

Improving Medicare's Data on Race and Ethnicity

By A. Marshall McBean

Summary

Medicare's databases provide a rich source of information about the program's 43 million beneficiaries. These data have played an important role in documenting racial, ethnic, and socioeconomic disparities in health and health care. Because they derive largely from administrative records that have been collected over many years using varying standards, however, they are not fully adequate for monitoring and reducing disparities.

The Centers for Medicare & Medicaid Services (CMS) has supported a number of initiatives to improve the quality of its data on race and ethnicity. Yet analyses of 2002 Medicare administrative data show that only 52 percent of Asian beneficiaries and 33 percent of both Hispanic and American Indian/Alaska Native beneficiaries were identified correctly. As CMS moves to reduce disparities, and as researchers strive to explain how and why disparities occur, further improvements in Medicare's data are essential. Health care organizations also need data on the race and ethnicity of the people they serve in order to improve the quality of care for minorities. This brief provides some recommendations for further efforts.

Research using Medicare's data on enrollment and utilization has enhanced our understanding of the Medicare population and allowed us to monitor beneficiaries' health, health care utilization, and quality of care. These databases have also been instrumental in highlighting racial, ethnic, and socioeconomic disparities, particularly in the use of preventive services, outpatient and inpatient surgical procedures, hospital admissions, and outcomes of treatment (Gornick *et al.* 1996; McBean 2005). Nevertheless, the quality of Medicare's data on race and ethnicity remains much better for whites and blacks than it is for Hispanics, Asians, and American Indians/Alaska Natives.

Medicare's data on race and ethnicity derive from Social Security's administrative records and are collected on a consistent basis only at the time of application for a Social Security number. This process does not allow for updating the data when the Office of Management and Budget promulgates new standards for categorizing race and ethnicity. Also, since the late 1980s, most applications for Social Security numbers are made through the Enumeration at Birth process, which does not provide data on race and ethnicity. Finally, Medicare health plans (which enroll 12 percent of beneficiaries) are not required to collect or report information regarding the race and ethnicity of plan members.

A. Marshall McBean is a professor in the School of Public Health at the University of Minnesota. This brief is drawn from his paper, *Medicare Race and Ethnicity Data*, which is available in its entirety on NASI's website

Medicare's Primary Datasets

Medicare's administrative databases comprise two types of information: enrollment data and utilization data. Enrollment data (consisting of eligibility and demographic information) are maintained for every person ever enrolled in Medicare. Utilization data are maintained for all claims sent for payment. As claims are processed, data from the enrollment database are incorporated into the claims record; these data make up the Medicare claims files. Annual utilization files based on claims files are created for each type of covered service. These files are available to and are used by researchers who study issues of race and ethnicity.

The enrollment database is also used to create an annual denominator file. This file facilitates the development of Medicare statistics by providing numbers of enrollees. For example, rates of health care utilization (or costs) can be calculated using the numerator counts of services (or costs) and the denominator counts of enrollees. The uniformity of the race/ethnicity information across Medicare datasets is assured because the enrollment database is used to populate both the denominator file and the utilization files. This process gives Medicare data an advantage over many other health datasets that use an external data source (such as Census data) as the denominator. It also provides an additional incentive for the Centers for Medicare & Medicaid Services (CMS) to improve the validity of the data.

How the Data Are Collected

When Medicare was established in 1965, the Social Security Administration (SSA) was given managerial responsibility. To this day, SSA remains responsible for certifying that an individual is eligible for Medicare and for transmitting demographic information about that individual to CMS. (Former railroad workers enroll through the Railroad Retirement Board, which does not collect information on race and ethnicity.)

The flow of information from SSA to CMS is as follows. When most current Medicare beneficiaries applied for a Social Security card and number, they completed Form SS-5, which included information on their date of birth, address, and race/ethnicity. When they applied for Social Security or Medicare benefits, SSA used the information on the Form SS-5 to create a file known as the Master Beneficiary Record (MBR). The MBR contains a record for every person awarded Social Security benefits. SSA uses the MBR to transmit electronically to CMS a record for each new Medicare beneficiary. In turn, CMS creates a record in its enrollment database for each beneficiary with eligibility and demographic information.

