Center for Social Development

GEORGE WARREN BROWN SCHOOL OF SOCIAL WORK

Validity of Infant Race/Ethnicity from Birth Certificates in the Context of U.S. Demographic Change

Lisa Reyes Mason
Center for Social Development

Yunju Nam
University at Buffalo, State University of New York

Youngmi Kim
Virginia Commonwealth University

Subsequent publication: Mason, L. R., Nam, Y., & Kim, Y. (2014). Validity of infant race/ethnicity from birth certificates in the context of U.S. demographic change. *Health Services Research, 49*(1), 249–267. doi:10.1111/1475-6773.12083

2012

CSD Working Papers
No. 12-40

Campus Box 1196 One Brookings Drive St. Louis, MO 63130-9906 • (314) 935.7433 • csd.wustl.edu

Acknowledgments

Support for this research comes from the Ford Foundation and Charles Stewart Mott Foundation. The authors thank former Oklahoma State Treasurer Scott Meacham; Tim Allen, Deputy Treasurer for Policy and Administration; and James Wilbanks, former Director of Revenue and Fiscal Policy who facilitated this research. We also appreciate contributions by staff at the Oklahoma State Department of Health, Kelly Baker and Sue Mallonee, and the survey research team at RTI International, Ellen Marks, Bryan Rhodes, and Jun Liu. We are grateful to Michael Sherraden and Margaret Clancy for designing the SEED OK study, and to Robert Zager for work with data management.

Validity of Infant Race/Ethnicity from Birth Certificates in the Context of U.S. Demographic Change

In this study, we examined consistency of infant race/ethnicity across two data sources (N=2,663) using measures of sensitivity and positive predictive value. First we created and compared conventional measures of infant race/ethnicity from 2007 Oklahoma birth certificates and SEED for Oklahoma Kids baseline survey data, classifying infants as White, African American, American Indian, Asian, or Hispanic. Then we created and tested alternative measures with a biracial classification, based on biological parentage from birth certificates or parent report of infant biracial identity in the survey. We find that, for conventional measures, sensitivity is highest for Whites and African Americans and lowest for Hispanics. Positive predictive value, meanwhile, is highest for Hispanics and African Americans, and lowest for American Indians. Alternative measures improve values for Whites, but yield mostly low values for minority and bi-racial groups.

Our main conclusions are that health disparities research should consider the source and validity of infant race/ethnicity data when creating sampling frames or designing studies that target infants by race/ethnicity. The common practice of assigning the maternal race/ethnicity as infant race/ethnicity should continue to be challenged.

Key words: *Vital statistics; racial/ethnic differences in health and health care; infant health; child and adolescent health; survey research and questionnaire design*

Introduction

Birth certificates are a valuable sampling frame for infant health studies, as they include almost everyone in the target population, one necessary condition for generating a representative sample. Birth certificates are especially valuable for health disparity studies since they include basic information on birth parents (e.g., race, ethnicity, and education), which facilitates stratified sampling and allows researchers to increase subsample sizes of under-represented groups (Schoendorf & Branum, 2006).

Since 1989, the National Center for Health Statistics (NCHS) has tabulated infant vital statistics primarily by race/ethnicity of the birth mother (National Center for Health Statistics, 2005). Subsequent studies have often followed this convention, using maternal race/ethnicity to describe infant race/ethnicity in health outcomes research (Ma, 2008). Although widespread, this practice is subject to critique for at least three reasons.

First, there is growing concern about the validity of race and ethnicity (i.e., Hispanic origin) data in birth certificates. Although racial/ethnic data are often consistent for White, Black, and Hispanic birth mothers when birth certificates are compared with other sources (e.g., hospital records) (Piper et al., 1993; Reichman & Hade, 2001; Vinikoor et al., 2010), similar data for American Indians are

often inconsistent and data for Asians have been little studied (Baumeister et al., 2000; Hahn & Stroup, 1994).

Second, exclusive use of maternal race/ethnicity ignores the potential role of paternal race/ethnicity in infant health outcomes. While the practice is partially explained by the fact that paternal information is often missing in birth certificates (National Center for Health Statistics, 2005), recent studies find that paternal characteristics, including missing race/ethnicity data, can have significant associations with infant outcomes (Getahun, Ananth, & Vintzileos, 2006; Ma, 2008).

Third, the practice of using maternal race/ethnicity for the infant may not adequately reflect changing U.S. demographics, particularly the increasing proportion of infants born to parents of different racial/ethnic backgrounds from each other. In studies that use the NCHS convention, a non-Hispanic infant reported on a birth certificate as having a White mother and African American father is categorized as White. The same infant, however, could subsequently be identified by a parent as White, African American, bi-racial, or some other race/ethnicity in health surveys or administrative records. The risk of mismatch between the current conventional approach using birth certificates and parent perception of infant race/ethnicity is likely to rise, as an increasing percentage of infants are born from interracial/interethnic relationships (Atkinson, 2001; Johnson & Lichter, 2010). Since parent perception of infant race/ethnicity may affect how the infant is raised through childhood and adolescence, it is likely to affect health and other developmental outcomes as well.

