

# **Medicare Race and Ethnicity Data**

December 2004

Prepared for the Study Panel on  
Sharpening Medicare's Tools to Reduce Racial and Ethnic Disparities  
National Academy of Social Insurance

By

Marshall McBean, M.D., M.Sc.  
3407 West 28<sup>th</sup> Street  
Minneapolis, MN 55416

## **Table of Contents**

### **Executive Summary**

### **1.0 The collection of data by the Centers for Medicare and Medicaid Services (CMS) with regard to racial and ethnic disparities**

## **1.1 Medicare's primary datasets**

### **1.2 The Medicare race variable**

**1.2.1 Improving information about American Indians/Alaskan Natives**

**1.2.2 Validity of the Medicare race/ethnicity information**

**1.2.3 Misclassification and its implications**

**1.2.4 Anomaly caused by the Sex and Race of Primary Beneficiary (SROP) variable in SSA's Master Beneficiary Record (MBR) file**

**1.2.5 Summary of current assembly of information on race/ethnicity in the Enrollment Database (EDB)**

**1.2.6 Race and managed care**

**1.2.7 Monitoring and improving race/ethnicity information at CMS and the Department of Health and Human Services (DHSS)**

**1.3 Other information in Medicare datasets; data documentation and data availability**

## **2.0 Linking Medicare datasets with other datasets**

### **2.1 Individual-level linking**

**2.1.1 EDB used as the sampling frame**

**2.1.2 Data from high quality surveys and registries**

**2.1.3 Collecting information from clinical settings or managed care organizations serving Medicare beneficiaries**

**2.1.4 Information from other administrative databases**

**2.1.5 Vital statistics (death data) link**

## **2.2 Group-level linking**

### **2.2.1 Group linking at the zip code or county-level**

### **2.2.2 Surname matching and imputation**

## **3.0 Use of Medicare disparity data**

## **4.0 Limitations of available data and potential opportunities**

## **5.0 Recommendations**

## Executive Summary

Eliminating racial and ethnic disparities in health and health care is one of the two overarching goals of Healthy People 2010. In order to achieve that goal accurate and complete race/ethnicity information is essential. This information is needed to set and monitor national objectives, to formulate policy, to write legislation, and to design programs to address the disparities. Medicare program staff has played a major leadership role in documenting existing disparities and implementing activities to eliminate them. A major reason for this has been the quality of the Medicare data, including race/ethnicity information. In spite of several improvements that have occurred in the past decade regarding Medicare race/ethnicity information, deficiencies remain, particularly with respect to the smaller minority groups (Asians, Hispanics and American Indians/Alaskan Natives). The primary structural issues affecting the Medicare data are (1) the use of race/ethnicity information from the Social Security Administration (SSA) as the primary source for Medicare data, and (2) the lack of a two-item designation method for race/ethnicity at SSA and the Centers for Medicare and Medicaid Services (CMS). In addition, the current issue regarding the quality of the race/ethnicity data is the plateauing of the validity of the existing race/ethnicity data. For example, for 2002, the most recent year that can be analyzed, only 52% of Asian, and 33% of Hispanic, and 33% of American Indian/Alaskan Native Medicare beneficiaries can be correctly identified in the Medicare data. While this is approximately a 3-fold increase since 1997 for Native Americans/Alaskan Natives, it is actually a slight decrease for Asians and Hispanics.

There are certain long-standing impediments to correcting these deficiencies in the near-term: the use of only three (White, Black, Other) race designations on the SSA's SS-5 form prior to 1980; the failure of SSA to capture ethnicity as a separate item on the SS-5 form when the race code options were expanded in 1980; the difficulty of systematically correcting the existing information by Medicare or SSA by contacting beneficiaries for that purpose; and the voluntary nature of the reporting of race. Some improvement in the quantity and quality of race/ethnicity information could be made through the addition to the Medicare databases of race/ethnicity information collected by surveys, disease registries, managed care plans, administrative databases of other organizations, or the imputation of race/ethnicity using surname lists. Medicare should strongly consider augmenting the information in the Enrollment Database (EDB), or its successor, as well as the derivative files, with information from these additional sources. Indicators of the sources of the information would have to be included in the EDB, and a hierarchy established in terms of the perceived validity of each source. Self-reported information from high quality surveys, and managed care, as well as imputation using lists of highly probable Spanish surname lists are recommended. In addition to work at CMS addressing the entire Medicare population, managed care plans have the opportunity to capture and use valid race/ethnicity information. This information should be obtained by the plans and shared with CMS to help the plans and Medicare understand many issues including racial disparities in health and health care among Medicare Advantage (MA) beneficiaries. This should be a very high priority. While improving the race/ethnicity data is very important, the absence of perfect race/ethnicity data must not, however, prevent

the use of Medicare data in studies and interventions designed to understand and alleviate existing disparities.

## **1.0 The collection of data by the Centers for Medicare and Medicaid**

### **Services (CMS) with regard to racial and ethnic disparities**

#### **1.1 Medicare's primary datasets**

Medicare's administrative datasets are comprised of enrollment data and utilization data. Enrollment data on all persons ever enrolled in the Medicare program are stored in the Enrollment Database (EDB). Utilization data are retained in claim-based files. For historical reasons the claims files are often categorized as institutional files and non-institutional files because institutions (hospitals, hospital out-patient departments, home health agencies, nursing homes, etc.) send their claims to one type of initial claims processor, Fiscal Intermediaries (FI), for payment. Other types of providers who are generally not institutions (physicians, ambulatory surgical centers, and others who generally provide Medicare Part B covered services) send their claims to another type of processor, the Carriers, except if they are suppliers of durable medical equipment (DME), in which case they send their claims to the Durable Medical Equipment Regional Carriers (DMERC). The FIs, Carriers and DMERCs, all of which are CMS contractors, enter the claims into the claim payment system, edit them for consistency and utilization errors, calculate payments and may deny claims based on payment rules. Information about the claims processed by the FIs, Carriers and DMERCs are sent to one of nine Common Working File (CWF) regional host sites as the next step of the claim adjudication process. There, the CWF host sites determine whether the beneficiary is entitled to receive the service and whether a deductible applies, check for duplicate claims, return denied claims to the claim processor, and authorize the claim processor to pay

appropriate claims. The claims are then sent on to CMS by the CWF regional host sites to be kept in the national claims history (NCH) repository, or file. While utilization information can be retrieved from the NCH file, most claims-based analyses exploring issues related to race/ethnicity are done using the seven Medicare Standard Analytic files (SAFs). Five of these have titles that generally reflect the origin of the claim, or place of treatment. These are the hospital in-patient file; the skilled nursing facility (SNF) file; the hospital out-patient file, the home health agency file and the hospice file. The last two SAFs, the Carrier file and the Durable Medical Equipment (DME) file are named after the processor of the claims or the materials provided, respectively. The annual SAFs are first created six months after the close of the calendar year, and then updated quarterly four times. In addition to the SAFs, the very frequently used annual Medicare Provider Analysis and Review (MedPAR) file is created by CMS, also. The inpatient MedPAR has an acute care hospital, also known as short-stay, as well as a long-stay hospital version. They are highly valued because they collapse the hospital claims for each hospitalization into one record per hospitalization. Similarly, the MedPAR SNF file has one record per SNF stay. The first versions of the annual MedPAR files are also made available approximately six months after the end of the calendar year. They are updated quarterly for the next two years. In addition to the SAFs which are categorized as research identifiable files, CMS has recently created analogous Limited Data Set (LDS) files for all the SAFs and MedPAR. These files retain the race information in the SAFs, but have missing data or ranged values entered for certain variables to prevent the identification of individual beneficiaries. These files, too, may also be accessed and used by researchers studying race/ethnicity related issues.



In addition to the EDB, enrollment information is also available in the annual Medicare Denominator files which contain demographic and enrollment information on all beneficiaries who were enrolled in Medicare at some time during that year. The Denominator file for 200X is created in March/April of 200X+1 based on information in the EDB as of the March/April of 200X+1 date of creation.

The various enrollment and claims files are available for 5% or 100% of the Medicare population. The 5% files are based on an essentially random sample of claims for 5% of the Medicare population, not a sample of 5% of the claims. This is done by selecting claims for beneficiaries whose claim account number (CAN) portion of the individual's Medicare Health Insurance Claim (HIC) number ends in 5 pre-selected pairs of the last two digits. See the following:

Example of Medicare Health Insurance Claim Number (HIC) = 123 45 67**89**A  
Claim Account Number (CAN) = 123 45 67**89**  
Digits used to select 5% sample = **89**

The digits in bold are used to select the 5% sample. Because the CAN is someone's nine-digit social security number (SSN) and the Social Security Administration (SSA) assigns random numbers for the last four digits of every SSN, the 5% sample is a systematic sample of a random set of numbers which makes it a random sample. In addition, the way the MedPAR files are stored on the CMS mainframe computer makes it easy to select records for 20%, 40%, 60% and 80% samples of Medicare beneficiaries based on similar selection criteria using the last two digits of the CAN, also. The Denominator and

MedPAR files have been created for 100% of the beneficiaries each year since 1986. They and the SAFs are available for 1991 onward.

Information about the race/ethnicity of the beneficiary is contained in each of the previously mentioned files. This information is the information known to CMS and contained in the EDB at the time that a claim is processed by the CWF regional host sites, or when the Denominator file is created. Thus, with only minor perturbations related to updates of the race/ethnicity information described below, the race/ethnicity of the individual indicated in the utilization data (often the numerator in analyses of access to care or the quality and cost of care) and the race/ethnicity in the enrollment data (often the denominator in such studies) are the same. This is a great advantage that the Medicare data has over other health and healthcare information systems such as the U.S. National Vital Statistics System of the National Center for Health Statistics, the National Inpatient Sample (NIS) of the Healthcare Utilization Project (HCUP) and the National Hospital Inpatient Survey (NHIS) of the Agency for Healthcare Research and Quality (AHRQ). All of these data sets or systems use the U.S. Census Bureau data, which is an external source, to estimate their denominator, including the racial and ethnic composition of the population of interest. Thus, because of this added precision in the Medicare data, there is even more incentive to have the correct race/ethnicity information in the Medicare administrative data files.

As will be further described below, the race/ethnicity information retained by CMS in the administrative data files is a one byte variable called “race”. No information is collected on other socio-demographic information such as income, educational attainment, country

of birth, language spoken at home, language of preference, etc. All of the above files, however, contain the county and zip code of residence of each beneficiary which allows the linking of files with zip code or county-level information to each Medicare beneficiary's information. This is frequently done and will be discussed more below. In a similar vein, matching of Medicare data at the individual level with other files is currently possible using the SSN. Individual-level matches using the name of the beneficiary, although performed in the past by CMS, are currently not done.

There are no CMS datasets containing reliable claims-based information regarding Medicare Advantage (MA) beneficiaries' health service utilization for use in studies of issues related to race and ethnicity. The Prospective In-patient Payment (PIP-DCG) file that was created for the years 2000-2003 had information on hospitalizations of MA beneficiaries, but it was not available to outside researchers. No Medicare Part B utilization data on MA enrollees is collected by CMS. The new (2004) CMS-Hierarchical Condition Categories methodology for reimbursing managed care plans only collects the diagnosis of each managed care enrollee, as well as source of the diagnoses and the date of one of the records that contains that diagnosis. Hence, studies regarding the use of services, quality of care, etc. among different race/ethnicity groups enrolled in managed care cannot be done using Medicare's claims data.

