

The Validity of Race and Ethnicity in Enrollment Data for Medicare Beneficiaries

Alan M. Zaslavsky, John Z. Ayanian, and Lawrence B. Zaboriski

Objective. To assess the validity of race/ethnicity in Medicare databases for studies of racial/ethnic disparities.

Data Sources. The 2010 Medicare Consumer Assessments of Healthcare Providers and Systems (CAHPS[®]) survey was linked to Medicare enrollment data and local area characteristics from the 2000 Census.

Study Design. Race/ethnicity was cross-tabulated for CAHPS and Medicare data. Within each self-reported category, demographic, geographic, health, and health care variables were compared between those that were and were not similarly identified in Medicare data.

Data Collection Methods. The Medicare CAHPS survey included 343,658 responses from elderly participants (60 percent response rate). Data were weighted for sampling and nonresponse to be representative of the national population of elderly Medicare beneficiaries.

Principal Findings. Self-reported Hispanics, Asians, Pacific Islanders, and American Indians were underidentified in Medicare enrollment data. Individuals in these groups who were identified in Medicare data tended to be more strongly identified with their group, poorer, and in worse health and to report worse health care experiences than those who were not so identified.

Conclusions. Self-reported members of racial and ethnic groups other than Whites and Blacks who are identified in Medicare data differ substantially from those who are not so identified. These differences should be considered in assessments of disparities in health and health care among Medicare beneficiaries.

Key Words. Medicare, race, ethnicity, disparities, CAHPS

Many analyses of racial/ethnic differences in use of services and quality of care under the Medicare program rely on the race/ethnicity variable in Medicare administrative files, linking them to measures of utilization and guideline-recommended care from claims data and Medicare quality measures to assess disparities in health and health care (Ayanian et al. 1993; Guadagnoli et al. 1995; Schneider, Zaslavsky, and Epstein 2002; Virnig et al. 2002, 2004;

McBean et al. 2003; Trivedi et al. 2005, 2006). Because current self-report is commonly accepted as defining racial and ethnic identity (Ulmer et al. 2009), the validity of Medicare's administrative data relative to self-reported race/ethnicity is of critical importance to the interpretation of study results.

As described by Arday et al. (2000) and McBean (2006), racial/ethnic variables in the Medicare Enrollment Database (EDB) were originally based on information from the Master Beneficiary Record of the Social Security Administration (SSA), which until 1980 used only the categories "White," "Negro," or "Other" (Scott 1999). These categories were expanded in SSA files in 1980 and in the Medicare EDB in 1994, but there have been only limited opportunities to update data for the current cohort of beneficiaries, most of whom would have registered with SSA before 1980. SSA race values for beneficiaries who completed new SS-5 registration forms to report a change in SSA status were incorporated into Medicare files in 1994, 1997, and annually since 2000, and since 1999 the Indian Health Service has provided the Medicare system with information on those it serves. The Centers for Medicare and Medicaid Services (CMS) also conducted a mail survey in 1997 targeted to beneficiaries identified as other or unknown race or with Spanish surnames. These measures have improved identification of groups not identified in the pre-1980 data, but substantial gaps remain. Furthermore, even the current coding used by SSA and CMS falls short of the level of detail mandated under the 1997 Office of Management and Budget (OMB) standards (Office of Management and Budget 1997) that are now used in federally sponsored surveys by combining Asians and Pacific Islanders and disallowing multiple racial/ethnic identifications.

A pioneering study by Arday et al. (2000) examined the relationship between race/ethnicity in administrative data and contemporary self-report data from the Medicare Current Beneficiary Survey (MCBS). Arday et al. found that the EDB had poor sensitivity for groups other than Whites and Blacks, although it was improved by updates over the 1990s. This study had limited precision for estimates involving smaller racial/ethnic groups, due to the modest size of the MCBS survey. Furthermore, the analysis was limited to tabulations of EDB by survey race/ethnicity classifications but did not analyze other characteristics of the groups falling into various cells of this cross-

Address correspondence to Alan M. Zaslavsky, Ph.D., Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, MA 02115; e-mail: zaslavsk@hcp.med.harvard.edu. John Z. Ayanian, M.D., M.P.P., and Lawrence B. Zaboriski, M.S., are with the Department of Health Care Policy, Harvard Medical School, Boston, MA.

classification. More recent studies by Waldo (2004) and Eicheldinger and Bonito (2008) have confirmed these findings. Underascertainment of a minority group may bias estimates for both that group and the majority White group to which some of its members are mistakenly assigned or vice versa. The effect of a given error on group estimates, however, would generally be greater for estimates concerning a relatively small minority than for the much larger White group.

In an analysis of racial/ethnic disparities using Medicare data on health care, we might ideally compare groups defined by a recent self-report of racial/ethnic identity using the more refined categories currently mandated under the 1997 OMB standards. However, analyses using administrative data are limited to the categories in use at the time racial/ethnic identification was most recently reported to the SSA or CMS. To evaluate the relationship between these two variables, we need a dataset that contains both types of information, such as the Consumer Assessments of Healthcare Providers and Systems (CAHPS) surveys. These surveys have been conducted in Medicare among managed care (currently Medicare Advantage [MA]) beneficiaries since 1998 and among fee-for-service (traditional Medicare or FFS) beneficiaries since 2001; the cumulative sample now exceeds 2 million respondents. While the primary purpose of the survey is to collect information on the quality of services provided, it also collects demographic data, including self-identified race/ethnicity according to current OMB standards.

The main objective of this study is to use CAHPS data to investigate the relationships between race/ethnicity as reported in administrative files and by self-report, extending the results of Arday et al. (2000). We first assess the concordance of EDB and CAHPS race/ethnicity variables. We also investigate the association of EDB race/ethnicity and survey responses, conditional on self-reported race/ethnicity, to better characterize the groups that are identified in the EDB and provide insight into the interpretation of patterns of health and health care detected in analyses that rely on the EDB data, thereby illuminating their utility for research on disparities.

METHODS

Data

Consumer Assessments of Healthcare Providers and Systems data were drawn from the 2010 Medicare CAHPS surveys, conducted in February–May 2010.

