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.1-5 Research that uses race and ethnicity data provides an important foundation for designing programs to reduce health disparities.6 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.7 Because we have previously documented problems with the completeness and accuracy of these data in our registry,8-34 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 procedure 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 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). 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.
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.17 Furthermore, previous studies have reported that racial/ethnic and birthplace misclassification exists in population-based cancer registry data,8–11,13,14,16 in surveillance data for AIDS patients,19 and in Medicare admissions data.20 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 Hispanic ethnicity as its own racial category.21 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,22 the 1977 standards were used at the time the study was conducted in 1993–1994.23) 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.5,22 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.

### 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

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

Note. NA = not applicable.

## 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