Existing studies suggest that researchers who use the conventional approach may produce a race/ethnicity variable inconsistent with parents' view of their children. Motivated primarily by U.S. demographic shifts, a few prior studies have examined how parents identify the race/ethnicity of their child when the parents are of different races. Using 1990 Census Bureau data, Qian (2004) provides one of the first analyses of this, focused on children under age five and born to couples wherein one parent is White and the other is not. Qian finds that African American-White couples are most likely to identify their child with a racial/ethnic minority label, while Asian American-White couples are least likely to do so. Previously, other studies had focused solely on Asian-White couples' reporting of their child's race/ethnicity, with similar results to Qian (2004) (Saenz et al., 1995; Xie & Goyette, 1997).

Brunsma (2005), meanwhile, uses data from a nationally representative sample of kindergarteners to find that while 10.4% are multiracial by birth based on parent race/ethnicity as reported in a survey, only 2.6% are designated by parents as multiracial when parents have the option of choosing a single race or "more than one race" to describe their child in the same survey. Interestingly, both Brunsma (2005) and Qian (2004) find that in many interracial couples, father's race may be more important than mother's in identifying the child, as fathers and children tend to have the same reported race/ethnicity in the datasets examined.

Most relevant to this study, Smith et al. (2010) compare child race/ethnicity based on birth certificates with another source of race/ethnicity for the child: administrative records for a major health plan. Using birth certificates as the "gold standard" or criterion data source, Smith et al. report that child race/ethnicity is misclassified in 23.1% to 33.6% of health plan records, primarily due to missing data, and that misclassification is low for Hispanic children but high for children of other racial/ethnic minority groups. While innovative as one of the first validity studies of child race/ethnicity data, limitations are that Smith et al. (2010) assume birth certificate data to be accurate

(despite limitations for certain groups, discussed above), exclusively use the convention of mother's race for child's race, and do not analyze data for infants or other children younger than kindergarten age.

The present study thus contributes to the literature by providing the first comparison of infant race/ethnicity based on birth certificates with parent report of infant race/ethnicity in a survey. Since the NCHS convention leads many researchers to classify infant race/ethnicity based on maternal race/ethnicity in birth certificates, it should be empirically tested whether this practice captures the dynamic nature of racial and ethnic distribution in the US. If infant health disparities are to be addressed and health services targeted to particular racial/ethnic groups in the context of U.S. demographic change, methodological aspects of measuring infant race/ethnicity must be considered and addressed.

Data and Sample

This study uses two data sources: 2007 Oklahoma birth certificates and the baseline survey for the SEED for Oklahoma Kids (SEED OK) experiment. Both data sets were collected for SEED OK, a longitudinal and statewide social experiment testing the policy concept of child development accounts (Zager et al., 2010).

SEED OK is unique in that birth certificates of all infants born in Oklahoma from April-June 2007 and August-October 2007 comprised the sampling frame. A sample of 7,297 infants was selected using stratified random sampling, with oversampling of non-Hispanic African American, non-Hispanic American Indian, and Hispanic infants (Marks, Rhodes, & Scheffler, 2008). Through oversampling, SEED OK aimed to obtain sizeable enough subsamples of the three minority groups to permit separate analyses by race/ethnicity. Birth mothers of the newborns were recruited and completed the SEED OK baseline survey between August-December 2007 and January-April 2008, respectively. If the birth mother did not live with the infant, the infant's main caretaker (e.g., father, grandparent, or sibling) was invited to participate in the study (Zager et al., 2010).

Of the 7,297 infants selected for SEED OK, 182 were ineligible (e.g., maternal or infant death). Of the remaining 7,115 infants, 2,704 had a birth mother or main caretaker who joined the study and completed the baseline survey, for an overall response rate of 38%. Study participation rates do not differ in a statistically significant way among four racial/ethnic groups: White (38.27%), African American (39.81%), American Indian (38.30%), and Hispanic (35.45%) (Nam et al., in press).

Among 2,704 SEED OK infants, this study includes 2,663 cases in which the birth mother joined the study and excludes 41 infants whose main caretakers were a father, grandparents, or siblings. This study thus provides a unique opportunity to compare infant race/ethnicity based on biological parentage as listed in birth certificates with birth mother's report of infant race/ethnicity within one year of the infant's birth.

