Discrepancies Between Published Data on Racial Classification and Self-Reported Race: Evidence from the 2002 North Carolina Live Birth Records

Paul A. Buescher, PhD[a]
Ziya Gizlice, PhD[a]
Kathleen A. Jones-Vessey, MS[a]

SYNOPSIS

Objectives. We compared data on race as reported by the mother on North Carolina birth certificates with data on race in officially reported statistics. We also determined to what extent differences in the classification of race affect measures of racial disparity in maternal and child health indicators.

Methods. We examined how data on race are collected, coded, and tabulated in North Carolina via live birth certificates, death certificates, the Behavioral Risk Factor Surveillance System (BRFSS) telephone survey, and the Central Cancer Registry case records. We showed how the data on race collected through North Carolina birth and death certificates are translated into 10 fixed racial categories designated by the National Center for Health Statistics (NCHS) for use in official vital statistics. We compared race as reported by the mother on birth certificates to racial tabulations used in the official published birth statistics. We also examined to what extent differences in the determination of race affect measures of racial disparity in maternal and child health indicators.

Results. Out of nearly 118,000 live births in North Carolina in 2002, mothers reported more than 600 different versions of race on birth certificates. These entries were collapsed into the 10 standard racial categories outlined in federal coding rules. Approximately two-thirds of mothers of Hispanic ethnicity report their race with a label that can be categorized as "Other" race, but nearly all of these births are re-coded to "white" for the official birth statistics. Measures of racial disparity vary depending on whether self-reported or officially coded race is used.

Conclusions. This study shows that, given the opportunity to report their own race, North Carolinians describe their race using a wide variety of terms and concepts. In contrast, health statistics are usually reported using a few standardized racial categories defined by federal policy. The NCHS rules for coding race should be reexamined. As the ethnic and racial diversity of the United States continues to increase, these rules will become increasingly antiquated.

[a]State Center for Health Statistics, North Carolina Division of Public Health, Raleigh, NC

Address correspondence to: Paul Buescher, PhD, State Center for Health Statistics, North Carolina Division of Public Health, 1908 Mail Service Center, Raleigh, NC 27699-1908; tel. 919-715-4478; fax 919-733-8485; e-mail <paul.buescher@ncmail.net>.

North Carolina health statistics are often tabulated by race as a means of measuring health disparities. Many of North Carolina's racial minority groups have worse outcomes than the majority white population on a number of health measures, such as low birthweight and infant mortality. Information on disparities can be used to target health programs to populations in need.

Published health data give the impression that racial categories are distinct, well defined, and homogeneous. However, there is a growing consensus in the scientific community that distinct human races do not exist.[1] But because of the historical social stratification role of race, particularly in the United States, categorization of people by race continues. Federal policy defines a limited number of discrete racial categories that must be used in reporting data for all federal programs.[2] These categories are used in North Carolina to tabulate health data by race. Our experience in North Carolina shows that self-reported race on birth certificates is quite different from the standard federal categories used to publish the data by race.

Measures of racial disparity may vary depending on whether self-reported or officially coded race is used. A Massachusetts study showed that the infant mortality rate in Boston for babies of mothers of black race was 17% lower when Cape Verdeans and Dominicans were included in the "black" racial category, as mandated by national vital statistics coding rules, compared to when they are not included in the black category.[3] Most people in these two groups self-report their race on birth certificates as "Other" rather than black.

The State Center for Health Statistics, North Carolina Division of Public Health, conducted a study to compare data on race as reported by the mother on North Carolina birth certificates with data on race in officially reported statistics. We also determined to what extent differences in the classification of race affect measures of racial disparity in maternal and child health indicators.