From 1936 (the beginning of the Social Security system) until 1980, Form SS-5 remained virtually unchanged; it had only three boxes to specify race: white, black, and other. If race was left unchecked, the MBR categorized the person as of unknown race. It is important to recognize that the vast majority of current Medicare beneficiaries applied for a Social Security card before 1980. For them, the race/ethnicity information in Medicare's enrollment database is generally limited to the information in the three boxes that were on Form SS-5 when they applied for a Social Security number. After being assigned a Social Security number, a person submits a new Form SS-5 only when seeking a replacement Social Security card or changing personal information (for example, changing a name by marriage).

In 1980 Form SS-5 was revised to comply with new Office of Management and Budget (OMB) standards promulgated in 1977. As a result of these standards, the "other" category was replaced by three new categories: (1) Asian, Asian American, or Pacific Islander; (2) Hispanic; and (3) American

Indian or Alaskan Native. However, agencies were not required to collect information on both race and ethnicity. Although other agencies captured and retained this information through the use of two separate questions (one about race and another about ethnicity), SSA did not.

Even though the revised Form SS-5 began to collect expanded race/ethnicity information in 1980, the MBR did not establish fields to accommodate this new information. Therefore, when SSA creates an MBR record from Form SS-5 (with the three new race/ethnicity fields), it collapses the new race/ethnicity data into a single “other” field.

SSA has also established another electronic file, the Master File of Social Security Number Holders and Applications file, known as the Numerical Identification (NUMIDENT) file, which uses information from Form SS-5 to create a record of each Social Security number ever issued. Fortunately, when Form SS-5 was revised in 1980, the NUMIDENT file was modified to accommodate the expanded race/ethnicity codes. Although the regular system for transferring information from SSA to CMS still uses the MBR (with its limited white, black, other, and unknown categories) to establish a record for each beneficiary in the Medicare enrollment database, the NUMIDENT file with its expanded race/ethnicity information can be used periodically to update CMS’ enrollment database.

Another recent development will create data problems for Medicare down the road. Since 1989, most newborn infants have been assigned Social Security numbers through a process known as Enumeration at Birth. With parental permission, the state’s vital statistics office transmits data from the child’s birth certificate to SSA, which issues a Social Security card and number. Under current arrangements, however, SSA receives no information on the child’s race and ethnicity, because that information is considered to be “for medical and health use only” (Scott 1999). In just a few years, some people enumerated at birth will become eligible for Medicare on account of disability, and information on their race and ethnicity will be lacking. Under current arrangements, SSA must negotiate separately with each state to change the Enumeration at Birth process.

In 1997 the Office of Management and Budget made further revisions to the federal standards for data on race and ethnicity. The new standards have five racial categories—white, black or African American, Asian, American Indian or Alaska Native, and Native Hawaiian or other Pacific islander—and a person may report more than one race. There are also two separate ethnic categories—Hispanic or Latino, and not Hispanic or Latino. The 2000 Census used these new categories, but the Social Security Administration has not yet incorporated them into Form SS-5 or its databases (Scott 1999).

In summary, the methods used to create the Medicare enrollment database have caused two underlying problems. First, because most Medicare beneficiaries filled out Form SS-5 before 1980, information from that source on race/ethnicity for most beneficiaries is limited to white, black, other, and unknown. Second, because the revised Form SS-5 does not have separate fields for race and ethnicity, precision regarding race and ethnicity is compromised.

