Published in final edited form as:

J Registry Manag. 2009 ; 36(4): 105–110.

Variability among Hospitals and Staff in Collection of Race, Ethnicity, Birthplace, and Socioeconomic Information in the Greater San Francisco Bay Area

Scarlett L. Gomez, PhD[a,b], **William Satariano, PhD**[c], **Gem M. Le, MHS**[a,d], **Patricia Weeks**[a,e], **Laura McClure, MSPH**[a], and **Dee W. West, PhD**[a,b]

[a]Northern California Cancer Center, Fremont, CA

[b]Division of Epidemiology, Department of Health Research and Policy, School of Medicine, Stanford, CA

[c]Division of Epidemiology, School of Public Health, University of California, Berkeley, CA

[d]Division of General Internal Medicine, University of California, San Francisco, CA

[e]Public Health Institute, Berkeley, CA

Abstract

Hospital data on race, ethnicity, birthplace, and socioeconomic status (SES) are important for identifying health disparities; however, little is known about the consistency across and within hospitals in the collection of these data. This study examined hospital practices and policies for the collection of these data and the variability across hospital staff and hospital characteristics. Surveys were mailed to selected hospital staff in all 59 hospitals in the San Francisco Bay Area, and completed questionnaires were received from 141 (of 367) staff from 41 hospitals. While most hospitals collect race/ethnicity (83% always collect) and birthplace (60% always or sometimes collect), few hospitals collect patient information on education (75% never collect) and income (55% never collect). There is vast variability in reported practices and policies across staff within hospitals, and variability across hospitals with regards to certain hospital characteristics. Nationally standardized policies, including standards for where, what, and when these data should be collected, are necessary for accurate and uniform data collection, and for effectively addressing health disparities.

Keywords

birthplace; ethnicity; hospital; race; socioeconomic status (SES)

Introduction

In the United States, hospital medical records data on patient sociodemographic characteristics serve as the foundation for identifying disparities in medical care and disease

---

Address correspondence to Scarlett Lin Gomez; Northern California Cancer Center, 2201 Walnut Avenue, Suite 300, Fremont, CA 94538. Telephone: (510) 608-5041, fax: (510) 608-5085, scarlett@nccc.org.

outcomes within and across medical systems. Medical records data are also the primary source of patient information for other data systems, including cancer registries, which are the basis for identifying disparities in cancer occurrence and survival.[1,2] Several studies, however, have shown that medical record data on race, ethnicity, and birthplace are subject to misclassification.[3–6] Furthermore, questions remain as to the consistency in collection of these data within and across hospitals.[3–6]

The aim of this study was to evaluate variation in practices among hospital staff and across hospitals for collecting information on patients' race, ethnicity, birthplace, and socioeconomic status (SES).

Methods

From October 2002 to October 2003, 59 hospitals within the San Francisco Bay Area were contacted to identify staff members who were thought to be knowledgeable regarding data collection procedures and/or policies. Staff included: 1) hospital administrators, 2) directors of admissions, 3) directors of medical records, 4) directors of outpatient services, 5) nurses, 6) admission clerks, and 7) chiefs of staff. Each identified staff member was mailed a questionnaire that included questions regarding whether information on race, Hispanic ethnicity, birthplace, education, and income/financial status (as markers of patient SES) were collected at the hospital, by whom, from whom, with what frequency, on which form(s), and with what level of specificity. We also asked about perceived barriers or concerns to collecting this information, suggestions for improvement of the completeness and quality of information, and hospital policies for collecting these data.

The distribution of reported practices and policies across hospitals were computed. Data were analyzed at the hospital level (N=41), which required consolidation of staff level information (N=141). This was done for hospitals with more than one staff responding, as follows: 1) if all responses agreed, that response was selected; 2) if disagreement, the most commonly reported response was selected; 3) if tie in responses, staff positions were prioritized based on who would have the most knowledge about actual practices: admissions clerk (highest priority), director of admission, director of medical records, hospital administrator, director of outpatient services, nurse, chief of staff, and other staff (lowest priority); and, 4) if responses differed across the same position, a response was randomly selected.