The instrument and survey methods have been described elsewhere (Goldstein et al. 2001; Zaslavsky and Cleary 2002). EDB race/ethnicity data were drawn from the EDB only a few months before administration of the survey and attached to the survey sample file. Every paper survey form included a printed and bar-coded identifying number that linked it to the sample file, as was required to direct follow-up mailings and telephone calls to initial nonrespondents. Population characteristics of ZIP code tabulation areas (ZCTAs) of residence were obtained from 2000 Decennial Census long form data.

Sample

Samples of respondents from MA risk contracts and free-standing prescription drug plans (PDPs) that serve traditional FFS Medicare beneficiaries were stratified by contract. Sample sizes were calculated to obtain a minimum of 390 respondents per risk contract and 580 per PDP based on historical response rates in each contract, with proportionally larger samples in the larger contracts. Within each contract, the sample was proportionately stratified by plan (benefit option), but members of some special needs plans (SNPs) were oversampled using an additional sample allotment. Similarly, samples of FFS beneficiaries not enrolled in a PDP were stratified by state and designed to obtain 500 responses in the smallest states and proportionally larger samples in larger states. The total sample size was 690,817 of which 1.2 percent was ineligible due to death or institutionalization; among the remaining 682,836 eligible subjects, the overall response rate was 59.8 percent. Analyses reported here are restricted to the 343,658 elderly (age \geq 65 years) respondents in the 50 states and District of Columbia, excluding younger beneficiaries eligible by reason of disability, because the latter subsample is selected by health conditions while elderly Medicare beneficiaries represent 97 percent of their age group, and also excluding 0.1 percent of cases for which EDB race/ethnicity was missing.

Variables

Respondents reported their race/ethnicity on the CAHPS surveys in conformity with the 1997 OMB standards using two questions. The first question asked, "Are you of Hispanic or Latino origin or descent?" and the second asked, "What is your race? Please mark one or more" with response options "White/Black or African American/Asian/Native Hawaiian or other Pacific Islander (NHOPI)/American Indian or Alaskan Native (AIAN)." Each

respondent could select any combination of the six race/ethnicity categories. Race/ethnicity in the EDB was measured by a single item, with options “White/Black/Hispanic/Asian or Pacific Islander/Native American/Other.” Thus, each respondent was coded with a single-race/ethnicity category. We use the term “race/ethnicity” consistent with the view that both race and ethnicity reflect social identities. In the surveys and statistical systems we consider here, “ethnicity” refers specifically to Hispanic/Latino origin or identification, while “race” represents the set of racial categories listed above.

Other demographic variables included age and sex (from the EDB), educational attainment from CAHPS, Medicaid dual eligibility from the EDB, and census area characteristics by ZCTA (racial/ethnic composition and poverty rate for over-65 population, median household income). Health variables included self-reported general health status and mental health status (both on 5-point scales from “poor” to “excellent”); health conditions included heart attack, angina or coronary heart disease, stroke, cancer other than skin cancer, emphysema/asthma/chronic lung disease, and diabetes, as well as the count of these conditions; number of limitations in activities of daily living (ADLs); and smoking status. Health care variables in the CAHPS survey are described in detail elsewhere (Zaslavsky and Cleary 2002; O’Malley et al. 2011). Of these, we selected MA enrollment status from the EDB and the following variables that were asked similarly in all forms of the survey: rating of personal doctor, rating of care received (both on 0–10 numerical scales), receipt of flu and pneumonia immunizations, how often the respondent received needed care and was seen within 15 min of appointment time (each on a 4-point “never” to “always” scale), a composite of four items on doctor communication (also on this 4-point scale), having a personal doctor, and number of visits with a personal doctor and with specialists. For comparability, survey variables assessing health and health care were rescaled to a 0–100 scale except where the original scale had a substantive interpretation (percentage or count variables).

Weighting

Weighting proceeded in three stages. First, sampling weights were calculated reflecting unequal sampling rates by stratum (contract, PDP enrollment, over-sampled SNP, or for beneficiaries not enrolled in a PDP or MA plan, state of residence). Next, initial nonresponse and poststratification weights were calculated by iterative proportional fitting (“raking”) (Deming and Stephan 1940; Purcell and Kish 1980) to make weighted survey totals consistent with

Medicare enrollment distributions (estimated from the entire CAHPS sample using sampling weights) by state and (for MA and PDP beneficiaries) contract. This step also matched distributions within contract or state by age, sex, EDB race/ethnicity, eligibility for Medicaid and the low-income subsidy (LIS) for a PDP, enrollment of MA beneficiaries in a SNP (dual-eligible, other, or none) and in a PDP; quartiles of median income and percents college educated, White, Black, and Hispanic by ZCTA; and within census region by additional interactions of these variables (Zaborski and Zaslavsky 2011). Finally, within each state by age (dichotomized at 75 years) by sex by EDB race/ethnicity cell, the weight of cases with no responses to the race or ethnicity items (representing 8.9 percent of the population) was proportionally redistributed to cases in the same cell that responded; most of these nonrespondents were also nonrespondents to adjacent demographic items. The final weights thus make proportions of EDB race/ethnicity consistent with population controls while preserving their interactions with self-reported race/ethnicity demonstrated in the survey data and closely matching population distributions of demographic and geographic variables. All analyses were made population-representative using these weights.

Analysis

We first cross-tabulated self-reported race/ethnicity with EDB race/ethnicity. Because the survey items allow respondents to select more than one category, we tabulated all those who selected a single category and then the frequency of selection of each category among those selecting multiple categories.

We summarized the associations of self-reported race/ethnicity by EDB race/ethnicity with sensitivities, specificities, and positive and negative predictive values of the various racial categories in the EDB for the corresponding self-reports. Definitions of sensitivity and specificity are complicated by multiple-race responses in the survey, including the many self-reported Hispanics who also report a race, as intended by the OMB standards. For this reason, we report sensitivity by both a narrow definition (limited to those who selected a single category) and a broad definition (including all those who checked the category, including those who checked multiple options). Specificity is defined over a denominator of all who did not check the category. We report positive and negative predictive values of EDB race by corresponding broad definitions; thus, any respondent whose EDB race matched one of his self-reported race/ethnicities would be considered a correct prediction. Similarly, we combined Asian and NHOPI self-report categories for calculation of specificity

and predictive values since these together most nearly correspond to the Asian or Pacific Islander category in the EDB.

