprised a probability subsample of PSU's used in the 1985-94 National Health Interview Survey. The second stage consisted of selecting non-certainty hospitals from the sampled PSU's. At the third stage a sample of discharges was selected by a systematic random sampling technique.

These changes in the survey may affect trend data. That is, some of the differences between NHDS statistics based on the 1965-87 sample and statistics based on the sample drawn for the new design may be due to sampling error rather than actual changes in hospital utilization.

Two data collection procedures were used for the survey. The first was a manual system of sample selection and data abstraction, used for approximately 61 percent of the responding hospitals.   The second was an automated method, used for approximately 39 percent of the responding hospitals. The automated method involved the purchase of computerized data files from abstracting service organizations, state data systems, or from the hospitals themselves.

In the manual system, the sample selection and the transcription of information from the hospital records to abstract forms were performed at the hospitals. Of the hospitals using this system in 2000, about 32 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.

For the automated system, NCHS purchased files containing machine-readable medical record data from which records were systematically sampled by NCHS.

The Medical Abstract Form (Appendix E) and the automated data contain items relating to the personal characteristics of the patient, including birth date or age, sex, race, and marital status, but not name and address; administrative information, including admission and discharge dates, discharge status, and medical record number; and medical information, including diagnoses and surgical and nonsurgical procedures. Since 1977, patient zip code, expected source of payment, and dates of surgery have also been collected. (The medical record number, date of birth, and patient zip code are confidential information and are not available to the public).

Medical Coding and Edits. The medical information that was recorded manually on the sample patient abstracts was coded centrally by NCHS staff. A maximum of seven diagnostic codes was assigned for each sample abstract. In addition, if the medical information included surgical or nonsurgical procedures, a maximum of four codes for these procedures was assigned. The system currently used for coding the diagnoses and procedures on the medical abstract forms as well as on the commercial abstracting services data files is the *International Classification of Diseases, 9th Revision, Clinical Modification,* or ICD-9-CM (4).

NHDS usually presents diagnoses and procedures in the order they are listed on the abstract form or obtained from abstract services; however, there are exceptions. For women discharged after a delivery, a code of V27 from the supplemental classification is entered as the first-listed code, with a code designating either normal or abnormal delivery in the second-listed position. In another exception, a decision was made to reorder some acute myocardial infarction diagnoses. If an acute myocardial infarction is listed with other circulatory diagnoses and is other than the first entry, it is reordered to first position. If a symptom appears as a first-listed code and a diagnosis appears as a secondary code, the diagnosis replaces the symptom which is moved back.

Following conversion of the data on the medical abstract 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 decision.

2000 NHDS DATA FILE DOCUMENTATION

A new edit program was developed for the NHDS and was implemented beginning in the 1996 data year. The updated edit program, while following the same general specifications as the previous edit program, was designed to make as few changes as possible in the data. Thus, there may be some minor anomalies in certain areas which would be apparent when examining data over time, performing trend analyses, or examining combinations of variables. Particular features of the new edit program which 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 which do not provide the exact date of admission and discharge.

Users of the National Hospital Discharge Survey (NHDS) diagnostic and/or procedure data, which is coded to the ICD-9-CM, must take into account annual ICD-9-CM addendum. The addendum lists new codes, new fourth or fifth digits to existing codes, as well as other modifications. Changes go into effect October 1 of the calendar year. Addendum changes for 1986 through 1998 are listed in Appendix B. Because of unforeseen or potential problems associated with the millennium change, no addendum was released in 1999. Therefore, coding of the 2000 data is consistent with the ICD-9-CM and the addendum which became effective October 1, 1998. For more information about the ICD-9-CM visit: http://www.cdc.gov/nchs/icd9.htm .

The Uniform Hospital Discharge Data Set (UHDDS). Starting with 1979 data, the 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 minimum data set of items uniformly defined (4). These items 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.

Population Estimates. Appendix C provides estimates of the civilian population of the United Stated as of July 1, 2000, which have been provided by the U.S. Bureau of the Census.   These estimates have been adjusted for net underenumeration using the 1990 National Population Adjustment Matrix. NOTE THAT PRIOR TO THE 1997 DATA YEAR, CENSUS ESTIMATES OF THE CIVILIAN POPULATION PROVIDED WITH THE NHDS PUBLIC USE DATA FILE DOCUMENTATION WERE NOT ADJUSTED FOR THE UNDERCOUNT.

