Inconsistencies in Coding of Race and Ethnicity Between Birth and Death in US Infants

A New Look at Infant Mortality, 1983 Through 1985

Robert A. Hahn, PhD, MPH; Joseph Mulinare, MD, MSPH; Steven M. Teutsch, MD, MPH

Objective.—To ascertain the consistency of the racial and ethnic classification of US infants between birth and death and its impact on infant mortality rates.

Subjects.—All US infants born from 1983 through 1985 who died within a year.

Design.—We used the national linked birth/infant-death computer tape, augmented with information on infants' race and ethnicity at death, to compare the coding of race and Hispanic ethnicity at birth and at death. We also assessed infant mortality rates by race and ethnicity as defined (1) by the standard algorithm and (2) by the rule that, beginning in published tabulations for 1989, assigns newborns the race of their mothers. Finally, we estimated infant mortality rates based on consistent coding of race and ethnicity at birth and death.

Results.—Inconsistency in the coding of race is low for whites (1.2%), greater for blacks (4.3%), and greatest for races other than white or black (43.2%). Most infants reclassified at death (87.3%) are classified as white at death. Inconsistency in coding is lower for non-Hispanic whites (3.5%) and non-Hispanic blacks (3.3%) than for Hispanic populations (30.3%). Compared with the standard algorithm for calculation of infant mortality, consistent definition at birth and death produces rates 2.1% lower for whites, and higher for all other groups—3.2% for blacks, 46.9% for American Indians, 33.3% for Chinese, 48.8% for Japanese, 78.7% for Filipinos, and 8.9% for Hispanics.

Conclusions.—The coding of race and ethnicity of infants at birth and death is remarkably inconsistent, with substantial impact on the estimation of infant mortality rates. A need exists to reconsider the nature and definition of race and ethnicity in public health.

(*JAMA*. 1992;267:259-263)

THE HEALTH status of minority racial and ethnic populations in the United States is a prominent focus of the national public health agenda. The second of three goals in *Healthy People 2000*[1] is the reduction of disparities in health among different segments of the population, including racial and ethnic groups. Achievement of this goal through public health surveillance, analysis, and program planning and implementation requires the valid and consistent classification of persons into racial and ethnic categories.

See also pp 268 and 275.

To ascertain the consistency of racial and ethnic classification in the United States, we compared the coding of race and ethnicity at birth with the coding of race and ethnicity at death for all infants born from 1983 through 1985 who died within a year. We then assessed infant mortality rates by race and ethnicity calculated by standard methods and estimated infant mortality rates based on consistent coding of race and ethnicity at birth and death.

METHODS

Data Sources

We used the newly available National Center for Health Statistics (NCHS) linked birth/infant-death computer data tapes for 1983 through 1985,[2-4] with additional information obtained from the NCHS on the race and ethnicity of infants recorded at death. The linked data tape includes information on infants born to US resident and nonresident parents who died in the United States before the age of 1 year and for whom birth certificates were linked with death certificates. (In 1983 and 1984, four states and Washington, DC, submitted only 50% of their birth certificate records to the NCHS. Published information and linked tape records estimate characteristics of the remaining 50% based on the available information.)

Most states use US Standard Certificates of Live Birth and Death and closely conform to standard certification procedures.[5] Birth registration instructions indicate that personal information should be elicited from mothers, fathers, or other knowledgeable persons at the time of birth; death registration instructions indicate that personal information should be elicited from next of kin.[6] On birth and death certificates, race and Hispanic ethnicity are regarded as separate, cross-cutting characteris-

From the Division of Surveillance and Epidemiology, Epidemiology Program Office, Centers for Disease Control, Atlanta, Ga.

Reprint requests to Division of Surveillance and Epidemiology, Epidemiology Program Office, Centers for Disease Control, Atlanta, GA 30333 (Dr Hahn).