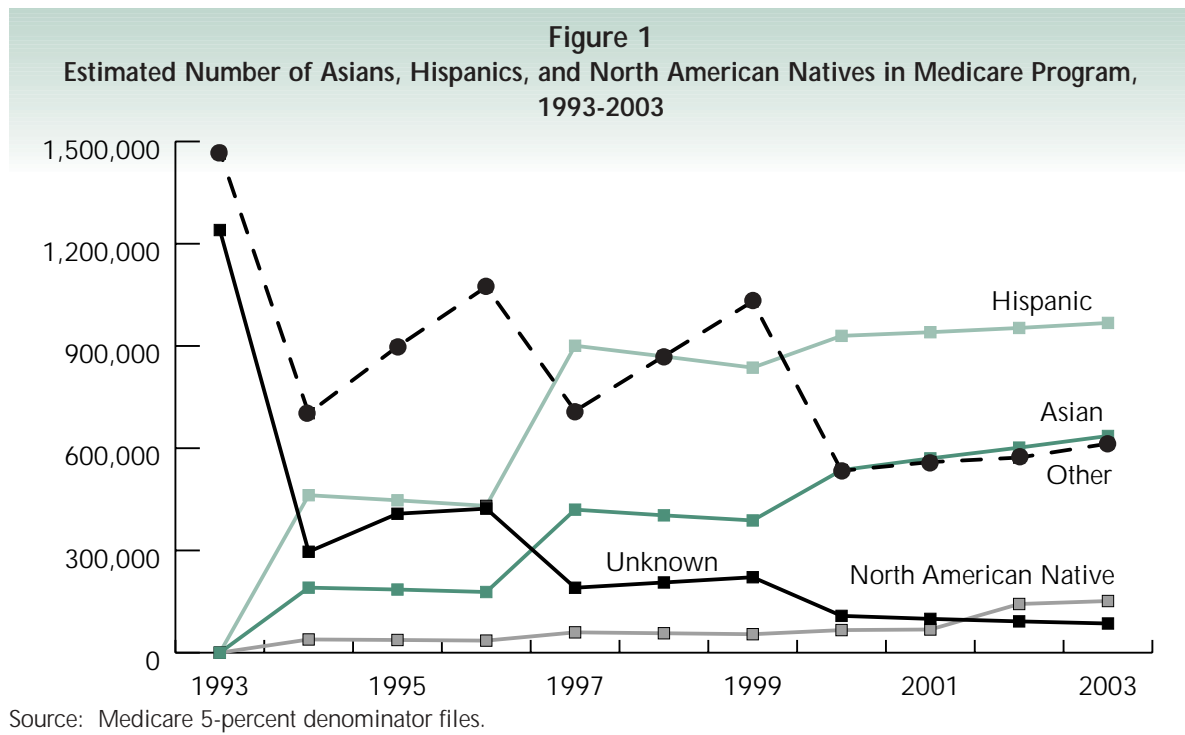
CMS Efforts to Improve the Quality of the Data

CMS has worked to improve the quality of the race/ethnicity data in its enrollment database through annual NUMIDENT updates, quarterly updates on American Indians and Alaska Natives from the files of the Indian Health Service, and mailings requesting a self-report of race. The results of these initiatives are evident in changes in the reported number of beneficiaries by race and ethnicity, as well as in comparisons of the enrollment database with data from the Census Bureau and the Medicare Current Beneficiary Survey.

CMS has made efforts to expand the race/ethnicity information in its databases using SSA's NUMIDENT information. In 1994, NUMIDENT race information for all current Medicare beneficiaries—35 million persons—was incorporated into the enrollment database. In 1997, the enrollment database was updated again using the NUMIDENT file.

In 1997 CMS also initiated a mailing to those beneficiaries classified as other or unknown, and to those with a Hispanic surname (as identified by a Census Bureau algorithm), requesting a self-report of race/ethnicity. CMS sent out 2.2 million letters, and 858,000 people responded. The mailing's impact is seen in the redistribution of persons from the other and unknown categories, which decreased by 55 percent and 34 percent, respectively, to the Asian, Hispanic, and American Indian/Alaska Native categories, which increased by 136 percent, 109 percent, and 68 percent from 1996 to 1997.

The Indian Health Service began providing CMS with information on the people it serves in 1999. In 1999, 54,000 beneficiaries were classified as American Indian/Alaska Natives; by 2003, 152,000 such beneficiaries were identified in the enrollment database. In addition, a third NUMIDENT update was conducted in late 2000, with updates done yearly since then. By now, few Medicare beneficiaries are recorded as having unknown race, although the other category remains stubbornly large (see Figure 1).