Analyses and Results

Conventional measure of infant race/ethnicity

First, we generated and compared conventional measures of infant race/ethnicity from 2007 Oklahoma birth certificate data and SEED OK survey data. Of note, in 2007, Oklahoma was using

the 1989 U.S. Certificate of Live Birth, as were at least 22 other states at the time (Martin et al., 2011). On this version, ethnicity of each birth parent is collected as: "OF HISPANIC ORIGIN? (Specify No or Yes—If yes, specify Cuban, Mexican, Puerto Rican, etc.)." The race of each birth parent is also recorded as: "RACE – American Indian, Black, White, etc. *(Specify below).*" Although more than one race could be written on the form, the Oklahoma State Department of Health (OSDH) entered only one race in its 2007 vital statistics dataset, following NCHS protocol for states at the time. Also, while OSDH recommended that the birth certificate worksheet be completed by birth mothers, it is possible that some hospitals used other procedures to collect birth certificate data (e.g., observation, extraction from medical records), as sometimes occurs in other states (Northam, Polancich, & Restrepo, 2003).

For this study, which collapses race/ethnicity into one measure, a conventional infant race/ethnicity variable based on birth certificate data (CBC) was created as follows:

1. If the mother was reported as Hispanic (Mexican, Puerto Rican, Cuban, Central or South American, or other and unknown Hispanic) on the birth certificate, the infant was coded as Hispanic.

2. If the mother's Hispanic origin was not recorded on the birth certificate, and the father was reported as Hispanic, the infant was coded as Hispanic.

3. Non-Hispanic infants were assigned to White, African American, American Indian, or Asian (Chinese, Japanese, Hawaiian, Filipino, or other Asian) categories based on the mother's race.

CBC thus consists of five groups: non-Hispanic Whites (Whites), non-Hispanic African Americans (African Americans), non-Hispanic American Indians (American Indians), non-Hispanic Asians (Asians), and Hispanics. Of note, CBC in this study is the same as mother's race/ethnicity since no cases in the SEED OK sample are identified as Hispanic based on father's Hispanic origin.

Next, we created an equivalent conventional infant race/ethnicity variable using the survey data (CS). The SEED OK baseline survey asked two questions about infant race/ethnicity: (1) "Is [*insert infant's name*] of Hispanic or [Latina/Latino] origin?"; and (2) "What is [*insert infant's name*] race?" For the race question, interviewers were instructed to, "Select all that apply. Prompt by reading categories if necessary." Available categories were: "White, African-American, American Indian/Alaska Native, Asian/Pacific Islander/Hawaiian native, and Other (Specify:_____)." The SEED OK baseline survey thus generated information on infant race/ethnicity by asking mothers' perception of their infant's identity.

As shown above, the baseline survey allowed multiple racial identities by permitting mothers to choose more than one answer for race. Since OSDH chose only one value for each parent's race information from infant birth certificates and CBC is categorized into five single racial/ethnic groups, we generated CS using only the first answer to the race question. We assumed that mothers mentioned their infant's primary racial identity first when asked. CS was generated as follows:

1. If the mother identified her infant as Hispanic, the infant was coded as Hispanic.

2. If the mother reported her infant as non-Hispanic or did not answer the infant's Hispanic origin question and if the mother identified her infant as White, African-American, American Indian/Alaska Native, or Asian/Pacific Islander/Hawaiian native, the infant was coded as the corresponding non-Hispanic racial group (White, African American, etc.).

3. If the mother identified her infant as "Other" (19 cases), we looked at her specified answer. Based on the answer, we coded nine cases as Hispanic and six as White. We coded four cases as unknown because answers could not be classified (e.g., "American").

4. We coded the remaining two cases as unknown because mothers refused to answer or chose "Do Not Know" for the race question, and also chose "Do Not Know" or reported "No" to the Hispanic origin question.

CS has six categories: White, African American, American Indian, Asian, Hispanic, and unknown.

Table 1 reports the distributions of infant race/ethnicity as generated from birth certificate and survey data. The percentages of Whites, African Americans, and Asians are comparable between the two datasets. The proportion of American Indians, however, is higher in CBC than CS, while that of Hispanics shows the opposite pattern.

Table 1. Infant race/ethnicity from birth certificates vs. survey data: Conventional approach

	Birth Certificate (CBC)	Survey (CS)
Whites	1,215	1,179
	(45.63%)	(44.27%)
African Americans	463	472
	(17.39%)	(17.72%)
American Indians	512	401
	(19.23%)	(15.06%)
Asians	26	23
	(0.98%)	(0.86%)
Hispanics	447	582
	(16.79%)	(21.86%)
Unknown	0	6
	(0.00%)	(0.23%)
Total	2,663	2,663
	(100.00%)	(100.00%)

