

RACE/ETHNICITY IN MEDICAL CHARTS AND ADMINISTRATIVE DATABASES OF PATIENTS SERVED BY COMMUNITY HEALTH CENTERS

Neil Maizlish, PhD; Linda Herrera, MS

Objective: The objective of this study was to measure the agreement in classification of patients' race/ethnicity in the medical charts and the automated practice management systems (PMSs) of seven community health centers.

Setting: Community health centers are on the frontlines of providing primary care to the under-served and racial/ethnic minorities. Public and private investments in information technology and the increasing use of automated disease registries hold promise to improve care and reduce ethnic and racial disparities. However, data quality may limit the accuracy of race/ethnicity classification and the ability to measure the effect of population-based clinical quality improvements.

Design/Participants: In a cross-sectional study, a probability sample of 947 patients with encounters in 2002 was selected from 79,119 patients. Each PMS used a single data field with a pick list that combined ethnicity and race. Race/ethnicity on registration forms completed by patients was abstracted from medical charts. Race/ethnicity classifications were aggregated into seven major categories: Asian/Pacific Islander, Black/African-American, Native American, White, Hispanic/Latino, Other, Missing/Unknown.

Outcome Measures: The sensitivity, positive predictive value, and proportion of agreement were outcome measures of agreement between information in the medical chart and PMS.

Results: The overall proportion of agreement (PA) between the medical chart (reference) and PMS was 87%. The PA varied significantly by health center (95%–74%). Hispanic/Latino had the highest sensitivity (91%) and positive predictive value (95%) and White the lowest (84% and 80%, respectively).

Conclusions: In broad categories, correspondence of race/ethnicity classifications in medical charts and PMS was good, although health centers varied. A careful appraisal of data quality of race/ethnicity is warranted before administrative databases are used in clinical quality improvement programs or research to assess health disparities. (*Ethn Dis.* 2006; 16:483–487)

Key Words: Clinical Information Systems, Community Health Center, Ethnicity, Race

INTRODUCTION

In 2004, \approx 13.1 million uninsured, poor Americans received health care at 914 community health centers funded by the US Bureau of Primary Health Care.¹ This number is a substantial share of the 43.1 million residents of the United States without health insurance in 2004.² Compared to the population of the United States, community health center patients are disproportionately African American (23%) and Latino (36%). Health outcomes are often poorer in these minorities compared to other racial and ethnic groups.³ The elimination of ethnic, racial, and other disparities in health is a national goal⁴ that is embraced by the community health center movement. Evidence shows that the health care provided by community health centers helps to reduce and eliminate disparities in access to care.⁵ Standardized collection of race and ethnicity is vital to measuring potential health disparities, and community health centers routinely report race and ethnicity statistics as part of the Uniform Data Set.⁶

The Community Health Center Network (CHCN) is a partnership of seven community health centers in the San Francisco Bay Area. The CHCN supports the health centers' practice management systems (PMSs), managed care contracting and claims processing,

From the Community Health Center Network, Alameda, California.

Address correspondence and reprint requests to Neil Maizlish, PhD; Community Health Center Network; 1320 Harbor Bay Parkway, Suite 250; Alameda, CA 94502; 510-769-2291; 510-769-2209 (fax); neilm@chcn-eb.org

Compared to the population of the United States, community health center patients are disproportionately African American (23%) and Latino (36%).

and clinical quality improvement programs. The health centers are staffed by 85 physicians and 40 midlevel practitioners at 20 different sites. Staff provides primary care in more than 30 languages for \approx 90,000 patients who are largely low-income Latinos, Asians, and African Americans. The CHCN also has nearly 30,000 managed care enrollees who are eligible for health services under Medicaid (Medi-Cal), State Child Health Insurance Program (Healthy Families), and related public programs. In addition to primary care, managed care patients have insurance coverage for specialists, emergency departments, and hospital services.