Further Improvements Needed

The accuracy of Medicare's race/ethnicity information can be measured at both the population level and the individual level. Aggregate measures compare Medicare race/ethnicity estimates with those of the Census Bureau. In 1998, there was high comparability for whites (99 percent of the Census Bureau estimates) and acceptable comparability for blacks (91 percent). However, Medicare's race/ethnicity estimates were much lower for Hispanics, Asians, and Native Americans. Medicare's estimate of the number of Hispanics was only 29 percent of that found by the Census Bureau, 42 percent for Asians, and 24 percent for Native Americans (Eggers and Greenberg 2000).

At the individual level, the standard of reference is the self-reported information from the Medicare Current Beneficiary Survey (MCBS), a continuous, multipurpose survey of a nationally representative sample of aged and disabled persons enrolled in Medicare. Table 1 compares the race/ethnicity information in the enrollment database with that of the MCBS. The validity of the individual data is high for whites and blacks but much lower for Asians, Hispanics, and American Indians/Alaska Natives.

Table 1 presents two measures for assessing the validity of the enrollment database's race/ethnicity data: the sensitivity (the probability that the enrollment database correctly identifies a person of a given race/ethnicity) and the positive predictive value, or PPV (the probability that a person identified by the enrollment database as being of a certain race/ethnicity is actually of that race/ethnicity).

Table 1
Validity of Enrollment Database Race/Ethnicity Classifications Compared to the Medicare Current Beneficiary Survey (in percents)

Race/Ethnicity	Sensitivity			Positive Predictive Value		
	1996	1997	2002	1996	1997	2002
White	97	97	97	98	98	98
Black	96	95	96	96	96	97
Asian	20	58	52	66	79	83
Hispanic	19	39	33	98	98	95
American Indian/ Alaska Native	3	11	33	57	78	92
Other	14	8	7	9	10	9

Sources: 1996 and 1997 estimates from Arday 2000; 2002 estimates from McBean 2005.

Prior to its updating in 1997, the sensitivity of the enrollment database for identifying people who were Asian, Hispanic, or American Indian/Alaska Native was less than or equal to 20 percent, meaning that less than one-fifth of people of these race/ethnicities were identified. In contrast, the sensitivity for whites and blacks was over 95 percent. Table 1 also shows the gains made in the data's validity between 1996 and 2002. Over that period, the sensitivity increased from 19 percent to 33 percent for Hispanics, from 20 percent to 52 percent for Asians, and from 3 percent to 33 percent for American Indian/Alaska Natives.

The high PPVs for whites, blacks, and Hispanics suggest there is not much opportunity for misclassification in these groups. In contrast, the lower PPV for Asians implies the possibility of significant errors in estimates of events (for example, health care utilization or post-treatment mortality) for this group.

Despite the importance of using the NUMIDENT and Indian Health Service files to update the enrollment database, it appears that a leveling off has been reached in the validity of race/ethnicity coding of Medicare beneficiaries. For Asians and Hispanics, no gains in validity have been achieved since 1997. Not surprisingly, additional matches to the same files have produced little additional improvement in the quality of the data.

CMS continues to look for ways of more accurately identifying the race and ethnicity of Medicare beneficiaries. One recent study funded by CMS, for example, developed algorithms to classify beneficiaries using surname and other information available on the enrollment database. It found that surname matching significantly improved the classification of Hispanic and Asian/Pacific Islander beneficiaries (Bonito, Eicheldinger, and Evensen 2005). Other studies of surname matching have had similar findings (Wei *et al.* 2006).

Racial misclassification matters because it can alter the results of studies that compare health care use or health outcomes between racial and ethnic groups. Escarce and McGuire have shown the potential bias in the results of studies comparing treatment rates (Escarce and McGuire 2003). Using

Medicare data, they calculated the rates at which members of different racial/ethnic groups used 15 cardiac services—with and without correction for misreporting. For whites, blacks, and Hispanics, the corrected rates were close to the reported rates. For Asians and Native Americans, however, the corrected rates were generally lower than the reported rates, in some cases substantially. “The main reason for this,” the authors state, “is the lower PPV for these groups. Non-Asians and non-Native Americans are more frequently mixed in with these groups, and the mix-ins have higher rates.” (Escarce and McGuire 2003). Using data from the Indian Health Service, Rhoades has illustrated how classification errors can conceal racial disparities in mortality (Rhoades 2005).