Table 1.—Race of Infants at Birth and Death, 1983 Through 1985, Linked Tape

	At Death (No./Row %)									
	White	Black	American Indian	Chinese	Japanese	Hawaiian	Filipino	Other Asian	Other	Total
At birth (No./row %)										
White	78 898 98.83	701 0.88	83 0.10	31 0.04	9 0.01	8 0.01	12 0.02	72 0.09	19 0.02	79 833
Black	1335 4.06	31 448 95.70	18 0.05	4 0.01	3 0.01	1 0.00	8 0.02	37 0.11	8 0.02	32 862
American Indian	574 34.81	25 1.52	1046 63.430	0 0.00	0 0.00	0 0.00	2 0.12	1 0.06	1 0.06	1649
Chinese	80 22.92	7 2.01	1 0.29	183 52.44	2 0.57	0 0.00	1 0.29	75 21.49	0 0.00	349
Japanese	74 40.44	3 1.64	0 0.00	3 1.64	84 45.90	2 1.09	1 0.55	16 8.74	0 0.00	183
Hawaiian	40 16.95	3 1.27	0 0.00	0 0.00	2 0.85	182 77.12	5 2.12	4 1.69	0 0.00	236
Filipino	220 44.90	9 1.84	1 0.20	1 0.20	3 0.61	4 0.82	233 47.55	19 3.88	0 0.00	490
Other Asian	548 36.36	35 2.32	32 2.12	44 2.92	9 0.60	1 0.07	10 0.66	823 54.61	5 0.33	1507
Other	56 70.89	13 16.46	2 2.53	0 0.00	0 0.00	0 0.00	0 0.00	5 6.33	3 0.80	79
Total	81 825	32 244	1183	266	112	198	272	1052	36	117 188

tics; persons filing certificates are asked to record both items. The race item on birth and death certificates provides a brief list of races as examples ("American Indian, Black, White, etc"); persons recording race are asked to write in the race of both parents (on birth certificates) or of the decedent (on death certificates). The Hispanic ethnicity item on birth and death certificates asks if the newborn or decedent is of Hispanic origin, and, if so, which Hispanic origin ("Cuban, Mexican, Puerto Rican, etc"). In this study, the coding of race is analyzed separately. Hispanic populations are compared with non-Hispanic whites and blacks. The racial and ethnic categories used are those coded in the linked data tape.

For the years 1983 through 1985, the race of an infant at birth was determined by an algorithm incorporating information on the race of the infant's parents as reported on birth certificates.[5,7,8] Because the algorithm was changed for published vital statistics beginning in 1989, we refer to the prior algorithm as "pre-1989." In the pre-1989 algorithm, if both parents were white, the child was white; if one parent was Hawaiian, the child was Hawaiian; if only one parent was white, the child was assigned the race of its other-than–white race parent; and if both parents were of races other than white, the child was assigned its father's race. If data on race were missing for either parent, the infant was assigned the race of the parent for whom information was available. If there was no information on the race of either parent (<0.1% of births in 1983 and <0.2% of births in 1984 and 1985),[6-8] the infant was assigned the race of the infant in the preceding record in the NCHS computer file. As of 1989, the infant's race in published, tabulated statistics is uniformly determined by the mother's race (although a public-use data tape is still available for assessment of the infant's race by the pre-1989 algorithm).

On death certificates, racial identification of the decedent as reported by next of kin to funeral directors (or medical examiners or coroners) is cited as the determinant of a decedent's race.[9-11] However, with unknown frequency, certifiers make independent assessments of race. In the absence of death certificate information on race (approximately 0.2% of all death certificates from 1983 through 1985),[10-12] the decedent is assigned white race if the race of the preceding decedent in the NCHS mortality computer file is white; otherwise, black race is assigned.

Parental Hispanic origin was reported on the birth certificates of 23 states and Washington, DC.[6-8] In published tabulations, infants are assigned the Hispanic origin of their mothers. (Thus, the 1989 revision of race coding would not affect the coding of Hispanic origin.) Infant mortality rates by Hispanic origin met NCHS criteria for tabulated publication (ie, reporting ≥90%) in 15 states in 1984 and 17 states and Washington, DC, in 1985.[9-11] This study analyzes the coding of Hispanic ethnicity in 1984 and 1985 in states meeting NCHS criteria.

In "standard" infant mortality rates, infant deaths in a given year are divided by births in the same year; since infants who die in a given year may have been born in the previous year and since infants born in a given year may die within the following year, standard infant mortality rates are ratios rather than true rates. The linked birth/infant-death tape allows calculation of true infant mortality rates, since it considers all deaths (over a 2-year span) occurring in a cohort of infants (born in a single year). (For simplicity, we employ the usual terminology, "infant mortality rate," to refer to both true rates and ratio estimates.)