## **1.2 The Medicare race variable**

Medicare entitlement and enrollment data, including race/ethnicity information, are transferred electronically from the Social Security Administration (SSA) and the Railroad

Retirement Board (RRB) to CMS and used in the creation of the Enrollment Database (EDB). The RRB does not collect information on race/ethnicity; therefore, many of the enrollment records involving Unknown race are RRB beneficiaries. Historically, the majority of enrollment record information transferred from SSA has been derived from the Master Beneficiary Record (MBR) File. As is explained below, this information is now supplemented with information from the SSA's Master File of Social Security Number Holders and Applications file, better known as the Numerical Identification (NUMIDENT) file, as well as information from the Indian Health Service (IHS).

The SSA collects race/ethnicity information on an application form called the SS-5. From 1936 to 1980, the SS-5 form included three categories for race (White, Black, and Other). Beginning in November 1980, the race choices on the SS-5 form were expanded in order to comply with Office of Management and Budget Directive 15, Race and Ethnic Standards for Federal Statistics and Administrative Reporting. As a result of Directive 15, the Other category on the SS-5 form was replaced by the following three categories: Asian, Asian American, or Pacific Islander; Hispanic; and Northern American Indian or Alaskan Native. Although preferred, Directive 15 did not require agencies to separately collect information on race and ethnicity. That is, Hispanic may be listed as a choice for a single race/ethnicity question on a questionnaire or housed as a single variable in a database, or it may be asked and retained separately as ethnicity information. Other agencies have chosen the preferred method of capturing and retaining race and ethnicity information through the use of two separate questions; one asking about Hispanic ethnicity and one asking about race. SSA has not. Finally, the response to the race

question on the SS-5 form has, except for a brief period in the 1980's, always been voluntary.

The SS-5 form information has been, and currently is, transferred to up to four SSA databases used for program administration depending on the SSA's needs. These are the Master Beneficiary Record (MBR) file, the NUMIDENT file, the Master Earnings file (MEF), and the Supplemental Security Record (SSR) file. The MEF and SSR files do not influence the race/ethnicity information in the CMS databases and are not discussed here.

The MBR is the principal SSA administrative file which has been used since 1936. A MBR record is created once a decision is made to award benefits to an individual, and it stores entitlement and payment data. There is one record for each Claim Account Number (CAN) in the MBR. It has always contained the three race categories of the SS-5 form (White, Black and Other), as well as the category Unknown for those on whom information was missing. Race information (White/Black/Other) derived from the SS-5 form is stored in the MBR variable "Sex and Race of Primary (SROP)". The MBR does not have a variable called "race" or "race of beneficiary". The only race information is in the SROP variable. The implications of this information being the race/ethnicity of the primary wage earner are discussed later.

Between 1973 and 1979, the SSA converted the SS-5 records it had into a computer file, the NUMIDENT. This file contains one record for every SSN ever issued. Updates of information (name change, etc.) for each SSN are stored with it as consecutively ordered

“entries”. According to Lauderdale and Goldberg (1996), and Schmulowitz (J. Schmulowitz, personal communication, September 2004) the conversion resulted in the loss of race information collected on the SS-5 form for up to 20% of the population (Lauderdale and Goldberg), approximately 80 million forms (Schmulowitz). However, because that information was in the MBR, and it occurred prior to the expansion of the race code, there was no loss to CMS, even though there are many records in the NUMIDENT file for which race/ethnicity information are not available.

The NUMIDENT file assumed great importance in 1980 because it was modified at the same time as the SS-5 form to accept the three additional race/ethnicity codes: Asian, Asian American or Pacific Islander; Hispanic; and American Indian or Alaskan Native. The NUMIDENT file also identifies the city/county, state and country of birth.

Unfortunately, while the SSA expanded the race codes in the NUMIDENT file to correspond to the expanded codes in the new SS-5 form introduced in 1980, the SSA did not restructure the Master Beneficiary Record database or the SROP variable to accommodate these new choices; instead the new race categories were collapsed into one of the original 3 choices (White/Black/Other), most frequently, Other. Thus, prior to 1980, Hispanics who did not think of their Hispanic identity as racial would most likely have checked White or Black. After 1980, when Hispanic became an option, they may have checked Hispanic, but it was collapsed into the Other category in the MBR. As a result, Hispanic individuals historically constituted a large portion of the Other category in the EDB. (Lauderdale and Goldberg, 1996).

In 1994, the Health Care Financing Administration (HCFA) expanded the race/ethnicity code information in its databases using the NUMIDENT information. NUMIDENT race information for all Medicare beneficiaries who were alive in October-November 1994, approximately 35 million persons, was incorporated into the EDB. The impact of this update can be seen in the number of people in the race groups Asian, Hispanic and American Indian/Alaskan Native in the 1994 Denominator file for 1994 (Table 1a).

**Table 1a. Number of Medicare beneficiaries, by race/ethnicity in July, 1995 through 2003**

Race/ethnicity	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
Unknown	296,180	407,740	422,680	190,600	206,060	221,420	108,040	99,240	91,860	85,640
White	31,967,020	32,240,340	32,545,880	32,711,520	32,898,200	33,028,240	33,726,340	33,999,660	34,262,700	34,700,200
Black	3,292,060	3,349,780	3,406,740	3,473,400	3,534,700	3,590,740	3,732,140	3,807,800	3,878,500	3,972,100
Other	702,660	898,080	1,073,480	709,480	870,140	1,030,960	533,780	559,300	572,540	613,320
Asian	190,820	185,300	178,040	419,740	402,960	388,020	535,700	570,140	601,740	635,880
Hispanic	462,080	447,100	430,780	900,700	869,320	836,400	929,660	940,360	952,860	967,640
American Indian/ Alaskan Native	38,800	37,280	35,440	59,680	57,020	54,240	66,340	68,020	142,880	151,600
Total	36,949,620	37,565,620	38,093,040	38,465,120	38,838,400	39,150,020	39,632,060	40,044,520	40,503,080	41,126,380

Source = 5% Denominator files; number = number in 5% file with Part A or Part B coverage in July of the year X 20.

**Table 1b. Percentage change in the number of Medicare beneficiaries, by race/ethnicity, 1994 through 2003**

Race/ethnicity	1994 to 1995	1995 to 1996	1996 to 1997	1997 to 1998	1998 to 1999	1999 to 2000	2000 to 2001	2001 to 2002	2002 to 2003
Unknown	37.7%	3.7%	-54.9%	8.1%	7.5%	-51.2%	-8.1%	-7.4%	-6.8%
White	0.9%	0.9%	0.5%	0.6%	0.4%	2.1%	0.8%	0.8%	1.3%
Black	1.8%	1.7%	2.0%	1.8%	1.6%	3.9%	2.0%	1.9%	2.4%
Other	27.8%	19.5%	-33.9%	22.6%	18.5%	-48.2%	4.8%	2.4%	7.1%
Asian	-2.9%	-3.9%	135.8%	-4.0%	-3.7%	38.1%	6.4%	5.5%	5.7%
Hispanic	-3.2%	-3.7%	109.1%	-3.5%	-3.8%	11.2%	1.2%	1.3%	1.6%
American Indian/ Alaskan Native	-3.9%	-4.9%	68.4%	-4.5%	-4.9%	22.3%	2.5%	110.1%	6.1%
Total	1.7%	1.4%	1.0%	1.0%	0.8%	1.2%	1.0%	1.1%	1.5%



In 1997, a second update of the EDB using the NUMIDENT file occurred. In addition, a mailing was sent to all those beneficiaries classified as Other or Unknown in the EDB, as well as those with a surname identified as Hispanic using the Census Bureau name algorithm asking them to self-report their race. Postcards were sent to 2.2 million persons. Approximately 858,000 responded. The impact of this mailing can be seen by the redistribution of persons primarily from the Other and Unknown categories which decreased by 54.9% and 33.9%, respectively, to primarily the Asian, Hispanic and American Indian/Alaskan Native categories which increased by 135.8%, 109.1% and 68.4%, respectively (Tables 1a and 1b).

The third NUMIDNET update occurred in late 2000, and annual NUMICENT updates have occurred at the end of each year, since then. The importance of the NUMIDENT file merge and the other activities with the EDB in 1997, 2000, and later can be seen in Tables 1a and 1b. Because the race/ethnicity categories of Asian, Hispanic and Native American are collapsed into the category Other in the MBR, there was an increase in the number of the elderly beneficiaries labeled as Other in 1995 and 1996, a drop in 1997, and an increase in 1998 and 1999 followed by a drop in 2000. Concurrently, in 1995 and 1996, and in 1998 and 1999, the number of Asians, Hispanics and American Indians/Alaskan Natives decreased. Since 2000, there has been a regular increase in the number of Asians and Hispanics because of the annual NUMIDENT updates.

### **1.2.1 Improving the information about American Indians/Alaskan Natives**

Because of an initiative of the Office of Research, Demonstrations and Information (ORDI) at CMS, the Indian Health Service (IHS) has provided information on people it serves to CMS since 1999 (Linda Greenberg, Agency for Healthcare Research and Quality, Personal communication, September 2004). Demographic information to allow matching of IHS information with the EDB (SSN, if available, name, date of birth and gender) are used hierarchically. If matches are found, the race is taken to be American Indian/Alaskan Native, and the race information in the EDB is overwritten. In July 1999, prior to the receipt of this information from IHS, there were 54,240 beneficiaries identified as American Indian/Alaskan Native in the Denominator file (Table 1a). Due to the NUMIDNET and IHS updates the number of American Indians/Alaskan Natives increased to 151,600 in 2003.

Table 2 shows the impact of the most recent IHS updates. The important point is that the incremental return from each update is now small.

**Table 2. Number of Medicare beneficiaries in Enrollment Database by race/ethnicity at time of Indian Health Service data updates, 2003-2004**

Race/ethnicity	Date			
	9/28/03	11/30/03	3/27/04	6/27/04
Unknown	109,140	111,446	95,667	98,582
White	35,711,212	35,759,876	35,944,885	36,010,572
Black	4,143,103	4,156,005	4,204,097	4,220,270
Other	885,006	940,473	766,236	856,827
Asian	558,366	555,424	664,513	659,868
Hispanic	939,354	933,095	975,053	967,010
American Indian/Alaskan Native	153,120	155,863	158,066	158,220
Total	42,499,201	42,612,182	42,808,517	42,971,349

Table 2 also shows how the number of Asians and Hispanics in the EDB decreases and the number with Other or Unknown increases throughout the year as CMS waits for the NUMIDENT update. There are two implications for researchers/analysts that derive from these race/ethnicity updates: 1) the race/ethnicity on claims of an individual with multiple claims may be different within a calendar year file because the race of the individual in the SAFs is the race in the EDB at the time the claim is processed; 2) the race/ethnicity of an individual may change from year to year. Researchers should be made aware of this issue, particularly as they create cohorts for longitudinal studies. One recommendation is that the race/ethnicity in the most recent Denominator file or SAF they are using be used to overwrite the race/ethnicity in all files.

### **1.2.2 Validity of the Medicare race/ethnicity information**

The evaluation of the validity of race/ethnicity information contained in administrative data is done by comparing the administrative data information with information from a reference population. This can be done at the aggregate- (population-)level or the person-level. Examples of population-level comparisons are those of Lauderdale and Goldberg (1996) and of Eggers and Greenberg (2000). Using 1993 data, from the U.S. Census Bureau estimates and the Numerical Identification (NUMIDENT) file, Lauderdale and Goldberg estimated that only 24% of Hispanics, 17% of Native Americans and 56% of Asians 65-94 years of age were in the NUMIDENT file. Eggers and Goldberg showed very similar results for 1998. They found that for the U.S. elderly population the number of Medicare Part A enrollees was between 92 and 97 % of the Census Bureau estimate

depending on the gender-age group examined. This was due primarily to a high comparability for Whites (95 to 101% of the Census Bureau estimates), and acceptable estimates for Blacks (87 to 105%). However, the estimate of the number of Hispanics using the Medicare data was only 29% (range for different gender-age groups = 18 to 34%) of that found by the Census Bureau. For Asians it was 42% (range = 36 to 57%, and Native Americans, 24% (range = 15 to 28%).

