Validity of Racial/Ethnic Classifications in Medical Records Data: An Exploratory Study

Susan Moscou, FNP, MPH, Matthew R. Anderson, MD, MS, Judith B. Kaplan, MS, and Lisa Valencia, CHES

Information on "race" and "ethnicity" is routinely collected in medical settings and used in research. Yet little is known about how patients are assigned to categories[1–6] or the validity of the racial/ethnic data in medical charts and clinical databases.[7] We conducted an exploratory study with 2 goals: to learn how racial/ethnic data were collected at 2 primary care clinics affiliated with Montefiore Medical Center, a large urban hospital in New York City, and to assess the accuracy of these data, with self-report as the standard.

METHODS

We first interviewed administrators and registration clerks about procedures used at the 2 clinics to assign race/ethnicity. We then conducted telephone interviews with a sample of patients from each of the 2 clinics.

Data Collection Procedures at the 2 Clinics

Clinic A, a federally funded community health center, opened in 1980. In conformity with federal requirements, the clinic collected racial/ethnic data on patients. In 1987, Montefiore implemented a computerized outpatient registration program that used 9 racial/ethnic categories: Alaskan, American Indian, Asian, Black, Filipino, Indochinese, Pacific Islander, White, and Other. A separate ethnicity field captured data on specific immigrant populations, including the "Hispanic" population. Clinic B opened in another medically underserved community in 1994. In 1996, the 9 categories were revised systemwide. American Indian/Alaskan Native became 1 category, and Hispanic was added to the list. Even though the categories were the same at all sites, the registration screens differed. At clinic A the clerk had to type "???" to see the 9 categories. At clinic B clerks saw 6 categories on their screens and had to type "???" to access the other 3 categories.

Although clerks were encouraged to request racial/ethnic information from patients, we learned that this rarely happened. Clerks noted that many patients became angry when asked about race/ethnicity, and some did not understand the question. Patients would sit with clerks during the registration process; 1 clerk indicated that patients would occasionally correct the race/ethnicity information entered. Several clerks told us that they assigned race/ethnicity based on the patient's last name or appearance. One, who described herself as a "Black Puerto Rican," commented that she used herself as a guide when assigning individuals to categories.

Interviews

We generated a random list of patients from each clinic using the hospital's computerized registration database. Because we lacked interviewers who spoke other languages, only English- and Spanish-speaking patients were contacted to participate in the study. A medical student and a bilingual college student conducted the patient interviews from November 2000 through March 2001.

Telephone interviews were conducted with 93 patients from clinic A and 68 patients from clinic B. After obtaining informed consent, interviewers asked the following questions:

1. "How would you describe your racial or ethnic background?" (Spanish speakers were asked, in Spanish, "What is your race or ethnic group?" On the basis of the results of pretesting, the interviewer added "or nationality" when interviewees hesitated or seemed to find the question confusing.)
2. "If you had to choose between the following options to describe yourself, which one would you choose?" For this question, we used lists of categories provided by clerks at the 2 sites. We later learned that these differed somewhat from the actual choices at the clinics, e.g., the categories provided to us by clinic B clerks were limited to the 6 options visible on the initial screen.

The interviewers also asked respondents whether they would prefer the option of choosing "more than 1 racial or ethnic category."

RESULTS

We found that many respondents had trouble identifying with the concepts of race and ethnicity as understood by health researchers, many respondents described themselves in ways that were inconsistent with the categories included in the registration database, and many respondents were assigned

TABLE 1—Correspondence Between Self-Identification in Response to Open-Ended Question and Racial/Ethnic Designation in a Patient Registration Database: New York City, 2000–2001

