Number 265 • July 6, 1995

Advance Data

From Vital and Health Statistics of the CENTERS FOR DISEASE CONTROL AND PREVENTION/National Center for Health Statistics

Underreporting of Race in the National Hospital Discharge Survey

by Lola Jean Kozak, Ph.D., Division of Health Care Statistics

Introduction

Race data from the National Hospital Discharge Survey (NHDS) has become increasingly incomplete in recent years. This report examines factors related to the underreporting of race and explores the effects of the underreporting on NHDS estimates of hospital use by race. A major concern is whether discharges for each racial group are equally likely to be underestimated. If underreporting of race is a general phenomenon, then the NHDS estimates of hospital use for each racial group are too low. However, if discharges for some racial groups are more likely to be underestimated than others in the NHDS, using the data to make comparisons across racial groups could be misleading.

The NHDS has been conducted continuously by the National Center for Health Statistics (NCHS) since 1965. The data for the survey come from a sample of inpatient records that are obtained from a national sample of non-Federal general and short-stay specialty hospitals located in the United States. In 1990, data were abstracted from medical records of 266,000 discharges from 474 hospitals. In 1991, 484 hospitals provided data from 274,000 medical records, and 494 hospitals provided data from 274,000 medical records in 1992.

Beginning in 1988, a 3-stage stratified sample design was put into operation for the NHDS. For the first stage, primary sampling units (PSU's) were sampled; in the second, hospitals were sampled from the PSU's, and the third stage consisted of sampling discharges within the selected hospitals. In addition, hospitals with 1,000 beds or more or 40,000 discharges or more per year were selected with certainty.

Since 1985, two data collection procedures have been used for the NHDS. One is a manual system in which data are abstracted from the face sheet or discharge summary of the medical record for each sampled discharge at the hospital, either by hospital staff or personnel of the U.S. Bureau of the Census, on behalf of NCHS. The other, an automated method, involves the purchase of data tapes from abstracting service organizations, State data systems, or hospitals.

Further information about the survey design, data collection procedures, sampling errors, and definition of terms used in this report can be found in the section entitled "Technical notes."

Data from the NHDS have been used to examine racial differences in patterns of hospital use that may reflect differences in access to care or in the distribution of health problems. Recent studies that have used race data from the NHDS have investigated a variety of topics, including hysterectomy (1), HIV (2), stroke (3), children's asthma (4), preeclampsia and eclampsia (5), appendicitis and appendectomy (6), coronary arteriography and coronary bypass surgery (7), hip fractures (8), and idiopathic cardiomyopathy (9).

Acknowledgments

This report was prepared in the Division of Health Care Statistics. Linda Lawrence and Chuck Dennison of the Hospital Care Statistics Branch provided operational data and assisted with the description of the survey. Michael J. Monsour, formerly of the Technical Services Branch, produced the parameters for relative standard error equations. The report was edited by Klaudia M. Cox and typeset by Annette F. Facemire of the Publications Branch, Division of Data Services.

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

CENTERS FOR DISEASE CONTROL AND PREVENTION

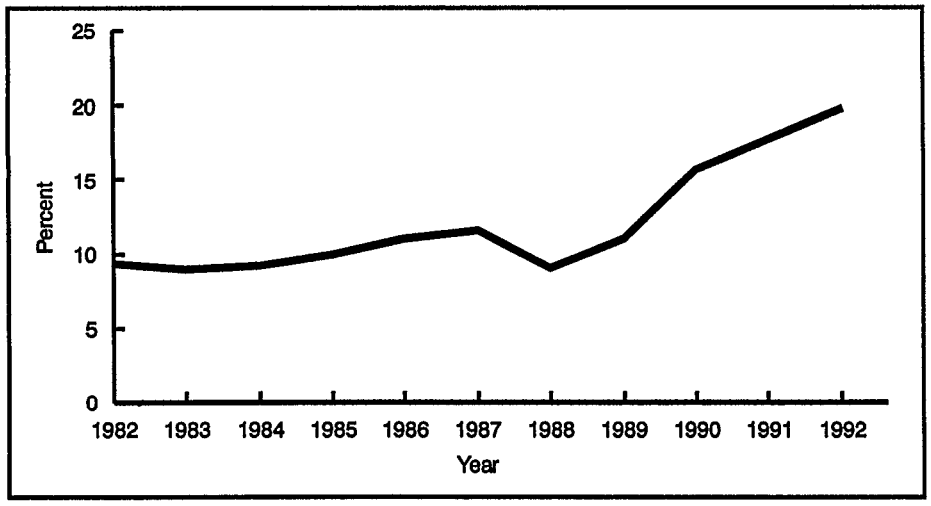

Figure 1. Percent of discharges with race not stated: United States, 1982–92

The NHDS race data have never been 100 percent complete. From 1982 through 1989, discharges with race not stated ranged from 8.9 to 11.5 percent of total discharges. Then, in 1990, the proportion of discharges with race not stated increased to 15.7 percent. In 1991, 17.8 percent of discharges were not identified by race, and in 1992, 19.8 percent of discharges were in the race-not-stated category (figure 1).

Two main factors were found to be related to the increasing underreporting of race in the NHDS. First, a growing number of the hospitals that participated in the NHDS in 1990–92 did not report race for any of their patients. In 1989, 17 hospitals did not report race, but in 1992 the number had increased to 63. Nonreporting hospitals made up only 12.8 percent of the hospitals participating in the survey in 1992, but accounted for 70.7 percent of the discharges with race not stated. Second, race was not reported for the majority of patients that were identified as Hispanic. In 1992, 16.0 percent of the discharges with race not stated were identified as Hispanic.