For Medicare race/ethnicity data comparisons at the individual level, the reference standard has been the self-reported race/ethnicity, and the source of this information in published reports has been the Medicare Current Beneficiary Survey (MCBS) (Adler, 1994). Arday, et al. (2000) compared the self-reported race of the MCBS participants to that race information contained in the Enrollment database (EDB) in 1996, two years after the first NUMIDENT merge with the EDB, as well as in 1997, after the second NUMIDENT update and the mailing to beneficiaries described earlier. See Tables 3a and 3b which are Tables 1 and 2 from Arday, et al. There are two questions regarding race/ethnicity in the MCBS. The initial one asks about Hispanic ethnicity, and the second asks about the race groups. Arday, et al. recoded these responses to conform with the one byte race variable in the EDB. Only community dwelling participants were included, and only those for whom the response was not give by a proxy. Central to this paper is the validity they reported, particularly for 1997. Of the four measures used to assess validity, the sensitivity (the probability that the EDB will correctly identify a person of a given race/ethnicity); the positive predictive value, PPV, (the probability that a person identified by the EDB as being of a certain race/ethnicity actually is of that

race/ethnicity); and specificity (the probability that the EDB will correctly identify those who are not of a particular race/ethnicity) are the most important. What is initially clear in Table 3a is that prior to the updating of the Medicare data in 1997, the sensitivity of the EDB for identifying people who were Asian, Hispanic, or American Indian/Alaskan Native were very low (< 20%). Thus, fewer than 1/5<sup>th</sup> of people of these race/ethnicities were being identified. The sensitivities for Whites and Blacks were acceptable. In 1997, Table 3b, the sensitivity for Hispanics doubled compared with 1996. It tripled for Asians, and increased by over 18-fold for American Indians/Alaskan Natives. However, the initial sensitivity among American Indians/Alaskan Natives was so low in 1996 (0.6%), that it was only 10.9% in 1997.

The PPV for Whites, Blacks and Hispanics was acceptable in 1996, 95.3 to 98.0%, and remained so in 1997. In 1996, the PPV for Asians was 66.0%, and only 14.3% for American Indians/Alaskan Natives. However, in 1997, it was almost 80% for both groups.

Tables 3c and 3d present information regarding the validity of the EDB race/ethnicity information in 2001 and 2002 using the same methods as Arday, et al., and the MCBS and Denominator files for those years. These years were chosen because they are the most recent for which MCBS information is available, and they are years in which the annual NUMIDENT updates of the EDB have been carried out.

**Table 3. Validity of Enrollment Data Base (EDB) Racial/Ethnic Classifications Using the Medicare Current Beneficiary Survey (MCBS) for Comparison: 1996, 1997, 2001 and 2002**

**Table 3a.**

Race/ethnicity	1996			
	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
White	96.7	88.5	98.2	80.2
Black	95.5	99.5	96.0	99.5
Asian	19.9	99.9	66.0	99.1
Hispanic	19.4	100.0	98.0	94.3
American Indian/ Alaskan Native	3.4	100.0	57.1	99.2
Other	13.9	98.3	9.2	98.9

**Table 3b.**

Race/ethnicity	1997			
	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
White	96.8	69.3	98.4	80.4
Black	94.8	99.6	96.1	99.4
Asian	57.6	99.9	79.2	99.6
Hispanic	39.2	99.9	97.7	96.4
American Indian/ Alaskan Native	10.9	100.0	77.8	99.2
Other	8.5	99.1	9.7	99.0

**Table 3c.**

<b>Race/ethnicity</b>	<b>2001</b>			
	<b>Sensitivity (%)</b>	<b>Specificity (%)</b>	<b>PPV (%)</b>	<b>NPV (%)</b>
White	97.1	90.6	98.5	82.5
Black	96.7	99.7	97.3	99.6
Asian	55.3	99.9	81.4	99.5
Hispanic	34.7	100.0	96.4	95.4
American Indian/ Alaskan Native	24.0	100.0	89.3	99.4
Other	3.5	99.2	4.8	99.0

**Table 3d.**

<b>Race/ethnicity</b>	<b>2002</b>			
	<b>Sensitivity (%)</b>	<b>Specificity (%)</b>	<b>PPV (%)</b>	<b>NPV (%)</b>
White	97.2	90.3	98.5	83.5
Black	96.5	99.7	96.9	99.1
Asian	52.4	99.9	82.7	99.5
Hispanic	32.8	99.9	95.2	95.2
American Indian/ Alaskan Native	32.7	99.8	92.1	99.4
Other	6.7	99.1	9.4	98.8

The first observation is that for the two largest race groups: Whites and Blacks, the sensitivity, specificity and PPV are all very acceptable in all four years, and essentially remain the same between 1997 and the latter two years.

The second observation is that the gains seen in the sensitivity between 1996 and 1997 among Asians and Hispanics have not increased in 2001 and 2002, in fact, they have drifted downward. Also, the specificity and PPV have remained essentially the same for these two groups.

The third observation is that the extra effort made by CMS to obtain information from the IHS has increased the sensitivity of the EDB information approximately 3-fold from 1997 to 2002, such that the sensitivity is equal to that of Hispanics. Also, the PPV is now over 90%.

In spite of the importance of the NUMIDENT and IHS file updates to the EDB, at this time, Fall 2004, it appears as if a “steady state” or plateau has been reached in the knowledge about the race ethnicity of Medicare beneficiaries. Tables 3c and 3d show no improvement between 2001 and 2002 in the validity of the race/ethnicity codes, except for American Indians/Alaskan Natives.

### **1.2.3. Misclassification and its implications**

The following two tables (4 and 5) show the way in which persons in the Enrollment Database (EDB) were misclassified in 1997 and 2002 according to the self-reported



race/ethnicity information in the Medicare Current Beneficiary Survey (MCBS). (A similar table for 1997 contributed by Susan Arday can be found in and Escarce and McGuire, 2003.) The diagonals (in bold) are the positive predictive values (PPVs) for each race expressed as probabilities rather than percentages.

**Table 4. Probability of Reported Race/Ethnicity on Medicare Current Beneficiary Survey (MCBS) by Reported Race/Ethnicity in Enrollment Database (EDB), 1997**

		MCBS						Row Total	# in EDB
		White	Black	Hispanic	Asian	Native Ams	Other/Unknown		
EDB	White	<b>0.949</b>	0.003	0.034	0.001	0.007	0.003	0.997	13,001
	Black	0.025	<b>0.936</b>	0.021	0.003	0.003	0.005	0.993	1,494
	Hispanic	0.017	0.001	<b>0.977</b>	0.000	0.001	0.005	1.000	351
	Asian	0.172	0.010	0.059	<b>0.745</b>	0.000	0.014	1.000	101
	Native Ams	0.107	0.108	0.056	0.000	<b>0.735</b>	0.000	1.006	18
	Other/Unknown	0.431	0.103	0.244	0.137	0.047	<b>0.053</b>	1.015	197
	Total	0.823	0.096	0.0576	0.008	0.008	0.004	0.9966	15,162

Data source: 1997 MCBS using RIC A (for EBD data) and RIC 1 of the Access to Care.

**Table 5. Probability of Reported Race/Ethnicity on Medicare Current Beneficiary Survey (MCBS) by Reported Race/Ethnicity in Enrollment Database (EDB), 2002**

		MCBS						Row Total	# in EDB
		White	Black	Hispanic	Asian	Native Ams	Other/Unknown		
EDB	White	<b>0.936</b>	0.002	0.048	0.001	0.005	0.005	0.996	10810
	Black	0.015	<b>0.945</b>	0.023	0.002	0.005	0.006	0.995	1263
	Hispanic	0.029	0.001	<b>0.952</b>	0.000	0.001	0.008	0.991	292
	Asian	0.136	0.020	0.075	<b>0.765</b>	0.010	0.000	1.006	98
	Native Ams	0.025	0.051	0.044	0.000	<b>0.880</b>	0.000	1.000	38
	Other/Unknown	0.262	0.071	0.309	0.272	0.025	<b>0.061</b>	1.000	137
	Total	0.808	0.097	0.070	0.01	0.007	0.006	0.998	12638

Data source: 2002 MCBS using RIC A (for EBD data) and RIC 1 of the Access to Care.

The first point is that because of the high positive predictive values for Whites, Blacks and Hispanics, there is not much opportunity for misclassification in these groups. In both years, the PPVs for these three race groups were  $\geq 0.94$ . The second point is that the lower PPVs for Asians and American Indians/Alaskan Natives, 0.74 for both groups in

1997, and 0.76 and 0.88, respectively, in 2002, means that there is the possibility of significant errors in estimates of events (healthcare utilization, post-treatment mortality, etc.) in these two groups. The third point is that for most minority groups, many are misclassified as Whites. In 1997, 17.3% of Asians were misclassified as White and in 2002, 13.6%. In 1997, 10.7% of American Indians/Alaskan Natives were misclassified as Whites. Partially because the PPV for American Indians/Alaskan Natives increased to 0.88 in 2002, only 2.5% were misclassified as Whites, however, 5.1% were misclassified as Blacks. The small number of American Indians/Alaskan Natives and the continuous changing of the sample may also have contributed to the instability of these relationships for American Indians/Alaskan Natives. Finally, of those who are not classified as Other/Unknown, approximately half of the Blacks and Native Americans/Alaskan Natives, and almost all of the Hispanics and Asians are misclassified as Whites.

What is the implication of this misclassification on research comparing treatment rates or quality of care measures between racial and ethnic groups? In most instances, these comparisons use Whites as the reference group and generally show the Whites have the preferred (usually higher) rate of surgery, test use, or quality of care (Bach, et al., 1999; Escarce, et al., 1993; McBean and Gornick, 1994; Gornick, et al., 1996; Mayberry, et al., 2000). More recent work has included Hispanics and Asians (Virnig, et al., 2002; Virnig, et al., 2003; McBean, et al., 2004; McBean, et al., in press; Escalante, et al., 2002), and Whites fare better than Hispanics. Escarce and McGuire (2003) have addressed the question of potential bias in the results of studies comparing outcomes between racial and ethnic groups caused by the misclassification of the CMS race variable. Using the 1997

Medicare Physician/supplier file (now the Carrier file) claims data for 15 different procedures, as well as information from the EDB and the MCBS, they have presented crude and corrected procedure rate estimates. The corrected rates adjusted for the rates among each of the race groups proportional to the amount of misclassification. In all cases except radionuclide stress tests, the crude and corrected procedure rates were higher in Whites than in any of the other race groups. Because the PPVs are so high for Whites, Blacks and Hispanics, the corrected rates are all within 10% of the crude rates. Among Asians and Native Americans/Native Alaskans, however, the uncorrected rates for 8 of the procedures were > 10% higher than the corrected rates, reflecting the lower PPVs in these two groups and the higher rates of the procedures among Whites. The implication of the work of Escarce and McGuire for studies comparing event rates between racial and ethnic groups is that the bias introduced by the misclassification is a “bias towards the null”. That is, the uncorrected rates among Asians and Native Americans/Alaskan Natives are inflated due to the misclassification and will be closer to that of the reference group, Whites, than if they were corrected. Thus, any significant differences in uncorrected rates between Whites and these groups should be considered valid. However, this bias may hide real differences between Whites and these minority groups. Also, the underlying assumption in this conclusion is that the misclassified beneficiaries are similar and have event rates similar to those of their true group. As pointed out by Escarce and McGuire, “those minorities who ‘misrepresent’ themselves to be White might have true rates closer to the White group”.