The last set of analyses focused on differences between self-reported members of a racial/ethnic group who are and are not so identified in the EDB. These analyses were conducted for Hispanics, Asians, NHOPI, and AIAN, using an inclusive criterion for membership in each group comprising all who checked that category alone or in combination with other categories. We distinguish Asians and NHOPI following previous findings of substantial health differences between these groups (Bitton, Zaslavsky, and Ayanian 2010). We first compared demographic characteristics (age, sex, education, Medicaid dual eligibility), location (regions and selected states, means by ZCTA of prevalence of the group, poverty rate, and median income), and use of a unique self-identification. We next compared health characteristics, case mix adjusted for age and sex using linear models. Finally, we compared health care variables, case mix adjusted for age, sex, Medicaid dual eligibility status, general and mental health status, and MA enrollment status. For comparison, we also reported on Whites and Blacks, each as a single group.

RESULTS

The percentage listed as Hispanic in the EDB (2.0 percent) approximated the percentage that self-reported Hispanic identity but did not report a race (1.5 percent) but is far less than the percentage of all self-reported Hispanics (6.2 percent) (Table 1). The percentage listed as Asian or Pacific Islander in the EDB (2.1 percent) fell short of the combined self-reports for these groups (3.5 percent, counting only once those selecting both categories). The proportion Native American in the EDB (0.39 percent) approximated the proportion self-reporting only AIAN (0.43 percent) but also fell far short of the more inclusive proportion that included multiple reports (1.8 percent). Whites, Blacks, and Asians usually (>90 percent for each) reported a single race/ethnicity, while single responses were less common for Hispanics (24 percent), NHOPI (44 percent), and AIAN (24 percent) (Table 2). Supporting Information Appendix Table A1 summarizes the frequencies of joint reporting of various races and Hispanic ethnicity. The weighted percentage of CAHPS respondents self-reporting as non-Hispanic with a single non-White race, or Hispanic with any racial combination, was between 89 and 96 percent of the percentage in each corresponding group according to Census Bureau

Table 1: Distributions of Self-Reported and Medicare Enrollment Database (EDB) Race/Ethnicity

	<i>Self-Reported Race</i>			<i>EDB Race</i>
	<i>Single-Race/Ethnic Checkmark</i>	<i>Multiple-Race/Ethnic Checkmarks</i>	<i>Total, Self-Report</i>	
White	80.52	5.82	86.35	85.18
Black	7.87	0.78	8.64	8.38
Hispanic	1.49	4.71	6.20	1.96
Asian	3.08	0.28	3.35	}2.09
NHOPI	0.17	0.22	0.38	
AIAN	0.43	1.38	1.81	0.39
Other				2.01
Unduplicated total	93.55	6.45	100.00	
Column sum		13.18	106.73	100.00

Note. Weighted percentages for CAHPS respondents ≥ 65 years old. EDB category “Asian and Pacific Islander” combines Asian and NHOPI.

AIAN, American Indian and Alaskan Native; NHOPI, Native Hawaiian and other Pacific Islander.

Table 2: Accuracy of Medicare Enrollment Database Race/Ethnicity Relative to Self-Report

	<i>White</i>	<i>Black</i>	<i>Hispanic</i>	<i>Asian</i>	<i>NHOPI</i>	<i>AIAN</i>
Sensitivity						
Narrow definition	99.16	97.91	40.57	59.61	38.55	56.99
Broad definition	97.31	93.81	30.19	56.74	26.33	17.60
Specificity	91.53	99.70	99.91	99.89		99.93
Predictive value positive	98.64	96.76	95.51	94.69		82.53
Predictive value negative	84.35	99.42	95.59	98.28		98.50
Reported single-race/ethnicity	93.26	91.01	24.00	91.77	43.51	23.78

Note. Weighted percentages for CAHPS respondents ≥ 65 years old. EDB category “Asian and Pacific Islander” combines Asian and NHOPI, which are therefore combined for some measures. For sensitivity, “narrow definition” is restricted to CAHPS respondents selecting only a single category and “broad definition” includes all respondents selecting the category singly or in combination.

AIAN, American Indian and Alaskan Native; NHOPI, Native Hawaiian and other Pacific Islander.

population estimates for the over-65 population in the month (April 2010) closest to the survey period (non-Hispanic single-race Whites 80.52 percent versus 80.07 percent, non-Hispanic single-race Blacks 7.87 percent versus 8.41 percent, Hispanics regardless of race 6.20 percent versus 6.91 percent, non-Hispanic single-race Asians 3.08 percent versus 3.45 percent, non-Hispanic

single-race Native Americans 0.43 percent versus 0.45 percent). The discrepancies are largely attributable to the higher rate of multiple-race selections in our data than in the Census estimates (1.74 percent versus 0.63 percent).

Sensitivities (by either definition) and specificities of Medicare EDB identification for Whites and Blacks were high, all exceeding 91 percent (Table 2). For Hispanics sensitivity was only 30.2 percent by the broad definition, but somewhat higher (40.6 percent) among the narrower group who reported Hispanic ethnicity without a race. Similarly, sensitivity was moderate (57.0 percent) among single-group AIAN identifiers but very low (17.6 percent) among the multiple identifiers. Sensitivity was also moderate even by the narrow definition for Asians (59.6 percent) and NHOPI (38.6 percent). Specificities were very high (>99.7 percent) for all groups except Whites (91.5 percent), indicating that non-White beneficiaries who did not self-identify at least partially with these categories were very unlikely to be linked to them in the Medicare EDB.

Positive predictive values for almost every group were also very high (>94.7 percent), by the broad definition that included both single and multiple identifiers as correct if they named the race to which they were linked in the Medicare EDB. The exception was Native Americans recorded in the Medicare EDB, of whom only 82.5 percent were self-reported as AIAN in the CAHPS survey.

More detailed cross-classifications of self-report by EDB race (Appendix Tables A2 and A3) show that those who self-report a single race/ethnicity of Black, Hispanic, or Native American but are not identified as such in the EDB are most likely to be listed as White; among the corresponding subsets of Asians and Pacific Islanders, the most likely identification is Other. Of those identified as "Other" in the EDB (2.0 percent), 57.4 percent self-reported Asian race, 29.3 percent checked White race, and 18.3 percent identified as Hispanic (with substantial overlap with White race).