These two main problems will be discussed further in the following sections of this report. Approaches to estimating the racial distribution of patients not reported by race will also be examined. To further evaluate the effects of the underreporting of race, the NHDS race data will be compared with data from other sources, including hospital use data from National Health Interview Survey, Medicare data from the Health Care Financing Administration, and natality data from the Division of Vital Statistics. Finally, adjustment of NHDS race data for underreporting will be discussed.

Highlights

- The number of hospitals in the NHDS that reported race for less than 3 percent of discharges increased from 17 in 1989 to 63 in 1992.
- Most of the hospitals that did not report race in the 1990–92 period used the automated data collection method.
- Based on data from previous years and county populations, hospitals that did not report race in 1990–92 were likely to have a higher proportion of white discharges than hospitals that reported race.
- A specific race was reported for only 25–35 percent of Hispanic patients in the 1990–92 period; most Hispanic patients were probably white.
- In comparison with the number of discharges estimated from the 1990–92 National Health Interview Survey, NHDS estimates of discharges were significantly lower for white patients, but not significantly different for black patients.
- Proportional adjustment of NHDS race data may produce more accurate estimates of white discharges, but it does not improve comparisons between racial groups.

Hospital reporting patterns

The majority of hospitals that participate in the National Hospital Discharge Survey (NHDS) report race for all or almost all discharges. In 1992, for example, 296 hospitals (59.9 percent) reported race on 97–100 percent of their sample records (table 1). The discharges not identified by race come primarily from a small group of hospitals that do not report race for any or almost any of their discharges. In 1992, 63 hospitals (12.8 percent) reported race on less than 3 percent of sampled records. These 63 hospitals accounted for 83.4 percent of the sampled records with race not stated in the 1992 NHDS, and 70.7 percent of the estimated number of discharges with race not stated.

Nonreporting hospitals, which are defined in this report as those providing race data for less than 3 percent of discharges, have increased in number in recent years (figure 2). In 1989 there were only 17 nonreporting hospitals, but the number jumped to 45 in 1990, increased to 50 in 1991, and to 63 in 1992. Nonreporting hospitals accounted for 50 percent of the sampled records with race not stated in 1989, but for 81–84 percent in 1990–92.

Most of the hospitals that did not report race in the 1990–92 period used the automated data collection method. As described earlier, the automated method refers to the purchase of medical record data in electronic form from abstracting service organizations, State data systems, or hospitals. In contrast, data collected via the manual method were transcribed from the medical record to abstract forms specifically for the NHDS.

As shown in figure 2, only 5 hospitals that did not report race used the automated data collection method in 1989, but in 1990, 31 nonreporting hospitals used the automated method. In 1991, 35 nonreporting hospitals used the automated data collection method, as did 48 nonreporting hospitals in 1992. The number of nonreporting hospitals using the manual method only increased from 12 in 1989 to 15 in 1992.

A concerted effort was made to improve the NHDS response rate in the 1990's, which resulted in a substantial

Table 1. Number and percent distribution of hospitals by proportion of sample records with race reported: United States, 1992

[Short-stay non-Federal hospitals that participated in the National Hospital Discharge Survey. Excludes newborn infants]

Proportion of records with race reported	Hospitals	
	Number	Percent
All records	494	100.0
0.0–2.9 percent	63	12.8
3.0–49.9 percent	15	3.0
50.0–74.9 percent	21	4.3
75.0–89.9 percent	41	8.3
90.0–96.9 percent	58	11.7
97.0–100.0 percent	296	59.9

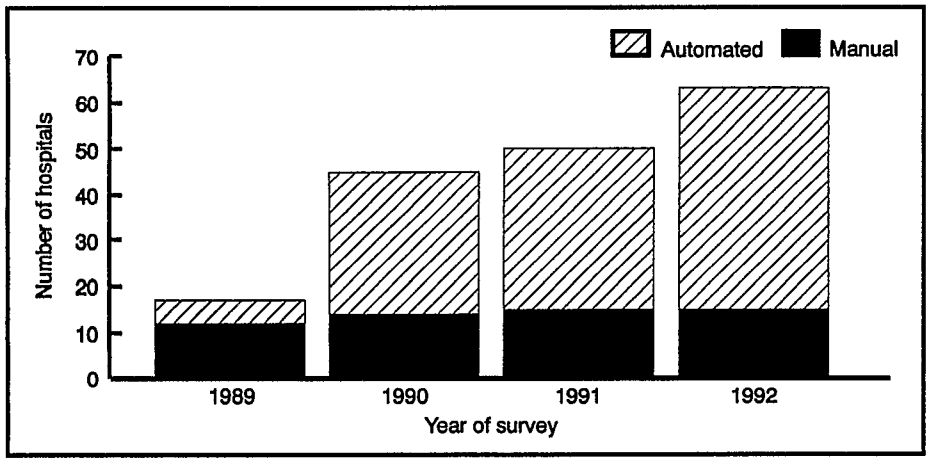

Figure 2. Number of hospitals not reporting race that used the automated and manual methods of data collection: 1989–92

Table 2. Percent distribution of discharges by race, for 26 nonreporting hospitals based on previous data and for reporting hospitals: United States, 1992

[Discharges of inpatients from non-Federal short-stay hospitals. Excludes newborn infants]

Race	26 nonreporting hospitals[1]	Reporting hospitals[2]
	Percent distribution	
All races	100.0	100.0
White	87.3	75.2
Black	8.3	13.9
All other races	2.1	4.2
Race not stated	2.3	6.7

[1]Nonreporting hospitals reported race for less than 3 percent of discharges. Racial distribution was based on the most recent data reported by each hospital.
[2]Hospitals that reported race for 3–100 percent of discharges.

increase in the number of hospitals participating in the survey, from 408 in 1989 to 474 in 1990, 484 in 1991, and 494 in 1992. Many hospitals that had not participated in previous years agreed to participate in the NHDS through the automated systems. Much of the data obtained using the automated method were from State data systems or other systems that used the National Uniform Bill, or UB-82 (and its successor, UB-92) for data collection. The National Uniform Bill was established for processing hospital bills and does not include race as a required item. Thus, the race of discharges was often not included in the data from such systems purchased for the NHDS.

