

Self-Reported vs Administrative Race/Ethnicity Data and Study Results

Ulrike Boehmer, PhD, Nancy R. Kressin, PhD, Dan R. Berlowitz, MD, MPH, Cindy L. Christiansen, PhD, Lewis E. Kazis, ScD, and Judith A. Jones, DDS, MPH, DSc

Concerns about administrative data on race/ethnicity have led some researchers to consider self-reported race/ethnicity as superior.^{1–5} However, few studies have examined the differential impact of the source of race/ethnicity data, that is, observed or self-reported, on study outcomes. We investigated whether differences in reporting of race/ethnicity led to different results with regard to the use of one therapeutic dental procedure, root canal therapy.

METHODS

From a retrospective secondary data study of Department of Veterans Affairs (VA) dental outpatients who underwent either root canal or tooth extraction between October 1, 1997, and September 30, 1998 (Jones et al., unpublished data), we selected the first treatment of the year for 15 137 patients, on whom we had both self-reported race/ethnicity from the 1999 Large Health Survey of Veteran Enrollees⁶ and administrative race/ethnicity data from the VA outpatient clinic files. Clinical information on the dental procedure performed as well as the severity of dental disease and the medical comorbidities was obtained from the VA administrative data files (Jones et al., unpublished data).

In the administrative data each patient was assigned a single race/ethnicity from among 6 categories: (1) Hispanic, (2) American Indian, (3) Black, (4) Asian, (5) White, (6) unknown or missing. Survey respondents were asked to describe their race/ethnicity by selecting all that applied from among 6 categories: (1) American Indian or Alaska Native, (2) Asian, (3) Black or African American, (4)

Spanish, Hispanic, or Latino, (5) Native Hawaiian or Pacific Islander, (6) White. Those who did not answer were coded as “missing.” We eliminated 35 who self-reported Native Hawaiian or Pacific Islander as their single race, since there was no comparable category in the administrative database. This reduced our sample to 15 102 patients with a single visit during which either a root canal or a tooth extraction was performed.

Using self-reported race/ethnicity as the gold standard, we calculated the proportion of each racial/ethnic category correctly recorded in the administrative database, once for those who self-reported a single race/ethnicity and once allowing for those who chose multiple responses. Using logistic regression, we estimated the probability of obtaining root canal therapy vs tooth extraction for patients of different race/ethnicity, calculating 3 models. The first model used administrative race/ethnicity data, the second self-reported single race/ethnicity, and the third used self-reported race/ethnicity by weighing multiple race/ethnicities by the number of categories.

RESULTS

We determined the amount of agreement between self-reported race/ethnicity and the administrative data, once using only the 82.4% of patients who reported a single race/ethnicity and once including the 4.9% who reported multiple (between 2 and 6) race/ethnicity categories. Following one of the US Census Bureau’s suggestions about the compilation of multiple race/ethnicity responses, we counted patients who reported combinations such as “White and Asian and African American” 3 times: (1) “White alone or in combination,” (2) “Asian alone or in combination,” and (3) “African American alone or in combination.”⁷ Thus the multiple counting increases the sample size from 15 102 patients to 15 906 race/ethnicity responses.

Table 1 summarizes the amount of agreement calculated both ways. Between 76.4% and 77.1% of self-reported Whites, between 68.4% and 68.9% of self-reported African Americans, between 57.1% and 61% of self-reported Hispanics, between 33.3% and 54% of Asians, and between 1.4% and 4.6% of self-reported American Indians were classified

TABLE 1—Agreement Between Self-Reported and Administrative Race/Ethnicity Data (n = 15 102)

| Race/ethnicity category | Self-Report | Administrative Data | | |
|--|-------------|-------------------------------|------------------|-----------------------------|
| | | Correctly Classified, No. (%) | Unknown, No. (%) | Falsely Classified, No. (%) |
| White | 8863 | 6832 (77.1) | 1898 (21.4) | 133 (1.5) |
| White or in combination | 9473 | 7235 (76.4) | 2018 (21.3) | 220 (2.3) |
| Black or African American | 2555 | 1760 (68.9) | 668 (26.1) | 127 (5.0) |
| Black or African American or in combination | 2721 | 1862 (68.4) | 704 (25.9) | 155 (5.7) |
| Spanish, Hispanic, or Latino | 766 | 467 (61.0) | 191 (24.9) | 108 (14.1) |
| Spanish, Hispanic, or Latino or in combination | 924 | 528 (57.1) | 216 (23.4) | 180 (19.5) |
| Asian | 87 | 47 (54.0) | 28 (32.2) | 12 (13.8) |
| Asian or in combination | 150 | 50 (33.3) | 43 (28.7) | 57 (38.0) |
| American Indian | 173 | 8 (4.6) | 43 (24.9) | 122 (70.5) |
| American Indian or in combination | 724 | 10 (1.4) | 148 (20.4) | 566 (78.2) |
| Missing | 1914 | NA | 531 (27.8) | 1383 (72.3) ^a |

Note. NA = not applicable.