METHODS

In North Carolina, data on race are collected, coded, and tabulated differently on live birth certificates, death certificates, the Behavioral Risk Factor Surveillance System (BRFSS) telephone survey, and records of new cases of cancer. Data on race from North Carolina birth and death certificates are translated into 10 fixed racial categories designated by the National Center for Health Statistics (NCHS) for use in official vital statistics. The information presented here on self-reported race is for all live births occurring in North Carolina in 2002 (nearly 118,000), in contrast to some other studies based on a small-scale sample.[4]

Data description

Birth certificates. Data on race from North Carolina live birth certificates are collected through a fill-in-the-blank box on the certificates. The mother usually fills out a birth certificate worksheet while in the hospital. The form includes a blank space to record her race and another blank space to record the race of the father ("Specify White, Black, American Indian, etc."). One race is usually entered, though multiple races are sometimes written in. This text, supplied by the mother, is then entered into the Electronic Birth Certificate (EBC) system by hospital staff. The instructions say, "Enter the color or race . . . of both parents as furnished by the mother or other informant. For Asians and Pacific Islanders, the national origin may be entered (e.g., Chinese, Japanese, Korean, Hawaiian, Vietnamese, etc.)." The NCHS considers the race of the child to be the same as the race of the mother. Prior to about 1990, the race of both the mother and father were considered in a complex algorithm to determine the race of the child. In North Carolina, approximately 16% of live birth records are missing information on the father's race.

These open-ended questions on race result in a variety of responses. The actual text of the mothers' responses is captured in the EBC system. For the 117,949 live births occurring in 2002 in North Carolina hospitals reporting through the EBC system, more than 600 different text versions of "race" of the mother were reported (counting different spellings and capitalizations). The most common text entries for race were white (74,789 or 63%), black (27,142 or 23%), Hispanic (9,746 or 8%), Asian (1,586 or 1%), and American Indian (1,512 or 1%). These five text entries accounted for 96% of all birth certificates. Specific nationalities were often reported as race (e.g., Cambodian, Dominican, Guatemalan, Hmong, Mexican, British). In many cases, racial combinations were reported (e.g., white/Mexican, Hispanic/black, Egyptian/Canadian, Mixed Indian-Italian, Biracial, and multiracial). On one birth certificate, the mother's race was reported as "Dominant White" and on another as "Son of God."

These racial entries on the birth certificate are prompted by the instructions ("Specify White, Black, American Indian, etc."), and therefore are not completely free of self-identification. A truly open-ended question on race would likely elicit an even greater variety of responses.

Death certificates. Data on race from North Carolina death certificates are also collected via a fill-in-the-blank box on the certificates. The box is labeled at the top, "RACE—American Indian, Black, White, etc. (Specify)." The instructions for the death certificate say, "Enter the race of the decedent as stated by the informant . . . For Asians and Pacific Islanders the national origin may be entered, such as Chinese, Japanese, Korean, or Vietnamese. Two races may be shown for persons of mixed racial heritage." On death certificates, responses to race are not captured electronically, but NCHS coding rules also apply to death certificates. A variety of racial labels are reported, most often elicited from family or friends of the decedent, but race is sometimes assigned by the funeral director based on physical appearance.[5] These racial labels are then converted into 10 fixed racial categories specified by the NCHS. All tabulated and published North Carolina mortality data are based on these 10 categories. For 2002 deaths occurring in North Carolina, 77.7% were coded as white, 21.2% as black, 0.9% as Indian, and 0.2% in the other seven categories combined.

BRFSS. North Carolina telephone survey data reveal a pattern similar to that found with the birth certificates. The BRFSS is a random telephone survey of persons aged 18 and

older. In 2002, approximately 6,700 interviews were completed and in 2003 more than 9,400 were completed. In addition to a variety of health-related questions, respondents are asked two separate questions about their ethnicity and race. They are asked, "Are you Hispanic or Latino? (Yes/No)." Then they are asked, "Which one of these groups would you say best represents your race?" and the following list is read over the phone: white, black or African American, Asian, Native Hawaiian or Other Pacific Islander, American Indian or Alaskan Native, or Other. Of the 634 respondents in 2002 and 2003 who indicated that they were Hispanic/Latino, three-fourths chose "Other" as their racial group. This is similar to data from a national survey of adolescents, where 46% of those who said that they were Latino or Hispanic chose "Other" as their only race.[6]

The federal policy on racial and ethnic classification[2] recommends that the question on Hispanic ethnicity be asked *before* the question on racial identity, in part to reduce the "Other" race responses for Hispanics. The BRFSS adheres to this recommended ordering of the ethnicity and race questions, but still two-thirds of Hispanics identify themselves as "Other" race.