Race for nonreporting hospitals

An important question to investigate is whether nonreporting hospitals have the same racial distribution of discharges as reporting hospitals. Two approaches were taken to answer this question. First, data from previous years were reviewed. The NHDS was redesigned in 1988, and with some exceptions (some hospitals went out of business, a few new hospitals were added to the sample in 1991), the same hospitals were in the sample from 1988 through 1992. Among the 63 hospitals that did not report race in 1992 were 26 that had reported race in at least 1 year from 1988–91. For each of these 26 hospitals, the percent distribution of the weighted number of discharges by race was obtained for the most recent year in which race was reported. This percent distribution was multiplied by the weighted number of discharges from the hospital in 1992.

The resulting distribution of discharges by race for the 26 hospitals combined is shown in table 2. The estimated proportion of patients who were white was significantly larger in the 26 nonreporting hospitals (87.3 percent) than in hospitals that reported race in 1992 (75.2 percent). The reporting hospitals had higher proportions of patients who were black, all other races, and race not stated.

Because the 26 hospitals that previously reported race may or may not be representative of all the hospitals that did not report race, a second approach to estimate the racial distribution of discharges from nonreporting hospitals was used. This involved an examination of the populations they served. The 1990 Census provided data on the racial distributions of counties (10) that can be used to approximate service areas of hospitals despite the fact that some hospitals may serve only part of a county or multiple counties.

A preliminary test was conducted to investigate whether it would be reasonable to use county population distributions by race as proxy measures for the distribution of discharges by race. A 20-percent stratified random sample was taken of the hospitals that

reported race for 97–100 percent of their discharges in 1992. The county in which each sampled hospital was located was identified and the percent distribution of the population of the county by race was obtained. This percent distribution was multiplied by the weighted number of discharges from the hospital in the county.

The resulting distribution of discharges by race for the sample hospitals was 76.8 percent white patients, 17.5 percent black patients, and 5.6 percent patients of other races. In comparison, the distribution reported for these sample hospitals was 80.0 percent white discharges, 15.2 percent black discharges, and 4.1 percent discharges of other races.

Although not exact, the population distribution of a county appeared useful as a general indicator of the racial distribution of discharges from a hospital in the county. Therefore, the procedure used in the test was applied to the nonreporting hospitals. The county in which each nonreporting hospital was located was identified and the percent distribution of the population of the county by race was obtained. This percent distribution was multiplied by the weighted number of discharges from the nonreporting hospital in the county.

The resulting distributions of discharges by race for nonreporting hospitals are shown in table 3. For each year 1990 through 1992, the proportion of discharges that were white calculated for nonreporting hospitals was significantly higher than the proportion in reporting hospitals. Conversely, the proportions of discharges that were black calculated for nonreporting hospitals were significantly lower than the proportions in reporting hospitals. The "all other races" category accounted for larger proportions of discharges for the nonreporting than for the reporting hospitals, but these proportions may have been somewhat overestimated, as in the preliminary test.

These findings are not definitive, but along with the data on racial distributions in previous years, they suggest that nonreporting hospitals may have a higher proportion of white discharges and a lower proportion of black discharges than reporting hospitals.

Hispanic patients

Race and ethnicity are separate data items for the NHDS. On the ethnicity variable, patients are identified as being of Hispanic origin, non-Hispanic origin, or not stated. Ethnicity, in general, is not well reported. For example, in 1992, only 24.6 percent of all NHDS discharges were identified as Hispanic or non-Hispanic. Because ethnicity data are not reliable, these data are not released from the NHDS.

Data on Hispanic origin are discussed here because patients identified as Hispanic usually have missing race data. As shown in table 4, more than half of Hispanic patients were in the race-not-stated category in 1990–92. Another 13–17 percent were reported in the "other" race category, so only 25–35 percent were identified as a specific race during this 3-year period.

The lack of race data for Hispanic patients is separate from the problem of hospitals not reporting race. Hospitals that do not report race almost never report ethnicity. However, certain other hospitals report all or almost all patients identified as Hispanic in the race not stated category. In 1992, for example, 112 hospitals reported race for less than 3 percent of their Hispanic discharges, and these hospitals accounted for two-thirds of the Hispanic discharges not identified by race.

These 112 hospitals were more likely to provide data through the manual data collection system than the automated system. In the manual system, NCHS staff are instructed to code records as "race not stated" when Hispanic is written in as a race. Automated system data are not coded by NCHS staff and do not necessarily follow this practice. In 1992, a total of 32 NHDS hospitals were found in which all Hispanic discharges were assigned to the "other" race category. These were predominantly hospitals using the automated data collection system.

Some of the hospitals that have all their Hispanic discharges assigned to the "race not stated" or "other" race categories are known to be using data collection forms that do not separate race and ethnicity. Other hospitals probably also use combined categories. The Federal standards for reporting race and ethnic statistics (11) allow race and ethnicity to be collected in a combined format in which Hispanics are not identified by race.