To see the relationship between the two conventional measures in detail, we run a cross-table of CBC and CS (Table 2). Based on the cross-table, we estimate the sensitivity score (or true positive rate) and positive predictive value (PPV), statistics frequently used in validation studies of birth certificate data (Baumeister et al., 2000; Piper et al., 1993; Reichman & Hade, 2001; Reichman & Schwartz-Soicher, 2007). The sensitivity score compares information in one data source against a criterion data source. Here, it is the percentage of infants in a given racial/ethnic group in the criterion data source (the SEED OK baseline survey), which are in the same group based on birth certificate data (e.g., the number of White infants based on birth certificates divided by the number of White infants as reported in the survey). In other words, this sensitivity score is the percentage of

infants whose race/ethnicity is correctly identified in birth certificate data, with the assumption that the birth mother's report in the survey accurately represents infant race/ethnicity. PPV is the percentage of infants in a given racial/ethnic group based on birth certificate data, which is in the same group in the survey data. In other words, PPV captures the ability of birth certificate data to correctly predict infant race/ethnicity as identified in the survey (Reichman & Hade, 2001; Reichman & Schwartz-Soicher, 2007).

As shown in Table 2, for the majority of cases, race/ethnicity is consistent between the two data sources, when single race/ethnic classifications are used (i.e., no consideration of bi-racial/multi-racial heritage). More specifically, sensitivity scores are highest for African Americans (89.83%) followed by Whites (86.68%), Asians (82.61%), American Indians (77.06%), and Hispanics (71.99%). These results suggest that birth certificates' ability to capture infant racial/ethnic identity that is coherent with mother's perception may be weaker among American Indians and Hispanics. PPV shows somewhat different patterns from sensitivity. PPV is highest among Hispanics (93.74%), followed by African Americans, Whites, Asians, and American Indians. It is of interest that Hispanics have the lowest sensitivity score but highest PPV among the five groups. These results along with a higher proportion of Hispanic infants in the survey data (21.86%) than based on birth certificates (16.79%), suggest that many mothers may identify their infant as Hispanic, even if mothers themselves are not Hispanic. It also should be noted that both sensitivity scores and PPV are very low among American Indian infants for the conventional measures in this study.

Table 2. Comparing two conventional infant race/ethnicity measures: Conventional birth certificate measure (CBC) vs. conventional survey measure (CS)

Conventional Survey Measure (CS)	Conventional Birth Certificate Measure (CBC)					
	White	African American	American Indian	Asian	Hispanic	Total
White	1022	20	116	6	15	1179
African American	21	424	23	1	3	472
American Indian	83	4	309	0	5	401
Asian	1	0	3	19	0	23
Hispanic	87	15	61	0	419	582
Unknown	1	0	0	0	5	6
Total	1,215	463	512	26	447	2663
Sensitivity	86.68	89.83	77.06	82.61	71.99	
Positive Predictive Value	84.12	91.58	60.35	73.08	93.74	

Looking for sources of inconsistency: The role of bi-racial heritage

In searching for sources of inconsistency between birth certificate and survey data, we examine the possible role of bi-racial/multi-racial heritage. We suspect that mothers' perception of infant race/ethnicity may be less stable if the infant has parents of two different racial/ethnic backgrounds. To test this hypothesis, we generated an indicator of infant bi-racial heritage using both parents' race/ethnicity data from birth certificates. First, we created each infant's race/ethnicity variable identical to CBC as described above. Second, we generated a racial/ethnic heritage variable using

mother's and father's race/ethnicity variables from birth certificates: single-racial heritage to those whose parents belong to the same racial/ethnic group, bi-racial heritage to those whose parents belong to different groups, and unknown to 78 cases whose father's information is missing.

Table 3 demonstrates that bi-racial heritage increases the risk of inconsistent information between birth certificate and survey data. In the top row, while over 90% of infants of single-racial heritage are reported to have the same racial/ethnic identity in the survey as they do based on birth certificates, only 62.46% of infants of bi-racial heritage have the same information across both data sources. The consistency rate is 88.46% among infants whose father's information is missing in birth certificates. The lower portion of Table 3 compares consistency rates between single- and bi-racial heritage groups within a racial/ethnic group. Analysis results indicate that the role of bi-racial heritage on consistency between the two data sources differs by race/ethnicity. Although the consistency rate is higher among infants with single- than bi-racial heritage across all five groups, African American and Hispanic infants with bi-racial heritage have much higher consistency rates (over 80%) than other bi-racial infants (around 50% or lower). These results suggest that African American and Hispanic mothers may identify their infants' racial/ethnic identity as the same as theirs even when fathers belong to different racial/ethnic groups than mothers, or when fathers' racial/ethnic identity is unknown.