We also computed the percentage of disagreement for each questioned practice and policy among staff within the 33 hospitals with more than one response. The effects of staff position on selected practices, taking into account within-hospital clustering using generalized estimating equations with an exchangeable covariance matrix, were also computed. We also examined the one-response-per-hospital distributions by hospital characteristics, based on hospital utilization data collected by the California office of Statewide Health Planning and Development (OSHPD),[7] as well as data based on the cancer registry patient population. Hospital-level data on patient characteristics are often incomplete in OSHPD data.

Results

Response Rates

Forty-six of the 59 (78%) hospitals provided names of one or more hospital personnel, for a total of 367 hospital staff; all were mailed a questionnaire. A total of 141 staff (38.4% of 367) from 41 hospitals (69.5% of 59) returned completed questionnaires. The highest response rates were from directors of medical records (58.8%) and outpatient services (44.4%); the lowest were from hospital administrators (20.0%) and chiefs of staff or other physicians (28.6%) (data not shown).

Characteristics of the 41 “responding” hospitals were compared to the 18 “non-responding” hospitals (data not shown). Fifty-six percent of responding hospitals were large (>250 beds) compared to 33% of the non-responding hospitals (p-value=0.11). Seventy and 17% of the responding hospitals were publicly owned and HMOs, respectively, compared to 33 and 44% of the non-responding hospitals (p-value=0.01). There were no differences between responding and non-responding hospitals by teaching status (p=0.59), racial/ethnic (p=0.54 for % White, non-Hispamc), or birthplace (p=0.65 for % US-born) characteristics of their cancer patients.

Reported Practices and Policies

Race/Ethnicity—Based on one response selected for each hospital, information on race and ethnicity were always collected at 83% of hospitals, and nearly all reported that race and ethnicity was likely to be collected by an admissions clerk (Table 1). Ninety-four percent of hospitals collected Spanish origin as a separate ethnicity code, 62% record specific Asian/ Pacific Islander subgroups, 36% record specific Spanish/Latino origin, and only 22% reported having procedures for recording more than one race (data not shown).

Birthplace—Forty percent of hospitals reported never collecting information on patient birthplace (Table 1). Of those that did, 93% reported that admissions clerks were most likely to collect this information, and 63% recorded the specific country of birth.

Socioeconomic Status (SES): Only 25% of hospitals reported collecting education while 45% “always” or “sometimes” collected patients’ income or other financial information (Table 1). Education information was most often found in physical exam and nurses’ and doctors’ notes, while income/financial information was most often found in insurance forms (data not shown).

Policies: Staff from 95% of hospitals indicated that it was required to collect data on race and ethnicity, and 92% reported that there was a standardized form, which was reportedly widely used by personnel (Table 1). Forty-five percent of hospitals reported that they were required to collect birthplace; 40% had a form for birthplace entry; and 80% reported that staff used the form. Considerably fewer staff reported the existence of policies for collection of SES data.

Variability Across Hospital Staff

However, there was generally vast disagreement among staff within those hospitals with more than one response (N=33) (Table 2). The least amount of disagreement was 19% on who is most likely to collect information on race/ethnicity and birthplace. The highest disagreement was on whether the hospital had a standard form for collecting income/financial information (83% disagreement) and whether personnel use the form (82% disagreement), whether the hospital records specific API information (84% disagreement), and whether the hospital has procedures for recording more than one race (85% disagreement). Additionally, about 55% of hospitals disagreed on whether Spanish origin is asked as a separate ethnicity code, 78% disagreed on the recording of specific Spanish/Latino origin, and 77% disagreed on the recording of specific country of birth (data not shown).