If identified by race, the evidence indicates that most Hispanics in the "race not stated" and "other" categories would be classified as white. Among

Table 3. Percent distribution of discharges by race for nonreporting hospitals based on county population and for reporting hospitals, according to year: United States, 1990–92

[Discharges of inpatients from non-Federal short-stay hospitals. Excludes newborn infants]

Race	1990		1991		1992	
	Nonreporting hospitals[1]	Reporting hospitals[2]	Nonreporting hospitals[1]	Reporting hospitals[2]	Nonreporting hospitals[1]	Reporting hospitals[2]
	Percent distribution					
All races	100.0	100.0	100.0	100.0	100.0	100.0
White	83.0	77.7	83.8	76.4	83.2	75.2
Black	11.4	13.1	10.5	13.6	10.6	13.9
All other races	5.6	3.5	5.7	3.8	6.2	4.2
Race not stated	–	5.7	–	6.2	–	6.7

[1]Nonreporting hospitals reported race for less than 3 percent of discharges. Race distribution was based on the population of the county where the hospital was located.
[2]Hospitals that reported race for 3–100 percent of discharges.

Table 4. Percent distribution of discharges by race for patients identified as Hispanic, according to year: United States, 1990–92

[Discharges of inpatients from non-Federal short-stay hospitals. Excludes newborn infants]

Race	1990	1991	1992
	Percent distribution		
All races	100.0	100.0	100.0
White	33.3	28.0	23.1
Black	0.8	*0.6	0.8
American Indian/Eskimo/Aleut or Asian/Pacific Islander	0.9	3.5	1.7
Other	13.3	14.4	17.2
Race not stated	51.8	53.4	57.3

Hispanic discharges reported as a specific race in 1992, 90.4 percent were identified as white, 3.1 percent as black, and 6.5 percent as American Indian/Eskimo/Aleut or Asian/Pacific Islander. In addition, the U.S. Bureau of the Census estimated that in 1990 the Hispanic population was 91.3 percent white, 5.4 percent black, and 3.3 percent American Indian/Eskimo/Aleut or Asian/Pacific Islander (12). Thus, the lack of information on the race of the majority of discharges identified as Hispanic is likely to affect NHDS estimates of white discharges disproportionately.

Comparisons

National Health Interview Survey

If discharges of white patients are underestimated to a greater extent than discharges of patients of other races in the NHDS, this should be evident in comparisons of estimates from the NHDS to data from other sources. A comparison was made of NHDS data with data from the National Health Interview Survey (NHIS), which also produces estimates of hospital use by race (13–15).

The NHIS is based on a different universe and uses different definitions and data collection procedures than the NHDS. The estimates of hospitalizations are obtained from interview questions about overnight stays in short-stay hospitals during the previous 6 months. Hospitalizations of persons who died or were institutionalized during the reference period are excluded, as are hospitalizations of healthy newborn infants.

Before comparing NHIS estimates with NHDS estimates of hospital use, the NHDS estimates were adjusted. Patients hospitalized for less than 1 day were excluded because the NHIS data were only for overnight stays. Persons discharged dead and those transferred to long-term care institutions were excluded, although the NHDS data would probably still include some persons who died or were institutionalized during a 6-month period. All newborn infants were excluded, as is usual for NHDS estimates, although some sick newborn infants may be included in the NHIS estimates. The adjustments do not make the NHDS and NHIS data completely alike, but they should be more comparable.

The adjusted NHDS estimates and the NHIS estimates of discharges from short-stay hospitals are shown in table 5 for 1990 through 1992. In each of the 3 years, the NHIS estimate of total discharges was not significantly different from the adjusted NHDS estimate. However, the estimated number of discharges for white patients from the NHIS was significantly higher than the adjusted NHDS estimate each year. The NHIS estimate was 22 percent higher in 1990, 26 percent higher in 1991, and 30 percent higher in 1992. The estimated number of discharges for black patients from the NHIS was not significantly different from the adjusted NHDS estimate of black discharges in any of the years. NHIS estimates of discharges for other racial groups were not available.

Medicare

Another source of information on the race of hospital discharges is the Health Care Financing Administration (HCFA), which obtains data on the hospitalizations of Medicare beneficiaries. In 1990, HCFA reported 10,522,000 discharges from short-stay hospitals for Medicare beneficiaries (16). Of these, 9,037,000 (85.9 percent) were identified as white, 1,185,000 (11.3 percent) were other than white, and 300,000 (2.9 percent) were not identified by race.

The NHDS estimate of discharges with Medicare as the principal expected

Table 5. Number of discharges estimated from the National Hospital Discharge Survey and the National Health Interview Survey, by year and race: United States, 1990–92

[Discharges of inpatients from short-stay hospitals]

Year and race	National Hospital Discharge Survey[1]	National Health Interview Survey
1990	Number in thousands	
All races[2]	27,250	27,058
White	18,713	22,821
Black	3,300	3,692
1991		
All races[2]	27,275	26,873
White	18,084	22,778
Black	3,395	3,420
1992		
All races[2]	27,289	27,039
White	17,429	22,607
Black	3,363	3,654

[1]Discharges from non-Federal hospitals. Excludes newborn infants, discharges to long-term care institutions, patients discharged dead, and discharges with inpatient stays of less than 1 day.
[2]Includes patients of all races and patients whose race was not stated.