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.

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. A very small proportion, (less than one-half of one percent) of the discharge records failed to include the sex, age, or date of birth of the patient. If the hospital record did not state either the age or sex of patient, it was imputed by

assigning an age or sex value according to the specifications described earlier. In a very few cases (about a quarter of a percent of the records), the age or sex was edited, because it was inconsistent with the diagnosis. In 2000, data for RACE were missing for 28.8 percent of the discharges, and no attempt was made to impute for these missing values.

Other edit and imputation procedures may have been applied to data in the NHDS collected 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 of the estimate is obtained by dividing the standard error by the estimate itself. The resulting value is multiplied by 100, so the relative standard error is expressed as a percent of the estimate. Estimates of sampling variability were 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 the approach it uses was published by Bieler and Williams (6).

Relative Standard Errors for Aggregate Estimates

Parameters for calculating approximate relative standard errors for aggregate estimates are presented in Table 1. To derive error estimates that would be applicable to a wide variety of statistics, numerous estimates and their variances were produced. A regression model was then used to produce best-fit curves, based on the empirically determined relationship between the size of an estimate X and its relative variance. The relative standard error of an estimate X [RSE(X)] is the square root of the relative variance and may be calculated from the formula:

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

with a and b provided in Table 1.   When multiplied by 100, the RSE(X) is expressed as a percent of X.

For example, in 2000 the estimated number of discharges from short-stay hospitals for children under age 15 with a first-listed diagnosis of asthma (ICD-9-CM code 493) was 203,000. Using the applicable constants from Table 1 for estimates by age produces:

$$RSE(203,000) = SQRT(.021332 + (284.11415/203,000))$$

$$RSE(203,000) = .151$$

When multiplied by 100, the relative standard error for the estimate of interest becomes 15.1 percent. The standard error of the estimate is obtained by multiplying the relative standard error by the estimate itself:

$$SE(203,000) = 203,000 * .151 = 30,653$$

The standard error can be used to generate confidence intervals for statistical testing. In this example, the 95% confidence interval for the estimate of children under age 15 with a first-listed diagnosis of asthma is:

$$(203,000 - 2*30,653) <-> (203,000 + 2*30,653)$$

$$141,694 <-> 264,306$$

Relative Standard Error for Estimates of Percents

Approximate relative standard errors for estimates of percents may also be calculated from Table 1.

2000 NHDS DATA FILE DOCUMENTATION

The relative standard error for a percent, 100p (0<p<1), may be calculated using the formula:

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

where 100p 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). The values for b are given in Table 1. When multiplied by 100, the RSE(p) is expressed as a percent of the estimate, p.

For example, in 2000 the estimated number of discharges from short-stay hospitals who were women was 19,192,000. This is 60.5 percent of the estimated 31,706,000 total discharges for that year. Using the applicable constants from Table 1 for estimates by sex produces:

$$RSE(.605) = SQRT(300.21724 * (1 - .605)/(.605 * 31,706,000))$$

$$RSE(.605) = .00249$$

When multiplied by 100, the relative standard error for the estimate of interest becomes .249 percent. The standard error is obtained by multiplying the relative standard error by the estimate itself:

$$SE(.605) = .605 * .00249 = .0015$$

The standard error can be used to calculate confidence intervals for statistical testing. In this example, the 95% confidence interval for the estimate of the percentage of female inpatients is:

$$(.605 - 2*.0015) <-> (.605 + 2*.0015)$$

$$.602 <-> .608$$

or, equivalently, **60.2% <-> 60.8%**

Relative Standard Error for Ratio Estimators

The approximate RSE of a ratio (X/Y) in which the numerator (X) and the denominator (Y) are both estimated from the same survey, but the numerator is not a subclass of the denominator, is calculated using the formula:

$$RSE(X/Y) = SQRT(RSE^2(X) + RSE^2(Y))$$

The approximation is valid if the RSE of the denominator is less than 5 percent or the RSE's of the numerator and denominator are both less than 10 percent. When multiplied by 100, the RSE(X/Y) is expressed as a percent of the ratio estimate, X/Y.

