

---

# Methods

## The Validity of Information on “Race” and “Hispanic Ethnicity” in California Birth Certificate Data

*Lisa Baumeister, Kristen Marchi, Michelle Pearl, Ronald Williams, and Paula Braveman*

---

**Objective.** To evaluate the validity of racial/ethnic information in California birth certificate data.

**Data Sources.** Computerized birth certificate data and postpartum interviews with California mothers.

**Study Design and Data Collection.** Birth certificates were matched with face-to-face structured postpartum interviews with 7,428 mothers to compare racial/ethnic information between the two data sources. Interviews were conducted in Spanish or English during delivery stays at 16 California hospitals, 1994–1995.

**Principal Findings.** The sensitivity of racial/ethnic classification in birth certificate data was very high (94 percent to 99 percent) for African Americans, Asians/Pacific Islanders, Europeans/Middle Easterners, and Latinas (Hispanics). For Native Americans, however, the sensitivity was only 54 percent. The positive predictive value of birth certificate classification of race/ethnicity was high for all racial/ethnic groups (96 percent to 97 percent).

**Conclusions.** Despite limited training of birth clerks, the maternal racial/ethnic information in California birth certificate data appears to be a valid measure of self-identified race and Hispanic ethnicity for groups other than Native Americans.

**Key Words.** Race, ethnicity, birth certificates, vital statistics

---

Racial/ethnic information in birth certificate data is used extensively in monitoring differences among population groups in infant mortality, low birth weight, and utilization of prenatal care. For example, racial/ethnic data in birth certificates will be used to monitor progress toward the goal of *The President's Initiative on Race* to eliminate racial/ethnic disparities in infant mortality (U.S. Department of Health and Human Services [HHS] 1998). The published literature, however, does not include studies that validate this information, and the quality of information on many items in birth certificates

has been questioned (Clark, Fu, and Burnett 1997; Emery et al. 1997; Green, Moore, Adams, et al. 1998; Piper, Mitchel, Snowden, et al. 1993; Woolbright and Harshbarger 1995).

In California birth certificates, "race" and "Hispanic ethnicity" are recorded as separate items. Although they receive no uniform training, birth clerks have access to the California Department of Health Service's birth registration handbook, which instructs them to obtain racial/ethnic information directly from the informant using a race identification worksheet (State of California, Department of Health Services 1994). The extent to which these guidelines are followed is unknown, and ongoing quality control or improvement measures do not exist. Furthermore, the ability of birth clerks to obtain accurate information from non-English-speaking persons is uncertain in that multilingual skills are not a requirement for the job. One concern is that birth clerks may at times base "race" and "Hispanic ethnicity" on their observations rather than on mothers' self-identification. Racial/ethnic information obtained by self-report may differ substantially from that obtained through observation, especially for some racial/ethnic subgroups. For example, one study showed that self-identified "race" corresponded well with observer-identified "race" for African Americans and whites, but that it showed significant discrepancies for Native Americans (Hahn, Truman, and Barker 1996).

This study was conducted to determine the validity of information in California birth certificate data on "race" and "Hispanic ethnicity." Because the "race" and "Hispanic ethnicity" assigned to infants in the United States since 1989 has been that of the mother, we focused on maternal "race" and "Hispanic ethnicity."

---

This project was supported by funds received from California Department of Health Services, Maternal and Child Health Branch. The interpretation of results and the conclusions stated in this article reflect the views of the authors and not necessarily those of the California Department of Health Services.

Lisa Baumeister, M.D., M.P.H. is a Family Physician at the Crownpoint Healthcare Facility in Crownpoint NM. At the time this study was conducted, Dr. Baumeister, Kristen Marchi, M.P.H., Michelle Pearl, Ph.D., and Paula Braveman, M.D., M.P.H. were with the Dept. of Family and Community Medicine, University of California, San Francisco. Ronald Williams, Ph.D. is with the Institute for Social, Behavioral, and Economic Research at the University of California, Santa Barbara. Address correspondence to Paula Braveman, M.D., M.P.H., Box 0900, University of California, San Francisco CA 94143-0900. This article, submitted to *Health Services Research* on October 30, 1998, was revised and accepted for publication on September 14, 1999.