Because of the highly accurate identification of Whites and Blacks in the EDB, we focused in the final analyses on the smaller, less completely ascertained groups, comparing Hispanics, Asians, NHOPI, and AIAN who are and are not identified as such in the EDB. The first line of Table 3 repeats the analysis of sensitivity of the EDB discussed above. The second line displays the percentage of those in each group who marked only a single race/ethnicity on the survey. This percentage was lowest for Hispanics (32.3 percent of those identified as Hispanic in the EDB, and only 20.4 percent of those not so identified), and greatest for Asians (96.4 and 85.7 percent, respectively, for those identified and not identified as Asian in the EDB).

Table 3: Demographic Characteristics and Geographic Distribution by Self-Reported and Medicare Enrollment Database Race/Ethnicity

	Self-Reported Race/Ethnicity											
	White		Black		Hispanic		Asian		NHOPI		AIAN	
	in EDB	Not in EDB	in EDB	Not in EDB	in EDB	Not in EDB	in EDB	Not in EDB	in EDB	Not in EDB	in EDB	Not in EDB
As percent of self-reported group	93.3	30.2	69.8	56.7	43.3	26.3	73.7	17.6	82.4			
Checked single-race/ethnic category (%)	93.3	32.3	20.4***	96.4	85.7***	63.7	36.3***	77.0	12.4***			
Demographics												
Age (years)	75.2	74.3	76.4	73.3***	76.4	73.1***	75.1	73.4*	72.9	73.2		
Female (%)	56.6	61.0	57.9	56.3	59.3	56.5	57.3	55.3	61.8	53.3***		
Education (years)	13.2	12.1	9.9	11.3***	12.7	14.0***	12.6	12.6	12.0	12.5**		
High school graduate (%)	82.7	63.3	27.8	52.6***	66.6	81.7***	70.6	71.2	67.6	70.8		
Medicaid dual-eligible (%)	8.1	24.7	59.6	29.3***	57.2	14.7***	36.8	21.5***	28.7	19.0***		
Area characteristics												
Group prevalence (%)	91.8	37.4	32.7	26.8***	19.3	16.5***	2.0	1.6	28.1	0.7***		
Poverty (%)	8.8	15.3	14.5	13.1***	9.2	8.3***	8.6	10.4**	18.1	10.8***		
Median income (\$1,000)	29.7	25.0	25.7	27.3***	34.1	35.4**	34.6	31.2**	22.4	27.0***		
Federal region (%)												
Northeast	5.6	2.3	2.3	2.3	2.6	2.5	0.0	4.3***	1.1	4.7***		
New York/New Jersey	9.1	12.2	12.7	12.2	12.2	11.8	3.6	12.1**	2.1	6.2**		
Mid-Atlantic	10.2	14.8	2.5	3.1	5.2	7.2*	1.2	6.4*	1.0	8.5***		
Southeast	20.8	31.5	18.7	17.9	5.9	8.8***	3.0	12.1**	4.7	21.9***		
Midwest	18.3	14.8	4.2	5.7*	6.8	9.7**	6.6	6.9	7.8	13.7**		
South Central	10.7	12.5	20.9	23.5*	4.9	5.8	6.0	5.8	33.9	17.0***		

continued

Table 3. Continued

	Self-Reported Race/Ethnicity																	
	White			Black			Hispanic			Asian			NHOPI			ALAN		
	White	Black	Hispanic in EDB	Black	Hispanic in EDB	Not Hispanic in EDB	Asian/PI in EDB	Not Asian/PI in EDB	Asian/PI in EDB	Not Asian/PI in EDB	Asian/PI in EDB	Not Asian/PI in EDB	Asian/PI in EDB	Not Asian/PI in EDB	Asian/PI in EDB	Not Asian/PI in EDB		
Central Plains	5.3	2.7	0.5	2.7	1.2**	0.9	0.9	0.9	0.2	1.0	0.6	4.3***						
Mountain	3.4	0.5	2.3	0.5	2.6	0.9	1.7*	0.9	0.9	2.0	10.2	2.6***						
Pacific	12.2	7.9	34.4	7.9	29.9***	56.8	46.9***	70.7	43.4***	25.2	15.9***							
Northwest	4.4	0.8	1.5	0.8	1.7	3.8	4.7	8.0	5.9	13.6	5.2***							
Specific states (%)																		
Florida	7.8	7.2	17.3	7.2	14.3***	3.0	4.5*	0.9	6.9**	0.8	6.0***							
Texas	6.4	6.7	18.0	6.7	18.9	4.0	4.4	3.8	2.2	2.1	8.4***							
California	9.1	6.8	30.5	6.8	25.6***	46.9	33.4***	43.9	21.2***	9.2	12.7**							
Hawaii	0.2	0.1	0.2	0.1	0.4*	7.4	10.9***	24.2	17.2*	0.0	0.4***							

Note. Weighted percentages or means for CAHPS respondents ≥ 65 years old. Respondents selecting multiple racial/ethnic categories are tabulated in more than one column.

* $p < .05$,

** $p < .01$,

*** $p < .001$ for comparisons of those identified and not identified with group in the Enrollment Database. ALAN, American Indian and Alaskan Native; NHOPI, Native Hawaiian and other Pacific Islander.

Enrollment Database Hispanics and Asians were slightly older than their non-EDB counterparts (Table 3). Among AIAN, the EDB group was substantially more likely to be female than the non-EDB group. EDB-identified Hispanics and Asians had substantially lower educational levels than their non-EDB counterparts. In each group, the EDB-identified members were much more likely to be dual-eligible for Medicaid, an indicator of low income; except for NHOPI, the mean poverty rates of their areas of residence are higher and their mean area median incomes are lower as well. EDB-identified members of each group also lived in areas with slightly higher mean concentrations of members of their own group (dramatically higher for AIAN). Hispanics, Asians, and NHOPI recorded in the Medicare EDB were more concentrated in California, while AIAN in the Medicare EDB were generally found west of the Mississippi.