Community Health Center Network (CHCN) has had an active clinical quality improvement program led by the medical directors at each community health center. Clinical quality is measured annually with data from computerized PMS and medical records review.⁷ Clinical quality performance measures have included the proportion of patients with diabetes who are monitored for glycemic control,⁸ the use of controller medications in patients with asthma, and percent of infants and children who have preventive medicine office visits. Clinical performance measures have been analyzed by age, sex,

race/ethnicity, insurance coverage, and other variables potentially related to health disparities.⁸ The objective of this study was to examine the accuracy of race/ethnicity classifications that are used to assess potential health disparities in clinical quality. This was part of a larger study to examine data quality of core data elements for CHCN's data warehouse, which includes patient demographics as well as clinical information.

METHODS

Practice Management System

Each health center had a computerized PMS that performed patient registration, appointment scheduling, reporting, and billing. In 2002, five different PMSs were used by the seven health centers in CHCN. Each PMS used a single data field with a list of options that combined the concepts of ethnicity and race.⁹ Lists averaged 22 categories per health center (range: 10 to 39). Registration staff could select only one category. Procedures for assigning race/ethnicity in the PMS varied within and between health centers. In general, registration staff made assignments that were entered into the PMS based on their directly observing patients, registration materials that indicated a patient's self-report of race/ethnicity, the patient's response to being read the race/ethnicity question on the registration form, or combinations.

Medical Chart

All but a few sites used paper forms to register new patients. The forms included patient name and residence and other contact information, income, number of family members, marital status, language, and race/ethnicity. These forms were available in several different languages. Forms differed within and between health centers, and several health centers used closed-ended race/ethnicity categories such as White, Latino, Asian, African American, and

Other. These forms were completed and signed by patients and then filed in the medical chart. Our abstraction protocol required that these patient registration forms be the primary source of information on race/ethnicity and they represent the data source for most patients in the study. If a patient registration form was not present, the medical chart was searched for other forms or documents completed or signed by the patient. The secondary data sources included health history forms and State of California forms for family planning services. For a small number of patients, the medical chart did not appear to have a document with an item on race/ethnicity completed by the patient. For these cases, we reviewed at least three other documents in the medical chart and selected the race/ethnicity most consistently mentioned. The sources of these documents included a photocopy of a state-issued driver license and clinical forms completed by providers at the health center, specialists, or hospitals.

Sample

The PMS at each health center was programmed to export electronic encounter data in a standard format for patient demographics, diagnoses, services, and encounter date. The demographics included the code for race/ethnicity. These computer files made up the sample frame for this study and accounted for 79,119 patients in 2002. Patients who received only limited services such as a laboratory test or who were served at special service sites for dental care or school-based clinics were excluded from the sample frame. This exclusion was done because the registration process was not comparable to full-service patients, and medical records were maintained physically apart. At each health center, a random sample of 135 patients who had one or more encounters from January 1, 2002, to December 31, 2002, was selected from the electronic encounter data. The

sample size at each health center was sufficient to measure a 10% prevalence of race/ethnicity misclassifications with a 95% confidence interval and a 5% margin of error.

Medical Chart Data Abstraction

A list of patients in the sample was provided to each health center, whose staff pulled medical charts. The race/ethnicity of the patient was abstracted on a laptop computer according to major categories that follow the federal format¹⁰ for race and ethnicity as single variable: Asian (including Pacific Islander), African-American/Black, Native American (American Indian/Alaskan Native), White, Hispanic/Latino, Other, and Unknown/Missing. The abstractor used the Bureau of Census¹¹ definition to make assignments to these major categories. A few instances of uncertainty, mostly involving Afghani and Middle Eastern ethnicity, were assigned to the White category. The abstractor was blinded to the race/ethnicity recorded in the PMS. To minimize disruptions to health center operations, medical chart abstraction was limited to five consecutive workdays at each health center.

Statistical Analyses

The race/ethnicity codes in the PMS were aggregated into the same seven categories (including Unknown/Missing) as those used in the medical chart abstraction, according to the standard census definition.¹¹ The agreement statistics for race/ethnicity coding in medical charts and the PMS were sensitivity, positive predictive value (PPV), and proportion of agreement. Sensitivity is a proportion (or percent) whose numerator is the number of patients of a given race/ethnicity as classified in the PMS and whose denominator is the total number of patients of the same race/ethnicity category using the medical chart. Positive predictive value is a proportion whose numerator is the number of

Table 1. Distribution of patients by race/ethnicity in medical charts and practice management system