## METHODS

### *Data Sources and Matching Methods*

Birth certificates were matched with data from face-to-face structured postpartum interviews to compare reported racial/ethnic information between the two data sources. The interviews were part of a larger statewide study of prenatal care use among mothers who gave birth in California hospitals from August 1994 through July 1995; survey methods have been described previously (Braveman, Pearl, Egerter, et al. 1998; Braveman, Egerter, and Marchi 1999). A total of 10,165 mothers were interviewed during their delivery stays at 19 randomly selected hospitals in California. We used 1991 statewide birth certificate data to select hospitals by stratified random sampling according to their delivery population characteristics (proportion of African American births, geographic region of the state, and prevalence of privately insured deliveries). Military hospitals, hospitals participating in California's Centers for Disease Control and Prevention-affiliated Pregnancy Risk Assessment Monitoring System, and hospitals with fewer than 600 deliveries in 1991 were excluded. These exclusions eliminated approximately 19 percent of all women older than 15 years who delivered statewide during the study period.

For 17 of the 19 hospitals in the larger survey sample, birth records were retrieved from the Automated Vital Statistics System (AVSS) (Williams, Marinko, and Shields 1983) and the California Automated Registration and Entry (CARE), two electronic birth certificate systems (Starr and Starr 1995). Two hospitals had to be excluded because they did not use AVSS or CARE and hence did not have complete electronic birth certificate data that were accessible for this study. Interview data were linked with electronic birth certificate files through use of a matching algorithm based on a combination of variables common to both data sources: delivery hospital, delivery date, mother's birth date, mother's first name, mother's maiden name, and child's last name. (A full description of the linkage algorithm is available on request.) The match rate of one of the survey hospitals was much lower than that of the other 16 hospitals, raising concerns that a significant number of false matches could result; therefore this hospital was excluded. The remaining 16 hospitals included in this validation study were diverse in ownership type. Semi-structured interviews were conducted with birth clerks at the 16 hospitals to obtain information on the procedures that they used for completing birth certificates.

A woman present on the postpartum ward in a selected hospital was eligible to participate if she had a live birth during this hospital stay, spoke English or Spanish, was at least 15 years old, was an "emancipated minor" or financially self-sufficient if she was 15–17 years old, and was not incarcerated then or during her pregnancy. She was ineligible if nursing staff believed that the interview would interfere with her care. Of the women on the postpartum wards during staffed hours in the 16 hospitals ( $n = 14,332$ ), 1,123 (7.8 percent) were ineligible: 458 (3.2 percent) because of language, 291 (2.0 percent) because of age/emancipated minor status, 280 (1.9 percent) because of nurse disapproval, and 94 (0.7 percent) for other reasons. Of the 13,209 women eligible to participate, 9,023 (68.3 percent) were approached for an interview, and 4,186 (31.7 percent) had been discharged before interviewers could approach them. A total of 7,712 women completed the interviews, representing 85.5 percent of the eligible women who were approached (58.4 percent of all eligible women). Of these women, 7,428 (96.8 percent) were successfully matched with corresponding birth certificates. Women with matched and non-matched records had similar sociodemographic characteristics (income, age, race/ethnicity, language, education, parity, and insurance).

As seen in Table 1, the validation study sample had a larger proportion of Latina women and fewer European/Middle Eastern women compared with all women who delivered in California in 1994. Other differences between the study sample and all women who gave birth in California that year were small (column 3 of Table 1).

### *Categorization of Race/Ethnicity*

Table 2 shows the exact terms, in both the birth certificate and the postpartum survey, that were used to define the racial/ethnic categories for comparison in the validation study. (A technical appendix [available on request] includes specifications for the recording of racial/ethnic information in birth certificates and the postpartum survey.)