#### **1.2.4 Misclassification of Hispanics - a special issue**

Another issue of misclassification, or perhaps better stated, of unknown classification is that of the sub-groups of Hispanics (Cubans, Mexican Americans and Puerto Ricans). The same issue exists for Asians and Pacific Islanders and Native Americans/Alaskan Natives, but will not be discussed because of the smaller numbers of individuals who would be in these subgroups even if they could be accurately identified. Lauderdale and Goldberg (1996) reported the distributions of race/ethnicity designation for elderly persons enrolled in Medicare in 1993 in the Numerical Identification (NUMIDENT) file who were reported being born in one of three Hispanic areas: Cuba, Mexico or Puerto Rico). They found that the majority of elderly Medicare enrollees born in Cuba, Mexico, and Puerto Rico were classified as White (74%, 58%, 55%, respectively), and not Hispanic. Only 18% of the Cuban-Americans, 29% of the Mexican-Americans and 25% of the Puerto Rican-Americans were reported as Hispanic. Of note in this study, also, is the number of persons indicated in 1993 who would be available for study if the country of birth information were available: 164,000 Cubans, 208,000 Mexicans; 366,000 Puerto Ricans, as well as 1,002,000 Filipinos, 963,000 Chinese, and 266,000 South Koreans, 194,000 Japanese, and 160,000 Indians.

### **1.2.5 Anomaly caused by the Sex and Race of Primary Beneficiary (SROP) variable in SSA's Master Beneficiary Record (MBR)**

One “anomaly” of the Medicare race information, indicated earlier, is the use of the race/ethnicity information of the person whose work history is the basis upon which a person receives SSA benefits and is entitled to Medicare as the race/ethnicity for the person with the work history, as well as others receiving benefits because of that work

history. This is caused by the use of the SROP variable in the Master Beneficiary Record (MBR) to create and update the Enrollment Database (EDB). Thus, in the Medicare databases a spouse or other dependent person (generally, the wife) who is receiving benefits under her spouse's work history will have the same race code as that of the husband. For those whose race/ethnicity information has been updated through the Numerical Identification (NUMIDENT) file, special mailings, or the Indian Health Service (IHS) mechanism, the new information overwrites the information from the MBR.

Table 6 presents the validity of the EDB information on race/ethnicity for beneficiaries who had a Beneficiary Identification Code (BIC) of A with those who had a BIC that is not A in 1997. As a reminder, the race of the person indicated in the MBR is the race of the primary beneficiary, the SROP variable. For persons with a BIC that is not A, their race in the MBR is the race of the primary beneficiaries. For all race groups the sensitivity and PPVs for the persons with the BICs that are not As are within 3 percentage points of those whose BICs are As, except for Asians and American Indians/Alaskan Natives. Somewhat unexpectedly, the sensitivity and PPV are higher for the Asians in the BIC not = A group, than the BIC = A group. For American Indians/Alaskan Natives, the expected higher sensitivity and PPV in the BIC = A group are seen. Thus, although the SROP variable in the MBR means that at the individual level one cannot be sure if the race/ethnicity of the dependent person is valid, the validity at the group level for the three largest race groups is similar in the BIC = A and the BIC not = A groups.

**Table 6. Validity of race/ethnicity information in Medicare EDB among those in MCBS who had a claim paid for a service with Beneficiary Identification Code (BIC) of A or not equal to A, 1997**

Race/ethnicity	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
BIC = A, N = 8,695				
White	99.0	79.4	96.0	94.1
Black	96.7	99.3	94.4	99.6
Hispanic	42.4	99.9	97.8	97.2
Asian	45.5	99.9	62.6	99.7
American Indian/Alaskan Native	14.7	100.0	71.4	99.3
BIC not = A, N = 3,065				
White	98.8	78.4	94.6	94.3
Black	95.4	99.6	97.1	99.4
Hispanic	39.4	99.9	97.6	95.9
Asian	71.2	99.6	79.2	99.4
American Indian/Alaskan Native	9.5	100.0	66.7	99.4

**1.2.6 Summary of current assembly of information on race/ethnicity in the Enrollment Database (EDB)**

On a daily basis, race/ethnicity information on new beneficiaries or those who have updates to their status received at SSA is passed electronically to CMS in the Master Beneficiary Record (MBR) with its four race/ethnicity categories. Thus, the race/ethnicity information for Asians, Hispanics, American Indians, Alaskan Natives, or others which is collected on the current SS-5 form, is collapsed into Other for this input to the EDB. This information is also sent by CMS daily to the nine Common Working File regional host sites that insert the EDB race information known on the day they process each claim onto the utilization claims. At the time of the Indian Health Service (IHS) quarterly updates, new information on any American Indian or Alaskan Native in the IHS file is corrected.

Finally, there is the annual NUMIDENT update which takes place in the last quarter of the calendar year. A finder file containing all persons who were added to the EDB from the MBR since the last NUMIDENT update, plus all living beneficiaries with a race/ethnicity designation of Other or Unknown is sent to SSA to obtain the race/ethnicity information contained in the NUMIDENT file under the expanded race/ethnicity code. This information is then returned to CMS which corrects the EDB race/ethnicity information for all who are not coded as Native American or Alaskan Native. Because this NUMIDENT and IHS updates take place a couple to a few months before the creation of the annual Denominator file, since 2000, those file have had essentially the best possible information known to SSA and CMS for the year of the Denominator file.

### **1.2.7. Race and managed care**

To date, there has been no requirement by CMS, the Department of Health and Human Services (DHHS), the National Committee for Quality Assurance (NCQA), or any other organization that requires managed care plans serving Medicare beneficiaries, collect or report information regarding the race/ethnicity of plan members. Several steps have been taken by CMS, by researchers and by foundations to encourage the use of available race/ethnicity information and the collection of more. The most comprehensive has been the post-hoc linking of the Health Plan Employer Data and Information Set (HEDIS) information submitted by the health plans to CMS to race/ethnicity information contained in the Enrollment Database (EDB). This has been done by two groups. Schneider and colleagues at Harvard linked the 1998 HEDIS information. The Research Data Assistance

Center (ResDAC) at the University of Minnesota has linked the 1999 through 2002 data as part of an effort by CMS to improve the quality of the data submitted by the plans to CMS via NCQA. These efforts have led to several publications describing the differences by race/ethnicity in the use of services and of reaching several quality indicators, including those for diabetes and mental health care (Schneider; et al., 2002, Virnig, et al., 2002; Virnig, et al., 2004; McBean, et al., 2004; McBean, et al., in press). As described in the publications, various errors by the plans prevented the researchers from using the information submitted by every plan for all beneficiaries for every quality of care indicator. However, all publications showed almost uniformly significantly lower rates of diabetes care, mental health, cancer screening, and other recommended services among Blacks compared with Whites, as well as frequently showing lower rates among Hispanics compared with Whites. While no validation of the EDB based racial coding for managed care members has been carried out, one can assume that the validity demonstrated by Arday, et al. (2000), and shown earlier in this paper, would be similar for the managed care plan enrollees as for the entire Medicare population.

Other than the linking of the HEDIS data with the EDB, described above, the most comprehensive effort to broadly identify minority group members and assess the frequency and quality of care in managed care plans has been the work of Nerenz, et al. (2002) supported by AHRQ and The Commonwealth Fund. In a pilot test to identify minority populations in managed care organization, these authors used three different methods to determine the race/ethnicity of members with the intent of eventually healthcare access and/or the quality of care delivered. The methods were: (1) self-



reported race/ethnicity on a survey about asthma, surveys about other chronic disease, or the Consumer Assessment of Health Plans Survey (CAHPS); (2) The Generally Useful Ethnicity Search System (GUESS) surname-recognition program for identifying Hispanic members of one plan; (3) information from providers' medical records and encounter databases. All methods were judged successful in obtaining accurate information. Of particular note to this report is that all of the expert panels they convened "consistently recommended that plans should collect data on the race and ethnicity of their members, assuming that the data will be used for public reporting and internal quality improvement purposes and not for any illegal or unethical discriminatory purposes".

### **1.2.8 Monitoring and improving race/ethnicity information at CMS and the Department of Health and Human Services (DHHS)**

The uniformity of the race/ethnicity information across Medicare datasets maintained at the CMS Data Center by the Office of Information Services (OIS) is assured because of the use of the information in the Enrollment Database (EDB) to populate the Denominator file and the utilization files. It is worth repeating that this is one of the greatest strengths of the data. Nonetheless, improving the race/ethnicity information has been a concern to CMS staff, particularly those in the Office of Research, Demonstrations and Information (ORDI), the Office of Information Services (OIS) and the Office of Clinical Standards and Quality (OCSQ). Ever since the number of categories in the race variable was increased in 1994, many discussions have been held regarding the constraints imposed by the dependence of the EDB on the SSA databases

and the preference on the part of CMS staff to have available and use a two-stage race/ethnicity coding scheme.

Currently, there are plans to replace the EDB with the Master Beneficiary Database (MBD) (Spike Duzor, ORDI, CMS, personal communication, September, 2004). One feature of the MBD being discussed would allow race/ethnicity fields in addition to the SSA-based field. Supplemental race/ethnicity information, particularly self-reported race information, collected by the Medicare Current Beneficiary Survey (MCBS), the Consumer Assessment of Health Plans Survey (CAHPS), or other surveys, could have its own field(s) that would include an indicator of the source of the information. Thus, CMS and researchers would have valid race/ethnicity information available on important subsets of the population included in the surveys.

One level above CMS is the Department of Health and Human Services' Data Council (HHS Data Council). The HHS Data Council coordinates all health and non-health data collection and analysis activities of the Department of Health and Human Services (DHSS), including an integrated health data collection strategy, coordination of health data standards and health information, and privacy policy activities. The Director of ORDI is CMS' representative. Of relevance to this document is the Racial and Ethnic Data Work Group. Richard Bragg of ORDI is the CMS representative. In addition to cataloging the race information available in Federally held databases, the HHS Data Council has issued in collaboration with the Data Work Group of the DHHS Initiative to Eliminate Racial and Ethnic Disparities in Health an extensive report "Improving the

Collection and Use of Racial and Ethnic Data in HHS” (available at <http://aspe.hhs.gov/datacncl/racerpt/index.htm> ), Appendix I, which resulted in the Policy Statement on Inclusion of Race and Ethnicity in HHS Data Collection Activities sent by Secretary Shalala to all HHS operating divisions on October 24, 1997 (<http://aspe.hhs.gov/datacncl/racerpt/appendg.htm>). See Appendix II.

In addition to its own Data Council, DHHS has been assisted for decades by the National Committee on Health and Vital Statistics (NCVHS) (<http://www.ncvhs.hhs.gov/>). The most recent Charter for the Committee is presented in Appendix III (available at <http://www.ncvhs.hhs.gov/charter06.pdf> ). Briefly stated, the NCVHS serves as the statutory [42 U.S.C. 242k(k)] public advisory body to the Secretary of Health and Human Services in the area of health data and statistics. In that capacity, the Committee provides advice and assistance to the Department and serves as a forum for interaction with interested private sector groups on a variety of key health data issues. CMS has a permanent liaison person, currently Judith Berek, and CMS staff routinely present to the Committee, and the Committee regularly makes recommendations to the Secretary. A recent important example is the letter from the Committee to the Secretary dated September 23, 2003 which recommends the collection and use of race/ethnicity information by managed care plans. The full letter is in Appendix IV and available at <http://www.ncvhs.hhs.gov/0309261tb.htm> .

### **1.3 Other information in Medicare datasets; data documentation and data availability**

General information about the Medicare datasets, the specific information about variables, data file documentation, and information about their availability to researchers is easily accessible on the Research Data Assistance (ResDAC) website at [www.resdac.umn.edu](http://www.resdac.umn.edu). Supported by Health Care Financing Administration (HCFA) and CMS contracts since 1996, ResDAC was created to assist researchers throughout the country in understand and use Medicare data, including helping them provide the necessary information to CMS to obtain access to the data.