Table 6. Number of discharges for women with deliveries and number of live births, by year and race: United States, 1990–92

Year and race	Discharges for women with deliveries [1]	Live births [2]
	Number in thousands	
1990		
All races	4,025	4,158
White	2,431	3,290
Black	584	684
All other races	262	184
Race not stated	748	–
1991		
All races	3,973	4,111
White	2,244	3,241
Black	557	683
All other races	289	187
Race not stated	883	–
1992		
All races	3,910	4,065
White	2,148	3,202
Black	511	674
All other races	334	190
Race not stated	916	–

[1] Discharges of inpatients from non-Federal short-stay hospitals estimated from the National Hospital Discharge Survey.
[2] Data from birth certificates. Race of mother assigned to child.

source of payment was 10,625,000 discharges in 1990. Of these, 8,135,000 (76.6 percent) were identified as white, and 1,037,000 (9.8 percent) were black and other races. The remaining 1,452,000 (13.7 percent) were in the race-not-stated category.

The estimates of total Medicare discharges from these two sources were not significantly different. Likewise, the estimate of discharges for Medicare patients of black and other races from HCFA was not significantly different from the NHDS estimate for this group. However, the HCFA estimate of white Medicare discharges was significantly higher than the NHDS estimate. Thus, these findings also suggest that white patients are more likely to be underestimated in the NHDS than patients of other races.

Birth certificates

Information about race from birth certificates was also compared to NHDS data on the race of women hospitalized for deliveries. Beginning with 1989, the data from birth certificates have been tabulated by the race of the mother. The number of births would be expected to be somewhat higher than the number of women with deliveries estimated from the NHDS because not all births take place in non-Federal short-stay hospitals, and because one delivery can result in multiple births. However, the differences in total number of deliveries and births were not significantly different in the 1990–92 period.

The racial distribution of live births reported from birth certificates (17) is compared to the racial distribution of women with deliveries from the NHDS in table 6. The number of live births identified as white was 35 percent higher than the number of white women with deliveries in 1990, 44 percent higher in 1991, and 49 percent higher in 1992. In 1990, the number of black live births was 17 percent higher than the NHDS estimate of black women with deliveries; it was 22 percent higher in 1991, and 32 percent higher in 1992.

The number of live births that were other races was not significantly different from the estimate of women with deliveries of other races in 1990. However, in 1991 and 1992, there were more women with deliveries in the "other races" category than live births. This was due to a large number of women with deliveries in the NHDS who were reported as an unspecific other race. These data suggest problems with NHDS estimates for all the racial categories of women with deliveries, but also support the thesis that white patients are markedly underestimated in the NHDS.

Adjustments for underreporting

Table 7 shows the number and rate of discharges by race as estimated from the NHDS in 1990–92. These estimates are compared with estimates produced by two types of adjustments. The first is proportional adjustment, a strategy used by researchers to compensate for the underreporting of race in the NHDS (1, 4, 8). In this approach, the discharges in the race-not-stated category are assigned to specific race categories based on the distribution of the discharges whose race is known.

For example, in the 1992 NHDS, 24,838,000 of the estimated 30,951,000 discharges were identified by race. Among the discharges identified by race, 80.6 percent were white, 14.9 percent were black, and 4.5 percent were other races. Distributing the 6,113,000 discharges in the race-not-stated category in the same proportions, 4,927,000 were added to the white category, 909,000 to the black category, and 278,000 to the other races category.

Proportional adjustment would be appropriate if the evidence suggested that patients of all races were equally underreported. Because white patients appear to be underreported to a greater extent than patients of other races in the NHDS, proportional adjustment would not be expected to correct accurately for nonresponse. Using this adjustment, white discharges would still be underestimated, and discharges of patients in the other race categories would be overestimated. The relative differences in discharge rates between racial groups are not affected by proportional adjustment. Thus, comparisons of proportionally adjusted rates across racial categories are no different than comparisons of unadjusted rates.

Numbers and rates of discharges resulting from a population-based adjustment are also shown in table 7. The population-based adjustment used the populations of the counties in which nonreporting hospitals were located to estimate the racial distribution of

Table 7. Number and rate of discharges, by year, race, and type of estimate: United States, 1990–92

[Discharges of inpatients from non-Federal short-stay hospitals. Excludes newborn infants]

Year and race	Discharges in thousands			Discharge rate per 1,000 population		
	NHDS estimate[1]	Proportional adjustment[2]	Population-based adjustment[3]	NHDS estimate[1]	Proportional adjustment[2]	Population-based adjustment[3]
1990						
White	21,376	25,366	24,995	102.8	122.0	120.2
Black	3,611	4,285	4,038	119.2	141.4	133.3
All other races	958	1,137	974	99.8	118.5	101.5
Race not stated	4,843	–	781	...	...	...
1991						
White	20,816	25,317	25,078	99.3	120.8	119.6
Black	3,717	4,521	4,184	120.5	146.5	135.6
All other races	1,036	1,260	1,046	103.0	125.2	104.0
Race not stated	5,528	–	790	...	...	...
1992						
White	20,018	24,945	24,778	94.6	117.8	117.1
Black	3,692	4,601	4,219	117.9	146.9	134.7
All other races	1,128	1,405	1,142	107.5	134.0	108.9
Race not stated	6,113	–	811	...	...	...

[1]Unadjusted race data from National Hospital Discharge Survey (NHDS).
[2]NHDS race data with discharges in race not stated category distributed to race categories in proportions of discharges with known race.
[3]NHDS race data with discharges from nonreporting hospitals distributed to race categories based on county populations and Hispanic discharges not identified as specific race distributed to race categories in the proportions of the Hispanic population.

discharges in those hospitals. This procedure is described in the section, "Race for nonreporting hospitals." In addition, Hispanic discharges in the race-not-stated category and the "other race" category were assigned a race using the racial distribution of the Hispanic population as estimated by the U.S. Bureau of the Census.