In comparisons adjusted for age and sex, EDB Hispanics and Asians report substantially worse general and mental health status than their non-EDB counterparts, but little difference was observed among AIAN (Table 4). Relative to their non-EDB counterparts, EDB AIAN reported significantly higher rates of diabetes and smoking but lower rates of all other conditions and fewer health conditions overall. Among the other groups, the majority of the significant differences indicated fewer health conditions for those not identified in the EDB, with the main exception being cancer for which rates were higher for the non-EDB subgroup of each self-reported group.

In each group, the MA enrollment rate for those identified in the Medicare EDB was lower than among those not identified in the EDB. After control for demographics, health status, and MA and Medicaid enrollment, Asians identified in the EDB gave worse ratings of care and less positive reports on communication with their doctors than those not so identified. EDB Hispanics and Asians generally report worse health care access than their non-EDB counterparts, reflected in items related to pneumococcal immunization, timely appointments, getting needed care, and having a personal doctor; the same applied to AIAN for this last item. However, among those who reported having a personal doctor, EDB Hispanics, Asians, and AIAN reported having more visits. Many of the differences between EDB and non-EDB subgroups of a racial or ethnic group on health and health care variables are substantial in magnitude relative to differences between Whites and other groups (including Blacks). In particular, non-EDB Asians report similar overall health status to Whites, and their health care assessments are about equidistant between EDB Asians and Whites. Likewise, non-EDB Hispanics are intermediate between Whites and EDB Hispanics on health status

Table 4: Adjusted Means of Health and Health Care Measures by Self-Reported and Medicare Enrollment Database Race/Ethnicity

	Self-Reported Race/Ethnicity											
	Hispanic			Asian			NHOPI			AIAN		
	White	Black	Hispanic in EDB	Not Hispanic in EDB	Asian in EDB	Not Asian in EDB	NHOPI in EDB	Not NHOPI in EDB	AIAN in EDB	Not AIAN in EDB	AIAN in EDB	Not AIAN in EDB
Health												
Health status	53.9	46.8	43.4	47.8***	47.3	54.0***	49.2	51.9	49.4	47.6		
Mental health status	72.3	65.9	58.8	65.7***	62.7	70.2***	66.1	69.0	66.2	67.0		
Conditions (count)	0.85	0.98	0.89	0.87	0.74	0.73	0.83	0.82	1.04	1.25***		
Heart attack	14.2	13.7	13.7	13.0	9.4	9.4	14.2	10.6	16.4	20.9**		
Angina/CHD	20.2	17.3	18.4	16.6**	14.5	14.8	16.5	15.2	18.9	23.0*		
Stroke	9.6	12.7	8.3	9.5**	9.6	8.7	18.6	10.2***	10.6	15.6***		
Cancer	19.8	17.6	11.0	14.3***	9.3	11.5**	8.7	14.1*	12.7	18.5***		
Emphysema/ asthma/COPD	16.9	16.2	14.2	13.3	12.3	11.8	8.3	16.1**	19.5	24.9**		
Diabetes	26.6	46.8	41.2	39.7	39.3	34.9***	41.5	37.4	42.3	36.7**		
ADL limitations	0.70	0.94	1.24	0.95***	0.98	0.60***	1.34	0.86***	0.98	1.06		
Smoker	9.4	13.2	7.0	7.8	5.4	5.0	5.6	7.1	18.2	14.0**		
Health care												
Enrolled in MA	23.7	30.2	31.7	38.1***	25.6	31.3***	28.8	34.8	11.7	26.1***		
Rating of doctor	90.2	90.1	90.5	90.6	85.8	87.2***	88.2	89.6	89.1	90.1		
Rating of care	85.7	83.1	84.1	83.7	77.8	81.5***	81.5	83.9	78.3	83.7***		
Communication with doctor	90.0	90.3	89.6	89.5	83.8	86.5***	86.3	88.8*	88.9	89.4		
Flu shot	70.4	55.5	68.3	64.9***	77.8	74.0***	67.4	69.0	78.5	63.2***		
Pneumonia shot	72.7	57.1	57.0	59.1**	59.5	67.0***	60.7	63.8	75.5	69.1***		

continued

Table 4. Continued

	Self-Reported Race/Ethnicity									
	Hispanic			Asian			NHOPI		AIAN	
	White	Black	Hispanic in EDB	Not Hispanic in EDB	Asian in EDB	Not Asian in EDB	NHOPI in EDB	Not NHOPI in EDB	AIAN in EDB	Not AIAN in EDB
Seen in 15 min	55.5	49.6	48.4	49.7*	48.3	52.6***	55.1	51.8	53.3	54.4
Get needed care	88.9	85.8	86.0	87.0	79.6	82.2***	80.6	85.4*	82.8	85.9
Have personal doctor	95.1	90.9	88.1	90.7***	90.1	92.6***	90.9	92.9	78.8	91.4***
Number of visits with personal doctor	1.92	2.09	2.26	2.12***	2.10	1.91***	2.02	2.06	2.22	2.03**
Number of visits with specialist	1.86	1.76	1.88	1.81*	1.75	1.70	1.87	1.71	1.59	1.89***

Notes: Weighted percentages (individual conditions, activities of daily living, smoking) or means for CAHPS respondents ≥ 65 years old. Survey items without a substantive quantitative interpretation are converted to a 0 (worst possible) to 100 (best possible) scale. Respondents selecting multiple racial/ethnic categories are tabulated in more than one column.

* $p < .05$,

** $p < .01$,

*** $p < .001$

for comparisons of those identified and not identified with group in the Enrollment Database. In each comparison with a clear direction, the more favorable outcome (better health, fewer conditions or limitations, more positive assessment of care, higher immunization rate) is bolded. ADL, activities of daily living; AIAN, American Indian and Alaskan Native; CHD, coronary heart disease; COPD, chronic obstructive pulmonary disease; NHOPI, Native Hawaiian and other Pacific Islander.

measures. NHOPI identified as Asian/Pacific Islander in the EDB have much higher rates of stroke and ADL limitations than Whites but their counterparts not identified in the EDB are similar to Whites on these measures.