Analysis

We assessed the consistency of infants' racial and Hispanic ethnic classification at birth and at death using the pre-1989 algorithm. We then examined the implications of inconsistent racial and ethnic classification at birth and death for the estimation of infant mortality rates. We first used pre-1989 definitions of infant race, published information, and standard methods to calculate infant mortality rates for racial and ethnic populations from 1983 through 1985. We then calculated infant mortality rates using the new 1989 NCHS rule that uniformly assigns infants the race of their mothers, while still using the traditional assignment of race at death. We ascertained births classified by the mother's race from the linked tape.

Because the mortality file in the linked tape includes only deaths for which matching birth certificates were found, and because matching was incomplete (2.5% incomplete overall), the mortality file represents fewer deaths than actually occurred in the cohort of newborns. In addition, linkage may differ by race but is not available in linked-tape documentation. To adjust for computer-file nonlinkage with recorded deaths, we used information about the proportion of infants who had died but whose records were not in the linked birth-

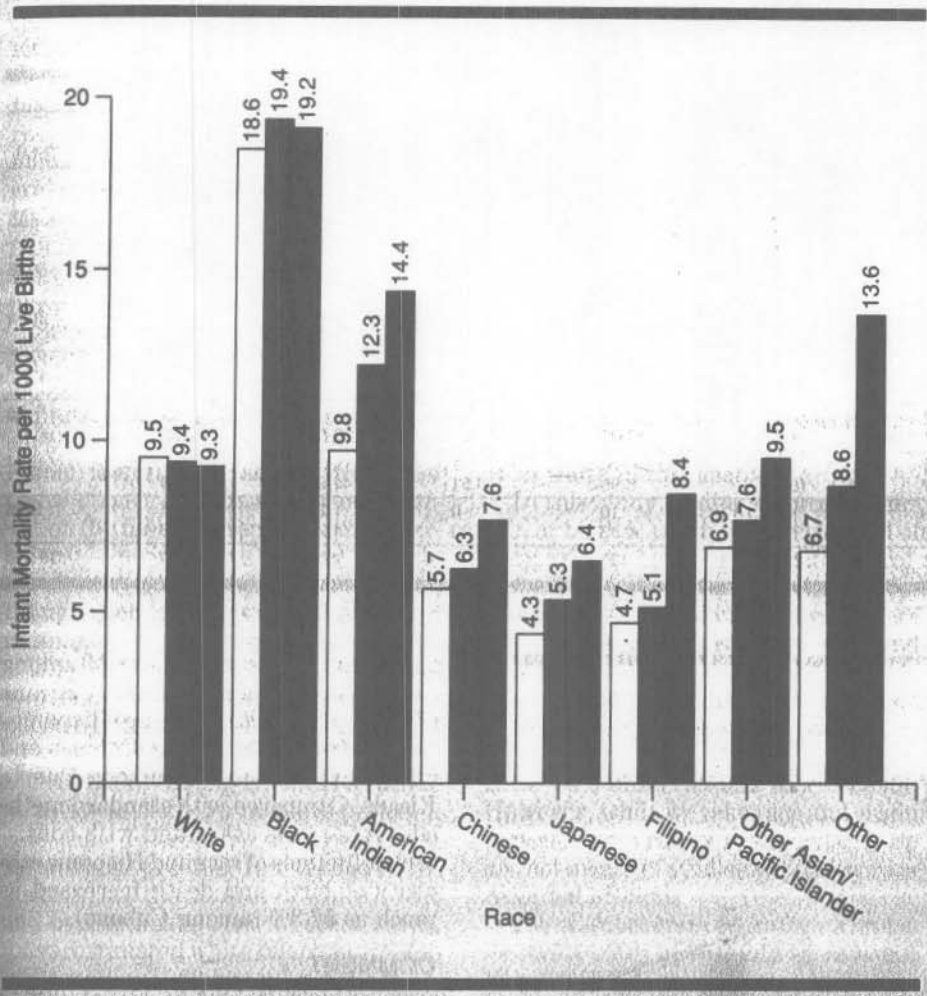

Fig 1.—Infant mortality rates by race, United States, 1983 through 1985, estimated from vital records and linked birth/infant-death tapes. Open bars indicate pre-1989 National Center for Health Statistics algorithm; shaded bars, new (1989) National Center for Health Statistics algorithm; and black bars, estimate from linked birth/infant-death tape (mother's race assigned to infant at birth and death).

death mortality file. In estimating linkage, we assumed that the number of deaths in the 1983 through 1985 birth cohort over 3 years was a good approximation of the number of deaths occurring in the same years. We also assumed that linkage by race at death was a good approximation of linkage by race at birth.