The population-based adjustment assumes that patients are hospitalized in the same proportions that they are in the population, which is probably not the case (13–15). In addition, the population-based adjustment does not result in an assignment of race to all discharges, only to those from the nonreporting hospitals and Hispanic patients. In 1990–92, 781,000 to 811,000 discharges remained in the race-not-stated category after the population-based adjustment. Thus, the numbers and rates of discharges for specific race categories, while probably more accurate, remain underestimated.

Using either method, adjusted numbers and rates of discharges for white patients were significantly higher than unadjusted estimates. Although calculated differently, the numbers and rates of white discharges produced by proportional adjustment and population-based adjustment were similar. It should be noted, however, that both of these adjustments are expected to underestimate white discharges to some extent. The adjusted estimates of black discharges and discharges of all other races were not significantly different from the unadjusted rates.

Discussion

Race data from the NHDS became increasingly incomplete in recent years primarily because a growing number of hospitals that participated in the survey did not provide racial data on any of their patients. Most of these hospitals used the automated data collection method. They submitted tapes of data that were usually collected for other purposes to the NHDS. These data were often collected using the National Uniform Bill (UB-82 and UB-92), which does not include an item on race.

One solution to this problem would be for the NHDS to stop using automated data collection. However, even before 1985, when all data were collected manually using NHDS abstracts, the proportion of discharges with no race reported was a concern, ranging from 9–14 percent. In addition, the automated data collection system has become an integral part of the NHDS survey design. Approximately one third of the hospitals that participate in the survey now provide data through the automated method, and many are unwilling to participate in a manual system.

Another solution would be to add race to the UB-92 form. This could greatly increase the amount of race data reported through the automated method. However, the NHDS is only a secondary user of UB-92 data. The principal users, insurance companies and the Health Care Financing Administration, do not view a billing form as the best place to collect race data. They have enrollment forms that provide information on the race of beneficiaries (18).

The other main problem with the NHDS race data, lack of racial information for Hispanic patients, is also related to data collection forms. Hospitals and data systems that use a combined race/ethnicity item cannot supply the NHDS with information on the race of Hispanic patients. The Federal standards for reporting racial and ethnic statistics have been undergoing a wide-ranging review (19). It is uncertain whether one standard approach to reporting will be

established, or whether it will continue to be acceptable to report race and ethnicity either separately or in a combined format.

In 1990–92, the hospitals that did not report race to the NHDS apparently had a larger proportion of white patients than the reporting hospitals. Estimates based on racial distributions of discharges in previous years and on racial distributions of county populations both suggested that white patients made up a larger share of discharges in nonreporting hospitals than in reporting hospitals. The Hispanic patients not reported by race were also likely to be white in larger proportions than all patients, based on the distribution of those with a reported race and on the racial distribution of the Hispanic population. Therefore, discharges of white patients were probably underestimated to a greater extent than discharges of other racial groups.

Comparisons of NHDS data with data from other sources supported the hypothesis that white patients were disproportionately underestimated. The National Health Interview Survey estimated significantly larger numbers of white discharges than the NHDS, but similar numbers of black discharges. A larger number of Medicare discharges were white according to data from the Health Care Financing Administration than estimated by NHDS. However, the number of Medicare discharges of other races reported by these two sources were not significantly different. The number of live births that were white or black were larger than the NHDS estimates of white or black women hospitalized for deliveries, but differences were greater for the white category.

Because white patients are probably underreported to a greater extent than patients of other races, proportional adjustment of NHDS data would not be expected to produce completely accurate estimates of the number of discharges in each race group. This adjustment would probably produce a more accurate estimate of white discharges, but it would overestimate discharges of other races and would not affect comparisons between racial groups. The population-based adjustment may also produce a more accurate estimate of white discharges, but because it is based on the assumption that all racial groups have the same discharge rates, the estimates of racial groups with higher rates would be underestimated to some extent, and comparisons between racial groups could be distorted.

At present, no ideal solution exists to eliminate the problem of underreporting of race in the NHDS. Therefore, the NHDS race data need to be used cautiously and not overinterpreted. The data can still be useful for some types of analyses. General inferences can be drawn if the differences between racial groups are large. For example, the rate of HIV hospitalizations for black patients was so much larger than the rate for white patients that even if all the patients in the race-not-stated category were added to the white category, the difference would remain highly significant (2).

When white patients have a higher rate than other racial groups despite the underestimate, such as for coronary artery bypass grafts (20), it is reasonable to conclude that the rate for white patients is higher. Research can also be done on hospital use patterns within racial groups, such as investigation of major diagnostic categories for black patients or sex differences in discharge rates for white patients. In all these areas, though, it must be recognized that the numbers and rates produced from the NHDS for specific racial groups will be underestimated to an unknown extent.