The 2002 North Carolina BRFSS included the question, "How do OTHER PEOPLE usually classify you in this country? Would you say White, Black or African American, Hispanic or Latino, Asian, Native Hawaiian or Other Pacific Islander, American Indian or Alaska Native, Multiracial, or some other group?" This question was preceded by the introductory statement, "Earlier you told me about your race. Now I will ask some questions about reactions to your race." Of the survey respondents who reported themselves as Hispanic and speak Spanish only, 93% reported that they were classified by others as being Hispanic or Latino and only 0.5% reported that they were classified by others as being white. Of the survey respondents who were Hispanic and speak English, 39% reported that they were classified by others as being Hispanic or Latino and 43% reported that they were classified by others as being white.

Central Cancer Registry. Records of new cases of cancer from the Central Cancer Registry show a different picture of race for Hispanics. Of the 929 cancer cases from 1996 through 2000 in which the patient was indicated to be of Hispanic ethnic origin, the race was coded as white for 83.5%. However, unlike the BRFSS survey and the mother's birth certificate worksheet where race is self-reported, race on the cancer records is usually based on visual criteria by the hospital clerk in the admissions office.

NCHS coding specifications

North Carolina submits its vital records data to the NCHS as part of the National Vital Statistics System. Under a contract where the NCHS provides funding for these data, states are required to incorporate NCHS coding specifications for demographic and other items on the vital records, including race. As of 2004, North Carolina had not adopted the new national model vital certificates, which allow for checking one or more racial categories, so North Carolina vital records are still coded to a single racial group. For birth and death certificates, all text entries for race are converted into one of the following 10 categories: white, black, Indian, Chinese, Japanese, Hawaiian, Filipino, Other Asian or Pacific Islander, Other Entries, and not reported. Some of the NCHS rules for this conversion are:[7]

- If Hawaiian is reported with any other race, code Hawaiian.
- If more than one race is reported (except Hawaiian), code the first race listed.
- If more than one race is reported with percentages or fractions given (except Hawaiian), code the race having the higher percentage or fraction.
- If entry is Col., N, Negro, Color(ed), B, Brown, A.A., Afro-American, or African American, code Black.
- States not mandated by law to code multi-racial as a separate category (North Carolina is not) may code entries such as multiracial, biracial, mixed, or other synonymous terms as Other Entries.

In addition, the NCHS Coding Instructions include an appendix that lists hundreds of terms that might be written in the race box on the vital certificates, with an indication of which of the 10 fixed race categories should be assigned to them. For example, Hispanic, Mexican, Puerto Rican, Egyptian, Cuban, Moroccan, Persian, Syrian, Turk, and Yugoslavian should be coded to white. African, Cape Verdean, Dominican, Jamaican, Liberian, Mulatto, Octaroon, Quadroon, and West Indies should be coded to black. Aleut, Eskimoan, Mexican Indian, Red, and Ute should be coded to Indian. Amerasian, Asian Indian, Burmese, Cambodian, Dutch East Indian, Eurasian, Hindu, Pakistani, Polynesian, Sikh, Thai, and Tibetan should be coded to Other Asian or Pacific Islander. Clearly, a lot of detail is lost in the conversion of these self-reported racial labels to the standard NCHS categories.

RESULTS

We compared race as reported by the mother on the birth certificate to racial tabulations used in the official NCHS published birth statistics. We also examined to what extent differences in the classification of race affect measures of racial disparity in maternal and child health indicators.