Hospital administrators and chiefs of staff or other physicians were four times less likely than hospital directors to report collecting data on race and ethnicity (Table 3). Admission clerks were two times, albeit not statistically significant, more likely than directors to report collecting multiple races. The other personnel group was three times more likely than directors to report collecting data on birthplace.

Variability Across Hospital Characteristics

Table 4 shows the one-per-hospital responses to selected questions according to hospital characteristics. Larger hospitals (ie, those with more than 250 beds) were less likely than smaller hospitals to collect multiple race data, but were more likely to collect specific country of birth. Hospitals with more than 50% of its cancer patients being non-Hispanic white were less likely to record specific API information. Hospitals with more cancer registry unknown race were more likely to collect multiple race data, and those with more foreign-born patients were more likely to collect birthplace as well as specific country of birth. Hospitals with more registry unknown birthplace were less likely to collect birthplace.

Discussion

Our study, set in a region of great racial/ethnic and socioeconomic diversity, adds to a growing body of literature on hospital practices and policies in the collection of patient sociodemographic information.[5,8,9] To our knowledge, these are the first findings to document the extent of hospital practices in the collection of patient data on education and income/financial information, which are two variables typically used to assess SES. Our study also contributes new information about the extent of variability in practices and policies across staff within hospitals, and variability across hospitals with regard to certain hospital characteristics. Most of the variability by hospital characteristics was consistent, such as hospitals with fewer foreign-born patients and those with greater percentage of registry data on unknown birthplace being less likely to record birthplace. However, some of results on the variability, specifically smaller hospitals being more likely to record multiple race, but less likely to record specific country of birth, may directly be used to target these hospitals in the region with efforts to improve collection of these data items.

The large degree of discordance in reported practices and policies among staff may reflect differences in actual practices or perceptions of practices. Hospital staff in management positions, such as administrators and directors, may be more aware of policies or may have different perceptions about actual practices compared to staff who collect these data directly from patients, such as admissions clerks.

This study is limited by its small sample, which is not unusual for surveys of hospital staff.[8,10] The response rates did not yield enough responses in each hospital position, which limited a thorough assessment of variations in practices and policies across staff positions. In addition, the prevalence measures are not based on report by the same type of staff at all hospitals, limiting comparisons to prevalence measures reported in other studies. There is a possibility that respondents may be more aware of their hospital's practices and policies; if this were the case, our results may overestimate the degree of reporting, and underestimate the degree of internal disagreement.

Collectively, despite the recognized importance of these patient sociodemographic data, our findings reflect the clear lack of standardization and awareness of how these data are collected within and across hospitals. These findings point to the need to educate hospital personnel to the importance of collecting these data and suggest the types of hospitals that could be targeted to determine barriers in collecting information and how to address them. Moreover, our results point to a critical need for a concerted national effort to systematize data collection across all hospitals. While these survey data were collected up to 7 years ago, only very recently have some facilities and health organizations began to systematize collection of patient sociodemographic data. For example, several health systems in the area recently began implementing a policy asking patients at admission to complete a one-page sheet of questions on race, ethnicity, birthplace, SES, and language and interpreter needs. Some information, such as language, income, and education may change over time, and thus would be useful to collect on multiple admissions. More facilities need to implement such policies, and furthermore, all facilities need to ensure that these data will be available electronically to maximize the probability that they will be used. A report from the Commonwealth Fund[8] offers useful recommendations for systematizing the collection of patient demographic data across hospitals, and such strategies will be most effective if mandated and implemented at a national level. As such data are critical for monitoring disparities in medical care and disease outcomes, and to the extent that these hospital data feed into surveillance systems such as cancer registry data, systematized collection of patient demographic data is needed for effectively addressing health disparities.