The classification of race and Hispanic ethnicity in 1994/1995 birth certificate data is compatible with the 1990 U.S. Census, which is based on the Office of Management and Budget's (OMB) Directive 15, *Race and Ethnic Standards for Federal Statistics and Administrative Reporting* (U.S. Department of Commerce [DoC] 1978). These guidelines provide for four racial categories (American Indian or Alaskan Native, Asian or Pacific Islander, Black, and White) and two ethnicities (Hispanic or non-Hispanic). If more than one "race" is listed in the birth certificate, California state coding standards specify that the first race listed shall be reported. The criteria for classification of race

Table 1: Sociodemographic Characteristics of the Validation Study Sample (1994/1995) and All Women Delivering in California (1994)

<i>Sociodemographic Characteristics*</i>	<i>All Women in Sample (N = 7,428) Percent</i>	<i>All Women Delivering in California† (N = 559,974) Percent</i>	<i>Difference in Proportions Percent (95% CI)‡</i>
<b>Primary insurance during pregnancy</b>			
Uninsured	1	3	-2 (-2, -2)
Medi-Cal	48	48	0 (-1, 1)
Private	48	46	2 (1, 3)
Other	3	3	0 (-4, 4)
<b>Age</b>			
17 years or younger	3	5	-2 (-2, -2)
18-19 years	8	7	1 (0, 2)
20-34 years	77	75	2 (1, 3)
35 years or older	12	13	-1 (-2, 0)
<b>Level of education</b>			
8th grade or less	19	16	3 (2, 4)
Some high school	18	19	-1 (-2, 0)
Completed high school/GED	31	30	1 (0, 2)
Some college	20	19	1 (0, 2)
Completed college/more	13	16	-3 (-4, -2)
<b>Parity</b>			
1st live birth	38	39	-1 (-2, 0)
2nd-4th live birth	56	55	1 (0, 2)
5th or greater	6	6	0 (-1, 1)
<b>Racial/Ethnic category</b>			
African American	8	7	1 (0, 2)
Asian/Pacific Islander	8	10	-2 (-3, -1)
European/Middle Eastern	30	36	-6 (-7, -5)
Latina	54	45	9 (8, 10)
Native American	1	1	0 (0, 0)
Other §	1	1	0 (0, 0)

\*Data source of sample and all births is birth certificate data.

†Includes births only to California residents.

‡95% confidence interval.

§Including unspecified ethnicity.

given in Directive 15 are inconsistent in that, although geographic region of origin is used to some extent for all groups, cultural identification is an added criterion for Native American “race,” and skin color is an additional criterion for Black “race” (which is defined tautologically as “a person having origins in any of the black racial groups of Africa”).

Table 2: Composition of Racial/Ethnic Categories Used in the Validation Study, by Data Source

<i>Racial/Ethnic Categories in the Validation Study</i>	<i>Categories in Birth Certificate</i>	<i>Categories in Postpartum Survey</i>
African American	Black	African American, Black, Negro, or African
Asian/Pacific Islander	Japanese, Chinese, Filipino, Korean, Vietnamese, Cambodian, Laotian, Thai, Asian Indian, Hawaiian, Guamanian, Samoan, or other Pacific Islander	Japanese, Chinese, Filipino, Korean, Vietnamese, other Asian or Asian American, (including people from India), or Pacific Islander
European/Middle Eastern	White/Caucasian	White, Caucasian, European (including from Spain), Arabic or Middle Eastern (including Jewish)
Latina	Hispanic,* including Mexican/Mexican-American/Chicana, Puerto Rican, Cuban, Spanish or other Hispanic	Hispanic, Latino, Mexican, Chicano, Mexican-American or other Latin American
Native American	American Indian, Eskimo, or Aleut	American Indian, Native American Indian, Eskimo or Aleut
Other†	Other, specified	Other, specified

\*Mothers classified as of "Hispanic ethnicity" in the birth certificate data were considered to be Latina regardless of race in the validation study.

†In the birth certificate data, "Other, unspecified" was coded as missing; "Other, unspecified" was not a category in the postpartum survey.

We have concluded, as have many researchers, that the term "race" is not scientifically based (Freeman 1998; Williams 1997)—hence our use of quotation marks. Consequently, the terminology we used for the comparison groups in this validation study is based on geographic region of origin (African American, Asian/Pacific Islander, European/Middle Eastern, Latina, and Native American). We refer to these as racial/ethnic categories because they are not based solely on either "race" or "Hispanic ethnicity" as used in vital statistics.