References

1. Wilcox LS, Koonin LM, Pokras R, et al. Hysterectomy in the United States, 1988–1990. Obstet-Gynecol. 83(4): 549–55. 1994.
2. Kozak LJ, McCarthy E, Moien M. Patterns of hospital use by patients with diagnoses related to HIV infection. Public Health Rep 108(5): 571–81. 1993.
3. Modan B, Wagener DK. Some epidemiological aspects of stroke: Mortality/morbidity trends, age, sex, race, socioeconomic status. Stroke 23(9): 1230–36. 1992.
4. Gergen PJ, Weiss KB. Changing patterns of asthma hospitalization among children: 1979 to 1987. JAMA 264(13): 1688–92. 1990.
5. Saftlas AF, Olson DR, Franks AL, et al. Epidemiology of preeclampsia and eclampsia in the United States, 1979–1986. Am J Obstet Gynecol 163(2): 460–65. 1990.
6. Addiss DG, Shaffer N, Fowler BS, Tauxe RV. The epidemiology of appendicitis and appendectomy in the United States. Am J Epidemiol 132(5): 910–25. 1990.
7. Ford E, Cooper R, Castaner A, et al. Coronary arteriography and coronary bypass survey among whites and other racial groups relative to hospital-based incidence rates for coronary artery disease: Findings from NHDS. Am J Public Health 79(4): 437–40. 1989.
8. Rodriguez JG, Sattin RW, Waxweiler RJ. Incidence of hip fractures, United States, 1970–83. Am J Prev Med 5(3): 175–81. 1989.
9. Gillum RF. Idiopathic cardiomyopathy in the United States, 1970–1982. Am Heart J. 111(4): 752–55. 1986.
10. U.S. Bureau of the Census. General population characteristics. 1990 Census of Population. vol CP 1-2–CP 1-52. Washington. U.S. Department of Commerce. 1992.
11. Office of Management and Budget. Directive no. 15: Race and ethnic standards for Federal statistics and administrative reporting. Statistical Policy Handbook. U.S. Department of Commerce, Office of Federal Statistical Policy and Standards. 1978.
12. U.S. Bureau of the Census. U.S. population estimates, by age, sex, race, and Hispanic origin: 1980 to 1991. Current population reports. series P-25, no. 1095. Washington: U.S. Department of Commerce. 1993.
13. Adams PF, Benson V. Current estimates from the National Health Interview Survey, 1990. National Center for Health Statistics. Vital Health Stat 10(181). 1991.
14. Adams PF, Benson V. Current estimates from the National Health Interview Survey, 1991. National Center for Health Statistics. Vital Health Stat 10(184). 1992.
15. Benson V, Marano MA. Current estimates from the National Health Interview Survey, 1992. National Center for Health Statistics. Vital Health Stat 10(189). 1994.

16. Helbing C. Hospital insurance short-stay hospital benefits. Health Care Finan Rev. 1992 annual supplement. 55–96. 1993.
17. Ventura SJ, Martin JA, Taffel SM, et al. Advance report of final natality statistics, 1992. Monthly vital statistics report; vol 43 no. 5, suppl. Hyattsville, Maryland: National Center for Health Statistics. 1994.
18. Moore R. Report on HCFA data. Presented to the Subcommittee on Health Statistics for Minority and Other Special Populations, National Committee on Vital and Health Statistics. Sept 14. Washington, D.C. 1994.
19. Office of Management and Budget. Standards for classification of Federal data on race and ethnicity. No. 59. Federal Register. 29831. 1994.
20. Graves EJ. National Hospital Discharge Survey: annual summary, 1992. National Center for Health Statistics. Vital Health Stat 13(119). 1994.
21. SMG Marketing Group, Inc. Hospital Market Database. Chicago, Illinois: Healthcare Information Specialists. 1987.
22. SMG Marketing Group, Inc. Hospital Market Database. Chicago, Illinois: Healthcare Information Specialists. 1991.
23. 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.
24. Haupt BJ, Kozak LJ. Estimates from two survey designs: National Hospital Discharge Survey. National Center for Health Statistics. Vital Health Stat 13(111). 1992.
25. Shah BV. SESUDAAN: Standard errors program for computing of standardized rates from sample survey data. Research Triangle Park, North Carolina: Research Triangle Institute. 1981.

Technical notes

Survey methodology

Source of data

The National Hospital Discharge Survey covers discharges from noninstitutional hospitals, except Federal, military, and Department of Veterans Affairs 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 fewer 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 beds or more staffed for patient use.

From 1988 through 1990, the NHDS sampling frame consisted of hospitals that were listed in the April 1987 SMG Hospital Market Database (21), met the above criteria, and began accepting patients by August 1987. In 1991 the sampling frame was updated to include hospitals from the 1991 SMG Hospital Database (22). The sample consisted of 542 hospitals in 1990 and 528 hospitals in 1991 and 1992. In 1990, 23 of the sample hospitals 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. Seven hospitals were out of scope in 1991, and 14 were out of scope in 1992. In 1990, 474 of the 519 in-scope (eligible) hospitals responded to the survey. In 1991, 484 of 521 in-scope hospitals responded, and 494 of 514 in-scope hospitals responded in 1992.

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. That sample was updated periodically with samples of newly opened hospitals. Sample hospitals were selected with probabilities ranging from certainty for the largest hospitals to 1 in 40 for the smallest hospitals. Within each sample hospital, a systematic random sample of discharges was selected. A report on the design and development of the original NHDS has been published (23).

Beginning in 1988, the NHDS sample included with certainty all hospitals with 1,000 beds or more or 40,000 discharges or more annually. The remaining sample of hospitals is based on a stratified three-stage design. The first stage consists of a selection of 112 primary sampling units (PSU's) that comprise a probability subsample of PSU's used in the 1985–94 National Health Interview Survey. The second stage consists of a selection of noncertainty hospitals from the sample PSU's. At the third stage, a sample of discharges was selected by a systematic random sampling technique. A detailed comparison of the old and new survey designs has been published (24).

Two data collection procedures are used for the survey. The first is a manual system of sample selection and data abstraction. The second is an automated method that involves the purchase of data tapes from abstracting service organizations, State data systems, or hospitals. Approximately one third of the respondent hospitals used the automated method in 1990 through 1992.