Table 3. Percentages of consistent infant racial/ethnic identity across two data sources: Infants with single-, bi-racial, and unknown heritage

	Single-racial Heritage	Bi-racial Heritage	Unknown (Father's Information Missing)	Total
Whole Sample	92.69%	62.46%	88.46%	82.31%
Infant's Race and Ethnicity (CBC)				
Whites	92.07%	45.10%	80.00%	84.03%
African Americans	97.50%	81.97%%	100.00%	91.58%
American Indians	80.12%	51.02%	33.33%	60.35%
Asians	92.86%	50.00%	NA	73.08%
Hispanics	98.50%	87.58%	80.00%	93.74%

Table 4 presents the number and percentage of infants of single- and bi-racial heritage in each racial/ethnic group based on birth certificate data (i.e., using the CBC measure). Combined with the information in Table 3, Table 4 helps explain why some racial/ethnic groups' sensitivity and PPV scores are higher than other groups when comparing CBC versus CS. Whites' scores appear higher because the proportion of infants with bi-racial heritage is much lower than other groups (16.79%, Table 4). To the contrary, African Americans' scores are high despite a high proportion of infants with bi-racial heritage (39.52%, Table 4), because the consistency rate is high even among bi-racial infants (81.97%, Table 3). For American Indians, a very high rate of bi-racial infants (66.99%, Table 4) and low consistency rates among both single- and bi-racial heritage groups help explain the low sensitivity score and PPV. Among Hispanics, the percentage of bi-racial infants is lower than in other groups, except Whites. In addition, consistency rates are higher among Hispanics than in other racial and ethnic groups for both single- and bi-racial heritage. A low sensitivity score and high PPV, along with these findings, indicate that a high proportion of mothers may identify their infants as Hispanics even if only one biological parent is Hispanic. Additional analyses confirm this: 85% of Hispanic mothers whose infant's father belongs to a non-Hispanic group identified their infants as

Hispanic (141 out of 159) while 91% of non-Hispanic mothers whose infant's father is Hispanic identified their infants as Hispanic (83 out of 91).

Table 4. Single-, bi-racial, and unknown heritage by infant's race/ethnicity as conventionally measured

Infant's Race/Ethnicity (CBC)	Single Heritage	Bi-racial Heritage	Unknown (Father's Information Missing)
Whites	996 (81.98%)	204 (16.79%)	15 (1.23%)
African Americans	240 (51.84%)	183 (39.52%)	40 (8.64%)
American Indians	166 (32.42%)	343 (66.99%)	3 (0.59%)
Asians	14 (53.85%)	12 (46.15%)	0 (0.00%)
Hispanics	266 (59.51%)	161 (36.02%)	20 (4.47%)
Total	1,682 (63.16%)	903 (33.91%)	78 (2.93%)

Seeking alternative measures

Having identified bi-racial heritage as a possible source of inconsistency between our initial measures, CBC and CS, we develop and test two alternative measures of infant race/ethnicity that include a bi-racial category. An alternative measure based on birth certificates (ABC) was created using birth father's race and Hispanic origin as well as birth mother's:

1. If both parents were reported as Hispanic on the birth certificate, the infant was coded as Hispanic.

2. Non-Hispanic infants were coded as White, African American, American Indian, or Asian if both parents belonged to the same racial group. If a parent's Hispanic origin information was missing but race was valid, infants were coded based solely on parents' race.

3. If parents' belonged to different racial groups, infants were coded as bi-racial. Those with one Hispanic parent and one non-Hispanic parent were also coded as bi-racial.

4. If either parent's race information was missing, we coded them as unknown.

ABC has seven categories: White, African American, American Indian, Asian, Hispanic, bi-racial, and unknown.

The alternative measure using survey data (AS) was generated using infant's Hispanic origin and race information reported by their mother as in the conventional survey measure (CS). Unlike CS, however, which considered only the first answer to the race question, AS used multiple answers if mothers provided more than one racial identity for the infant and/or if mothers indicated the infant as of Hispanic origin and with a non-Hispanic racial identity. AS was generated as follows:

1. If the infant was identified as non-Hispanic or Hispanic origin information was missing (13 cases), and if the infant was classified as single racial heritage (the mother gave only

one answer for infant race) the infant was coded based on race: White, African American, American Indian, or Asian.

2. If the infant was identified as Hispanic, and the mother identified the infant's race in a way that can only be classified as Hispanic (e.g., mother's response to infant race question was "Other" and the specified response was upcoded to Hispanic), the infant was coded as Hispanic.

3. If the infant was identified as Hispanic, and the mother identified the infant's race as one or more of White, African American, American Indian, Asian, or "Other" where the specified response was not upcoded as Hispanic or unknown, the infant was coded as bi-racial.

4. If the infant was identified as non-Hispanic or Hispanic origin information was missing, and if the infant was reported to have more than one race (the mother gave two or more different answers for infant race), the infant was coded as bi-racial.

5. If the infant was identified as non-Hispanic or Hispanic origin information was missing, and if race was not reported, the infant was coded as unknown.