The Denominator file contains individual-level demographic information as well as information on Medicare entitlement and eligibility status and mortality, as well as Medicaid (state buy-in) and managed care enrollment by month. The utilization data files include the dates of service, the type and place of service, diagnoses, procedures, reimbursement and other payment-related information (e.g. Diagnostic Related Group), admission type, discharge destination, information about the provider that can be linked to other provider-specific files, as well as more recently, clinical and functional status information in the Minimum Data Set (MDS).

The purposes to which the data have been applied include the understanding of the Medicare population, reporting of Medicare activities, monitoring the health of the Medicare population and sub-populations of special interest, quality of care studies, policy documents, the understanding of health care costs, program planning and evaluation, and waste, fraud and abuse.

The datasets are well documented. As stated above, data documentation is currently available through the ResDAC website for the Denominator file, the Medicare Provider and Analysis Review (MedPAR) file, the Standard Analytic Files (SAFs), Medicare limited data sets, the Medicare Current Beneficiary Survey (MCBS), MDS, Outcome and Assessment Information Set (OASIS), Medicaid data sets, etc. For example, the data dictionaries for the institutional SAFs are available at [http://www.resdac.umn.edu/ddvib/dd\\_via2.asp](http://www.resdac.umn.edu/ddvib/dd_via2.asp), the non-institutional SAFs at [http://www.resdac.umn.edu/ddvib/dd\\_vib.asp](http://www.resdac.umn.edu/ddvib/dd_vib.asp), etc.

The ease of working with the data sets depends on the skill of the investigators. Most requests for data from CMS result in large amounts of data arriving in formats that are unfamiliar to those not used to working in the IBM mainframe environment. Again, information is available on the ResDAC website or from ResDAC Technical Advisors to assist researchers understand the data and how to prepare it for analysis in their local environment.

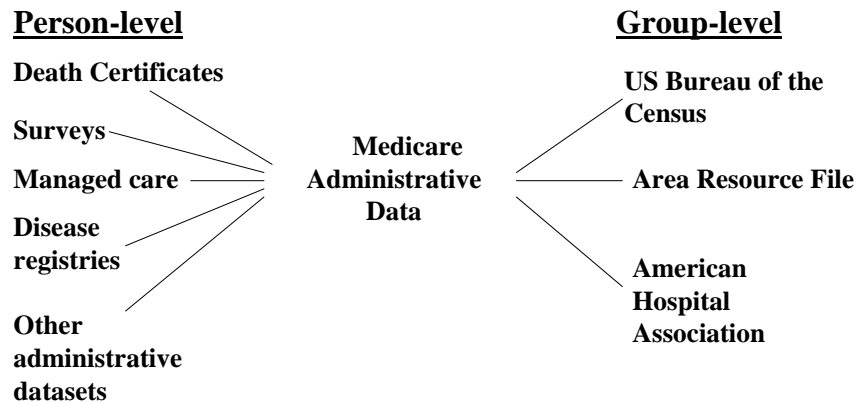
As mentioned earlier, the SAFs and MedPAR for calendar year 200X are generally available in the summer of 200X+1. The Denominator for 200X is available in March/April of 200X+1. Information in the more recently created files such as the MDS and OASIS are available three months after data collection. The MCBS data are generally available two to three years after the end of the year of interest. ResDAC has helped simplified the path for researchers to take in order to obtain Medicare data by having the necessary information on the website

([http://www.resdac.umn.edu/Medicare/requesting\\_data.asp](http://www.resdac.umn.edu/Medicare/requesting_data.asp)) and by preparing and distributing Data Request packages for the Research Identifiable files (RIFs) and the Limited Data Sets (LDS).

## 2.0 Linking Medicare administrative datasets with other datasets

Summary - It is very appealing to think of merging Medicare administrative data with other data collected by CMS, other Federal agencies or Federally sponsored projects, health care providers (particularly, managed care organizations), and others, especially if the data include race/ethnicity information obtained from the individual. Datasets can be linked at either the person-level or the group-level; for example, zip code.

### Linking Medicare Administrative Data to Other Data



The most efficient person-level matches to Medicare administrative data are numeric matches, using the Medicare Health Insurance Claim (HIC) number or the Social Security number (SSN). Name matches have been possible in the past, but they are no longer available. Examples of different categories of actual or possible person-level links

include: (1) merging of data when the Enrollment Database (EDB) was used as the sampling frame: for example, the Medicare Current Beneficiary Survey (MCBS), Fee-for-Service Consumer Assessment of Health Plans Study (FFS CAHPS); (2) merging of data collected by other Federal agencies, their agents or research projects when EDB is not the sampling frame: for example, the Surveillance, Epidemiology and End-Results (SEER) program, the Medicare managed care CAHPS (MMC CAHPS), the Health and Retirement Study (HRS), the Asset and Health Dynamics Among the Oldest-old survey (AHEAD); (3) health care providers, primarily managed care; 4) other administrative databases: for example, the Indian Health Service, Medicaid, and the Department of Veterans Affairs; 5) vital statistics data.

Population-level links have occurred with Census Bureau data and other files that provide zip code or county level data, such as the Area Resource File (ARF) of the Health Services and Resource Administration (<http://www.arfsys.com/>). A final option is the imputation of race or ethnicity using surname or similar matches.

## **2.1 Individual-level linking**

### **2.1.1 EDB used as the sampling frame**

The Enrollment Database (EDB) has been used as the sampling frame for the Medicare Current Beneficiary Survey (MCBS), the Fee-for-service CAHPS (FFS CAHPS), and the Health Outcomes Survey (HOS). The major advantage of this strategy is the new race/ethnicity and survey information collected by the survey can be applied to the

appropriate Medicare beneficiary with 100% reliability. In addition these surveys collect race/ethnicity using two questions (all three), other socioeconomic information such as education level (all three), and income (MCBS and HOS), as well as many variables that can be used for stratification or case mix adjustment (general and mental health status, chronic diseases, supplemental insurance, etc.). The MCBS has provided large amounts of information in its reports (<http://www.cms.hhs.gov/mcbs/PublDT.asp>), as well as in scientific publications regarding the use of common services, as well as health expenditures, particularly out-of-pocket expenditures. These have included studies of the use and costs of services by race/ethnicity. Unfortunately, the sample (approximately 15,000 persons in a 3-year panel design) has too few minority populations to repopulate the race variable in the EDB, even when multiple years are combined.

The FFS CAHPS, which has had a much larger sample size collected information on 107,348 beneficiaries who responded in 2000 and 117,830 in 2001, the last year of the survey. Reports developed from the survey information are available at <http://www.cms.hhs.gov/researchers/projects/consumers/cahps.asp> and [http://www.cms.hhs.gov/researchers/projects/consumers/ffs\\_finalreport2.pdf](http://www.cms.hhs.gov/researchers/projects/consumers/ffs_finalreport2.pdf).

The HOS, originally the Health of Seniors survey and now the Medicare Health Outcomes Survey, is another source of self-reported race/ethnicity information available for additional Medicare + Choice/Medicare Advantage (M+C/MA) beneficiaries (<http://www.cms.hhs.gov/surveys/hos/hosoverview.asp>) that has been linked to the EDB. In addition, education, income, and many health related questions are asked including the



full Medical Outcomes Study Short Form-36 (SF-36) questionnaire designed to measure both physical and mental health in both a baseline survey and follow-up surveys. Over 900,000 managed care enrollees have been queried between 1998 and 2003, and HOS will continue under Section 722 of the Medicare Modernization Act of 2003.

### **2.1.2 Data from high quality surveys and registries**

Perhaps the most frequently used linked race/ethnicity data are the Surveillance, Epidemiology and End-Results - Medicare (SEER-Medicare) datafiles (available at <http://healthservices.cancer.gov/seermedicare/>). Since 1973, the 11 cancer registries of the SEER program, sponsored by the National Cancer Institute, have collected cancer incidence and initial treatment information for the first four months after a person is diagnosed with cancer from areas in the United States containing approximately 14.5% of the U.S. population (New areas were added to SEER in 2001 that will increase the coverage of the U.S. population to 25%, but these new areas have not yet been merged with CMS data.) SEER data have been merged 4 times with Medicare data (1991, 1995, 1999, 2003) using a probability match based on the social security number and demographic information available in the SEER database. The match rates were high, 93%, in all years. As of 1996, the total number of Medicare beneficiaries in the SEER-Medicare data is over 1.1 million (Klabunde, et al., 2002). The file which has information from both sources, the Patient Enrollment and Diagnosis Summary file (PEDSF), contains race/ethnicity data from each, and their origin is identified. This allows the researcher to decide which race/ethnicity information to use, and serves as a model for similar datasets that combine Medicare administrative data information from the EDB

and self-reported information from a survey or other sources. At present, the SEER race variable includes 26 unique categories, as well as the categories Other and Unknown (<http://seer.cancer.gov/>). Bach, et al. (2002) have stated that the SEER data regarding race has not been externally validated. The following, taken from Bach, et al. (2002), summarizes the derivation of the Hispanic ethnicity variable in the SEER data and compares it with the Medicare information

“In the SEER program, Hispanic ancestry is an ethnic characteristic, captured by its own variable. Whereas race in SEER is determined from medical records and registration information, Hispanic ancestry in SEER is identified through algorithms that identify Spanish surnames. The categories of Hispanic ancestry generated by this algorithm closely match those on the 1990 and 2000 US Census. Although, in contrast to the EDB, SEER obtains race and ethnic information for each individual based on their own records, one of the major limitations of the use of a surname-matching algorithm is the tendency to misclassify females more often than males due to surname changes that can occur with marriage. However, because the SEER program uses a separate variable to identify Hispanic ancestry, and because it uses a Hispanic-surname algorithm, it is likely that Hispanics are more accurately identified in the SEER program. For reference, the SEER ethnicity variable is very likely to be more sensitive for Hispanic ancestry than is the EDB race variable, the latter of which has only a 39% sensitivity (Arday, et al., 2000). Only 20% of individuals coded as having a Hispanic surname in SEER were coded as Hispanic according to the Medicare race variable. Given these results, one should also assume that the increased sensitivity in SEER is accompanied by a decrease in specificity, although the magnitude of either effect is unknown.”

“At present, most analyses employ the recoded SEER race variable (Race Recode B), which combines information from the race and Hispanic-surname variables”. The categories in the SEER Race Recode B variable are White, White-Hispanic, Black, Chinese, Japanese, Filipino, Puerto Rican, American Indian/Eskimo/Aleutian, Hawaiian, Other. “Virtually all persons with a Hispanic surname in SEER (98%) are captured as “white Hispanic” in this recoded variable.”

In merged databases such as the SEER-Medicare PEDSF, there is a very small, but worrisome, possibility that the race/ethnicity information that supplements the Medicare

information may be inaccurate. This leads to the requirement that information from each data source be stored as separated variables.

Predating the FFS CAHPS, the Medicare Managed Care CAHPS (MMC CAHPS) has collected information on approximately 150,000 Medicare managed care (MMC) beneficiaries annually since 1998. The samples of 600 beneficiaries per plan are drawn by the health plans and, except for the smaller plans, different members have been selected each year. Thus, at this time self-reported race/ethnicity information is available on over 1 million MMC beneficiaries. This information could be made available to CMS to update the EDB or the future MBD. However, as noted in one of their reports, 15% of women did not indicate a race

(<http://www.cms.hhs.gov/researchers/projects/consumers/mc1overview.pdf> ).

Two other well-known surveys the Health and Retirement Study (HRS) and Asset and Health Dynamics Among the Oldest-old (AHEAD) that have been linked to Medicare data could also provide race and ethnicity information for their study members. By the end of this year, approximately 80%-90% of the 12,000 persons in the initial HRS cohort (birth cohort from 1931-41 and spouses of any age) would be expected to have aged in to Medicare. All 7,447 AHEAD members were 70+ years of age in 1993, the start of the survey. In spite of the high likelihood that the race/ethnicity information has a high validity, the problem with these surveys is their small size.