We validated birth certificate information against postpartum survey information on primary race/ethnicity. Survey respondents were asked by interviewers: "What racial or ethnic group do you consider yourself?" Their responses were coded to one or more of eight racial/ethnic categories. The categories, which were based on a combination of geographic origin (of a

person or her ancestors) and identification with a socially recognized cultural group, were mutually exclusive and included a separate Latina category. Interviewers, who were bilingual in English and Spanish, were explicitly instructed to base a mother's race/ethnicity on her own self-identification rather than on the interviewer's observations. If respondents had difficulty specifying a racial/ethnic category, a list of racial/ethnic categories was read to them until they identified at least one category.

If more than one racial/ethnic group was selected, the respondent was subsequently asked to select one category with which she identified most; this was considered to be her primary racial/ethnic identification. If a mother was unable to choose a primary racial/ethnic category ( $n = 28$ ), or if no primary category was recorded ( $n = 29$ ), the mother's race/ethnicity was recorded as "missing" in the postpartum survey.

Hence, the classification algorithms for race/ethnicity were generally equivalent in the birth certificate and the postpartum survey, with two exceptions. First, as noted earlier, the birth certificate allowed for identification as "Hispanic" in combination with any "race," while in the survey, "Hispanic/Latina" was a racial/ethnic group mutually exclusive of others such as "European American/Middle Eastern" or "African American." In this validation study, all mothers classified as "Hispanic, Latina, Mexican, Chicano, Mexican American, or other Latin American" in the postpartum survey are compared with all of those classified as "Hispanic" regardless of "race" as recorded in the birth certificate. Discrepancies in the analysis attributable to the different coding schemes are probably slight, because 98 percent of the mothers classified as both "White/Caucasian" and "Hispanic" in the birth certificate self-identified as "Latina" in the postpartum survey. Second, in the birth certificate, mothers identified as "Spanish" or "Spaniard" (who comprised fewer than one percent of Hispanics in the birth certificates) were classified as Hispanic; however, in the postpartum survey, mothers with origins in Spain were classified as "White, Caucasian, European (including from Spain)."

## STATISTICAL ANALYSIS

Because the postpartum survey interviewers were specifically trained to elicit racial/ethnic information based on the mother's self-identification, the racial/ethnic classification in the survey was considered to be more accurate than that of the birth certificate. Using SAS version 6.12, we thus examined

the sensitivity and positive predictive value of racial/ethnic classification in birth certificate data for each of the five racial/ethnic groups as compared with racial/ethnic classification in the survey (SAS Institute 1996). The sensitivity was defined as the percentage of mothers in each racial/ethnic category in the survey who were classified within the same racial/ethnic category in birth certificate data. The positive predictive value was defined as the percentage of women classified as a specific race/ethnicity in birth certificate data who were identified as the same race/ethnicity in the postpartum survey.

## RESULTS

The sensitivity of racial/ethnic classification in birth certificate data was high ( $\geq 94$  percent) for the African American, Asian/Pacific Islander, European/Middle Eastern, and Latina categories (Table 3). For the Native American category, the sensitivity was only 54 percent. No Alaskan Natives were found in either the birth certificate or the postpartum survey. In 15 of the 16 hospitals, at least one mother self-identified as Native American in the postpartum survey (data not shown). The positive predictive value of birth certificate classification of race/ethnicity was high for all five racial/ethnic groups ( $\geq 96$  percent). Fewer than one percent of mothers were classified as "other" or "missing" in either the postpartum survey or the birth certificates.

The racial/ethnic category for 220 mothers (3 percent) differed in the birth certificate from that in the survey (Table 4). Of these discrepant cases, 19 percent initially identified with more than one racial/ethnic category in the postpartum survey, compared to only one percent of non-discrepant cases. Of the 113 mothers who initially identified with more than one racial/ethnic category in the postpartum survey, 37 percent were discrepant cases with regard to birth certificate race/ethnicity; of the 7,221 mothers who identified with only one racial/ethnic category in the postpartum survey, only 2 percent were discrepant cases. As Table 5 shows, those identified as Asian/Pacific Islander or African American in the birth certificate appeared slightly more likely to report more than one race/ethnicity in the postpartum survey compared to other racial/ethnic categories, with the exception of "other." Nearly half of the mothers identified as "other" in the birth certificate had initially specified more than one racial/ethnic category in the postpartum survey.