Like ABC, AS has seven categories: White, African American, American Indian, Asian, Hispanic, bi-racial, and unknown.

We compare the alternative measures ABC and AS to test whether they are more consistent to each other than the two conventional measures. We also estimate the sensitivity score and PPV. We exclude 83 cases with missing values in ABC or AS, reducing the analysis sample size to 2,580.

Comparison between the alternative measures shows a complicated relationship between biological heritage and mother's perception of infant race/ethnicity, as shown in Table 5. The majority of infants identified as single race/ethnicity in ABC are reported to have the same single racial heritage in AS. With the exception of American Indians, PPV is higher than 80% for all single-heritage groups. Among American Indians in ABC, almost 63% are identified as in the same racial/ethnic group as AS, while about 31% (51 out of 166) are identified in AS as bi-racial. In contrast, over half of infants classified as bi-racial in ABC are reported as a single heritage group in AS, while about 47% (425 out of 901) are identified as bi-racial. These results suggest that many mothers may identify their infants as belonging to a single racial/ethnic group even when parents belong to different racial/ethnic groups from each other. Similarly, sensitivity scores are low for all groups except Whites, indicating low probabilities of having accurate racial/ethnic identities in ABC if the information in AS is correct.

Table 5 demonstrates that, in this study, ABC does not improve infants' chance of having identical racial/ethnic identity as that reported in the survey when AS is used for comparison. In contrast with conventional measures (Table 2), sensitivity scores and PPVs are often much lower when the alternative measures are used. When alternative measures do have a better estimate, they show little improvement (e.g., sensitivity score for Whites improves from 87% to 90%). To the contrary, alternative measures show dramatic declines in sensitivity score or PPV in comparison to

conventional measures for several infant groups: sensitivity score drops to 61% from 90% for African Americans, to 37% from 77% for American Indians, and to 68% from 72% for Hispanics.

Table 5. Comparing two alternative infant race/ethnicity measures: Alternative birth certificate (ABC) vs. alternative survey (AS) measures

	Alternative Birth Certificate Measure (ABC)						
Alternative Survey Measure (AS)	White	African American	American Indian	Asian	Hispanic	Bi-racial	Total
White	857	0	7	1	2	85	952
African American	2	201	1	0	0	128	332
American Indian	15	0	104	0	0	161	280
Asian	1	0	2	13	0	3	19
Hispanic	6	0	1	0	228	99	334
Bi-racial	114	39	51	0	34	425	663
Total	995	240	166	14	264	901	2580
Sensitivity	90.02	60.54	37.14	68.42	68.26	64.10	
PPV	86.13	83.75	62.65	92.86	86.36	47.17	

Discussion

This study contributes to research on infant health disparities by examining the validity of infant race/ethnicity based on birth certificates, an important data source for sampling frames in disparities research. Taking advantage of a unique social experiment with both birth certificate and survey data for the same infants, this study adds to the critique of the conventional approach of creating infant race/ethnicity from maternal race/ethnicity in birth certificates (Ma, 2008). This study also contributes more broadly to discussions of measuring race/ethnicity in health disparities research given changing U.S. demographics (Mays et al., 2003; Ramirez et al., 2005; Zaslavsky, Ayanian, & Zaborski, 2012). Overall, this study finds that, for conventional measures of infant race/ethnicity, sensitivity is highest for Whites and African Americans and lowest for Hispanics. Positive predictive value, meanwhile, is highest for Hispanics and African Americans, and lowest for American Indians. Alternative measures improve values for Whites, but yield mostly low values for minority and bi-racial groups.

Taken together, results suggest that neither single-race classification from birth certificates, nor multiple-race classification based on biological parentage from birth certificates, are reliable measures of infant race/ethnicity for all racial/ethnic groups. Instead, a more nuanced understanding of identity among different groups must be sought. For example, a revised birth certificate could begin to collect parents' self-report of infant race/ethnicity, in addition to each birth parent's race/ethnicity.

For African American infants in this study, the finding that single-race conventional measures are more consistent than alternative measures allowing bi-racial classification is consistent with prior research on African American identity. Prior studies find persistence of the so-called *one-drop rule* wherein having any African American ancestry increases the likelihood that an individual will self-identify or be identified by others as African American (Bratter, 2007; Perez & Hirschman, 2009). At

the same time, however, findings are contrary to some research which finds that having an African American mother and White father increases the odds of a parent identifying the child as White (Brunsma, 2005).