Acknowledgments

The authors thank Sushma Balakrishna, David Purdie, Erin Eberle, Lyndsey Darrow, Cammie d'Entremont, Michael Mendoza, and Sarah Shema for their contributions. The design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, and approval of the manuscript was supported by the SEER Rapid Response Surveillance Study supplement to contract N01-CN-65107 with the National Cancer Institute, National Institutes of Health. The collection of cancer incidence data used in this study was supported by the California Department of Health Services as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885; the National Cancer Institute's Surveillance, Epidemiology and End Results Program under contract N01-PC-35136 awarded to the Northern California Cancer Center, and contract N02-PC-15105 awarded to the Public Health Institute; and the Centers for Disease Control and Prevention's National Program of Cancer Registries, under agreement #U55/CCR921930-02 awarded to the Public

Health Institute. The ideas and opinions expressed herein are those of the author(s) and endorsement by the State of California, Department of Health Services, the National Cancer Institute, and the Centers for Disease Control and Prevention or their contractors and subcontractors is not intended nor should be inferred.

References

1. Cockburn, M., Deapen, D. Cancer Incidence and Mortality in California: Trends by Race/Ethnicity, 1988–2001. Los Angeles, CA: Los Angeles Cancer Surveillance Program, University of Southern California; 2004.
2. Glaser SL, Clarke CA, Gomez SL, O'Malley CD, Purdie DM, West DW. Cancer surveillance research: a vital subdiscipline of cancer epidemiology. Cancer Causes Control. 2005; 16:1009–1019. [PubMed: 16184466]
3. Gomez SL, Glaser SL. Quality of cancer registry birthplace data for Hispanics. Cancer Causes Control. 2005; 16(6):713–723. [PubMed: 16049810]
4. Gomez SL, Kelsey JL, Glaser SL, Lee MM, Sidney S. Inconsistencies between self-reported ethnicity and ethnicity recorded in a health maintenance organization. Ann Epidemiol. 2005; 15(1): 71–79. [PubMed: 15571996]
5. Kressin NR, Chang BH, Hendricks A, Kazis LE. Agreement between administrative data and patients' self-reports of race/ethnicity. Am J Public Health. 2003; 93(10):1734–1739. [PubMed: 14534230]
6. 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(11):1063–1071. [PubMed: 10355383]
7. Office of Statewide Health Planning and Development (OSHPD) Website. Annual Utilization Report of Hospitals–2001. Available at: http://www.oshpd.cahwnet.gov/. Accessed May 2004
8. Hasnain-Wynia, R., Pierce, D., Pittman, MA. Who, Where, and How: the Current State of Race, Ethnicity, and Primary Language Data Collection in Hospitals. New York, NY: The Commonwealth Fund; May. 2004
9. Polednak AP. Collecting information on race, Hispanic ethnicity, and birthplace of cancer patients: policies and practices in Connecticut hospitals. Ethn Dis. 2005; 15(1):90–96. [PubMed: 15720054]
10. Kellerman SE, Herold J. Physician response to surveys. A review of the literature. Am J Prev Med. 2001; 20(1):61–67.

Table 1

Relative Distributions of Reported Practices and Policies Regarding Collection of Patient Data on Rice, Ethnicity, Birthplace, and Socioeconomic Status among Hospitals (N=41) in the San Francisco Bay Area, 2002–2003

Question (N reporting)	**Race and Ethnicity**	**Birthplace**	**Socioeconomic Status – Education**	**Socioeconomic Status – Income/Financial Situation**
Practices:	% distribution[1]	% distribution[1]	% distribution[1]	% distribution[1]
Does hospital collect: (N)	41	40	40	38
% Always	82.9	30.0	2.5	15.8
% Sometimes	14.6	30.0	22.5	29.0
% Never	2.4	40.0	75.0	55.3
Who is most likely to collect?[2] (N)	40	29	–[3]	–
% Registration/adm clerk	97.5	93.1		
% Other	2.5	6.9		
Most likely source? (N)	39	25	–	–
% Patient’s own verbal or written report	97.4	92.0		
% Another source	2.6	8.0		
Policies:				
Is hospital required to collect? (N)	39	38	38	37
% Yes	94.9	44.7	5.3	35.1
Does hospital have standardized form for collecting data? (N)	38	38	37	38
% Yes	92.1	39.5	8.1	31.6
Do personnel use form? (N)	34	20	21	22
% Yes	94.1	80.0	23.8	54.6