Collection of Data by Health Care Organizations

Health plans, hospitals, and other health care organizations need data on the race and ethnicity of the people they serve in order to monitor and improve the quality of care for minority groups. Yet processes for collecting these data are inconsistent across health care organizations, and some organizations mistakenly believe that it is illegal to collect information on race and ethnicity from patients (Lurie and Fremont 2006; Hasnain-Wynia and Baker 2006). An analysis by legal scholars at the George Washington University affirms that collecting such data—when conducted as part of a program to raise the quality of health care—does not violate state or federal laws or increase the risk of race-based malpractice claims. The authors of the study call on the Department of Health and Human Services (HHS) to issue guidelines for the uniform and appropriate collection of data on race and ethnicity (Rosenbaum *et al.* 2006).

To date, neither CMS, HHS, the National Committee for Quality Assurance, nor any other organization has required Medicare Advantage plans to collect or report information on the race/ethnicity of their members. Nevertheless, CMS, researchers, and foundations have taken steps to encourage the collection and use of race/ethnicity information. The most comprehensive effort has been the linking of the Health Plan Employer Data and Information Set (HEDIS) information submitted by plans with the race/ethnicity information in the enrollment database.

As with analyses of the fee-for-service population, studies of members of Medicare Advantage plans have found significantly lower rates of diabetes care, mental health, cancer screening, and other recommended services for blacks compared with whites, as well as frequently showing lower rates among Hispanics compared with whites (Schneider, Zaslavsky, and Epstein 2002; Virnig *et al.* 2002; McBean 2005). While no validation of the race/ethnicity coding in the enrollment database has been carried out for members of Medicare health plans, it seems plausible that its validity for this group is similar to that for the entire Medicare population, as shown in Table 1.

Recommendations

Currently, Medicare’s identification of minority beneficiaries and the validity of this information have reached a plateau. While analyses of Medicare’s databases have proved extremely useful, better data are essential for understanding and reducing racial, ethnic, and socioeconomic disparities in health and health care.

Many recommendations have been made for improving the quantity and quality of Social Security and Medicare’s data on race and ethnicity (General Accounting Office 2002; National Research Council 2004; McBean 2005; Wei *et al.* 2006; Rosenbaum *et al.* 2006). The following deserve particular consideration:

- The Social Security Administration should collect information on race and ethnicity on Form SS-5 and through the Enumeration at Birth process using the 1997 OMB standards.

- The Social Security Administration and the Centers for Medicare & Medicaid Services should develop a program to collect data on race, ethnicity, and socioeconomic position at the time of enrollment in Medicare and for current enrollees.
- CMS should reexamine the feasibility, costs, and effectiveness of conducting another mailing to persons with race/ethnicity coded as other or unknown, as well as to beneficiaries selected on the basis of surname or place of residence.
- CMS should evaluate the findings of recent studies of the effectiveness of using lists of surnames to identify the race and ethnicity of enrollees.
- HHS should issue guidelines for the uniform and appropriate collection of data on race and ethnicity by health care organizations.
- Medicare Advantage plans should be required to collect and report to CMS the race and ethnicity of all enrolled Medicare members.

Medicare's data are a powerful tool for reducing racial and ethnic health disparities. As gains in the quality of Medicare's data on race and ethnicity have stalled, however, new tactics are needed. In the meantime, the imperfections of the current data should not prevent their use in studies and interventions designed to understand and alleviate racial and ethnic disparities.

