

# Hospital Policy and Practice Regarding the Collection of Data on Race, Ethnicity, and Birthplace

Scarlett Lin Gomez, PhD, Gem M. Le, MHS, Dee W. West, PhD, William A. Satariano, PhD, and Lilia O'Connor, MBA, CTR

The US population has experienced dramatic increases in racial/ethnic diversity over the last several decades, particularly with immigration from Latin America and Asia.<sup>1-5</sup> Research that uses race and ethnicity data provides an important foundation for designing programs to reduce health disparities.<sup>6</sup> Birthplace, which serves as an indicator of migrant status, can be used to further identify subpopulations to be targeted for disease control and to provide more specific information on disease patterns.

The Surveillance, Epidemiology, and End Results cancer registries obtain data on race, Hispanic ethnicity, and birthplace primarily from hospital records.<sup>7</sup> Because we have previously documented problems with the completeness and accuracy of these data in our registry,<sup>8-14</sup> we were interested in assessing the policies and practices at the hospital level

in the collection of patient data on race, ethnicity, and birthplace.

## METHODS

In late 1993, we identified 70 hospitals in the San Francisco Bay Area, Calif, which comprises the counties of San Francisco, Alameda, Contra Costa, San Mateo, Marin, Santa Clara, Monterey, Santa Cruz, and San Benito. These facilities included all of the inpatient hospitals in this region that had at least 1 cancer admission in 1993. Self-administered questionnaires were mailed to the hospital administrators, and 60 (86%) of these were returned in 1994. Repeated efforts were made to enhance participation rates. On occasion, questionnaires were completed by hospital administration staff, rather than by the administrators themselves.

Hospitals were surveyed on the frequency with which data on race, ethnicity (Spanish origin or surname), and birthplace were collected; the time of collection; the usual data sources; where information on race, ethnicity, and birthplace can be found in the hospital records; and whether the hospital has a pro-

cedure for recording the race or ethnicity for patients with mixed ancestry.

We used 1994 data from the Office of Statewide Health Planning and Development's Hospital Data Query System<sup>15</sup> to classify hospitals by ownership (private vs public/other), teaching status (teaching vs nonteaching/other), and size (fewer than 200 beds vs 200 beds or more).<sup>12</sup> These characteristics were selected for analysis because they were the only data available from Office of Statewide Health Planning and Development.

## RESULTS

Table 1 shows the frequency of collection of data on race, ethnicity, and birthplace. The vast majority (85%) of hospital administrators reported always collecting data on race, but 55% reported never collecting ethnicity data. Only 38% reported always collecting birthplace data. The practices of collecting these data items varied somewhat by hospital characteristics; for example, more large hospitals reported collecting data on race, ethnicity, and birthplace than did small hospitals. We

did not assess whether variability in collection practices was greater within these groups defined by the hospital characteristics than between them.

Table 2 shows the percentage of hospitals reporting "yes" to collection of these data items, by the time of collection and source and location of data. Almost all hospitals obtained data on race at the time of inpatient and outpatient admissions; however, only about one third reported collecting data on ethnicity, and one-half obtained birthplace data during these times. Approximately one-half of the hospitals also obtained data on race by observing a patient's physical appearance. Eighty percent of the hospitals used at least 3 sources to obtain race data. The patterns seen for birthplace and ethnicity were similar to those seen for race. Notably, only 7% and 13% of the hospitals reported recording ethnicity data on the basis of birthplace and surname, respectively. Twenty-five percent of the hospitals reported using at least 3 sources to obtain ethnicity or birthplace data.

In response to the question, "Does your hospital have a procedure for recording the race and/or ethnic information of a patient who has mixed ancestry?," only 12% responded "yes," 77% responded "no," and 10% responded "don't know."

## DISCUSSION

By surveying hospitals in the diverse San Francisco Bay Area, we found variations in the practices of collecting race, ethnicity, and birthplace information for cancer patients. The fact that not all hospitals collect this information all the time suggests that these hospitals do not have established policies regarding the collection of these data. Hospitals may not have incentives for collecting these data because it is not mandated by payers. Collecting such data also may be perceived as being too sensitive and possibly irrelevant. This position is most recently reflected in the Racial Privacy Initiative in California, which proposes to "end the governmental practice of classifying and tracking individuals by race, ethnicity, color or national origin" and to eliminate racial check-off boxes on governmental forms.<sup>16</sup>

**TABLE 1—Relative Distribution, %, of Collection of Data on Race, Ethnicity, and Birthplace, by Hospital Characteristic: Greater San Francisco Bay Area, 1994**