Table 1 shows the percentages of 2002 live births received through the North Carolina EBC system that fall into four broad racial groupings, comparing what is self-reported by the mother to what is coded for NCHS purposes. A major reason for the difference in self-reported race and the NCHS racial coding is that many mothers in North Carolina list their "race" as "Hispanic" in the blank on the birth certificate worksheet (which would be counted in the "All Other" category in the first column of Table 1). However, NCHS considers Hispanic to be an ethnic group rather than a racial group and includes a separate ethnicity variable for capturing Hispanic origin on the vital statistics files. According to NCHS coding specifications, if "Hispanic" is listed as a race on the birth certificate, race should be recorded as "white." Among the 2002 live birth records received through the North Carolina EBC system, 15,074 included a designa-

Table 1. Percentages of live births in four broad racial groups: race as self-reported by the mother compared to race as determined by NCHS coding rules

	Self-reported	*NCHS coding*
White	63.4%	72.7%
Black	23.0%	23.4%
Indian	1.3%	1.4%
All other categories and combinations	12.3%	2.4%

SOURCE: 2002 North Carolina Live Births, State Center for Health Statistics, North Carolina Division of Public Health

NCHS = National Center for Health Statistics

tion by the mother that she was of Hispanic ethnic origin. However, among these 15,074 live births, 10,361 (or 69%) had an "Other" race written in by the mother (with 9,445 of these listing "Hispanic" as their race), and 30% listed their race as "white." By contrast, in the official North Carolina live birth statistics, among the 2002 live births occurring in North Carolina where a Hispanic ethnic origin was indicated, 98.5% were recorded as white, according to NCHS coding rules.

Reclassifying some of the live birth records into a standard racial category that is different from what the mother self-reported affects measures of racial disparities in health. For example, Hispanics have a relatively low percentage of low birthweight and maternal smoking during pregnancy and a relatively high percentage of births with prenatal care starting after the first trimester or with no prenatal care, compared to the percentages for all North Carolina resident live births. Counting Hispanics mostly in the white racial category influences the rates for whites. Table 2 shows how the two methods of racial classification affect measures of racial disparity in three health indicators from the birth certificate. In the "self-reported" rows, births are counted in the category written by the mother on the birth certificate worksheet. So only births where the mother wrote "white" or "black" are counted in those categories. More than 9,500 mothers reported "Hispanic" as their race (and some mothers reporting Hispanic as their ethnicity reported "Other" as their race), and these are excluded from the self-reported white category in Table 2. In the "NCHS Coding" rows, births are counted in the racial categories as determined by the NCHS coding rules (more than 98% of Hispanics are counted in the white category).

The top half of Table 2 illustrates the effects of the different classification methods on black/white ratios. The percentage of low birthweight for Hispanics is similar to that for whites, so leaving births where the mother reported her race as Hispanic out of the white category does not have much effect on the white percentage or on the black/white ratio. However, Hispanics have a much lower rate of smoking during pregnancy than whites and a much higher rate of late or no prenatal care, so the effect on these measures is greater. The black/white ratio for maternal smoking is 0.70 when race as reported by the mother is used and 0.79 when race is coded according to the NCHS rules. The black/white ratio for late or no prenatal care changes from 2.42 to 1.94.

The bottom half of Table 2 shows the effects of the different classification methods on minority/white ratios. "Minority" is any race other than white. Under the self-reported classification method, a birth by a mother reporting her race as Hispanic would be counted in the minority category. The effects are similar to those observed for the black/white ratios, but more pronounced. For low birthweight, the minority/white ratio changes from 1.55 for self-reported race to 1.81 for race according to the NCHS coding rules. For maternal smoking during pregnancy and late or no prenatal care, the ratios change from 0.54 to 0.78 and from 2.47 to 1.87, respectively.