In the manual system, the sample selection and the transcription of information from the hospital records to abstract forms are performed at the hospitals. The completed forms, along with sample selection control sheets, are forwarded to NCHS for coding, editing, and weighting. Of the hospitals using the manual system, about two-thirds had the work performed by their own medical records staff in 1990 and 1991 and 58 percent in 1992. In the remaining hospitals using the manual system, personnel of the U.S. Bureau of the Census do the work on behalf of NCHS. For the automated system, NCHS purchases tapes containing machine-readable medical record data that are systematically sampled by NCHS.

The medical abstract form and the automated data tapes contain items relating to the personal characteristics of the patients, including birth date, 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 operations or procedures. Since 1977, patient ZIP Code, expected source of payment, and dates of surgery have also been collected. (The medical record number, birth date, and patient ZIP Code are confidential information and are not available to the public.)

Presentation of estimates

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) are used to identify estimates with relatively low reliability.

Because of the complex sample design of the NHDS, estimates of less than 5,000 are not presented; only an asterisk (*) appears in the tables. These estimates generally have a relative standard error of more than 30 percent or are based on a sample of fewer than 30 cases. Estimates based on fewer than 60 cases are preceded by an asterisk (*) to indicate that they should not be assumed to be reliable. These estimates are generally 5,000 to 9,000.

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 and is expressed as a percent of the estimate. 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 SESUDAAN software, which computes standard errors by using a first-order Taylor approximation of the deviation of estimates from their expected values. A description of the software and the approach has been published (25).

The constants for relative standard error curves for estimates of discharges by race from the 1990–92 NHDS are presented in table I. The relative

Table I. Estimated parameters for relative standard error equations for number of discharges, by race: National Hospital Discharge Survey, 1991–92

Race	1990		1991		1992	
	a	*b*	*a*	*b*	*a*	*b*
All races	0.00213	228.834	0.00101	546.321	0.00097	449.059
White	0.00212	298.564	0.00234	927.094	0.00241	419.274
Black	0.00537	264.999	0.00569	273.368	0.00740	363.901
All other races	0.02899	119.661	0.02889	280.075	0.02271	182.649
Race not stated	0.02252	226.201	0.01666	427.619	0.01496	301.892

standard error [RSE(X)] of an estimate X may be estimated from the formula:

$$\mathrm{RSE}(X) = 100\sqrt{a + b/X}$$

where X, a, and b are defined in table I.

Estimates have been rounded to the nearest thousand. For this reason, figures within tables do not always add to the totals. Rates and percents were calculated from original, unrounded figures and will not necessarily agree precisely with rates or percents calculated from rounded data.

Tests of significance

In general, statistical inference was based on the two-tailed t-test using the Bonferroni critical values for post-hoc multiple comparisons (0.05 level of significance). Critical values were determined for each set of comparisons, that is, within each table. For comparisons of NHDS estimates with the Medicare data from the Health Care Financing Administration (HCFA) and numbers of live births from birth certificates, confidence intervals at the 95 percent level (plus and minus 1.96 times the standard error) were constructed around the NHDS estimates. If the number of HCFA Medicare discharges or live births fell outside the confidence interval, it was reported as significantly different from the NHDS estimate.

In this report, terms such as "higher" and "less" indicate that differences are statistically significant. Terms such as "similar" or "no difference" mean that no statistically significant difference exits between the estimates being compared. A lack of comment on the difference between any two estimates does not mean that the difference was tested and found to not be significant.

Definitions of terms

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 a year to the number of persons in the civilian population on July 1 of that year.

Ethnicity—In the NHDS, the ethnicity of discharges can be reported in three categories, which are Hispanic origin, non-Hispanic, and not stated.

Hospital—All hospitals with an average length of stay for all patients of less than 30 days or hospitals whose specialty is general (medical or surgical) or children's general are eligible for inclusion in the National Hospital Discharge Survey except Federal hospitals, hospital units of institutions, and hospitals with fewer than six beds staffed for patients' use.

- *Reporting hospital*—In this report, a reporting hospital is one that reported race for 3–100 percent of discharges.
- *Nonreporting hospital*—In this report, a nonreporting hospital is one that reported race for less than 3 percent of discharges.

Live birth—A live birth is the complete expulsion or extraction from its mother of a product of conception, irrespective of the duration of the pregnancy, which, after separation, breathes or shows any evidence of life.

Newborn infant—A newborn infant is a patient admitted by birth to the hospital.

Patient—A person who is formally admitted to the inpatient service of a short-stay hospital for observation, care, diagnosis, or treatment is a patient. The terms "patient" and "inpatient" are used synonymously.

Population—The U.S. civilian population, which is the resident population of the United States, excluding members of the Armed Forces, was used to compute rates. The U.S. resident population was used to make population-based adjustments in estimates of discharges by race.

Race—In the NHDS, the race of discharges can be reported in six categories, which are white, black, American Indian/Eskimo/Aleut, Asian/Pacific Islander, other, and not stated.

Symbols

- - -	Data not available
. . .	Category not applicable
–	Quantity zero
0.0	Quantity more than zero but less than 0.05
Z	Quantity more than zero but less than 500 where numbers are rounded to thousands
*	Figure does not meet standard of reliability or precision (see Technical notes)
#	Figure suppressed to comply with confidentiality requirements

Suggested citation

Kozak LJ. Underreporting of race in the National Hospital Discharge Survey. Advance data from vital and health statistics; no 265. Hyattsville, Maryland: National Center for Health Statistics. 1995.

DEPARTMENT OF
HEALTH & HUMAN SERVICES

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

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300

To receive this publication regularly, contact the National Center for Health Statistics by calling 301-436-8500

DHHS Publication No. (PHS) 95-1250
5-1327 (6/95)

BULK RATE
POSTAGE & FEES PAID
PHS/NCHS
PERMIT NO. G-281