^aThese patients were not necessarily falsely classified. This number indicates the number of patients who provided no self-reported race/ethnicity information but had race/ethnicity recorded in the administrative data.

TABLE 2—Odds Ratio (With 95% Confidence Interval) for Obtaining a Root Canal vs Tooth Extraction, by Administrative and Self-Reported Race/Ethnicity (n = 15 102)

| | Administrative | Single self-reported race/ethnicity ^a | Multiple self-reported race/ethnicity ^b | Differences between parameter estimates (P values), t test order | | |
|--|--------------------------------|--|--|--|--------|--------|
| | | | | 1 vs 2 | 1 vs 3 | 2 vs 3 |
| Black | | | | | | |
| Unadjusted | 0.53 ^c (0.46, 0.61) | 0.49 ^c (0.43, 0.56) | 0.51 ^c (0.45, 0.58) | 0.48 | 0.66 | 0.74 |
| Adjusted | 0.57 ^c (0.49, 0.66) | 0.53 ^c (0.46, 0.61) | 0.54 ^c (0.47, 0.63) | 0.44 | 0.67 | 0.75 |
| Hispanic | | | | | | |
| Unadjusted | 0.74 ^c (0.59, 0.92) | 0.68 ^c (0.56, 0.84) | 0.72 ^c (0.59, 0.87) | 0.65 | 0.85 | 0.80 |
| Adjusted | 0.83 (0.64, 1.07) | 0.74 ^c (0.58, 0.94) | 0.76 ^c (0.61, 0.94) | 0.44 | 0.60 | 0.87 |
| Asian | | | | | | |
| Unadjusted | 2.45 ^c (1.50, 4.02) | 1.98 ^c (1.26, 3.10) | 1.80 ^c (1.24, 2.61) | 0.55 | 0.38 | 0.77 |
| Adjusted | 1.99 ^c (1.11, 3.57) | 1.83 ^c (1.06, 3.15) | 1.81 ^c (1.16, 2.81) | 0.84 | 0.81 | 0.58 |
| American Indian | | | | | | |
| Unadjusted | 0.19 (0.03, 1.39) | 0.76 (0.51, 1.15) | 0.78 ^c (0.63, 0.96) | 0.33 | 0.32 | 0.93 |
| Adjusted | 0.18 (0.02, 1.47) | 0.93 (0.59, 1.49) | 0.82 (0.65, 1.03) | 0.13 | 0.18 | 0.50 |
| Unknown | | | | | | |
| Unadjusted | 0.96 (0.87, 1.07) | 0.59 ^c (0.52, 0.69) | 0.60 ^c (0.52, 0.69) | 0.00 | 0.00 | 0.53 |
| Adjusted | 0.95 (0.84, 1.07) | 0.60 ^c (0.51, 0.70) | 0.60 ^c (0.51, 0.71) | 0.00 | 0.00 | 0.52 |
| Multiple race/ethnicity including White | | | | | | |
| Unadjusted | — | 0.79 ^c (0.63, 0.98) | — | — | — | — |
| Adjusted | — | 0.80 (0.62, 1.03) | — | — | — | — |
| Multiple non-White race/ethnicity | | | | | | |
| Unadjusted | — | 0.65 (0.40, 1.06) | — | — | — | — |
| Adjusted | — | 0.62 (0.36, 1.08) | — | — | — | — |

Note. Odds ratios are shown both unadjusted and adjusted for dental and medical disease.

^aTo keep the sample size consistent, we included in the analysis 2 variables that reflected the 4.9% of patients who reported more than one race/ethnicity.

^bRace/ethnicity is weighted by number of responses, in that "Asian and white" is weighted as 0.5 Asian and 0.5 White.

^cOdds significantly different from those of the reference group of White patients.

as such in the observer-based administrative data. Self-reported Whites had the fewest “unknown” classifications and the fewest incidences of being classified as something other than White in the administrative database, whereas self-reported Asians had the most administrative classifications of “unknown” and self-reported American Indians the most occurrences of being falsely classified in the administrative data.