	Registration Database						
Self-Reported Race/Ethnicity	American Indian/ Alaskan Native	Black	Hispanic	White	Other	None	Total
Clinic A							
African American, Afro American, *or* Black	…	32	…	…	1	1	34
Belizean	…	1	…	…	…	…	1
Black/Puerto Rican/Indian	…	1	…	…	…	…	1
Celtic from Scotland	…	…	…	…	1	…	1
Dominican	…	…	1	…	2	…	3
Ecuadoran	…	…	2	…	…	…	2
Filipino	…	…	…	1	…	…	1
Hispanic *or* Latino	…	…	14	5	2	1	22
Indian/Guyanese Indian[a]	…	…	3	…	…	…	3
Jewish	…	…	…	1	…	…	1
Mexican	…	…	1	1	…	…	2
Puerto Rican	…	…	7	7	…	1	15
South American	…	…	…	1	…	…	1
White	…	…	…	2	…	…	2
Other responses							
Beautiful	…	1	…	…	…	…	1
Other	…	…	…	…	…	1	1
Declined to answer	1	…	…	…	…	1	2
Total Clinic A	1	35	28	18	6	5	93
Clinic B							
African American, Afro American, *or* Black	…	11	…	…	…	…	11
American	…	…	2	…	…	…	2
Dominican	…		14	…	…	…	14
Ecuadoran	…	1	…	…	…	…	1
Guatemalan	…	…	1	…	…	…	1
Hispanic *or* Latina(o)[b]	…	…	9	…	…	…	9
Hispanic/Black	…	1	…	…	…	…	1
Hispanic/Dominican	…		1	…	…	…	1
Honduran	…	1	…	…	…	…	1
Mexican	…		1	…	…	…	1
Native American	…	1	…	…	…	…	1
Puerto Rican	…	…	23	…	…	…	23
Other responses							
Human race	…	1	…	…	…	…	1
Struggling and hard worker	…	1	…	…	…	…	1
Total Clinic B	…	17	51	…	…	…	68

[a]Two of these respondents answered "Indian" and 1 answered "Guyanese Indian" on the open-ended question; all 3 chose the "Indian/East Asian" category on the forced-choice question.
[b]Includes 1 respondent who self-identified as "Spanish" on the open-ended question and as "Hispanic" on the forced-choice question; we thus considered the term "Spanish" to be a synonym for "Hispanic."

categorizations in the database that were inconsistent with their self-reported identities (Tables 1 and 2).

Characterizations in the database were inconsistent with forced choices for 27 of 81 respondents at clinic A (33%) and 13 of 59 respondents at clinic B (22%; κ value 0.5 at both sites after exclusion of patients with missing data and patients who gave responses that were not in the registration database).

Comments in response to the forced-choice question included: "What does it matter?" "We are all the same." "I am not a racist." Several respondents declined to select a specific race/ethnicity from the choices offered.

Seventeen percent (16 of 93) at clinic A and 37% (25 of 68) at clinic B said they preferred the option of choosing multiple races.

DISCUSSION

This small-scale, exploratory study found that 33% of 81 respondents in one setting and 22% of 59 respondents in another setting saw themselves differently from the way they were categorized in a clinical database. The database categories failed to capture the full range of ways in which people self-identified (e.g., many respondents identified with a national origin instead of a race or ethnicity). A quarter of respondents indicated a preference for being allowed to identify multiple races/ethnicities.

It might be argued that our results simply reflect sloppy clerical work in 1 health care system, yet we suspect that this is not the case. Administrators, motivated in part by a federal requirement to collect racial/ethnic data and in part by a desire to better characterize the patient population, made a complex set of decisions regarding the collection of racial/ethnic data. The result was a system that was not uniform and that lacked clear guidelines, which may be typical for racial/ethnic data collection systems.