We then assessed rates of infant mortality by Hispanic origins as calculated by standard methods for 1984 and 1985, the first years in which they were published. Finally, we calculated infant mortality rates by race and by Hispanic origin, consistently defining an infant's race and Hispanic origin as that of its mother both at the infant's birth and at its death, adjusting for nonlinkage in the mortality portion of the linked tape. In assessing linkage by Hispanic origin and race, we excluded records in which origin was not stated.

RESULTS

Of all US infants who died from 1983 through 1985 and whose records were successfully linked to birth certificates, 4288 (3.7%) were classified as having a different race at death from that at birth (Table 1). Inconsistent racial classification varied widely by race: 1.2% of infants classified as white at birth, 4.3% of infants classified as black at birth, and 43.2% of infants classified as members of all other races at birth had a different race at death (Table 1). More infants were classified as white at death than at birth (Table 1). Whites "gained" 2.5% in population from birth to death, while all other races "lost" in population, from 1.9% (among blacks) to 44.5% (among Filipinos). Of 3353 infants not classified as white at birth who were classified differently at birth and death, 2927 (87.3%) were classified as white at death. Almost as many infants classified as Filipino and Japanese at birth were classified as white at death (44.9% and 40.4%, respectively) as were consistently classified as Filipino and Japanese at both birth and at death (47.6% and 45.9%, respectively).

Overall, an estimated 97.5% of infant death certificates were linked with birth certificates; linkage varied by race, from 97.7% for whites to 93.3% for Japanese and 81.8% for "all other" races (data not shown). Approximately 87% of infants who died within a year died in the same calendar year in which they were born (data not shown). Thus, over 95% of infants of the 1983 through 1985 cohort who died died from 1983 through 1985. Comparison of the consistency of racial classification by time interval from birth to death revealed no remarkable pattern.

From 1983 through 1985, infant mortality rates estimated by pre-1989 definitions and standard procedures range from 4.3 per 1000 among Japanese live births to 18.6 per 1000 among blacks (Fig 1). American Indians have approximately the same rate of infant mortality as whites. Compared with whites, rate ratios range from 0.45 among Japanese to 1.95 among blacks.

Compared with standard rates calculated with pre-1989 definitions, estimates of infant mortality using maternal race at the infant's birth (the new 1989 NCHS rule) and the traditional assignment of infant race at death are lower for whites and higher for all other races (Fig 1). Correspondingly, rate ratios are thus also greater than rates calculated by standard procedures, ranging from 0.54 among Filipinos to 2.07 among blacks.

Finally, estimates of infant mortality using maternal race both at the infant's birth and at death, adjusted for linkage, are again lower than standard rates for whites (2.1%) and higher for members of all other racial groups—3.2% for blacks, 46.9% for American Indians, 33.3% for Chinese, 48.8% for Japanese, and 78.7% for Filipinos. Infant mortality rates ranged from 6.4 per 1000 among Japanese to 19.2 per 1000 live births among blacks. Compared with standard rates calculated with pre-1989 definitions, rate ratios calculated with consistent definition of race at birth and death are also substantially higher, from 0.70 among Japanese to 2.08 among blacks.

Among infants from the states analyzed for Hispanic populations, 69.7% of infants were assigned the same specified Hispanic origin (Cuban, Mexican, or Puerto Rican) at birth and death; an additional 13.4% were assigned a different Hispanic origin at death (Table 2). Inconsistency in coding is less for non-Hispanic whites (3.5%) and non-Hispanic blacks (3.3%) than for Hispanic populations (30.3%). Linkage of birth and death certificates was complete for 95.5% of Hispanic records overall: 94.2% for Mexicans, 94.0% for Puerto Ricans, and 100% for Cubans (data not shown).