References

- Arday, Susan L., David R. Arday, Stephanie Monroe, and Jianyi Zhang. 2000. "HCFA's Racial and Ethnic Data: Current Accuracy and Recent Improvements," *Health Care Financing Review*, 21(4): 107-16.
- Bonito, Arthur J., Celia R. Eicheldinger, and Christian Evensen. 2005. *Health Disparities: Measuring Health Care Use and Access for Racial/Ethnic Populations*. Research Triangle Park: RTI International. http://www.cms.hhs.gov/reports/downloads/Bonito_Final.pdf#search=%22bonito%20eicheldinger%20rti%22. Accessed September 27, 2006.
- Eggers, Paul W., and Linda G. Greenberg. 2000. "Racial and Ethnic Differences in Hospitalization Rates Among Aged Medicare Beneficiaries," *Health Care Financing Review*, 21(4): 91-105.
- Escarce, Jose J., and Thomas 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.
- General Accounting Office. 2002. *SSA Disability Decision Making: Additional Measures Would Enhance Agency's Ability to Determine Whether Racial Bias Exists*. Washington. Report GAO-02-831.
- Gornick, Marian E., Paul W. Eggers, Thomas W. Reilly, Renee M. Mentnech, Leslye K. Fitterman, Lawrence E. Kucken, and Bruce C. Vladeck. 1996. "Effects of Race and Income on Mortality and Use of Services among Medicare Beneficiaries," *New England Journal of Medicine*, 335(11): 791-9.
- Hasnain-Wynia, Romana, and David W. Baker. 2006. "Obtaining Data on Patient Race, Ethnicity, and Primary Language in Health Care Organizations: Current Challenges and Proposed Solutions," *Health Services Research*, 41(4): 1501-18.
- Lurie, Nicole, and Allen Fremont. 2006. "Looking Forward: Cross-Cutting Issues in the Collection and Use of Racial/Ethnic Data," *Health Services Research*, 41(4): 1519-33.
- McBean, A. Marshall. 2005. *Medicare Race and Ethnicity Data*. Washington: National Academy of Social Insurance. http://www.nasi.org/usr_doc/McBean.pdf.
- National Research Council. 2004. *Eliminating Health Disparities: Measurement and Data Needs*. Washington: National Academies Press.

References continued on back

References continued

Rhoades, Dorothy A. 2005. "Racial Misclassification and Disparities in Cardiovascular Disease among American Indians and Alaska Natives," *Circulation*, 111(10): 1250-56.

Rosenbaum, Sara, Taylor Burke, Sonia W. Nath, Jennifer Santos, and Dana Thomas. 2006. *The Legality of Collecting and Disclosing Patient Race and Ethnicity Data*. Washington: George Washington University for the Robert Wood Johnson Foundation. <http://www.rwjf.org/files/publications/other/RaceEthnicDisparitiesData06222006.pdf>. Accessed September 27, 2006.

Scott, Charles G. 1999. "Identifying the Race or Ethnicity of SSI Recipients," *Social Security Bulletin*, 62(4): 9-20.

Schneider, Eric C., Alan M. Zaslavsky, and Arnold 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.

Virnig, Beth A., Nicole Lurie, Zhen Huang, Dorothea Musgrave, A. Marshall McBean, and Bryan Dowd. 2002. "Racial Variation in Quality of Care among Medicare+Choice Enrollees," *Health Affairs*, 21(6): 224-30.

Wei, Iris I., Beth A. Virnig, Dolly A. John, and Robert O. Morgan. 2006. "Using a Spanish Surname Match to Improve Identification of Hispanic Women in Medicare Administrative Data," *Health Services Research*, 41(4): 1469-81.

This brief was prepared for the National Academy of Social Insurance's study panel on Medicare and racial and ethnic disparities. The Robert Wood Johnson Foundation provided the primary financial support for the project. Additional funding was provided by The California Endowment and the Joint Center for Political and Economic Studies.



NATIONAL
ACADEMY
OF SOCIAL
INSURANCE

1776 Massachusetts Avenue, NW Suite 615
Washington, DC 20036-1904
(202) 452-8097
(202) 452-8111 Fax
nasi@nasi.org
www.nasi.org

The National Academy of Social Insurance is a nonprofit, nonpartisan organization made up of the nation's leading experts on social insurance. Its mission is to promote understanding and informed policymaking on social insurance and related programs through research, public education, training, and the open exchange of ideas.