DISCUSSION

In this comparison of self-reported race/ethnicity to codes in the EDB, we found that all codings in the database were highly specific and had high positive and negative predictive power. The sensitivity of the Medicare EDB measure was high for White and Black respondents, consistent with prior findings of Arday et al. (2000). However, the EDB sensitivities were low (ranging from 39 to 60 percent by the most focused definition) for all other groups. In short, Whites and Blacks recorded in the Medicare EDB closely matched those identified by self-report, but Hispanics, Asians, Pacific Islanders, American Indians, and Alaska Natives identified in the Medicare EDB constituted only subgroups of their corresponding self-reported populations.

A number of historical factors may account for under-identification of these groups. Most important, Medicare relies primarily on SSA records for identification, which were limited to categories of White, Black, and Other until 1980, with only limited updates since then. Although the rule adopting the 1997 OMB racial/ethnic standard called for its implementation across federal agencies by 2002, the SSA was granted extensions by OMB through 2009. Beginning on January 1, 2010, SSA forms completed by new adult registrants and individuals requesting a replacement Social Security card use the 1997 OMB categories, but applications for Social Security numbers issued at birth, which are taken by state agencies, do not ask for racial/ethnic identity. It remains unclear whether and when CMS will start to receive and incorporate data in the new format (Manuel de la Puente, Associate Commissioner, Office of Retirement and Disability Policy, Social Security Administration, Personal communication).

Furthermore, with changing social understandings of race and ethnicity, racial and ethnic identifications declared in the past may be inconsistent with current self-identification. Between the 1990 and 2000 decennial censuses, for example, the numbers of self-identified AIAN increased substantially within the same birth cohorts, likely due to a more positive image of identification with that group and the growing prosperity of some tribes (Appendix Table A4).

The structure of the categories used by the SSA also could create inconsistencies between Medicare EDB and self-reported race/ethnicity. The SSA system requires selection of a single category. The CAHPS items, consistent with current OMB standards, allow selection of multiple categories and, in particular, are designed to distinguish Hispanic ethnicity and to encourage Hispanic respondents to select a race as well as Hispanic ethnicity. Among self-reported Hispanics in CAHPS, 76 percent checked at least one race as well, and likewise 76 percent of AIAN and 57 percent of NHOPI selected multiple categories; it is unknown how many of these would have reported these categories if they had to choose only one.

Underascertainment of some groups in the EDB would be less concerning if the characteristics of members of a group who are so identified are equivalent to those of members who are not. In that case, the main effect would be to reduce the sample size of the under-ascertained group and classify the unrecognized members into a larger group (notably, Hispanics, AIAN, or NHOPI are often classified as White) on whose estimates they would have little statistical impact. We found, however, that there were substantial differences in geographic, demographic, health, and health care variables. Members of each group identified in the Medicare EDB were more likely to select the group as a single category and to live in areas with higher concentrations of the group (at both regional and local levels), as well as to have less education, be Medicaid dual-eligible, and to live in lower-income areas, all compared to those not identified in the EDB. These differences indicate that those identified in the EDB populations have a stronger group identification and greater socioeconomic disadvantage. The striking differences in geographic distribution for AIAN might be attributed to the major contribution to their identification in the EDB from enrollment with the Indian Health Service, which is most active on Indian reservations in the western states. Swan et al. (2006) report findings parallel to ours in comparisons of American Indians reporting a single versus multiple racial/ethnic identifications. Similarly, EDB-identified group members tended to report worse health. The main exception was the prevalence of a history of cancer; because most reports of cancer were obtained from cancer survivors, this paradoxical finding might reflect better ascertainment and survival in the non-EDB subgroups due to superior screening and treatment. Access to health care, as measured by CAHPS items, also appeared to be superior for most of the non-EDB subgroups.

Because of these differences, for example, estimates of White-Hispanic differences in health or health care experiences relying on racial/ethnic

identifications in the Medicare EDB reflect the experiences of a select subgroup of Hispanics but are not representative of all those who currently self-identify as Hispanic. This might be regarded as more or less problematic depending on the objectives of the analysis, since the boundaries of racial/ethnic subgroups are not immutable or unambiguous. Although we do not know exactly what historical or social factors determined classification as Hispanic in the EDB, we have no reason to think it was a consequence of health status at the time of registration. Therefore, we might regard the group so identified as an appropriate subject for an analysis of social disparities in health or health care, in a way that would not be appropriate for a group identified through its use of health care services (such as Hispanics served at a clinic). On the other hand, a disparity finding for EDB-identified Hispanics would be less generalizable and possibly less actionable than one that is more broadly representative of those who self-identify as Hispanic since the latter group is more readily identifiable in the community and in health care settings (although not within Medicare itself) and is recognizable as a social and political entity.

The differences we found by EDB identification within groups are challenging for methods that attempt to recover means or proportions for self-reported groups by backing out the relationship between self-report and EDB coding. Escarce and McGuire (2003) proposed such a method, using the matrix cross-classifying the two race/ethnicity codings reported by Arday et al. (2000). Their method assumed that each group defined by self-report was homogeneous with respect to the variable under study. Regarding each EDB-based group as a mixture in known proportions of the self-report groups, equations can be written expressing means for the former in terms of means for the latter, and solved to estimate self-report means from the more readily observable EDB-based means. On the other hand, the rationale for this method breaks down when, as we found for many variables, means differ by EDB category within each self-report group. Indeed in that case, the proposed method could increase bias of estimates for some groups, relative to taking EDB groups at face value, rather than reducing it.

Researchers have made considerable progress over recent years on methods for assigning racial/ethnic categories to individuals, or more precisely estimating probabilities that an individual would report each of the possible racial-ethnic categories, using information such as surnames (Morgan, Wei, and Virnig 2004; Wei et al. 2006), given names, or racial/ethnic composition of the area of residence. Each of these racial/ethnic correlates contributes distinctive information and is most informative for identifying particular

groups, so the results may be most useful or accurate when several variables can be combined (Fiscella and Fremont 2006; Elliott et al. 2008, 2009). We anticipate (based on Eicheldinger and Bonito 2008; and our own exploratory analyses) that including EDB race/ethnicity codes as an additional predictor will further contribute to the reliability of these “indirect estimation” methods. In related work, the National Center for Health Statistics (NCHS) used multivariate Bayesian models to bridge racial/ethnic identification in their surveys from the older (1977) to the new (1997) OMB-mandated system by imputing a primary race for those reporting multiple races (Schenker and Parker 2003). On the other hand, regression analyses proposed for use with such estimated probabilities (McCaffrey and Elliott 2008) similarly rely on a homogeneity assumption, namely that within each group the outcome under study is unrelated to the probabilities assigned by the model. This assumption could fail, for example, if Hispanics with Spanish surnames or living in areas with high proportions of Hispanics differ systematically from those with non-Spanish surnames or in non-Hispanic neighborhoods. Although validation studies of indirect estimation have had promising results, we recommend continued monitoring of the validity of the underlying assumptions.