Interviews with a birth clerk at each of the 16 hospitals in this study revealed that, although most birth clerks used a worksheet other than that supplied by the California Department of Health Services, almost all birth

Table 3: Distribution of Racial/Ethnic Categories According to Birth Certificate and Survey Data, with Sensitivity and Positive Predictive Value and Corresponding 95% Confidence Intervals

Survey Racial/Ethnic Category	Birth Certificate Racial/Ethnic Category (N)							Total
	African American	Asian/Pacific	European/Middle	Latina*	Native American	Other	Missing	
African American	539	0	8	12	0	6	2	567
Asian/Pacific Islander	0	538	6	9	0	9	7	569
European/Middle Eastern	4	15	2171	90	1	6	21	2308
Latina§	3	1	25	3841	0	1	7	3878
Native American	1	2	12	6	25	0	0	46
Other	0	1	0	2	0	0	0	3
Missing**	8	4	16	22	0	6	1	57
Total	555	561	2238	3982	26	28	38	7428
Sensitivity* (95% CI)	95.1 (93.0, 96.6)	94.6 (92.4, 96.2)	94.1 (93.0, 99.3)	99.0 (98.6, 99.3)	54.3 (35.5, 71.9)			
Positive Predictive Value† (95% CI)	97.1 (95.3, 98.2)	95.9 (93.9, 97.3)	97.0 (96.2, 97.6)	96.4 (95.8, 96.9)	96.1 (80.6, 99.3)			

\*Sensitivity is defined as the percent of women identified as a specific race/ethnicity in the postpartum survey who were similarly identified in birth certificate data.

†Positive predictive value is defined as the percent of women identified as a specific race/ethnicity in birth certificate data who were similarly identified in the survey.

‡Mothers identified as of “Hispanic ethnicity” in birth certificates, regardless of “race.”

§Mothers who identified their primary race/ethnicity as Latina, regardless of additional racial/ethnic identification they indicated.

\*\*In the survey, mothers who could not select a primary identification among 2 or more racial/ethnic groups (N = 28) were classified as “missing” information on race/ethnicity along with those (N = 29) for whom there was no racial/ethnic information.

**Table 4: Discrepancy Between Data Sources by Number of Racial/Ethnic Categories Specified by Respondents in Postpartum Survey ( $N = 7,334$ )**

<i>Number of Racial/Ethnic Categories Specified by Mothers in Survey</i>	<i>Discrepant* Cases</i>		<i>Nondiscrepant Cases</i>		<i>Total</i>	
	<i>N</i>	<i>(Column %)</i>	<i>N</i>	<i>(Column %)</i>	<i>N</i>	<i>(Column %)</i>
More than one	42	(19.1)	71	(1.0)	113	(1.5)
Only one	178	(80.9)	7043	(99.0)	7221	(98.5)
Total	220	(100.0)	7114	(100.0)	7334	(100.0)

*Note:* Sample does not include 94 mothers whose racial/ethnic information was missing in at least one data source.

\*"Discrepant" refers to cases that were identified as one racial/ethnic group in the birth certificate and another racial/ethnic group in the postpartum survey.

**Table 5: Number of Racial/Ethnic Categories Specified by Mothers in the Postpartum Survey, by Race/Ethnicity as Recorded in Birth Certificates ( $N = 7,428$ )**

<i>Racial/Ethnic Category Used in Validation Study Based on Birth Certificate Data</i>	<i>Postpartum Survey</i>			
	<i>Mother Specified Greater than One Race/Ethnicity</i>		<i>Mother Specified only One Race/Ethnicity</i>	
	<i>N</i>	<i>(Row %)</i>	<i>N</i>	<i>(Row %)</i>
African American	18	(3.2)	537	(96.8)
Asian/Pacific Islander	23	(4.1)	538	(95.9)
European American	34	(1.5)	2204	(98.5)
Latina	66	(1.7)	3916	(98.3)
Native American	0	(0.0)	26	(100.0)
Other	12	(42.9)	16	(57.1)
Missing	2	(5.6)	36	(94.4)
Total	155	(2.1)	7273	(97.9)

clerks reported using self-identification rather than their own observations to determine the mother's "race" and "Hispanic ethnicity" (data not shown). At one hospital, however, a birth clerk indicated that "race" and "Hispanic ethnicity" of mothers was often based on the clerk's observations. (The population at this hospital was relatively homogeneous, with 92 percent of the mothers self-identifying as Latina in the postpartum survey.) There appeared to be no uniform method of recording race when the mother self-identified

with more than one "racial" group. Some clerks reported listing both "races" in the "other" category; others said they asked the mother to choose one of the two (or more) "races."