For American Indians, consistency rates are also higher for conventional than alternative measures in this study, although sensitivity and PPV values in the conventional approach are overall low for this group. That neither approach in this study seems to consistently capture racial/ethnic identification of American Indian infants may reflect the somewhat contradictory and evolving findings on American Indian identity in prior research. Eschbach (1995), for example, finds in early research on this topic that children born from interracial unions wherein one parent is American Indian tend to be identified with the non-American Indian group. However, both Eschbach (1995) and Liebler (2010) suggest that geography may matter. As Oklahoma is home to a sizeable and more concentrated American Indian population than many other parts of the US, higher rates of infant identification as American Indian as opposed to multiracial or non-American Indian might be expected in the present study, potentially explaining the higher consistency rates found for the conventional versus alternative measures of infant race/ethnicity for this group.

For Hispanic infants, racial/ethnic identification is complicated by separation in vital statistics data of Hispanic origin from racial identity, as prescribed by the Office of Management and Budget's Standards for the Classification of Federal Data on Race and Ethnicity. As other studies have shown for non-infant populations, the common protocol of assigning Hispanics the racial category of White can artificially inflate or deflate health disparities outcome measures (Buescher, Gizlice, & Jones-Vessey, 2005). In this study, this complication also seems present, as there is moderate to high consistency when Hispanic infants of any racial origin in birth certificates are classified solely as Hispanic (conventional measure) versus when Hispanic infants identified with at least one other racial origin are classified as bi-racial (alternative measure). The alternative assignment of some Hispanic infants as bi-racial may not appropriately reflect that many Hispanics see their Hispanic identity as primary and central, over any other racial origins they may also have (Perez & Hirschman, 2009).

Finally, for bi-racial infants in this study, as classified in the alternative measures created, low overall consistency rates may be explained by the fact that each of the minority groups above may respond differently to the option of bi-racial classification. Thus, forcing infants into bi-racial categories based on biological parentage alone may not be an appropriate response to the problem of conventional reliance on maternal race/ethnicity for infant race/ethnicity either. While the recent tendency to describe infants in terms of birth parent couplings based on birth certificates (*White/White, White/Black, Black/Asian,* etc.; e.g., Ma 2008) certainly provides useful information, our results suggest that researchers should use this approach cautiously: a substantial proportion of mothers in SEED OK identify their infant as a single racial/ethnic group even when the infant has parents of different racial/ethnic backgrounds. It is necessary for future disparities research to consider and test alternate classifications when analyzing outcomes for racial/ethnic disparities, to see how different classification schemes may affect study results.

Despite the unique contributions and strengths of this study, study limitations must be considered. While the survey was chosen as the criterion source over birth certificates, we recognize that birth parent report of infant race/ethnicity may be influenced by numerous factors, including gender as the birth parent in this study was the mother, and that such identification may change over time.

Also, as noted previously, in 2007 Oklahoma had neither adopted the 2003 version of the live certificate of birth (which allows for multiple parent races to be marked), nor entered multiple race information optionally provided by parents into its dataset. If multiple race data for parents had been available, sensitivity and positive predictive values for some groups in this study may have been different.

The study points to important recommendations for health services policy, practice, and research. First, policymakers and practitioners should consider the source of infant race/ethnicity data when targeting services to particular infant groups, and research studies should clearly identify how infant race/ethnicity is being measured and used. Second, the convention of assigning birth mother's race/ethnicity to the infant should continue to be challenged, with movement toward adding direct collection of parent report of infant race/ethnicity when possible. Third, all remaining states should move toward adopting the 2003 version of the live certificate of birth, which allows birth parents at minimum the option of self-reporting their own multiple races.

References

Atkinson, J. O. (2001). Trends in births to parents of two different races in the United States: 1971-1995. *Ethnicity & Disease, 11*(2), 273-285.

Baumeister, L., Marchi, K., Pearl, M., Williams, R., & Braveman, P. (2000). The validity of information on "race" and "Hispanic ethnicity" in California birth certificate data. *Health Services Research, 35*(4), 869-883.

Bratter, J. (2007). Will "multiracial" survive to the next generation?: The racial classification of children of multiracial parents. *Social Forces,* 86(2), 821-849.

Brunsma, D. L. (2005). Interracial families and the racial identification of mixed-race children: Evidence from the Early Childhood Longitudinal Study. *Social Forces, 84*(2), 1131-1157.

Buescher, P. A., Gizlice, Z., & Jones-Vessey, K. A. (2005). Discrepancies between published data on racial classification and self-reported race: Evidence from the 2002 North Carolina live birth records. *Public Health Reports, 120*(4), 393-398.

Eschbach, K. (1995). The enduring and vanishing American Indian: American Indian population growth and intermarriage in 1990. *Ethnic and Racial Studies, 18,* 89-108.

Getahun, D., Ananth, C. V., & Vintzileos, A. M. (2006). Uteroplacental bleeding disorders during pregnancy: Do missing paternal characteristics influence risk? *BMC Pregnancy and Childbirth, 6*(2). doi:10.1186/1471-2393-6-2

Hahn, R. A., & Stroup, D. F. (1994). Race and ethnicity in public health surveillance: Criteria for the scientific use of social categories. *Public Health Reports, 109*(1), 7-15.