Nonetheless, we believe self-report is the preferred method of racial/ethnic identification for studies of health and health care disparities, as underscored by the 2009 Institute of Medicine report *Race, Ethnicity and Language Data: Standardization for Health Care Quality Improvement* (Ulmer et al. 2009). This approach does not rely on the particular features of any administrative system, although the sensitivity of survey responses to details of item wording, response options, or context is also of concern (Bates et al. 1995; Martin et al. 2001). For studies in which self-report is unavailable, other information sources or indirect estimation methods should be used, while recognizing that the groups identified in this way may not be consistent with those identified by self-report. Finally, recognition of the fluidity (different reporting over time or in different contexts) of race/ethnicity in no way detracts from the reality of disparities in health and health care and the importance of detecting and addressing them.

Our study has a number of limitations. Nonresponse to the CAHPS survey could bias our results, although we speculate that this might have greater effects on marginal distributions of race/ethnicity than on associations between variables. While we weighted our data to match population distributions of EDB race and numerous individual and area characteristics, we were unable to calibrate CAHPS racial identification against a strictly comparable census self-reported race distribution due to differences in response options

(inclusion of an “other race” category) in the American Community Survey and census. Furthermore, the most recent data tabulated by ZCTA are from the 2000 Census, but these geocoded measures are probably stable enough over 10 years to give useful information on differences in area characteristics.

In summary, our findings highlight the ambiguity and fluidity of racial/ethnic identification. Research on racial identity highlights how self-reported race/ethnicity can change even for the most reliably distinguished groups, Blacks and Whites, with changing social or family circumstances (Lieberson and Waters 1993) or even in a retest after a few months time (Elwert and Christakis 2006). This fluidity may be even greater for some of the smaller groups within the Medicare population. In particular, Hispanic identity may be variously regarded as an alternative to or distinct categorization from race (Morning 2005), an ambiguity perhaps reflected in the selection of “other race” by many Hispanics (37 percent in the 2010 census), an option not offered in the CAHPS survey. Furthermore, identification with multiple races is gradually increasing, from 2.4 percent in 2000 to 2.9 percent in 2010 (Humes et al. 2011). On the other hand, the cohorts born in the United States since 1989 were registered with Social Security at birth, a process that provides no racial identification because racial data on birth certificates are not transmitted to the SSA (McBean 2006). Thus, the complexities of racial and ethnic identification in the Medicare population can only be expected to grow as a more diverse population ages into Medicare.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statements: This research was supported by a subcontract to Harvard Medical School under contract 9920070047 between the Center for Medicare and Medicaid Services (CMS) and the RAND Corporation. Dr. Ayanian is supported by the Health Disparities Research Program of Harvard Catalyst/The Harvard Clinical and Translational Science Center (NIH grant 1 UL1 RR 025758 and financial contributions from Harvard University and its affiliated academic health centers). The authors acknowledge the efforts of the entire Medicare CAHPS implementation team at CMS and collaborating organizations for their efforts, which made this research possible.

Disclosures: None.

Disclaimers: None.

REFERENCES

- Arday, S. L., D. R. Arday, S. Monroe, and J. Zhang. 2000. "HCFA's Racial and Ethnic Data: Current Accuracy and Recent Improvements." *Health Care Financing Review* 21 (4): 107–16.
- Ayanian, J. Z., I. S. Udvarhelyi, C. A. Gatsonis, C. L. Pashos, and A. M. Epstein. 1993. "Racial Differences in the Use of Revascularization Procedures after Coronary Angiography." *Journal of the American Medical Association* 269 (20): 2642–6.
- Bates, N., E. A. Martin, T. J. DeMaio, and M. de la Puente. 1995. "Questionnaire Effects on Measurements of Race and Spanish Origin." *Journal of Official Statistics* 11 (4): 433–59.
- Bitton, A., A. M. Zaslavsky, and J. Z. Ayanian. 2010. "Health Risks, Chronic Diseases, and Access to Care among US Pacific Islanders." *Journal of General Internal Medicine* 25 (5): 435–40.
- Deming, W. E., and F. F. Stephan. 1940. "Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals Are Known." *Annals of Mathematical Statistics* 11 (4): 427–44.
- Eicheldinger, C., and A. Bonito. 2008. "More Accurate Racial and Ethnic Codes for Medicare Administrative Data." *Health Care Financing Review* 29 (3): 27–42.
- Elliott, M. N., A. Fremont, P. A. Morrison, P. Pantoja, and N. Lurie. 2008. "A New Method for Estimating Race/Ethnicity and Associated Disparities Where Administrative Records Lack Self-Reported Race/Ethnicity." *Health Services Research* 43 (5): 1722–1736.
- Elliott, M. N., P. A. Morrison, A. Fremont, D. F. McCaffrey, P. Pantoja, and N. Lurie. 2009. "Using the Census Bureau's Surname List to Improve Estimates of Race/Ethnicity and Associated Disparities." *Health Services Outcomes and Research Methodology* 9 (2): 69–83.
- Elwert, F., and N. A. Christakis. 2006. "Online Supplement to: Widowhood and Race." *American Sociological Review* 71: 16–41.
- Escarce, J. J., and T. G. McGuire. 2003. "Methods for Using Medicare Data to Compare Procedure Rates among Asians, Blacks, Hispanics, Native Americans, and Whites." *Health Services Research* 38 (5): 1303–17.
- Fiscella, K., and A. M. Fremont. 2006. "Use of Geocoding and Surname Analysis to Estimate Race and Ethnicity." *Health Services Research* 41 (4 Pt 1): 1482–500.
- Goldstein, E., P. D. Cleary, K. M. Langwell, A. M. Zaslavsky, and A. Heller. 2001. "Medicare Managed Care CAHPS: A Tool for Performance Improvement." *Health Care Financing Review* 22 (3): 101–7.
- Guadagnoli, E., J. Z. Ayanian, G. Gibbons, B. J. McNeil, and F. W. LoGerfo. 1995. "The Influence of Race on the Use of Surgical Procedures for Treatment of Peripheral Vascular Disease of the Lower Extremities." *Archives of Surgery* 130 (4): 381–6.
- Humes, K. R., N. A. Jones, and R. R. Ramirez. 2011. *Overview of Race and Hispanic Origin: 2010*. Census Brief C2010BR-02. Suitland, MD: US Bureau of the Census.
- Liebersohn, S., and M. C. Waters. 1993. "The Ethnic Responses of Whites: What Causes Their Instability, Simplification and Inconsistency?" *Social Forces* 72: 421–50.