## DISCUSSION

The results of this study show that the racial/ethnic information in California birth certificate data generally reflects a valid measure of racial/ethnic self-identity when compared to information obtained from a face-to-face survey that relies on carefully trained and closely supervised bilingual interviewers. For most racial/ethnic categories, the sensitivity of classification in birth certificate data was quite high.

Some amount of the discrepancy between racial/ethnic identification in the postpartum survey and the birth certificate may have been due to the differences in categorization of race/ethnicity in the two data sources. As noted earlier, the birth certificate uses two questions to identify "race" and "Hispanic ethnicity" as separate items for each mother, whereas in the postpartum survey, mothers who identified as Hispanic/Latina would not also be assigned to a separate "racial" category. Despite this difference, there was excellent agreement between the two data sources for the category "Latina."

The issue of whether to elicit information on racial and ethnic identity as one or two questions relates to the conceptual validity of the term "race." Because use of the term can erroneously imply biological differences and lead to misinterpretation of data, some researchers have suggested alternatives to the terms "race" and "ethnicity" (U.S. Department of Health and Human Services [HHS] 1993). Some have proposed using the term "race/ethnicity" or simply "ethnicity" instead of using the separate terms "race" and "ethnicity," both of which are social constructs (Hahn and Stroup 1994; Cooper 1994; Warren et al. 1994). A one-question format, such as that used in the postpartum survey, would be consistent with the concept of "ethnicity" or "race/ethnicity" as a single category.

Although the sensitivity of racial/ethnic classification in birth certificate data was high for most racial/ethnic categories, it was relatively poor for Native Americans. This discrepancy may be attributable to a variety of factors. First, although most birth clerks interviewed stated that racial/ethnic identification was obtained from the mother, it is possible that the "race" and "Hispanic ethnicity" identification found on the birth certificate was sometimes assigned by observation rather than by asking the mothers. This

is consistent with our finding of a positive predictive value of 96 percent for Native Americans, versus a sensitivity for that race/ethnicity of only 54 percent. In fact, several studies have reported substantial misclassification of Native Americans in vital statistics when racial identification is determined by observation (Frost and Shy 1980; Frost, Tollestrup, Ross, et al. 1994; Hahn, Mulinare, and Teutsch 1992; Hahn, Truman, and Barker 1996; Watson, Bennett, Reed, et al. 1993).

Some amount of discrepancy in the classification of Native Americans may have resulted from the different categorization methods used in the two data sources. If some of the mothers classified as both Native American and Hispanic in the birth certificate data actually self-identify more as Native American than as Hispanic, then the validation analysis (which classified as Latina all mothers identified as Hispanic in the birth certificate regardless of "race") may have underestimated the percentage agreement between the two data sources. However, a reanalysis that classified as Native American all mothers who were identified as Native American in the birth certificate, regardless of the "Hispanic ethnicity" recorded in the birth certificate, revealed only a modest increase in sensitivity (from 54 percent to 61 percent).

It was not surprising to find that those who specified more than one racial/ethnic category were much more likely to be discrepant cases. Other studies have shown that identification with more than one racial/ethnic group is associated with less consistency in self-identification over time (Hahn, Truman, and Barker 1996). If the postpartum survey were regarded as the standard for self-identification of race/ethnicity, one-fifth of "inaccuracies" in the birth certificates might be accounted for by mothers identifying with more than one racial/ethnic category. Only 2 percent of mothers in the postpartum survey identified with more than one racial/ethnic category; nevertheless, as California becomes increasingly diverse, this phenomenon may become more common. In anticipation of this, the 2000 U.S. Census employs OMB's recently revised Directive 15, which allows for more than one self-identified "racial" group (MB 1997). To maintain consistency between the "race"/"Hispanic ethnicity" classification and that of the census, California birth certificates have been revised to allow for up to three choices for race.