DISCUSSION

The new national model birth certificate, scheduled to be adopted by most states in the next few years, presents 15 discrete racial groups with a check-box next to each. The instructions are to "check one or more races to indicate what the mother considers herself to be." Only those checking "Other Asian," "Other Pacific Islander," or "Other" race are allowed to write in an entry for race. So, in the future, there will be much less opportunity to collect self-reported,

Table 2. Racial disparity measures using birth certificate data: race as self-reported by the mother compared to race as determined by NCHS coding rules

	Number of births	*Percent low birth-weight*	*Percent maternal smoking*	*Percent late or no prenatal care*
Black/white				
Self-reported				
Black	27,142	14.2	11.2	24.4
White	74,789	7.5	15.9	10.1
Black/white ratio	—	1.89	0.70	2.42
NCHS coding				
Black	27,579	14.2	11.1	24.3
White	85,423	7.4	14.1	12.5
Black/white ratio	—	1.92	0.79	1.94
Minority/white				
Self-reported				
Minority	43,160	11.6	8.6	24.9
White	74,789	7.5	15.9	10.1
Minority/white ratio	—	1.55	0.54	2.47
NCHS coding				
Minority	32,209	13.4	11.0	23.4
White	85,423	7.4	14.1	12.5
Minority/white ratio	—	1.81	0.78	1.87

SOURCE: 2002 North Carolina Live Births, State Center for Health Statistics, North Carolina Division of Public Health

NCHS = National Center for Health Statistics

open-ended designations of race on the birth certificates, as was done for this study using 2002 North Carolina data.

Hispanics in North Carolina are predominantly recent immigrants from Mexico and so their reporting of race on birth certificates, as described above, may be different from that in states with more established Hispanic populations like California, Texas, and Florida. Also, the pattern of reporting race by Hispanics in North Carolina may change over time as they become more acculturated and as they begin to represent a wider variety of national backgrounds. The findings from this study point to the need for comparable investigations in other geographic areas with greater racial and ethnic diversity.

This study shows that, given the opportunity to report their own race, North Carolinians describe their race using a wide variety of terms and concepts. In contrast, health statistics are usually reported using a few forced-choice racial categories defined by federal policy.

For statistical purposes, a limited number of standard racial categories are desirable for at least two reasons: (*1*) aggregating detailed racial groups makes the numbers of health events in each racial category larger, which promotes statistically reliable rates and measures, and (*2*) analyses done across diverse geographic areas and by different agencies will be more comparable. However, it should be recognized that a great deal of detail is lost in the process of aggregation. As stated by Moscou et al., "The commonly used racial/ethnic categories are at best approximations of broad and overlapping groups defined by society according to shifting criteria."[4] In addition, some of the standard racial categories may be so heterogeneous as to have little meaning. For example, the umbrella term "Asian" includes a wide variety of nationalities, cultures, ancestries, and backgrounds.[8]

The findings of this study are not surprising considering the frequent confusion of the terms "race" and "ethnicity." Many immigrants come from countries that do not use racial classifications and they find it hard to fit into one of the prescribed racial categories used in the United States. New immigrants may identify with their country of origin or tribal affiliation rather than a particular "race." These ethnic identities based on nation of origin may be more meaningful than a limited number of official racial groups.

Ultimately, "*people are who they say they are.* This requires a recognition that such definitions change over time, and that they may not correspond to any of the choices that researchers have fixed in advance."[9] Even the federal policy that establishes the fixed racial categories recognizes that "respect for individual dignity should guide the processes and methods for collecting data on race and ethnicity; ideally, respondent self-identification should be facilitated to the greatest extent possible."[2]

Due to the different methods of collecting racial data in various types of health records, comparisons of racial tabulations across various data sources is problematic. For example, race in the birth and BRFSS data are self-reported (although the birth data get reclassified into the standard federal categories), while race on the death and cancer case records is often determined by third-party observation.