When infants were assigned mater-

Table 2.—Mother's Hispanic Origin at Infant's Birth and Infant's Hispanic Origin at Death From 15 Reporting States, 1984, and 17 Reporting States, 1985, Linked Tape*

	Hispanic Origin Coded at Death (No./Row %)							
	NH White	NH Black	Mexican American	Puerto Rican	Cuban	Other SA	Others	Total
Hispanic origin coded at birth (No./row %)								
NH white	16 003 96.46	95 0.57	279 1.68	41 0.25	3 0.02	124 0.75	45 0.27	16 590
NH black	195 2.42	7779 96.68	12 0.15	19 0.24	1 0.01	17 0.21	23 0.29	8046
Mexican American	348 11.40	11 0.36	2445 80.09	5 0.16	2 0.07	234 7.66	8 0.26	3053
Puerto Rican	108 20.04	29 5.38	16 2.97	282 52.32	1 0.19	100 18.55	3 0.56	539
Cuban	11 36.67	2 6.67	1 3.33	3 10.00	10 33.33	3 10.00	0 0.00	30
Other SA	184 23.56	31 3.97	152 19.46	70 8.96	4 0.51	330 42.25	10 1.28	781
Others	312 25.89	19 1.58	22 1.83	10 0.83	0 0.00	21 1.74	821 68.13	1205
Total	17 161	7966	2927	430	21	829	910	30 244

*NH indicates non-Hispanic; and SA, South American.

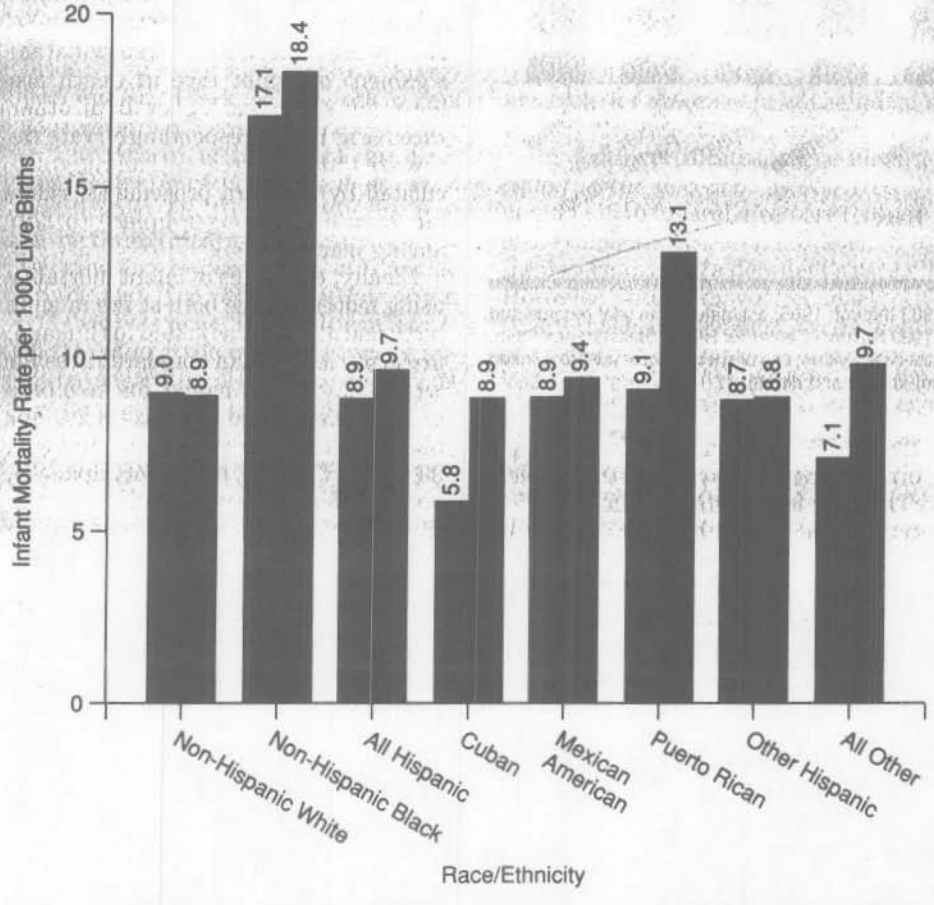

Fig 2.—Infant mortality rates by race and ethnicity, United States, 1984 and 1985, estimated from vital records and linked birth/infant-death tapes. Shaded bars indicate National Center for Health Statistics algorithm; and black bars, estimate from linked birth/infant-death tape (mother's race assigned to infant at birth and death).

nal race and origin at both birth and death, adjusted for linkage, infant mortality declined for non-Hispanic whites and increased for non-Hispanic blacks and for all Hispanic groups, compared with infant mortality calculated by standard methods (Fig 2). For Hispanics overall, rates were generally higher than those for non-Hispanic whites; rates ranged from 8.8 per 1000 live births among "other Hispanics" and 8.9 among Cubans to 13.1 among Puerto Ricans (Fig 2). Rate ratios among Hispanics ranged from 0.99 among Cubans and other Hispanics to 1.47 among Puerto Ricans. Compared with standard methods, rate ratios calculated with consistent definitions of race and Hispanic ethnicity at birth and death increased as much as 52.3% (among Cubans).