This study of the validity of racial/ethnic information in California birth certificate data had several limitations. Three different sample exclusions may have led to an overestimate of validity. First, although the 19 hospitals selected for inclusion were chosen randomly within strata, three hospitals were excluded from the validation study. The exclusion of one of these three

hospitals based on a low rate of birth certificates matched to survey data may reflect poor quality of birth certificate or survey data overall in that hospital, including racial/ethnic information. Second, an estimated 3 percent of the target study sample were excluded from postpartum interviews because they did not speak English or Spanish; racial/ethnic information in the birth certificates of these mothers was likely to be of lower quality than the information elicited from English- or Spanish-speaking women because birth clerks were more likely to speak only English and/or Spanish. Third, of the women included in the validation study, 3.2 percent were excluded because their records could not be matched with data from birth certificates; non-matched cases may have had poorer-quality birth certificate or survey data in general, including racial/ethnic information. Our study sample included more Latina women and fewer European/Middle Eastern women than did the statewide maternity population in 1994. However, this difference is unlikely to have affected the validation study, as the numbers in each of these racial/ethnic subgroups were sufficient to allow adequate comparisons with birth certificate data.

The validity of racial/ethnic information in birth certificate data is of importance since it is used extensively by state and federal agencies and by researchers to monitor health and health care, and can be the basis for resource allocation to some health programs. Despite the limitations of this study, the results indicate that the maternal racial/ethnic information in California birth certificate data is likely to be a valid measure of self-identified race/ethnicity for groups other than Native Americans; the accuracy of identification of Native Americans, however, is questionable.

The generalizability of these results to other states may depend in part on the composition of racial/ethnic groups in a state's population. In addition, the generalizability depends on the extent to which other states collect racial/ethnic information similarly. In California, despite the lack of uniform training or quality control measures, our interviews with birth clerks indicated that they generally use self-identification by the mother to determine her "race" and "Hispanic ethnicity" as instructed by the local registrars and the birth registration handbook. These findings should, on the whole, be encouraging to other states. All states should clearly instruct birth clerks to use parents' self-identification, rather than observation, in recording information on race/ethnicity. Extensive training may not be required to achieve an acceptable quality of birth certificate information that identifies most racial/ethnic groups.

## ACKNOWLEDGMENTS

We are grateful to Susan Egerter, Ph.D. for her helpful comments and to Peter Chen for programming the matching algorithm.

## REFERENCES

- Braveman, P., M. Pearl, S. Egerter, K. Marchi, and R. Williams. 1998. "Validity of Insurance Information on California Birth Certificates." *American Journal of Public Health* 88 (5): 813-16.
- Braveman, P., S. Egerter, and K. Marchi. 1999. "The Prevalence of Low Income Among Childbearing Women in California: Implications for the Private and Public Sectors." *American Journal of Public Health* 89 (6): 868-74.
- Clark, K., C. Fu, and C. Burnett. 1997. "Accuracy of Birth Certificate Data Regarding the Amount, Timing, and Adequacy of Prenatal Care Using Prenatal Clinic Medical Records as Referents." *American Journal of Epidemiology* 154 (1): 68-71.
- Cooper, R. S. 1994. "A Case Study in the Use of Race and Ethnicity in Public Health Surveillance." *Public Health Reports* 109, no. 1 (January/February): 46-52.
- Emery, E. S., A. Eaton, J. K. Grether, and K. B. Nelson. 1997. "Assessment of Gestational Age Using Birth Certificate Data Compared with Medical Record Data." *Paediatric and Perinatal Epidemiology* 11 (3): 313-21.
- Freeman, H. P., President's Cancer Panel. 1998. "The Meaning of Race in Science: Considerations for Cancer Research." *Cancer* 82 (1): 219-225.
- Frost, F., and K. K. Shy. 1980. "Racial Differences Between Linked Birth and Infant Death Records in Washington State." *American Journal of Public Health* 70, no. 9 (September): 974-76.
- Frost, F., K. Tollestrup, A. Ross, E. Sabotta, and E. Kimball. 1994. "Correctness of Racial Coding of American Indians and Alaska Natives on the Washington State Death Certificate." *American Journal of Preventive Medicine* 10, no. 5 (September/October): 290-94.
- Green, D. C., J. M. Moore, M. M. Adams, C. J. Berg, L. S. Wilcox, and B. J. McCarthy. 1998. "Are We Underestimating Rates of Vaginal Birth After Previous Cesarean Birth? The Validity of Delivery Methods from Birth Certificates." *American Journal of Epidemiology* 147 (6): 581-86.
- Hahn, R. A., and D. F. Stroup. 1994. "Race and Ethnicity in Public Health Surveillance: Criteria for the Scientific Use of Social Categories." *Public Health Reports* 109, no. 1 (January/February): 7-15.
- Hahn, R. A., J. Mulinare, and S. M. Teutsch. 1992. "Inconsistencies in Coding of Race and Ethnicity Between Birth and Death in U.S. Infants: A New Look at Infant Mortality, 1983 through 1985." *Journal of the American Medical Association* 267, no. 2 (January): 259-63.
- Hahn, R. A., B. I. Truman, and N. D. Barker. 1996. "Identifying Ancestry: The Reliability of Ancestral Identification in the United States by Self, Proxy, Interviewer, and Funeral Director." *Epidemiology* 7, no. 1 (January): 75-80.