What are the implications of these findings? First, data collection processes must be continuously validated. Second, these findings reinforce suggestions that a third party should not determine an individual's race/ethnicity, be it registration clerk, clinician, or researcher.[8–10] Finally, decisions about how to conceptualize racial/ethnic categories

TABLE 2—Correspondence Between Self-Identification in Response to Forced-Choice Question and Racial/Ethnic Designation in Registration Database: New York City, 2000–2001

	Registration Database								
Self-Reported Race/Ethnicity	Asian	American Indian/ Alaskan Native	Black	Hispanic	Pacific Islander	White	Other	None	Total
Clinic A									
Black	. . .	. . .	26	**1**	. . .	. . .	**1**	2	30
Filipino	. . .	. . .	. . .	. . .	. . .	**1**	. . .	. . .	1
Hispanic	. . .	. . .	**1**	22	. . .	**11**	**4**	2	40
Indian/East Asian	. . .	. . .	. . .	**3**	. . .	. . .	. . .	. . .	3
White	. . .	. . .	. . .	**2**		6	. . .	. . .	8
Other		1	**5**	. . .	. . .	. . .	1	1	8
Declined to answer	. . .	. . .	3	. . .	. . .	. . .	. . .	. . .	3
Total Clinic A	. . .	1	35	28		18	6	5	93
Clinic B									
Black	. . .	. . .	8	**1**	. . .	. . .	. . .	. . .	9
Hispanic	. . .	. . .	**1**	38	. . .	. . .	**2**		41
White	. . .	. . .	. . .	**4**	. . .	. . .	. . .	. . .	4
Other	. . .	. . .	**3**	**2**	. . .	. . .	. . .	. . .	5
Declined to answer	. . .	. . .	4	5	. . .	. . .	. . .	. . .	9
Total Clinic B	. . .	. . .	16	50	. . .	. . .	2	. . .	68

Note: Boldface indicates discordance between self-identification and registration data.

carry with them important implications for health practitioners and researchers.

The commonly used racial/ethnic categories are at best approximations of broad and overlapping groups defined by society according to shifting criteria. Given the central role of race/ethnicity in health disparities, it is important to ensure that the data used to study these disparities represent as accurately as possible the populations in question. ■

About the Authors

Susan Moscou is a PhD candidate at the Heller School, Brandeis University, Waltham, Mass. Matthew R. Anderson is with the Department of Family Medicine and Community Health, Albert Einstein College of Medicine, Bronx, NY. Judith B. Kaplan is a public health editor in Medford, Mass. At the time of this study, Lisa Valencia was a student at the State University of New York at Old Westbury.

Requests for reprints should be sent to Matthew R. Anderson, MD, Montefiore Family Health Center, 360 E 193rd St, Bronx, NY 10458 (e-mail: andersonma@aol.com).

This brief was accepted August 21, 2002.

Contributors

S. Moscou and M. Anderson planned the study. L. Valencia assisted in creation of the data collection instrument and interviewed most of the study subjects. M. Anderson, S. Moscou, and J. Kaplan performed the data analysis. All authors contributed to the writing of the brief.

Human Participant Protection

This study was approved by the Montefiore Medical Center institutional review board.

References

1. Finucane TE, Carrese JA. Racial bias in presentation of cases. *J Gen Intern Med.* 1990;5:120–121.

2. Moscou S. Racial and ethnic identifications. *Nurse Pract.* 1996;21:8,11–12.

3. Parker DL. Medicalization of race. *Ann Intern Med.* 1997;126:669–670.

4. Anderson MR, Moscou S, Fulchon C, Neuspiel DR. The role of race in the clinical presentation. *Fam Med.* 2001;33:430–434.

5. Huth EJ. Identifying ethnicity in medical papers [editorial]. A*nn Intern Med.* 1995;122:619–621.

6. South-Paul JE. Racism in the examination room: myths, realities, and consequences. *Fam Med.* 2001;33:473–475.

7. Blustein J. The reliability of racial classifications in hospital discharge abstract data. *Am J Public Health.* 1994;84:1018–1021.

8. Use of race and ethnicity in public health surveillance: summary of the CDC/ATSDR Workshop. *MMWR Morb Mortal Wkly Rep.* 1993;42(RR):16.

9. Office of Management and Budget (US). Revisions to the Standards for Classification of Federal Data on Race and Ethnicity. 62 *Federal Register* 58781 (1997).

10. Kaufman JS. How inconsistencies in racial classification demystify the race construct in public health statistics. *Epidemiology.* 1999;10:101–113.