[1] % distribution excludes missing and “don t know” responses

[2] Question allowed one response

[3] Question not asked

Table 2

Percent Disagreement among Hospital Staff on Reported Practices and Policies Regarding Collection of Patient Data on Race, Ethnicity, Birthplace, and Socioeconomic Status among Hospitals (N=33) in the San Francisco Bay Area, 2002–2003

Question (N reporting)	**Race and Ethnicity**	**Birthplace**	**Socioeconomic Status – Education**	**Socioeconomic Status – Income/Financial Situation**
Practices:	Hospitals disagreeing (%)[1]	Hospitals disagreeing (%)[1]	Hospitals disagreeing (%)[1]	Hospitals disagreeing (%)[1]
Does hospital collect? (N)	33	33	33	33
	45.5	33.3	75.8	72.7
Who is most likely to collect? (N) (% Registration/admin. clerk vs. % Other)	32	16	–[2]	–
	18.8	18.8		
Most likely source? (N) (% Patient’s own verbal or written report vs. % Another source)	32	16	–	–
	34.4	62.5		
Policies:				
Is hospital required to collect? (N)	30	30	30	29
	40.0	70.0	63.3	75.9
Does hospital have standardized form for collecting data? (N)	29	29	29	29
	65.5	79.3	69.0	82.8
Do personnel use form? (N)	21	15	15	17
	42.9	73.3	73.3	82.4

[1] Among hospital staff, for hospitals with more than one staff member reporting; includes unknown but does not include missing responses

[2] Question not asked

Table 3

Odds Ratios (OR) and 95% Confidence Intervals (CI) of Staff Position on Selected Hospital Practices in the Collection of Data on Race/Ethnicity, Birthplace, and Socioeconomic Status in the San Francisco Bay Area, 2002–2003

Staff Position[1]	**Response = Yes (n)**	**Response = No (n)**	**OR**[2]	**95% CI**[2]
Hospital “always” collects race/ethnicity?[3]				
Directors[4]	34	5	1.00	–
Hospital admin. + chief of staff or other physician	10	6	0.24	0.07–0.89
Nurse	27	5	0.77	0.22–2.70
Admissions clerk	38	4	1.20	0.38–3.79
Other personnel[5]	8	3	0.34	0.08–1.53
Procedures exist for recording multiple race?[6]				
Directors[4]	6	27	1.00	–
Hospital admin. + chief of staff or other physician + other staff	1	14	0.69	0.07–6.57
Nurse	1	8	0.55	0.05–5.81
Admissions clerk	12	24	2.56	0.87–7.52
Hospital “always/sometimes” collects birthplace?				
Directors[4]	20	15	1.00	–
Hospital admin. + chief of staff or other physician	8	5	1.13	0.35–3.61
Nurse	17	11	1.45	0.77–2.74
Admissions clerk	20	17	1.32	0.70–2.48
Other personnel[5]	8	2	3.33	1.05–10.53

[1] Referent category = hospital administrator + chief of staff or other physician

[2] OR for positive response, adjusted for clustering effect within hospitals using exchangeable covariance matrix in generalized estimation equations (GEE)

[3] Yes = “Always”, No = “Sometimes”/”Never”

[4] Includes director of registrations/admissions, director of medical records, and director of outpatient services

[5] Other personnel = other personnel, not specified; medical affairs; privacy official; regulatory compliance; certified tumor registrar

[6] Yes = “Always”/”Sometimes”, No = “Never”

Table 4

Frequency Distributions of Selected Practices and Policies on the Collection of Patient Information on Race, Ethnicity, and Birthplace, by Hospital Characteristics (N=41 hospitals) in the San Francisco Bay Area, 2002–2003