Medical Chart	Practice Management System								
	Asian	African American/Black	Native American	White	Hispanic/Latino	Other	Missing	Total	(%)†
Race/Ethnicity									
Asian	193	1	1	4	4	30	6	239	(27.3)
African American/Black		103	2	2			9	116	(9.6)
Native American		1	35	3	1			40	(1.2)
White				94	6	5	7	112	(9.1)
Hispanic/Latino	2	2	1	10	358	6	13	392	(46.9)
Other		1				4	2	7	(0.7)
Missing	1	5	1	5	7	3	19	41	(5.3)
Total	196	113	40	118	376	48	56	947	(100.0)
(Row %)*†	(23.8)	(9.4)	(1.3)	(9.4)	(46.1)	(3.3)	(6.7)	(100.0)	
Proportion of agreement (%)†									
Total									(86.8)
Excluding Other and Missing									(93.7)

* Percentages difference at $P < .001$.

† Weighted for probability of selection.

patients of a given race/ethnicity as classified in the medical chart and whose denominator is the total number of patients of the same race/ethnicity category classified using the PMS. For the purpose of these analyses, the true classification or “gold standard” was assumed to be that in the medical chart. Sensitivity and PPV were calculated in two-by-two tables for single categories of race/ethnicity versus all other categories, and the proportion of agreement was the sum of the diagonal elements of a seven-by-seven table divided by the total of cases classified. The race/ethnicity categories most commonly mismatched (false positives, false negatives) were listed, including those for a known race/ethnicity in one data source but missing in the other. The overall proportion of agreement was calculated as the weighted average of the seven health center-specific proportion of agreements, using the number of patients at each health center as the weights.

RESULTS

The sample included 1076 patients. Of these, 947 had charts available on the day of data abstraction and were

eligible for the study. After a review of medical charts, 13 patients were excluded because they were ineligible (limited services patients not previously excluded) or did not have the correct medical chart pulled.

Compared to the distribution based on PMS data (Table 1), we found proportionately more Asian and Hispanic/Latino patients and fewer White, Other or Missing patients in the medical chart data ($P < .01$). With the seven broad categories, the overall proportion of agreement was 86.8% and varied significantly by health center (95%–74%) (Table 2). Hispanic/Latino had the highest sensitivity (91%) and PPV (95%) (Table 3). White had lower sensitivity (84%) and PPV (80%). Other race/ethnicity had a low sensitivity (57%) and the lowest PPV (8%). Among the most common false negative classifications, patients coded as Asian in the medical chart tended to be classified as Other in the PMS, and White, Hispanic/Latino, and African American in the medical chart were coded as Missing in the PMS. The most common false-positive coding in PMS data was assigning Hispanic/Latino or Missing to patients whose medical chart indicated Asian, Native American, White, Other, or Missing.

DISCUSSION

The distributions of race/ethnicity of patients from medical charts and the PMS are similar and are consistent with communities in Alameda County,¹² where most CHCN’s patients reside. For ≈13% of patients, information on race/ethnicity in the medical chart disagreed with the PMS. The correspondence of medical charts and the PMS was greatest for Hispanic/Latino and lowest for White (exclusive of Other or Missing). The results varied by health center. Larger health centers and those with consistent technical support services for their PMSs appeared to have

Table 2. Proportion of agreement in race/ethnicity in medical chart (gold standard) versus practice management system

Health Center	Matches <i>n</i>	Proportion of Agreement (%)
A	123	95*
B	118	92
C	113	90
D	136	89
E	108	80
F	108	76
G	100	74

* Proportions vary significantly by health center ($P < .05$).

Table 3. Sensitivity and positive predictive value (PPV) of race/ethnicity in medical chart (gold standard) versus practice management system

Race/Ethnicity	Matches <i>n</i>	Sensitivity %	PPV %	Most Frequent Misclassification	
				False Negative (<i>n</i>)	False Positive (<i>n</i>)
Asian/Pacific Islander	193	81*	98*	Other (30)	Hispanic/Latino (2)
African American/Black	103	89	91	Missing (9)	Missing (5)
Native American	35	88	88	White (3)	African American/Black (2)
White	94	84	80	Missing (7), Hispanic/Latino (6)	Hispanic/Latino (6)
Hispanic/Latino	358	91	95	Missing (13)	Missing (7), White (6)
Other	4	57	8	Missing (2)	Hispanic/Latino (6)
Missing	19	46	34	Hispanic/Latino (7)	Hispanic/Latino (13)

* *P* < .05.

a higher proportion of agreement. The overall sample size was large, and the subgroup analyses cannot be discounted because of sample variability. However, within a health center, the sample variability was larger, especially for small race/ethnicity categories.