- Martin, E., M. de la Puente, and C. Bennett. 2001. *The Effects of Questionnaire and Content Changes on Responses to Race and Hispanic Origin Items: Results of Replication of the 1990 Census Short Form in Census 2000*. ASA Proceedings of the Joint Statistical Meetings, American Statistical Association, Alexandria, VA.
- McBean, A. M. 2006. *Improving Medicare's Data on Race and Ethnicity*. Medicare Brief. Washington, DC: National Academy of Social Insurance.
- McBean, A. M., Z. Huang, B. A. Virnig, N. Lurie, and D. Musgrave. 2003. "Racial Variation in the Control of Diabetes among Elderly Medicare Managed Care Beneficiaries." *Diabetes Care* 26 (12): 3250–6.
- McCaffrey, D. F., and M. N. Elliott. 2008. "Power of Tests for a Dichotomous Independent Variable Measured with Error." *Health Services Research* 43 (3): 1085–101.
- Morgan, R. O., I. I. Wei, and B. A. Virnig. 2004. "Improving Identification of Hispanic Males in Medicare: Use of Surname Matching." *Medical Care* 42 (8): 810–6.
- Morning, A. 2005. "Multiracial Classification on the United States Census: Myth, Reality, and Future Impact." *Revue Européenne des Migrations Internationales* 21: 111–34.
- Office of Management and Budget. 1997. *Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity*. Washington, DC: US Government Printing Office.
- O'Malley, A. J., K. Walsh, and A. M. Zaslavsky. 2011. "Chapter 5: Outline of Medicare CAHPS Statistical Analyses for 2010 Reports." In *2010 Medicare CAHPS Technical Report*, edited by N. Orr, pp. 75–104. Santa Monica, CA: The RAND Corporation.
- Purcell, N. J., and L. Kish. 1980. "Postcensal Estimates for Local Areas (or Domains)." *International Statistical Review* 48 (1): 3–18.
- Schenker, N., and J. D. Parker. 2003. "From Single-Race Reporting to Multiple-Race Reporting: Using Imputation Methods to Bridge the Transition." *Statistics in Medicine* 22 (9): 1571–87.
- Schneider, E. C., A. M. Zaslavsky, and A. M. Epstein. 2002. "Racial Disparities in the Quality of Care for Enrollees in Medicare Managed Care." *Journal of the American Medical Association* 287 (10): 1288–94.
- Scott, C. G. 1999. "Identifying the Race or Ethnicity of SSI Recipients." *Social Security Bulletin* 62 (4): 9–20.
- Swan, J., N. Breen, L. Burhansstipanov, D. E. Satter, W. W. Davis, T. McNeel, and C. M. Snipp. 2006. "Cancer Screening and Risk Factor Rates among American Indians." *American Journal of Public Health* 96 (2): 340–50.
- Trivedi, A. N., A. M. Zaslavsky, E. C. Schneider, and J. Z. Ayanian. 2005. "Trends in the Quality of Care and Racial Disparities in Medicare Managed Care." *New England Journal of Medicine* 353 (7): 692–700.
- . 2006. "Relationship between Quality of Care and Racial Disparities in Medicare Health Plans." *Journal of the American Medical Association* 296 (16): 1998–2004.

- Ulmer, C., B. McFadden, and D. T. Nerenz. 2009. *Race Ethnicity, and Language Data: Standardization for Health Care Quality Improvement*. Washington, DC: National Academies Press.
- Virnig, B. A., N. Lurie, Z. Huang, D. Musgrave, A. M. McBean, and B. Dowd. 2002. "Racial Variation in Quality of Care among Medicare+Choice Enrollees." *Health Affairs* 21 (6): 224–30.
- Virnig, B., Z. Huang, N. Lurie, D. Musgrave, A. M. McBean, and B. Dowd. 2004. "Does Medicare Managed Care Provide Equal Treatment for Mental Illness across Races?" *Archives of General Psychiatry* 61 (2): 201–5.
- Waldo, D. R. 2004. "Accuracy and Bias of Race/Ethnicity Codes in the Medicare Enrollment Database." *Health Care Financing Review* 26 (2): 61–72.
- Wei, I. I., B. A. Virnig, D. A. John, and R. O. Morgan. 2006. "Using a Spanish Surname Match to Improve Identification of Hispanic Women in Medicare Administrative Data." *Health Services Research* 41 (4 Pt 1): 1469–81.
- Zaborski, L. B., and A. M. Zaslavsky. 2011. "Appendix 2.01: Individual-Level Respondent/Nonrespondent Weight Construction." In *2010 Medicare CAHPS Technical Report*, edited by N. Orr, pp. 133–146. Santa Monica, CA: The RAND Corporation.
- Zaslavsky, A. M., and P. D. Cleary. 2002. "Dimensions of Plan Performance for Sick and Healthy Members on the Consumer Assessments of Health Plans Study 2.0 Survey." *Medical Care* 40 (10): 951–64.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Table A1: Joint Appearance of Racial/Ethnic Selections in CAHPS Responses.

Table A2: Distribution of CAHPS Race/Ethnicity Selections by EDB Race/Ethnicity.

Table A3: Distribution of EDB Race/Ethnicity by CAHPS Race/Ethnicity Selections.

Table A4: American Indian and Alaskan Native Population for Age-Matched Cohorts in U.S. Census, 1980–2000.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.