For example, average length of stay (ALOS) is considered a ratio estimator since it is the ratio of days of care to the number of discharges. In 2000, the estimated number of days of care for inpatients with a first-listed diagnosis of septicemia (ICD-9-CM code 038) was 2,597,000. The estimated number of discharges for inpatients with a first-listed diagnosis of septicemia was 326,000. The ALOS for inpatients with a first-listed diagnosis of septicemia was 2,597,000/326,000 = 8.0.

To compute the RSE for ALOS, first compute the RSE for the estimated number of days of care and the RSE for the estimated number of discharges. See the section above on **Relative Standard Errors for Aggregate Estimates** for computation of these RSE's.

$$RSE(2,597,000) = .0618$$
$$RSE(326,000) = .0568$$

Next, substitute those RSE's into the formula above to approximate the RSE for the ALOS estimate:

$$RSE(8.0) = SQRT((.0618)^2 + (.0568)^2)$$

$$RSE(8.0) = .0839$$

The standard error of the estimate is obtained by multiplying the relative standard error by the estimate itself:

$$SE(8.0) = .0839 * 8.0 = .671$$

The standard error can be used to generate confidence intervals for statistical testing. In this example, the 95% confidence interval for the estimate of the ALOS for inpatients diagnosed with septicemia is:

$$(8.0 - 2*.671) <-> (8.0 + 2*.671)$$

$$6.7 <-> 9.3$$

2000 NHDS DATA FILE DOCUMENTATION

Table 1. Parameter values for relative standard errors for National Hospital Discharge Survey aggregate statistics by statistic type: United States, 2000

Characteristic	FIRST-LISTED DIAGNOSIS		DAYS OF CARE		ALL-LISTED DIAGNOSES		ALL-LISTED PROCEDURES	
	a	B	a	b	a	b	a	b
TOTAL	0.002287	307.96962	0.003503	824.08156	0.002846	279.63119	0.003840	309.28733
SEX								
Male	0.002492	330.84852	0.004031	1524.79149	0.004294	453.83201	0.004677	284.75142
Female	0.002296	300.21724	0.003590	758.48227	0.002176	277.66480	0.002855	296.90783
AGE GROUP								
Under 15 years	0.021332	284.11415	0.035212	520.36022	0.023027	249.07632	0.031270	242.65411
15-44 years	0.002306	309.63184	0.003788	846.80014	0.002161	303.96264	0.003001	307.89369
45-64 years	0.002495	294.03933	0.003990	1245.18684	0.003046	312.39495	0.004791	287.43793
65 years and over	0.003283	334.40841	0.003590	2018.69120	0.003000	362.13321	0.004734	311.25612
REGION								
Northeast	0.013516	176.21901	0.017328	581.83012	0.018498	189.26179	0.017236	293.50715
Midwest	0.009542	213.25373	0.012172	428.23450	0.010369	231.86476	0.013317	120.09107
South	0.003554	363.25616	0.006166	1223.94608	0.004845	352.50429	0.005879	261.57132
West	0.005469	351.03280	0.010168	1046.78781	0.005785	346.80421	0.008474	316.77742
RACE								
White	0.004180	346.49528	0.006018	919.13719	0.007361	352.42351	0.004811	317.21730
Black	0.005455	240.64604	0.008475	859.96136	0.005781	209.24930	0.010493	177.67685
All other	0.021333	222.49920	0.038480	476.52066	0.028589	285.70901	0.028614	267.60494
Race not stated	0.018744	278.76061	0.019217	837.45019	0.021136	256.13906	0.025611	174.22370
EXPECTED SOURCE OF PAYMENT								
Medicare	0.003730	347.59293	0.004077	1885.28816	0.003649	315.73286	0.004973	325.03095
Medicaid	0.006329	323.72265	0.009505	1113.96754	0.005271	363.82164	0.007488	305.73437
Worker's compensation & Other govt payments	0.010432	337.34361	0.022387	1100.21733	0.017122	259.22962	0.015112	274.59336
HMO/PPO	0.004596	267.09076	0.008327	467.29188	0.004956	235.70957	0.005690	231.02811
BC/BS & Other private insurance	0.005157	295.04612	0.007509	828.24439	0.005760	296.04960	0.009506	334.25082
Self pay	0.004822	314.42800	0.007746	1005.64303	0.004554	312.44459	0.006655	260.43225
No charge and other	0.038141	171.15139	0.051189	506.79249	0.037292	200.24927	0.043448	144.74118