Some forms in the registration process incorporated the six major categories used in the Uniform Data Set. However, the PMSs at the seven health centers had a larger number of detailed categories. The degree of misclassification in this study appears to be less than that reported by others.^{13,14} Moscou et al¹³ compared patient self-reports in telephone interviews with the registration database at two community health centers. Reclassifying their 22 self-reported categories into major groupings used in the Uniform Data Set, the proportion of agreement was 56%. A similar proportion of agreement (60%) was found comparing veterans' self-reports of race/ethnicity in a written survey with the patient database of the

federal Department of Veterans Affairs.¹⁴

Several possible explanations exist for the higher proportion of agreement in this study compared to those found in other studies. We relied mainly on written records in the medical chart that often had closed-ended responses and were generated at the time of patient registration. These records come closest to a self-report, but a small percentage of classifications was based on documents in which healthcare providers designated the patient's race/ethnicity. Also, registration staff helped some patients complete forms and may have entered the results into the PMS. Therefore, the information in the medical chart and that in the PMS may not have always been independent. Another possibility is that our health centers' process of patient registration results in more accurate data.

Missing race/ethnicity occurred in ≈5%–6% of patients. This percentage is much lower than other studies and surveillance data.⁹ "Missing" was included as a category in statistical analyses. Exclusion of "Missing" often improves the proportion of agreement,¹⁴ so our results may be conservative. Indeed, when we repeated the analyses excluding "Missing," the overall proportion of agreement increased by ≈7% to 93.7%. "Other" was a notably small percentage (<1%) in the medical charts and PMS (5%). Interestingly, 9% of Alameda County residents character-

ized themselves as "other race" in the 2000 Census, and 6% characterized themselves as multiple races. The single-variable format used at our health centers did not permit us to examine how forced single-choice categories influence the results for patients who designate multiple races, but researchers report that smaller single-race groups (Native American, Asian/Pacific Islander) can be affected more than the Black or White groups by multiple-race classifications.¹⁵

Despite calls for consistency,⁹ federal agencies and researchers¹⁶ use at least three different formats for race and ethnicity based on the federal standard:¹⁰ single variable (eg, Uniform Data Set), two separate variables, and three variables for race and a separate variable for Hispanic/Latino ethnicity (eg, 2000 Census). The single variable classification has a well-known impact on Hispanic/Latino and White/African American categories that vary by US region (Southwest, New York metropolitan area, southern Florida) based on immigration patterns of persons of African or European descent from the Caribbean Basin, Mexico, and other countries in Latin America. Given that some community health centers are adopting the multiple-race classifications from the 2000 Census, guidance is warranted on how this may be adapted to existing PMSs. Technical assistance could also include providing rules on aggregating detailed race/eth-

For ≈13% of patients, information on race/ethnicity in the medical chart disagreed with the practice management system.

nicity classifications for vendors of PMSs and information technology staff of health centers and education for executive leadership of health centers.

We recognize the practical difficulties in collecting accurate demographic and clinical information that forms the basis of a data warehouse. Like Moscou et al,¹³ we have anecdotal evidence of patients who do not welcome questions on race/ethnicity, even in anonymous patient-satisfaction surveys. Registration staff anticipating hostile responses may simply use their own judgment in selecting race/ethnicity in the PMS without asking patients.

Medical charts were not available for 12% ($n=129$) of the sample on the days data were abstracted. Medical charts were typically unavailable because the chart was needed for a patient office visit or was being shelved. Some charts may have been unavailable because they had been misfiled. None of these reasons appear to be related to race or ethnicity. Thus, missing observations in the sample were unlikely to be sufficiently large or biased to alter the findings of this study. In order to minimize disruptions to the operations of the health centers, we did not make repeated requests or return visits for medical charts that were not available during the week data were abstracted at each health center.