### **2.1.3. Collecting information from clinical settings or managed care organizations serving Medicare beneficiaries for CMS**

Collecting information from clinical settings, although part of the Surveillance, Epidemiology, and End-Results (SEER) program strategy, does not appear to be a good option for CMS, and will not be discussed. As described earlier, based on our experience with the Health Plan and Employer Data Information Set (HEDIS) data, all the data files from all plans would not be expected to be 100% linkable with the EDB. However, 90+% would be likely be linkable initially, and it would improve over time. (Virnig, et al., 2000; Virnig, et al., 2003; McBean et al., 2003; McBean et al., in press). Thus, in spite of some shortcomings, requiring Medicare Advantage (MA) plans to collect race/ethnicity information and to submit it to CMS is recommended. See section 1.2.7 on pages 31-33. In addition to improving the information in the EDB, it would allow MA plans and/or the National Committee for Quality Assurance (NCQA) to conduct their own analyses using this information.

### **2.1.4. Information from other administrative databases from organizations that provide services to Medicare beneficiaries in addition to what they receive from Medicare: for example, the Department of Veterans Affairs (VA) and Medicaid**

Dartmouth researchers (Fleming, et al., 1992) and others have pointed out the importance of linking Medicare and VA information in order to have a complete picture of health services use on the part of persons eligible for both programs. The recent merge of the two databases managed by CMS and the VA Information Resource Center (VIREC) was accomplished using the SSNs of VA users and persons eligible to use VA services

between 1997-99 (<http://www.virec.research.med.va.gov/DataSourcesName/VA-MedicareData/VA-Medicare.htm> ). Medicare data are available for 1999 through 2001. The VA patient treatment file and other databases would be of use to Medicare oriented researchers. However, regarding the race/ethnicity information in the VA data, 30% of VA beneficiaries had missing race information and, “The Medicare Denominator file appears to be a more complete source for race data for those dually enrolled” (<http://www.virec.research.med.va.gov/References/VirecInsights/Insights-V04n3.pdf> ).

The linking of Medicare and Medicaid databases has special importance because dually eligible beneficiaries are among the most vulnerable no matter what their race/ethnicity. Currently, linking with either “state owned” or CMS-housed Medicaid databases using the SSN of the Medicaid beneficiary and performing a “SSN link” to the Medicare data. If only Medicaid eligibility is needed to be known, that can be determined from the “state buy-in” variable in the Denominator file.

CMS is in the process of creating and validating a Medicaid “link key” which will facilitate the linking of Medicare and Medicaid data. This link key is an identifier that is unique to each Medicaid program participant, no matter the state in which they were a beneficiary. The interest in such a identifier has been enhanced by the recently created Medicaid Analytic Extract (MAX) files at CMS (<http://www.cms.hhs.gov/researchers/max/default.asp> ) and the availability of Medicaid State Information System (MSIS) files for almost all states. MSIS files contain paid Medicaid claims data for billed Medicaid inpatient, long-term care, and other (physician,

prescription drug, etc.) services. These claims contain information on the types of services provided, providers of services, service dates, types of reimbursement, and service costs. MSIS files are submitted quarterly. Prior to 1999, MSIS claims data were available for only a subset of states, but they are available for all states (except Hawaii) in CY 1999. For CY 2003 data for over 40 states are available.

While knowledge of Medicaid program use by Medicare beneficiaries and vice versa is important, the race/ethnicity data in the Medicaid files have not been validated and would likely vary by state. Until they are validated it is not recommended that this information be added to the EDB.

#### **2.1.5. Vital statistics (death data) link**

Researchers and policymakers at CMS have long been interested in linking CMS data to the mortality data from the National Center for Health Statistics. CMS knows the data of death of almost all beneficiaries, and knowing the cause of death would be very useful. Promoting that linking is beyond the scope of this document, but it should be a high priority for CMS, as well as NCHS, which would benefit by knowing more about the events contributing to death. Linking of these files to obtain race/ethnicity information would not likely be worth the effort. As pointed out by Rosenberg, et al. (2002), comparison at the individual-level between the race on the death certificate and other sources for Asians (and sub-populations of Asians), Hispanics and American Indians/Alaskan Natives is inadequate.

## **2.2 Group-level Linking**

### **2.2.1. Group linking at the zip code or county level**

Many researchers and agencies have used the zip code and/or county of residence in the Medicare databases to link with Census Bureau, Department of Health Resources and Services Administration, American Hospital Association and other databases. The primary purpose has been to obtain additional socioeconomic data (income, education) or information about the healthcare infrastructure, not race/ethnicity. No change in this approach is recommended.

One of the first examples of combining zip code income information to the Medicare databases was the study of Gornick, et al. (1996) in which they linked 1990 Census Bureau zip code median household income for Black and White Medicare beneficiaries with Medicare data. The impact of this study and its approach has been seen in the work of many other researchers who have subsequently incorporated zip code-level income and other socioeconomic information into their studies. The most “insitutionalized” example of this is the Patient Entitlement and Diagnosis Summary file (PEDSF) of the linked Surveillance, Epidemiology, and End Results program (SEER)-Medicare datasets. The seminal work of Gornick, et al. would be well worth repeating to determine the impact of the disparities initiatives undertaken since that time.

### **2.2.2. Surname matching and imputation**

A few researchers have used Spanish surname lists (Escalante, et al., 2002; Morgan et al., 2004; Nerenz, et al., 2002) to increase the number and the validity of the information of their study populations. They have used the Generally Useful Ethnicity Search System (GUESS) or the Census Bureau's 1990 Spanish surname list (<http://www.census.gov/genealogy/www/spanname.html> and <http://www.census.gov/population/documentation/twpno13.pdf>). By using the surnames that are "heavily Hispanic" (in > 75% of the households in a subset of the 1990 census identified as Hispanic the householder was Hispanic), Escalante, et al. estimated that the prior probability of correctly identify a Hispanic male beneficiary was between 0.79 and 0.97 depending on the state. It would be expected to be lower among women, 0.86 of the value for men. In a small population of male joint Medicare-VA beneficiaries in 5 counties in South Florida, Morgan, et al. increased the sensitivity of the Hispanic race designation in the Medicare data from 0.43 to 0.87 by including a Spanish surname match. They also, compared the number of Hispanic males identified in 16 other widely scattered counties in the U.S. using the Medicare race information with the number identified after the addition of the surname identifier, the 1990 Spanish surname list. The difference in the average difference between census-based and Medicare-based estimates of the Hispanic population greatly improved from 11% ( $p \leq 0.0001$ ) to < 1% (p value not significant). Also, the percent of White males also more closely matched the census estimate, 9.3% before and < 1.0% after. Thus, improvement in the validity of the race information for both Hispanics and for Whites was achieved.



More sophisticated imputation is frequently used by the U.S. Census Bureau and the National Center for Health Statistics (NCHS) largely in working with their survey results, but this is not recommended for the administrative data at this time.

### **3.0 Use of Medicare disparity data**

The information derived from the Medicare datasets regarding racial/ethnic disparities has been enormous. The identical benefit package among fee-for-service beneficiaries and the relative comparability of total insurance coverage among all racial/ethnic groups has made the Medicare population very attractive for the studies of Medicare beneficiaries, and the Medicare program, as well as for descriptive analyses of health and health service utilization, the evaluation of broader policy decisions and health services interventions. In addition, over 98% of the elderly are entitled to Part A Medicare, 96% of these are enrolled in Part B, and only 11.5% (in 2003) of the beneficiaries are in MA programs. Thus, studies of racial disparities using the fee-for-service population have tremendous generalizability to the entire elderly population, and strong implications for other age groups as well.

The leadership and program managers at CMS have been well aware of the disparity data because much of the early work on health disparities in health service utilization between race groups was done when CMS was the Health Care Financing Administration (HCFA) by HCFA researchers (Health Care Financing Administration volume 1 and 2, 1990; McBean, et al., 1991; McBean and Gornick, 1994; Gornick, et al., 1996), or by outside researchers using Medicare data, (Kasiske BE, et al., 1991; Escarse, et al., 1993;

Udvarhrlyi, et al., 1992; Ayanian, et al., 1993), often in collaboration with HCFA researchers (Javitt, et al., 1991). Medicare data and the researchers who used it led the way in highlighting major discrepancies in the use of preventive services, outpatient and in-patient surgical procedure rates, and hospital admission rates, as well as treatment outcomes. The availability and utility of the Medicare race/ethnicity data enhanced its attractiveness to yet other researchers and funding agencies, which in turn led to more useful information. Thus, an interaction was established between the data, including the enhancement of the data, and the understanding of disparities.

General awareness of the disparity issue in health and healthcare expanded through additional studies, as well as reviews and monographs supported by foundations, the Institute of Medicine, and DHHS. Racial and Ethnic Differences in Access to Medical Care: A Synthesis of the Literature (Mayberry, et al., 1999) appeared in 1999; Healthy People 2010 (DHHS, 2000) in 2000, and Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care (Institute of Medicine, 2002) and Diverse Communities, Common Concerns: Assessing Health Care Quality for Minority Americans – Findings from The Commonwealth Fund 2001 Health Care Quality Survey (Collins, et al., 2002) appeared in 2002. Congress increased its requirements on Medicare to address racial/ethnic disparities through the Benefits Improvement and Protection Act of 2000 (BIPA) and the Medicare Modernization Act of 2003 (MMA).

Foundations, especially the Henry J. Kaiser Family Foundation (KFF) (<http://www.kff.org/minorityhealth/disparities.cfm>) and The Commonwealth Fund

(<http://www.cmf.org/> ) have focused on disparities in health and health care, particularly among vulnerable populations such as those in the Medicare and Medicaid programs. Both foundations regularly support research and publish information regarding health care and racial/ethnic disparities. KFF supported the very important summary monograph by Mayberry, et al., (2002), Racial and Ethnic Differences in Access to Medical Care ([http://mcr.sagepub.com/cgi/reprint/57/suppl\\_1/108](http://mcr.sagepub.com/cgi/reprint/57/suppl_1/108)), mentioned earlier. Part of its web site is devoted to racial disparities, and it has publications such as Racial/Ethnic Differences in Cardiac Care: The Weight of the Evidence (<http://www.kff.org/uninsured/20021009c-index.cfm> ) and Key Facts: Race, Ethnicity and Medical Care (<http://www.kff.org/minorityhealth/6069-index.cfm> ) readily available on its web site. The Commonwealth Fund has major interests in health care coverage and access, as well as health care quality, and devotes a great deal of effort to these topic in relationship to Medicare beneficiaries, especially those in managed care, and other vulnerable populations ([http://www.cmf.org/topics/topics.htm?attrib\\_id=9087&portal=yes](http://www.cmf.org/topics/topics.htm?attrib_id=9087&portal=yes) ). It too, has sponsored research by many groups in these areas (for example: Schneider, et al., 2002; Nerenz, et al., 2002 ; Virnig, et al., 2001). As mentioned, many of the reviews and the researchers have relied heavily on Medicare data and/or studies which have used Medicare data. This has lead to more and more sophisticated analyses attempting to better understand and correct the disparities in health and healthcare. Cooper et al., 2004; Gornick, et al., 2004; Epstein, et al., 2003; are a few examples.

In addition to using the Medicare administrative data itself and encouraging others to use it by funding contractors such as the Research Data Assistance Center (ResDAC – [www.resdac.umn.edu](http://www.resdac.umn.edu) ), CMS is currently requiring Medicare managed care (MMC) organization and Quality Improvement Organizations (QIOs) to conduct activities to reduce racial/ethnic disparities within their organizations (MMC) or states (QIO).