As shown in Table 2, including almost all Hispanic births in the white racial category (based on NCHS coding rules) affects measures of racial health disparity. Most official releases of state and national birth and infant death data by race do not exclude persons of Hispanic ethnicity from the racial tabulations. In North Carolina, Hispanic live births have increased from less than 2% of total live births in 1990 to nearly 13% in 2002. The percentage increase in the Hispanic population in North Carolina during the 1990s was the largest in the nation. If this growth continues, measures of racial health disparity will be even more strongly affected in the future. Other researchers studying child health disparities should closely examine the methods by which information on race is originally collected in their data and determine how this information gets translated into the racial categories used for analysis.

The results from our study indicate that the NCHS rules for coding race should be reexamined. These rules imply a "one drop" criterion whereby someone who is any part Hawaiian is categorized as Hawaiian and someone who reports "octoroon" or "quadroon" (one-eighth or one-fourth black) as their race is categorized as black. As the ethnic and racial diversity of the United States continues to increase, these rules will become increasingly antiquated.

"Race" in the mind of an individual may be quite different from fixed statistical categories determined by governmental agencies. Some people do not understand the concept of race,[4] and others do not want to be categorized by race. A broadly defined racial group is at best a crude marker for particular health problems, and certainly not a risk factor or cause.[1,10] Racial discrimination, however, may account for part of the observed differences between racial groups in some health indicators.[11–14]

The authors would like to thank Trude Bennett, Jay Kaufman, and Emmanuel Ngui for reviewing this paper and providing helpful comments.

REFERENCES

1. Kaplan JB, Bennett T. Use of race and ethnicity in biomedical publication. JAMA 2003;289:2709-16. Erratum in: JAMA 2004;292:1022.
2. Office of Management and Budget (US). Revisions to the standards for the classification of federal data on race and ethnicity. Federal Register 1997;62:58781-58790. Also available from: URL: http://www.whitehouse.gov/omb/fedreg/ombdir15.html
3. Friedman DJ, Cohen BB, Averbach AR, Norton JM. Race/ethnicity and OMB Directive 15: implications for state public health practice. Am J Public Health 2000;90:1714-9.
4. Moscou S, Anderson MR, Kaplan JB, Valencia L. Validity of racial/ethnic classifications in medical records data: an exploratory study. Am J Public Health 2003;93:1084-6.
5. Hahn RA. Why race is differentially classified on U.S. birth and infant death certificates: an examination of two hypotheses. Epidemiology 1999;10:108-11.
6. Udry JR, Li RM, Hendrickson-Smith J. Health and behavior risks of adolescents with mixed-race identity. Am J Public Health 2003; 93:1865-70.
7. National Center for Health Statistics. Instruction manual, part 3a: classification and coding instructions for live birth records, 1999. Hyattsville (MD): Centers for Disease Control and Prevention, U.S. Department of Health and Human Services; 1998 December. Also available from: URL: http://www.cdc.gov/nchs/data/dvs/3amanual.pdf
8. Sheldon TA, Parker H. Race and ethnicity in health research. J Public Health Med 1992;14:104-10.
9. Kaufman JS. How inconsistencies in racial classification demystify

the race construct in public health statistics. Epidemiology 1999; 10:101-3.
10. Use of race and ethnicity in public health surveillance: summary of the CDC/ATSDR workshop. MMWR Morb Mortal Wkly Rep 1993;42(RR-10). Also available from: URL: http://www.cdc.gov/mmwr/preview/mmwrhtml/00021729.htm
11. Krieger N. Analyzing socioeconomic and racial/ethnic patterns in health and health care. Am J Public Health 1993;83:1086-7.
12. Freeman H, Payne R. Racial injustice in health care. N Engl J Med 2000;342:1045-7.
13. Thomas SB. The color line: race matters in the elimination of health disparities. Am J Public Health 2001;91:1046-48.
14. Mofidi M, Rozier RG, King RS. Problems with access to dental care for Medicaid-insured children: what caregivers think. Am J Public Health 2002;92:53-8.