COMMENT

Particularly for infants not classified as white or black at birth, the classification of the race of US infants at birth and at death is remarkably inconsistent. While 1.2% of infants classified as white at birth were assigned a different race at death, 4.3% of infants classified as black at birth were assigned a different race at death—3.6 times as many as whites. Infants who were classified as neither white nor black at birth were little more than half as likely to have the same racial classification at birth and at death. Among infants classified inconsistently at birth and death, 87% were classified as white at death. Classification of Hispanic origins at the infant's birth and at death is also substantially inconsistent and varies greatly by Hispanic group. Even in states that satisfy NCHS reporting criteria, 30% of infants assigned a specific Hispanic origin at birth were assigned a different origin at death.

This national study confirms earlier findings from linked birth-death certificate studies in three states.[12-14] From 1968 through 1977 in Washington State, 4.2% of infants were classified with a different race at death than at birth. Discrepancies were greatest for nonwhites and nonblacks; most (87.2%) nonwhite infants with discrepant racial clas-

sifications at birth and death were classified as white at death.[13] Comparable findings have been reported in California in the years 1965 through 1967.[14] A recent study of infant mortality in Oklahoma from 1975 through 1988 noted increasing racial misclassification since 1975 and estimated that reported infant mortality rates for American Indians were 41% below rates calculated from linked records.[15]

Several factors may account for differential classification of race and origin at birth and death. The algorithm and definition of race and origin differ for each event, one based on the racial or ethnic identity of parents, the other based on ascribed identity or on observation (by funeral directors or other certifiers). And individuals' racial and ethnic identity may also change over time, though such change is unlikely to affect infants.

The NCHS has recently addressed two important issues previously complicating the study of infant mortality by race. First, in published tabulations, the rule determining an infant's race at birth by its mother's race will replace a rule that differed for different races. Because the pre-1989 rule tended to underestimate the number of white births and overestimate those of other races, it overestimated white infant mortality, while minimizing infant mortality among races other than white. Although perhaps not fully in accord with popular conceptions of race, the new rule is uniform for all races and will remove this systematic error.

Second, the recently introduced linked birth/infant-death tape allows assessment of (*a*) true infant mortality rates rather than the ratios estimated by standard methods and (*b*) infant mortality rates by race and origin with consistent definitions at birth and death. Consistent definition will eliminate the problem demonstrated in this report, namely, the inconsistency of racial and ethnic classification of infants at birth and death.

Two studies have recently used mother's race and origin at the infant's birth and death to estimate infant mortality rates from 1983 through 1984 by Hispanic origins[15] and race and Hispanic origins.[16] In addition to infants of mothers who report Hispanic origins, Becerra et al[15] classify among Hispanic births and deaths infants whose mothers report being born in Mexico, Puerto Rico, or Cuba, on the assumption that women born in these countries are Hispanic. The present study differs from the studies of Becerra et al and Kleinman[16] in four ways: (*a*) it adjusts for linkage in the linked birth/infant-death computer tape, (*b*) it adjusts for differential linkage by race and Hispanic origins, (*c*) it includes for analysis of rates by Hispanic origin only states that meet NCHS publication criteria, and (*d*) it includes data for 1983 through 1985. Adjustment for linkage is expected to produce higher infant mortality rates in the present analysis compared with those generated in other studies; lower rates of linkage among populations other than white produces greater rate ratios for races other than white in this analysis.

In this study, coding of race and origin at birth is taken as the standard of comparison, because information collected at birth, in the presence of the mother or other family members, is more likely to be accurate. When race and origin are recorded at death, it is possible that neither parent is present, and funeral directors (or others) who record demographic characteristics may make their own observational assessments. However, while we take race and origin recorded at birth as the standard, we are not aware of studies validating race recorded at birth.