MMC plans that participated in M+C program in 2003 were required to participate in a special project on either racial and ethnic disparities in care, a cultural and linguistically appropriate services carried out under the Quality Improvement System for Managed Care (QISMC) (See <http://www.cms.hhs.gov/cop/2d.asp> and <http://www.cms.hhs.gov/cop/2d1.asp> ). In addition to the individual plan efforts, CMS has had conducted the MMC CAHPS surveys since 1998, described in section 2.1.2, in order to better understand racial/ethnic disparities among managed care enrollees.

Numerous reports and publications based on these surveys are available (<http://www.cms.hhs.gov/researchers/projects/consumers/cahps.asp> ), and several focus on issues related to health disparities: for example, a comparison between racial/ethnic groups with respect to smoking (<http://www.cms.hhs.gov/researchers/projects/consumers/mc1overview.pdf> ); the health status of American Indians compared with non-Hispanic Whites ([http://www.cms.hhs.gov/researchers/projects/consumers/2mc\\_overview.pdf](http://www.cms.hhs.gov/researchers/projects/consumers/2mc_overview.pdf) ); and a discussion of methodological issues regarding studying health disparities in Medicare managed care plans ([http://www.cms.hhs.gov/researchers/projects/consumers/4mc\\_overview.pdf](http://www.cms.hhs.gov/researchers/projects/consumers/4mc_overview.pdf) ).

Currently, the QIOs are operating under the guidance of the 7<sup>th</sup> Scope of Work (SOW) which began in August 2002, in 19 of the 53 QIOs and in approximately the same number 6 and 12 months later. Task 1e of the SOW required 52 of the QIO to identify and carry out a project to address a health disparity involving on of the following underserved populations: African Americans, Hispanics, American Indians/Alaskan Natives, or Asian Americans and Pacific Islanders; beneficiaries who are also enrolled in Medicaid; or beneficiaries who live in rural areas. The goal is to improve the delivery and appropriate use of health services, eliminate barriers to care, improve data sources, and improve the understanding of the problems of health disparities among the QIOs and others. The outcome or performance measures for the interventions were the same measures used for the seven hospital and physician office clinical conditions: Acute myocardial infarction, congestive heart failure, pneumonia, surgical infection, adult immunizations, diabetes and breast cancer screening. They included: improving the rates of immunization with influenza or pneumonia vaccines, increasing mammography rates, and increasing the percent of the population with diabetes who receive appropriate care, etc. QIOs selected the clinical condition and the performance measure based on the disparity in the state or territory. A minimum of a 7% disparity in the relevant outcome measurement between the Caucasian Medicare population and an underserved/rural Medicare population was required.

## QIO 7<sup>th</sup> Scope of Work Projects to deal with Disparities

### *Target Population*

African Americans	22
American Indians	3
Dually Enrolled	3
Hispanics	4
Pacific Islander	1
Rural	19

### *Clinical Condition*

AMI	1
Breast Cancer	9
Diabetes	23
Heart Failure	7
Adult Immunization	5
Pneumonia (inpt)	6
In-pt septicemia	1

Reference: [http://www.medqic.org/assets/iom/QI\\_Underserved\\_Rural.pdf](http://www.medqic.org/assets/iom/QI_Underserved_Rural.pdf)

Because no QIO has reached the end of the 7<sup>th</sup> SOW, the results of these interventions are pending for public review. Information on the progress of the individual projects is available to those with access to the Standard Data Processing System (STPS) supported by the Office of Clinical Standards and Quality of CMS.

A final CMS activity of particular interest and use to the minority research community trying to address health disparities has been the research grant program for minority researchers in the Office of Research and Demonstrations (ORDI), header by Richard Bragg. Since 1998, this program has award grants to researchers from Historically Black Colleges and Universities (HBCUs) and Hispanic Serving Institutions <http://www.cms.hhs.gov/researchers/priorities/grants.asp#HBCU> . Annual conferences have been held to present the research findings. In addition, the same Office has funded the Morehouse School of Medicine to conduct introductory workshops for researchers from these institutions on the use of Medicare and Medicaid data.

#### **4.0 Limitations of available data and potential opportunities**

The limitations of the existing databases have been presented in Section 1, and many of the opportunities for improvement were discussed then or in Section 2 in describing the potentials for linking databases and surname matching. Because of the interaction of race/ethnicity with education, income and other measure of socioeconomic status at the individual level, the more information of this type that can be included in the EDB, the better.

One such addition is available from the SSA. In addition to the Master Beneficiary Record (MBR) and the Numerical Identification (NUMIDENT) files mentioned frequently earlier, the SSA also maintains the Master Earnings File (MEF), mentioned briefly (<http://policy.ssa.gov/poms.nsf/lnx/0103907001>). A record is established for a person in the MEF at the same time a NUMIDENT record is created. The record is a running tally of earnings information received from the W-2 form and updated weekly. It includes both FICA earnings and total earnings compensation “to date” in its summary segment, as well as the required linking information (SSN) to transfer all the information to the EDB. While the earnings information would not capture all income, it would enhance the information currently in the EDB.

The addition of geographic-based group-level socioeconomic information to individual-level Medicare records has been carried out for the merged Surveillance, Epidemiology, and End-Results program (SEER) – Medicare data, mentioned earlier. The SEER-Medicare Patient Entitlement and Diagnosis Summary Files (PEDSF) provides Census Bureau information regarding population density, income, education, percent of English

speaking households, etc. for the census tract and zip code of residence of each beneficiary, as well as the Health Services Area from the Area Resource file ([http://healthservices.cancer.gov/seermedicare/aboutdata/pedsf\\_2003.pdf](http://healthservices.cancer.gov/seermedicare/aboutdata/pedsf_2003.pdf) ). Medicare could perform the same, or a similar service, for researchers and append such information to the Denominator file. This would encourage, facilitate, and to some extent, standardize the use of such information. For these reasons, it is recommended.

A question could be raised whether to focus efforts to improve the quality of the race/ethnicity information and its application on a sample of beneficiaries, say the routine Medicare 5% sample of beneficiaries, or to work to improve the data for 100%. In July 2003, the 5% Denominator file contained approximately 1.74 million Whites and 199,000 Blacks. However, there were only 32,000 Asians, 48,000 Hispanics and 7,600 American Indians/Alaskan Natives. While these sample sizes would be adequate to study the use of such things as preventive services, certain disease incidence rates and overall hospital mortality rates, they are inadequate if the population needs to be divided into two gender groups and three age groups, for example. Studies of particular diseases, except the most common, or the outcomes of disease-specific hospitalizations or particular procedures, would also be difficult to impossible using the 5% sample except for studies of Whites and Blacks. Studies of state-specific events and outcomes would not be possible. Therefore, efforts should continue to focus on improving the quality of the race/ethnicity data for the entire Medicare population.



Beyond the scope of this document, is a presentation of the limitations of the Medicare data with respect to information not uniquely related to understanding issues concerning race and ethnicity and reducing disparities in health and healthcare use. However, a few examples include health behaviors, height and weight information, employment history, and laboratory test values. The first two would include historic information, as well as information about change over time. They are not likely to be collected in such a dynamic way. The inability to get adequate work history information, particularly with respect to information that directly affects health such as chemical or radiation exposure, is well known and would not be fruitful for CMS to pursue at this time. However, the values of laboratory tests are communicated electronically between many of the providers of these services and the users of this information in a standardized nomenclature and format, the Logical Observation Identifiers Names and Codes (LOINC®) (C.J. McDonald, Regenstreif Institute for Health Care, Indiana University, personal communication, and <http://www.regenstrief.org/>). Thus, the possibility exists for CMS to obtain these data, and this should be pursued.

## **5.0 Summary and recommendations**

Currently, the collection of race/ethnicity information is reaching a plateau regarding the number of minority group beneficiaries identified and the validity of the race/ethnicity information retained by Medicare. While studies using previous and current race/ethnicity information have been extremely useful, improving the race variable will enhance future studies of racial and ethnic disparities in health and healthcare, as well as

the policies and programs developed from them. Various options exist for improving the quantity and quality of the race/ethnicity data, and the following are recommended:

1. Have the Social Security Administration (SSA) replace or update the SS-5 form used to collect race/ethnicity information. The new form must use two questions, one regarding ethnicity, and one regarding race. Modify the Enrollment Database (EDB), and downstream Centers for Medicare and Medicaid Services (CMS) files to accommodate this change. The full impact of this modification will be long in coming. Also, because the entry of race/ethnicity information on the SS-5 form voluntary, it will never reach 100%. In spite of its deficiencies, the use of the SSA-based race information as the basis for the race in the utilization files should be maintained. In addition, the insertion of the race/ethnicity information contained on the EDB on to claim records by the Common Working File regional host sites should be continued. As has been stated previously, one of the great strengths of the Medicare databases is the consistency of the race/ethnicity information between the enrollment and the utilization files.
2. Consistent with current plans, the new Master Beneficiary Database (MBD) must have separate variables for ethnicity and for race. In addition, the MBD, or Denominator and utilization files derived from it, should allow race/ethnicity information from other sources to be included. Because there are many potential sources of self-reported or surname based information, the number of times this information is stored would have to be limited to say, four or five.
3. Managed care plans must be required to collect and report to CMS the race and ethnicity of all enrolled members.

4. Managed care plans must report Health Plan and Employer Data Information Set (HEDIS) information to the National Commission for Quality Assurance (NCQA) and CMS by race and ethnicity. Even those plans with small numbers of minority group members need to report so that their information combined with that from other plans.
5. In anticipation of storing race/ethnicity information from multiple sources in the MBD, CMS should review the larger surveys indicted in this report, evaluate the quality of the race/ethnicity data, and develop a mechanism to transfer the data to CMS.
6. At the time of the creation of new Denominator files, the Denominator files for the previous 5 years should be updated with the most current race/ethnicity data, and be made available to CMS and other investigators.
7. Surname matching of Hispanics should be done, and that information stored in the MBD and future Denominator files.

Other possible activities include:

1. Repeat mailing(s) to persons with race/ethnicity of Other and Unknown, as well as other beneficiaries selected on surname or place of residence, similar to that done in 1997. It will unlikely have a greater response rate than the 1997 survey, 39%, without major pre-survey publicity directed at minority groups regarding the importance of the information.

2. Creating special research databases to study minority populations. CMS could create databases from existing enrollment and utilization files containing a recommended number of one or all minority groups, as well as Whites as a reference population. Creating a set of static data sets with defined numbers or percentages of beneficiaries of each race/ethnicity might be of assistance to beginning researchers or for the generation of descriptive statistics. However, it is an inferior choice to allocating the necessary resources to allow CMS to maintain the data and the extract systems to meet the needs of government and non-government researchers as they arise.

## References

Adler GA. A profile of the Medicare Current Beneficiary Survey. *Health Care Financing Rev* 2000;15:153-163

Arday SL, Arday DR, Monroe A, Zhang MD. HCFA's racial and ethnic data: current accuracy and recent improvements. *Health Care Financing Rev* 2000;21:107-115.

Ayanian JZ, Udvardelyi IS, Gastonis CA, Pashos CL, Epstein AM. Racial differences in the use of revascularization procedures after coronary angiography. *JAMA* 1993;269:2642-2646.

Brown ML, Riley GF, Schussler N, Etzioni R. Estimating Health Care Costs Related to Cancer Treatment From SEER-Medicare Data. *Medical Care* 2001;40:Supplement:IV-104-IV-117.

CMS. Medicare Quality Improvement Program, 8<sup>th</sup> Scope of Work. Available at <http://www.cms.hhs.gov/qio/2s.pdf>.