Johnson, K. M., & Lichter, D. T. (2010). Growing diversity among America's children and youth: Spatial and temporal dimensions. *Population and Development Review, 36*(1), 151-176.

Liebler, C. A. (2010). Homelands and indigenous identities in a multiracial era. *Social Science Research, 39*(4), 596-609.

Ma, S. (2008). Paternal race/ethnicity and birth outcomes. *American Journal of Public Health, 98*(12), 2285-2292.

Marks, E. L., Rhodes, B. B., & Scheffler, S. (2008). *SEED for Oklahoma Kids: Baseline analysis*. Research Triangle Park, NC: RTI International.

Martin, J. A., Hamilton, B. E., Ventura, S. J., Osterman, M. J. K., Kirmeyer, S., Mathews, T. J., & Wilson, E. C. (2011). Births: Final data for 2009. *National Vital Statistics Reports, 60*(1), 1-70.

Mays, V. M., Ponce, N. A., Washington, D. L., & Cochran, S. D. (2003). Classification of race and ethnicity: Implications for public health. *Annual Review of Public Health, 24*, 83-110.

Nam, Y., Mason, L. R., Kim, Y., Clancy, M., & Sherraden, M. (In press). Survey response in a statewide social experiment: Differences in being located and collaborating by race and Hispanic origin. *Social Work Research.*

National Center for Health Statistics. (2005). *Technical Appendix Form: Vital statistics of the United States: 2003: Natality.* Hyattsville, MD: National Center for Health Statistics, Centers for Disease Control and Prevention.

Northam, S., Polancich, S., & Restrepo, E. (2003). Birth certificate methods in five hospitals. *Public Health Nursing, 20*(4), 318-327.

Perez, A. D., & Hirschman, C. (2009). The changing racial and ethnic composition of the US population: Emerging American identities. *Population and Development Review, 35*(1), 1-51.

Piper, J. M., Mitchel, E. G., Snowden, M., Hall, C., Adams, M., & Taylor, P. (1993). Validation of 1989 Tennessee birth certificates using maternal and newborn hospital records. *American Journal of Epidemiology, 137*(7), 758-768.

Qian, Z. (2004). Options: Racial/ethnic identification of children of intermarried couples. *Social Science Quarterly, 85*(3), 747-766.

Ramirez, J., Ford, M. E., Stewart, A. L., & Teresi, J. A. (2005). Measurement issues in health disparities research. *Health Services Research, 40*(5, Part II), 1640-1657.

Reichman, N. E., & Hade, E. M. (2001). Validation of birth certificate data: A study of women in New Jersey's HealthStart program. *Annals of Epidemiology, 11*(3), 186-193.

Reichman, N. E., & Schwartz-Soicher, O. (2007). Accuracy of birth certificate data by risk factors and outcomes: Analysis of data from New Jersey. *American Journal of Obstetrics and Gynecology, 197*(1), 32.e1-32.e8.

Saenz, R., Hwang, S., Aguirre, B. E., & Anderson, R. N. (1995). Persistence and change in Asian identity among children of intermarried couples. *Sociological Perspectives, 38*(2), 175-194.

Schoendorf, K. C., & Branum, A. M. (2006). The use of United States vital statistics in perinatal and obstetric research. *American Journal of Obstetrics and Gynecology, 194*(4), 911-915.

Smith, N., Iyer, R. L., Langer-Gould, A., Getahun, D. T., Strickland, D., Jacobsen, S. J., Chen, W., Derose, S. F., & Koebnick, C. (2010). Health plan administrative records versus birth certificate records: Quality of race and ethnicity information in children. *BMC Health Services Research, 10*, 316.

Vinikoor, L. C., Messer, L. C., Laraia, B. A., & Kaufman, J. S. (2010). Reliability of variables on the North Carolina birth certificate: A comparison with directly queried values from a cohort study. *Paediatric and Perinatal Epidemiology, 24*(1), 102-112.

Xie, Y., & Goyette, K. (1997). The racial identification of biracial children with one Asian parent: Evidence from the 1990 Census. *Social Forces, 76*, 547-570.

Zager, R., Kim, Y., Nam, Y., Clancy, M., & Sherraden, M. (2010). *The SEED for Oklahoma Kids Experiment: Initial account opening and savings* (CSD Research Report 10-14). St. Louis, MO: Washington University, Center for Social Development.

Zaslavsky, A. M., Ayanian, J. Z., & Zaborski, L. B. (2012). The validity of race and ethnicity in enrollment data for Medicare beneficiaries. *Health Services Research, 47*(3, Part II), 1300-1321.