Few organizations take a systematic approach to data quality improvement and consciously consider how data quality affects organizational culture (and the converse). Measurement of data quality and feedback to the collectors, analysts, and users of information are key elements of a data quality program that includes standardized procedures for collecting data, staff training, supervision, setting standards for error tolerance, corrective action,

positive reinforcement, and remonitoring. The ability to leverage administrative data for measurement of clinical quality and reduce health disparities depends on the quality of data. A careful appraisal of data quality is warranted before administrative databases are used for research or clinical quality improvement.

ACKNOWLEDGMENTS

Staff at the health centers are gratefully acknowledged for their assistance with medical charts. This project was supported by grant number 1-R21-HS013543 from the Agency for Healthcare Research and Quality. The Bureau of Primary Health Care provided additional support.

REFERENCES

1. Bureau of Primary Health Care. Uniform Data System National Summary for 2004 (914 grantees)—UDS national trend data. Health Resources and Services Administration. Available at: bphc.hrsa.gov. Accessed on 8/17/05.
2. Cohen RA, Coriaty-Nelson Z, Ni H. Health insurance coverage: estimates from the National Health Interview Survey, January–September 2003. Centers for Disease Control and Prevention. Available at: www.cdc.gov. Accessed on 3/20/04.
3. Agency for Healthcare Research and Quality. *2004 National Health Disparities Report*. Rockville, Md: AHRQ; 2004. AHRQ Publication No. 05-0014.
4. US Department of Health and Human Services. *Healthy People 2010: Understanding and Improving Health*. 2nd ed. Washington, DC: US Govt. Printing Office; 2000.
5. Politizer RM, Yoon J, Shi L, Hughes RG, Regan J, Gaston MH. Inequality in America: the contribution of health centers in reducing and eliminating disparities in access to care. *Med Care Res Rev*. 2001;58:234–248.
6. Office of Policy Evaluation and Data. Data from the Uniform Data System. Bureau of Primary Health Care. Available at: bphc.hrsa.gov. Accessed on: 6/24/04.
7. National Committee for Quality Assurance. *HEDIS 2004 (Health Plan Employer Data and Information Set). Technical Specifications*. Vol 2. Washington, DC: National Committee for Quality Assurance; 2004.
8. Maizlish NA, Shaw B, Hendry K. Glycemic control in diabetic patients served by community health centers. *Am J Med Qual*. 2004; 19:172–179.
9. Centers for Disease Control and Prevention. Use of race and ethnicity in public health surveillance. Summary of the CDC/ASTDR workshop. *Morb Mortal Wkly Rep*. 1993; 42(RR-10):1–17.
10. Office of Management and Budget. Standards for maintaining, collecting, and presenting Federal data on race and ethnicity. Available at: www.whitehouse.gov. Accessed on: 6/24/04.
11. Bureau of the Census. Race, definitions used in the US Census, 2000. Available at: quickfacts.census.gov.
12. US Bureau of Census. State and county quickfacts. Data derived from Population Estimates, 2000 Census of Population and Housing. Available at: quickfacts.census.gov. Accessed on: 6/24/04.
13. Moscou S, Anderson MR, Kaplan JB, Valencia L. Validity of racial/ethnic classifications in medical records data: an exploratory study. *Am J Public Health*. 2003;93: 1084–1086.
14. Kressin NR, Chang B-H, Hendricks A, Kazis LE. Agreement between administrative data and patients' self-reports of race/ethnicity. *Am J Public Health*. 2003;93:1734–1739.
15. Sondik EJ, Lucas JW, Madans JH, Smith SS. Race/ethnicity and the 2000 Census: implications for public health. *Am J Public Health*. 2000;90:1709–1713.
16. Comstock RD, Castillo EM, Lindsay SP. Four-year review of the use of race and ethnicity in epidemiologic and public health research. *Am J Epidemiol*. 2004;159: 611–619.

AUTHOR CONTRIBUTIONS

Design concept of study: Maizlish, Herrera
Acquisition of data: Maizlish, Herrera
Data analysis interpretation: Maizlish
Manuscript draft: Maizlish, Herrera
Statistical expertise: Maizlish
Acquisition of funding: Maizlish
Administrative, technical, or material assistance: Maizlish, Herrera
Supervision: Maizlish