- Office of Management and Budget. 1997. *Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity*. Available from URL: <http://www.whitehouse.gov/WH/EOP/OMB/html/fedreg/Ombdir15.html>
- Piper, J. M., E. F. Mitchel, Jr., M. Snowden, C. Hall, M. Adams, and P. Taylor. 1993. "Validation of 1989 Tennessee Birth Certificates Using Maternal and Newborn Hospital Records." *American Journal of Epidemiology* 137 (7): 758–68.
- SAS Institute, Inc. 1996. *SAS: Proprietary Software Release 6.12*. SAS Institute, Inc., Cary, NC.
- Starr, P., and S. Starr. 1995. "Reinventing Vital Statistics: The Impact of Changes in Information Technology, Welfare Policy, and Health Care." *Public Health Reports* 110 (5): 534–44.
- State of California, Department of Health Services. 1994. *Handbook for Birth and Death Registration*. Sacramento: Office of Vital Records and Statistics.
- U.S. Department of Commerce. 1978. *Directive No. 15: Race and Ethnic Standards for Federal Statistics and Administrative Reporting*. Statistical Policy Handbook. Washington, DC: Office of Federal Statistical Policy and Standards.
- U.S. Department of Health and Human Services. 1998. "Eliminating Racial and Ethnic Disparities in Health." Available from URL: <http://raceandhealth.hhs.gov/>
- . 1993. "Use of Race and Ethnicity in Public Health Surveillance: Summary of the CDC/ATSDR Workshop." *Morbidity and Mortality Weekly Report* 42 (RR-10) (June): 1–17.
- Warren, R. C., R. A. Hahn, L. Bristow, and E. S. H. Yu. 1994. "The Use of Race and Ethnicity in Public Health Surveillance." [editorial] *Public Health Reports* 109, no. 1 (January/February): 4–6.
- Watson, C. C., T. Bennett, F. W. Reed, W. H. McBroom, and S. D. Helgeson. 1993. "Classification of American Indian Race on Birth and Infant Death Certificates: California and Montana." *Morbidity and Mortality Weekly Reports* 42, no. 12 (April): 220–22.
- Williams, D. R. 1997. "Race and Health: Basic Questions, Emerging Directions." *Annals of Epidemiology* 7, no. 5: 322–33.
- Williams, R. L., J. A. Marinko, and M. L. Shields. 1983. *An Automated Vital Statistics System. Proceedings of the 19th National Meeting of the Public Health Conference on Records and Statistics*. DHHS Pub. No. PHS 81-1214. Washington, DC: National Center for Health Statistics.
- Woolbright, L. A., D. S. Harshbarger. 1995. "The Revised Standard Certificate of Live Birth: Analysis of Medical Risk Factor Data from Birth Certificates in Alabama, 1988–92." *Public Health Reports* 110, no. 1 (January/February): 59–63.