Question (N reporting)	Hospital Size		Hospital Ownership		% White, non-Hispanic[1]		%Unknown Race[1]		% Foreign-born[1,2]		% Unknown Birthplace[1,2]	
	<250 beds (n=18)	>=250 beds (n=23)	Public (n=29)	Private/HMO (n=12)	<50 (n=10)	>=50 (n=31)	<=1 (n=27)	>1 (*n*=14)	<35 (n=21)	>=35 (*n* = 20)	<50 (n=17)	>=50 (*n* = 24)
Does hospital collect race and ethnicity? (N)	18	23	29	12	10	31	27	14	21	20	17	24
Always	83.3	82.6	79.3	91.7	100.0	77.4	88.9	71.4	85.7	80.0	82.4	83.3
Sometimes	16.7	13.0	17.2	8.3	0	19.4	7.4	28.6	9.5	20.0	17.7	12.5
Never	0	4.4	3.5	0	0	3.2	3.7	0	4.8	0	0	4.2
p-value[3]	0.65		0.60		0.26		0.16		0.42		0.64	
Record specific API[4] subgroup? (N)	18	19	27	10	10	27	24	13	19	18	15	22
Yes	50.0	73.7	66.7	50.0	90.0	51.9	58.3	69.2	52.6	72.2	73.3	54.6
p-value[3]	0.18		0.45		0.03		0.51		0.22		0.25	
Record specific Spanish/ Latino origin? (N)	18	18	26	10	10	26	24	12	19	17	14	22
Yes	27.8	44.4	30.8	50.0	50.0	30.8	25.0	58.3	36.8	35.3	42.9	31.8
p-value[3]	0.49		0.44		0.28		0.05		0.92		0.50	
Record specific AI/AN[5] tribe? (N)	18	18	26	10	10	26	24	12	19	17	14	22
Yes	33.3	33.3	30.8	40.0	50.0	26.9	29.2	41.7	31.6	35.3	35.7	31.8
p-value[3]	1.00		0.70		0.19		0.45		0.81		0.81	
Procedure for recording more than one race? (N)	18	18	26	10	10	26	23	13	19	17	14	22
Yes	38.9	5.6	15.4	40.0	10.0	26.9	8.7	46.2	26.3	17.7	21.4	22.7
p-value[3]	0.04		0.18		0.27		0.01		0.53		0.93	
Does hospital collect birthplace? (N)	18	22	28	12	10	30	26	14	20	20	17	23

Question (N reporting)	Hospital Size		Hospital Ownership		% White, non-Hispanic[1]		%Unknown Race[1]		% Foreign-born[1,2]		% Unknown Birthplace[1,2]	
	<250 beds (n=18)	>=250 beds (n=23)	Public (n=29)	Private/HMO (n=12)	<50 (n=10)	>=50 (n=31)	<=1 (n=27)	>1 (*n*=14)	<35 (n=21)	>=35 (*n* = 20)	<50 (n=17)	>=50 (*n* = 24)
Always	22.2	36.4	35.7	16.7	30.0	30.0	34.6	21.4	10.0	50.0	58.8	8.7
Sometimes	38.9	22.7	25.0	41.7	30.0	30.0	26.9	35.7	25.0	35.0	29.4	30.4
Never	38.9	40.9	39.3	41.7	40.0	40.0	38.5	42.9	65.0	15.0	11.8	60.9
p-value[3]		0.47		0.41		1.00		0.67		0.01		0.01
Record specific country of birth? (N)	13	14	20	7	8	19	18	9	9	18	14	13
Yes	38.5	85.7	70.0	42.9	62.5	63.2	72.2	44.4	33.3	77.8	71.4	53.9
p-value[3]		0.02		0.37		0.97		0.16		0.02		0.35

[1] Based on the cancer registry patient population diagnosed in 2002–2003

[2] Among Asians and Hispanics

[3] P-value from chi-square statistic

[4] API = Asian or Pacific Islander

[5] AI/AN = American Indian/Alaskan Native