Users of NHDS data are cautioned that computed estimates based on fewer than 30 unweighted records are not reliable and should not be reported. Because these estimates are based on so few data points, they are excluded from the calculation of the generalized variance curves. Thus, application of generalized variance curves is appropriate only for estimates based on at least 30 records.

2000 NHDS DATA FILE DOCUMENTATION

Presentation of Estimates. Publication of estimates for the NHDS is based on the relative standard error of the estimate and the number of sample records on which the estimate is based (referred to as the sample size). Estimates are not presented in NCHS reports unless a reasonable assumption regarding the probability distribution of the sampling error is possible.

Based on consideration of the complex sample design of the NHDS, the following guidelines are used for presenting the 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 is 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 present 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.

Monthly and Seasonal Estimates Under the New Design. An important difference between the old and new designs is the method used to adjust for nonresponse. 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 partial respondent. A nonresponding hospital was one which failed to provide at least half of the expected number of discharges for at least half of the months for which it was in-scope. 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 a second 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 which 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. As a result monthly and seasonal estimates may be skewed. While the effect is believed to be small, it is recommended that partial year estimates NOT be produced. In the 2000 NHDS, 87 percent of the 434 responding hospitals provided data for all twelve months, and 97 percent provided at least nine months of data.

How to Use the Data File. The NHDS records are weighted to allow inflation to national or regional estimates. The weight applied to 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 are summed. Average length of stay data can be obtained by dividing the days of care by the number of discharges as calculated above.

Appendix D contains unweighted and weighted frequencies for selected variables. These may be used as a cross-check when processing NHDS data.

2000 NHDS DATA FILE DOCUMENTATION

Diagnosis-Related Groups (DRGs). Many users of the NHDS data have expressed an interest in converting the medical data to DRGs. This has been done using DRG Grouper Programs obtained from the Centers for Medicare and Medicaid Services (formerly HCFA). The DRGs and the DRG Grouper Programs were developed outside of the National Center for Health Statistics; any questions about DRGs, other than specific questions about how they relate to NHDS data, should be addressed elsewhere.

Questions. Questions concerning NHDS data should be directed to:

<div align="center">

Centers for Disease Control and Prevention
National Center for Health Statistics
Division of Health Care Statistics
Hospital Care Statistics Branch
6525 Belcrest Road, Room 956
Hyattsville, Maryland 20782
Phone: 301.458.4321
Fax: 301.458.4032
email: NHDS@cdc.gov

</div>

For more information about the NHDS, visit our website:
http://www.cdc.gov/nchs/about/major/hdasd/nhds.htm

For email discussions and dissemination of NHDS data, join the Hospital Discharge and Ambulatory Surgery Data listserv (HDAS-DATA). In the body of an email message (leaving the subject line blank), type:

subscribe hdas-data Your Name

Send this message to:
listserv@cdc.gov

2000 NHDS DATA FILE DOCUMENTATION

REFERENCES

[1]Dennison C, Pokras R. Plan and Operation of the National Hospital Discharge Survey. National Center for Health Statistics. Vital Health Stat 1 (39). 2000. http://www.cdc.gov/nchs/data/sr1_39.pdf

[2]SMG Marketing Group, Inc. Hospital Market Database. Healthcare Information Specialists, 1342 North LaSalle Drive, Chicago, IL. 1987, April 1991, April 1994, April 1997, April 2000.

[3]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.

[4]International Classification of Diseases, 9th Revision, Clinical Modification, 6th edition. U.S. Department of Health and Humans Services, National Center for Health Statistics, Health Care Financing Administration. 1998.

[5]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.

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