The most accurate estimates of infant mortality rates by race and by race and Hispanic origin are produced by a combination of (*a*) the new 1989 NCHS rule for assigning race at birth, (*b*) the assignment of an infant's race or origin at death from birth certificate records, and (*c*) adjustment for computer-file linkage. Application of this procedure confirms lower rates for whites and non-Hispanic whites, slightly higher rates for blacks and non-Hispanic blacks, and higher rates for all other groups than indicated by both pre-1989 and, except for blacks, new 1989 NCHS procedures. Compared with those for whites and non-Hispanic whites, rate ratios are thus also increased by this method, particularly for races other than black and for Hispanics.

The effective collection of vital statistics depends on the collaboration of national and state public health agencies and the medical personnel who witness vital events. An effective vital records system requires the selection of categories of information, such as "race" and "ethnicity," that are valid (ie, that accurately represent the concepts being evaluated) and that allow reliable ascertainment by designated personnel; it also requires specification of procedures for the collection of vital information. This study indicates problems in the classification of race and ethnicity at death. It also suggests that increased uniformity in approaches among the states to the recording of Hispanic origins could greatly improve our knowledge of the health status of Hispanic populations in the United States. Overall, this study reinforces other findings that suggest a need to reconsider definitions and procedures for the classification of race and origin in federal health statistics, an issue currently being addressed by the Public Health Service Task Force on Minority Health Data.

The authors are grateful for helpful comments from Carol Hogue, PhD, National Center for Chronic Disease Prevention and Health Promotion, and the late Joel Kleinman, PhD, National Center for Health Statistics, Centers for Disease Control.

References

1. *Healthy People 2000*. Conference edition. Washington, DC: US Dept of Health and Human Services, Public Health Service; 1990.
2. National Center for Health Statistics. *Public Use Data Tape, 1983*. Hyattsville, Md: US Dept of Health and Human Services; 1989.
3. National Center for Health Statistics. *Public Use Data Tape, 1984*. Hyattsville, Md: US Dept of Health and Human Services; 1989.
4. National Center for Health Statistics. *Public Use Data Tape, 1985*. Hyattsville, Md: US Dept of Health and Human Services; 1989.
5. National Center for Health Statistics. *Vital Statistics of the United States, 1983: Natality*. Washington, DC: US Dept of Health and Human Services, Public Health Service; 1987;1. Publication (PHS)87-1113.
6. National Center for Health Statistics. *Hospitals' and Physicians' Handbook of Birth Registration and Fetal Death Reporting*. Washington, DC: US Dept of Health and Human Services, Public Health Service; 1987.
7. National Center for Health Statistics. *Vital Statistics of the United States, 1984: Natality*. Washington, DC: US Dept of Health and Human Services, Public Health Service; 1988;1:7. Publication (PHS)88-1100.
8. National Center for Health Statistics. *Vital Statistics of the United States, 1985: Natality*. Washington, DC: US Dept of Health and Human Services, Public Health Service; 1988;1. Publication (PHS)88-1113.
9. National Center for Health Statistics. *Vital Statistics of the United States, 1983: Mortality, Part A*. Washington, DC: US Dept of Health and Human Services, Public Health Service; 1987;2. Publication (PHS)87-1101.
10. National Center for Health Statistics. *Vital Statistics of the United States, 1984: Mortality, Part A*. PHS, Washington, DC: US Dept of Health and Human Services, Public Health Service; 1988;2. Publication (PHS)88-1100.
11. National Center for Health Statistics. *Vital Statistics of the United States, 1985: Mortality, Part A*. PHS, Washington, DC: US Dept of Health and Human Services, Public Health Service; 1989;2. Publication (PHS)88-1101.
12. Frost F, Shy KK. Racial differences between linked birth and infant death records in Washington State. *Am J Public Health*. 1980;70:974-976.
13. Norris FD, Shipley PW. A closer look at race differentials in California's infant mortality, 1965-67. *HSMHA Health Rep*. 1971;86:810-814.
14. Kennedy RD, Deapen RE. Differences between Oklahoma Indian infant mortality and other races. *Public Health Rep*. 1991;106:97-99.
15. Becerra JE, Hogue CJR, Atrash HK, Perez N. Infant mortality among Hispanics. *JAMA*. 1991;265:217-221.
16. Kleinman JC. Infant mortality among racial/ethnic minority groups, 1983-1984. *MMWR*. 1990;39(suppl SS-2):31-39.