	Hospital Size		Ownership		Teaching Status		Total (N = 60)
	<200 beds (n = 23)	≥ 200 beds (n = 37)	Private (n = 31)	Public/Other (n = 29)	Teaching (n = 7)	Nonteaching/Other (n = 53)	
<b>Race</b>							
Always	82.6	86.5	90.3	79.1	85.7	85.0	85.0
Sometimes	4.4	13.5	9.7	10.3	0.0	11.3	10.0
Never	13.0	0.0	0.0	10.3	14.3	3.8	5.0
Don't know <sup>a</sup>	0.0	0.0	0.0	0.0	0.0	0.0	0.0
<b>Ethnicity</b>							
Always	17.4	18.9	9.7	27.6	0.0	20.8	18.3
Sometimes	4.4	24.3	9.7	24.1	14.3	17.0	16.7
Never	65.2	48.7	67.7	41.4	85.7	50.9	55.0
Don't know	13.0	8.1	12.9	6.9	0.0	11.3	10.0
<b>Birthplace</b>							
Always	34.8	40.5	38.7	37.9	57.1	35.9	38.3
Sometimes	17.4	27.0	32.3	13.8	28.6	22.6	23.3
Never	43.5	24.3	25.8	37.9	14.3	34.0	31.7
Don't know	4.4	8.1	3.2	10.3	0.0	7.6	6.7

Note. Percentages may not add to 100 because of rounding.

<sup>a</sup>Missing values were coded as unknown.

**TABLE 2—Percentage of Hospital Administrators Responding “Yes” to Collection of Data on Race, Ethnicity, and Birthplace, by Collection Time, Source, and Location (N = 60): Greater San Francisco Bay Area, 1994**

	Race	Ethnicity	Birthplace
Collection time			
Inpatient admission	96.5	36.0	54.7
Outpatient admission	90.9	27.7	51.0
Radiation therapy	36.1	5.9	18.9
Source			
Patient self	84.2	43.8	61.1
Patient's family/relative	77.2	37.5	53.7
Patient's friend	59.6	26.1	35.3
By noting patient's birthplace	15.6	7.1	NA
By observing patient's physical appearance	52.1	7.1	NA
By observing patient's language	32.7	13.3	8.3
By seeing or hearing patient's surname	20.8	13.3	4.0
From existing medical records	71.1	35.6	47.1
Location			
Face sheet	87.7	25.0	49.1
Physical examination report	61.8	33.3	23.1
Discharge summary	46.3	20.0	10.0
Nurse's notes	27.8	17.0	5.9

Note. NA = not applicable.

Our results of an underreporting of race/ethnicity and variability in practices on the collection of race/ethnicity data are similar to those found in the National Hospital Discharge Survey.<sup>17</sup> Furthermore, previous studies have reported that racial/ethnic and birthplace misclassification exists in population-based cancer registry data,<sup>8–11,13,14,18</sup> in surveillance data for AIDS patients,<sup>19</sup> and in Medicare admissions data.<sup>20</sup> All of these databases obtain their race/ethnicity data primarily from hospital records, thus supporting our conclusions of lack of systematic procedures in hospitals to collect patient information on race, ethnicity, and birthplace. Although we have studied this issue in the context of a regional Surveillance, Epidemiology, and End Results registry, the lack of a standardized approach to collecting these data has implications on other national data collection efforts that depend on data collected at the local hospital level.

The relatively low frequency of collection of ethnicity data may partly reflect inconsistencies in the definitions of *ethnicity* because hospitals and their patients may consider His-

panic ethnicity as its own racial category.<sup>21</sup> We suspect that this is the case because tumor registrars abstract information on Hispanic ethnicity from elsewhere in the medical records to ensure complete data on ethnicity in the registry. (Although the Office of Management and Budget recently revised their standards,<sup>22</sup> the 1977 standards were used at the time the study was conducted in 1993–1994.<sup>23</sup>) Nevertheless, a lack of a clear and uniform definition for ethnicity hampers consistent collection and limits the validity of research using these data.

The primary limitation of our study was that it was conducted in 1993 to 1994, and practices and policies of collecting race, ethnicity, and birthplace data may have changed over time. In particular, the issue of multiple race has received much focus in the past several years primarily because of the latest decennial census, which allowed respondents to identify with more than 1 race.<sup>5,22</sup> It would be of value and interest to evaluate whether hospitals have followed the practices of the census and have updated their procedures regarding multiple race. A more contemporary

evaluation of hospitals' policies and practices, across the United States, is warranted. Finally, because hospital administrators or their staff, who do not generally have direct patient contact, reported these practices, our results may reflect hospital policies but not actual practices. It would be useful in future studies to obtain reports from other personnel in the hospital.

Our results indicate that inconsistent hospital policies and practices contribute to the lack of completeness and accuracy of data on race, ethnicity, and birthplace in cancer registries. Given the importance of these data for public health and policy research, a concerted effort should be made to systematize the collection of these patient data across all facilities. Such an effort could begin with a national task force of hospital representatives to agree on a policy for the collection of these data. ■

#### About the Authors

Scarlett Lin Gomez, Gem M. Le, Dee W. West, and Lilia O'Connor are with the Northern California Cancer Center, Union City. Scarlett Lin Gomez is also with the Division of Epidemiology, Department of Health Research and Policy, School of Medicine, Stanford University, Stanford, Calif. William A. Satariano is with the Division of Epidemiology, School of Public Health, University of California, Berkeley.

Requests for reprints should be sent to Scarlett Lin Gomez, PhD, Northern California Cancer Center, 32960 Alvarado-Niles Rd, Suite 600, Union City, CA 94587 (e-mail: slin@nccc.org).