Table 2 presents odds ratios for obtaining root canal therapy for different groups, using different sources of race/ethnicity data, both unadjusted and adjusted for dental and medical disease. Regardless of the source of race/ethnicity data, African Americans were significantly less likely and Asians significantly more likely than Whites to obtain root canal therapy. Source of race/ethnicity data significantly changed the adjusted odds ratios for Hispanics and persons of unknown race/ethnicity. Parameter estimates for the model using self-reported data and the model using administrative data showed no significant differences for any known race/ethnicity category. However, when administrative data were used, the effect of unknown race/ethnicity on the odds of obtaining root canal therapy was similar to the effect of White race/ethnicity, whereas when self-reported data were used, the effect of unknown race/ethnicity was statistically different from the effect of White race/ethnicity.

DISCUSSION

Race/ethnicity in administrative data were more frequently incorrect for patients whose race/ethnicity was other than White, as indicated by other studies.^{8,9} Source of race/ethnicity data influenced the assessment of the outcome, root canal therapy, in that results for patients with unknown race/ethnicity differed significantly by data source. The level of disagreement and the differences in assessment suggest that estimates of racial/ethnic differences are dependent on the source of race/ethnicity data. Our results suggest that future studies of racial/ethnic variations should be sensitive to the source of race/ethnicity data. Studies that depend on administrative race/ethnicity should note the limitations of this approach. ■

About the Authors

The authors are with the Department of Health Services, Boston University School of Public Health, Boston, Mass, and the Center for Health Quality, Outcomes and Economic Research, Bedford, Mass. Judith A. Jones is also with the Department of General Dentistry, Boston University Goldman School of Dental Medicine.

Requests for reprints should be sent to Ulrike Boehmer, PhD, Center for Health Quality, Outcomes and Economic Research, 200 Springs Rd, Bldg 70, Bedford, MA 01730 (e-mail: boehmer@bu.edu).

This brief was accepted December 28, 2001.

Note. The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.

Contributors

U. Boehmer designed the study, analyzed the data, and wrote the brief. N.R. Kressin, D.R. Berlowitz, and J.A. Jones assisted in the study design, data analyses, and writing. C.L. Christiansen provided guidance on the data analyses and assisted in the writing. L.E. Kazis designed and conducted the 1999 survey that provided the self-reported data and reviewed the manuscript.

Acknowledgments

This research was supported by funding from the Department of Veterans Affairs Health Services Research and Development Service (HSR&D IIR 98.161) and by the Veterans Health Administration, Office of Quality and Performance.

The authors gratefully acknowledge A. Pitman for providing programming support.

Human Participant Protection

The study was approved by the institutional review board of the Edith Nourse Rogers Memorial Veterans Hospital, Bedford, Mass.

References

- Williams DR. Race/ethnicity and socioeconomic status: measurement and methodological issues. *Int J Health Serv.* 1996;26:483–505.
- Amaro H, Zambrana RE. Criollo, mestizo, mulato, LatiNegro, indigena, white, or black? The US Hispanic/Latino population and multiple responses in the 2000 census. *Am J Public Health.* 2000;90:1724–1727.
- 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;8:218–27.
- Williams DR. Race and health: basic questions, emerging directions [see comments]. *Ann Epidemiol.* 1997;7:322–333.
- Hahn RA, Stroup DF. Race and ethnicity in public health surveillance: criteria for the scientific use of social categories. *Public Health Rep.* 1994;109:7–15.
- Health Status and Outcomes of Veterans: Physical and Mental Component Summary Scores Veterans SF-36. 1999 Large Health Survey of Veteran Enrollees. Executive Report.* Washington, DC: Veterans Health Administration, Dept of Veterans Affairs; May 2000.
- US Census Bureau. Questions and answers for Census 2000 data on race. March 14, 2001. Available at: <http://www.census.gov/Press-Release/www/2001/raceqandas.html>. Accessed June 28, 2002.
- Hahn RA, Truman BI, Barker ND. Identifying ancestry: the reliability of ancestral identification in the United States by self, proxy, interviewer, and funeral director. *Epidemiology.* 1996;7:75–80.
- Hahn RA, Mulinare J, Teutsch SM. Inconsistencies in coding of race and ethnicity between birth and death in US infants. A new look at infant mortality, 1983 through 1985. *JAMA.* 1992;267:259–63.