This brief was accepted February 14, 2003.

**Note.** The content of this publication does not necessarily reflect the views or policies of the US Department of Health and Human Services or the California Department of Health Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US government or state of California.

#### Contributors

S.L. Gomez supervised the data analysis and wrote the brief. G.M. Le analyzed the data and contributed to the writing of the brief. D.W. West, W.A. Satariano, and L. O'Connor planned the study, collected the data, and contributed to the writing of the brief.

#### Acknowledgments

Cancer data were collected by the Northern California Cancer Center under contract N01-CN-65107 with the National Cancer Institute, National Institutes of Health, and with support of the California Cancer Registry, a project of the Cancer Surveillance Section, California Department of Health Services, under subcontract 1000891 with the Public Health Institute.

The authors thank the reviewers for their contributions.

### Human Participant Protection

No protocol approval was needed for this study because the participants were hospital administrators.

### References

1. US Dept of Commerce. *We the American Foreign Born*. Washington, DC: Economics and Statistics Administration, Bureau of the Census; September 1993.
2. US Dept of Commerce. *We the American Hispanics*. Washington, DC: Economics and Statistics Administration, Bureau of the Census; September 1993.
3. US Dept of Commerce. *We the American Asians*. Washington, DC: Economics and Statistics Administration, Bureau of the Census; September 1993.
4. *Historical Census Statistics on the Foreign-Born Population of the United States: 1850–1990*. Washington, DC: Population Division, US Bureau of the Census; February 1999.
5. Bureau of the Census, American FactFinder. Census 2000 redistricting (Public Law 94-171) summary file. Available at: <http://factfinder.census.gov/home/en/pldata.html>. Accessed January 2002.
6. *Healthy People 2010: Understanding and Improving Health*. Washington, DC: US Dept of Health and Human Services; 2001. Also available at: <http://web.health.gov/healthypeople/document>. Accessed August 2002.
7. National Cancer Institute. *Surveillance, Epidemiology, and End Results*. Available at: <http://seer.cancer.gov>. Accessed August 2003.
8. Stewart SL, Swallen KC, Glaser SL, Horn-Ross PL, West DW. Adjustment of cancer incidence rates for ethnic misclassification. *Biometrics*. 1998;54:774–781.
9. Stewart SL, Swallen KC, Glaser SL, Horn-Ross PL, West DW. Comparison of methods for classifying Hispanic ethnicity in a population-based cancer registry. *Am J Epidemiol*. 1999;149:1063–1071.
10. Swallen KC, West DW, Stewart SL, Glaser SL, Horn-Ross PL. Predictors of misclassification of Hispanic ethnicity in a population-based cancer registry. *Ann Epidemiol*. 1997;7:200–206.
11. Swallen KC, Glaser SL, Stewart SL, West DW, Jenkins CN, McPhee SJ. Accuracy of racial classification of Vietnamese patients in a population-based cancer registry. *Ethn Dis*. 1998;8:218–227.
12. Lin SS, O'Malley CD, Lui SW. Factors associated with missing birthplace information in a population-based cancer registry. *Ethn Dis*. 2001;11:598–605.
13. Lin SS, Clarke CA, O'Malley CD, Le GM. Studying cancer incidence and outcomes in immigrants: methodologic concerns. *Am J Public Health*. 2002;92:1757–1759.
14. Lin SS, O'Malley CD, Clarke CA, Le GM. Birthplace and survival among Asian women diagnosed with breast cancer in cancer registry data: the impact of selection bias [letter]. *Int J Epidemiol*. 2002;12:511–513.
15. Office of Statewide Health Planning and Development (OSHDP) Web site. Available at: <http://www.oshpd.cahwnet.gov>. Accessed 2000.
16. Racial Privacy Initiative. Available at: <http://www.racialprivacy.org>. Accessed August 2002.
17. Kozak LJ. Underreporting of race in the National Hospital Discharge Survey. *Vital Health Stat*. 1995; No. 265:1–12.
18. Frost F, Taylor V, Fries E. Racial misclassification of Native Americans in a surveillance, epidemiology, and end results cancer registry. *J Natl Cancer Inst*. 1992;84:957–962.
19. Kelly JJ, Chu SY, Diaz T, Leary LS, Buehler JW. Race/ethnicity misclassification of persons reported with AIDS. The AIDS Mortality Project Group and the Supplement to HIV/AIDS Surveillance Project Group. *Ethn Health*. 1996;1:87–94.
20. Arday SL, Arday DR, Monroe S, Zhang J. HCFA's racial and ethnic data: current accuracy and recent improvements. *Health Care Finance Rev*. 2000;21:107–116.
21. McKenney NR, Bennett CE. Issues regarding data on race and ethnicity: the Census Bureau experience. *Public Health Rep*. 1994;109:16–25.
22. Office of Management and Budget. *Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity*. Washington, DC: Executive Office of the President, Office of Information and Regulatory Affairs; 1997.
23. Office of Management and Budget. *Directive No. 15: Race and Ethnic Standards for Federal Statistics and Administrative Reporting*. Washington, DC: Executive Office of the President, Office of Information and Regulatory Affairs; 1977.