Collins KS, Hughes DL, Doty MM, Ives BL, Edwards JN, Tenny. Diverse communities, common concerns: assessing health care quality for minority Americans --- findings from The Commonwealth Fund 2001 Health Care Quality Survey, New York, NY: The Commonwealth Fund, 2002; publication number 523.

Cooper GS, Koroukian SM. Racial disparities in the use and indications for colorectal procedures in Medicare beneficiaries. *Cancer* 2004;100:418-424.

DHHS. Healthy People 2010. (conference ed., in 2 vols). Washington, D.C.: US Department of Health and Human Services, 2000.

Eggers PW, Greenberg LG. Racial and ethnic differences in hospitalization rates among aged Medicare beneficiaries, 1998. *Health Care Financing Rev* 2000;21:91-105.

Eppig FJ, Chulis GS. Matching MCBS (Medicare Current Beneficiary Survey) and Medicare data: the best of both worlds *Health Care Financing Review* 18 3 1997 211 229

Epstein AM, Weissman JS, Schneider EC, Gastonis C, Leape LL, Piana RN. Race and gender disparities of cardiac revascularization: do they reflect appropriate use of procedures or problems in quality of care. *Med Care* 2003;41:1240-1255.

Escarce JJ and McGuire TG. Methods for using Medicare data to compare procedure rates among Asians, Blacks, Hispanics, Native Americans and Whites. *Health Services Research* 1003;35:1304-17.

Fleming C, Fisher ES, Cheng CH, Bubolz TA, Malenka DJ. Studying outcomes and hospitalization in the elderly: the advantages of a merged database for Medicare and Veterans Affairs hospitals. *Medical Care* 1992;37:601-614.

Gornick ME, Eggers PW, Riley, GF. Associations of race, education and patterns of preventive service use with stage of cancer at time of diagnosis. *Health Services Research* 2004;39:1403-1427.

Health Care Financing Administration. Special Report: Hospital Data by Geographic Area for Aged Medicare Beneficiaries: Selected Diagnosis Groups, 1986. Volume 1. HCFA Pub 03300, June 1990.

Health Care Financing Administration. Special Report: Hospital Data by Geographic Area for Aged Medicare Beneficiaries: Selected Procedures, 1986. Volume 2. HCFA Pub 03301, June 1990.

Institute of Medicine. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. National Academy Press. Washington D.C. 2000.

Javitt JC, McBean AM, Nicholson GA, Babish JD, Warren JL, Krakauer H. Undertreatment of glaucoma among black Americans. *JEJM* 1991;325:1418-1422.

Kasike BE, Newlan JF, Riggio RR, Danovich GM, Kahana L, Alexander SR, White MG. The effect of race on access and outcomes of transplantation. *N Eng J Med* 1991;324:302-307.

Lauderdale DS, Goldberg J. The expanded racial and ethnic codes in the Medicare data files: their completeness of coverage and accuracy. *AJPH* 1996;86:712-716.

Lurie N, Zhen C, Sangl J, Bierman AS, Sekscenski S. Variations in racial and ethnic differences in consumer assessments of health care. *Am J Manag Care* 2003;9:502-509.

Mayberry RM, Mili F, Ofili E. Racial and ethnic differences in access to medical care. *Medical Care Research and Review* 2000;57:108-145.

McBean AM, Babish JD, Prihoda R. The utilization of pneumococcal polysaccharide vaccine among elderly Medicare beneficiaries, 1985 through 1988. *Arch Intern Med* 1991;151:2009-2016.

McBean AM, Haung Z, Virnig BA, Lurie N, Musgrave D. Racial variation in the control of diabetes among elderly Medicare managed care beneficiaries. *Diabetes Care* 2003;26:2350-2356.

McBean AM, Jung K, Virnig BA. Improved care and outcomes among elderly Medicare managed care beneficiaries with diabetes. *Am J Manag Care*, in press.

Nerenz DR, Bohnham YL, Green-Weir R, Joseph C, Gunter M. Eliminating racial/ethnic disparities in health care: can health plans generate reports? *Health Affairs* 2002;21:259-263; Also, available from The Commonwealth Fund as Publication #557.

Office of Management and Budget. Revisions to the standards for the classification of Federal data on race and ethnicity. *Federal Register* 1997;62:58781-58790.

Panis C, Euller R, Grant C, et al. *SSA Program Data User's Manual*. Rand Corporation, June 2000.

Rosenberg HM, Maurer JD, Sorlie PD, Johnson NJ, et al.. Quality of death rates by race and Hispanic origin: A summary of current research, 1999. National Center for Health Statistics. *Vital Health Stat* 2(128). 1999.

Schneider EC, Zaslavsky AM, Epstein AM. Racial disparities in the quality of care for enrollees in Medicare managed care. *JAMA* 2002;287:1288-94.

Udvarhelyi IS, Gastonis C, Epstein AM, Pashos CL, Newhouse JP, McNeil BJ. Acute myocardial infarction in the Medicare population. Process of care and clinical outcomes. *JAMA* 1992;268:2530-2536.

Virnig BA, Fisher ES, McBean AM, Kind S. Hospice use in Medicare managed care and fee-for-service systems. *Am J Manag Care* 2001;7:831-833.

Virnig BA, Lurie N, Huang Z, Musgrave D, McBean AM, Dowd B, Racial variation in quality of care among Medicare+Choice enrollees. *Health Affairs* 2002;21:224-230.

Virnig B, Huang Z, Lurie N, Musgrave D, McBean AM, Dowd B. Does Medicare managed care provide equal treatment for mental illness across races? *Arch Gen Psychiatry* 2004;61:201-205.

# APPENDICES

## Appendix I

---

### **Policy Statement on Inclusion of Race and Ethnicity in HHS Data Collection Activities**

#### **Summary**

This document describes the Department of Health and Human Services (HHS) policy on the inclusion of racial and ethnic categories in HHS funded and sponsored data collection and reporting systems. This inclusion policy covers all programs of the Department, including both health and human/social services. It is consistent with existing inclusion policies that have been implemented by the National Institutes of Health (NIH) and the Centers for Disease Control and Prevention (CDC). This inclusion policy clearly states that the minimum standard categories of racial and ethnic groups specified in the Office of Management and Budget (OMB) Directive 15 and future revisions thereof should be collected and reported in all HHS data systems except those exempted by this policy. The need for the HHS-wide policy is in part caused by an incomplete adherence to use of the standard categories in the OMB Directive 15 by HHS agencies, even when their use is feasible and appropriate. OMB Directive 15 specifies the minimum racial and ethnic categories that are to be used when race and ethnicity are included in data collection and reporting; it does not *require* that race and ethnicity be included in data collection and reporting. This inclusion policy does require the collection and reporting of racial and ethnic groups in HHS data collection activities.

This policy reaffirms the HHS commitment to the appropriate inclusion of data on minority groups in its research, services, and related activities. This inclusion policy is needed to help monitor HHS programs to determine that Federal funds are being used in a nondiscriminatory manner and to promote the availability of standard racial and ethnic data across various agencies when the Department is required to make a coordinated response to major health and human services issues. Implementation of this policy will help to identify major health conditions of minority populations, monitor progress in meeting their needs, and help to ensure nondiscrimination in access to and provision of appropriate HHS services for various racial and ethnic groups. HHS encourages the expanded collection of data that will improve research on disparities in health status and social services needs between minority groups and the general population. This inclusion policy applies to HHS program administrative records as well as research and survey data but does not cover HHS personnel employment data. Programs that are directed to minority populations have exemptions; but they are encouraged to collect and report data on groups within their target minority populations.

#### **I. INTRODUCTION AND POLICY RATIONALE**

##### **A. Background**

The Department of Health and Human Services is the United States government's principal agency for protecting the health of all Americans and providing essential human services, especially for those who are least able to help themselves. Consistent with its mission, the Department of Health and Human Services is



committed to the following goals: (1) addressing racial and ethnic disparities in health; (2) ensuring that members of all racial and ethnic groups are provided appropriate levels of health and social services, and (3) assuring nondiscriminatory health care access and treatment and access to other Departmentally funded or directly operated services.

A recent review of HHS data systems found that: (1) some HHS data systems do not collect data on the race and ethnicity of its respondents or participants, (2) in some cases, racial and ethnic data have been collected but not reported, and (3) minimum racial and ethnic categories reported are sometimes not consistent across the various HHS agencies. While the available data indicate important disparities in health conditions between minority groups and the general population, recent reports to HHS identified major omissions of the racial and ethnic data necessary to address the minority populations' specific health and social services needs (See References).

## **B. Rationale for an HHS Policy on the Inclusion of Racial and Ethnic Data in HHS Data Collection and Reporting Activities**

The need for an HHS-wide policy on the inclusion of data on racial and ethnic groups in HHS data collection and reporting activities is based on the following:

1. Many Departmental and Public Health Service reports on the health status of minorities during the past decade have indicated that the limited data available show racial and ethnic minorities have significant health disparities compared with the rest of the population. Consistent, reliable racial and ethnic data are needed to develop and implement effective prevention, intervention, treatment, and other needed health programs, policies, and services.
2. The Department needs standard and reliable racial and ethnic data across the various HHS agencies and major operating components when it is necessary for the Department to make a coordinated response to major health and social services issues.
3. The Department is committed to developing meaningful standards and criteria that would improve its ability to determine and analyze the efficacy of HHS data collection activities for ensuring nondiscrimination in all HHS funded and directly operated programs.
4. Section 80.6(b) of CAR 45 Part 80 implementing Title VI of the Civil Rights Act of 1964 requires each recipient of Federal financial assistance to keep such records as "the responsible Department official or his designed may determine to be necessary to enable him to ascertain whether the recipient has complied or is complying with this part. For example, recipients should have available for the Department, racial and ethnic data showing the extent to which members of minority groups are beneficiaries of and participants in Federally assisted programs." A policy on the inclusion of racial and ethnic categories would provide uniform guidance to all HHS programs regarding data that may be used to help determine their compliance with Title VI.
5. A recent review of HHS data systems has found that not all HHS data systems collect data on the race and ethnicity of its respondents or participants when the subject matter is relevant to important HHS goals. This is partly because OMB Directive 15 provides the minimum racial and ethnic categories to be used in Federal

data collections but does not require that all relevant data systems must collect data on race and ethnicity.

6. Recent HHS task forces and initiatives on minority health have identified major omissions of needed data on racial and ethnic groups to address their specific health and social services needs. These task forces have recommended to several agencies that they need both improved and more consistent collection and reporting of health and social services data on racial and ethnic groups. The number and range of these recommendations covering data collection, analysis, and dissemination suggest the need for a Department-wide policy.

The purposes of this inclusion policy, therefore, are: (1) to ensure that data on race and ethnicity are collected in all HHS systems obtaining information relevant to the Department's goals, (2) to ensure that such data are collected and reported in a standardized manner, and (3) to address the various major health data omissions identified for minority racial and ethnic groups.

Agencies and Operating Divisions of HHS are expected to develop any mechanisms needed to implement this policy. When implemented, the policy will enable the Department to more effectively fulfill its mission to assure the health and well-being of the Nation and to ensure, on an ongoing basis, that Federal funds are being used in a nondiscriminatory manner. Moreover, the implementation of this policy would help satisfy the Department's need for consistent and relevant racial and ethnic data when its response to major health and social services issues must be based on data from various of its components.

## **II. POLICY ON THE COLLECTION AND REPORTING OF RACE AND ETHNICITY**

The HHS-wide policy on the inclusion of race and ethnicity in the data collection and reporting for programs, research, and survey activities funded or sponsored by HHS or its Agencies or other major operating components is as follows:

**A. Inclusion of Race and Ethnicity:** Data on race and ethnicity will be included in all data collection and reporting activities covered by this policy.

**B. Minimum Standard of Racial and Ethnic Categories:** For Federal systems of records, the minimum standard for the basic racial and ethnic categories will be OMB Directive 15 and any subsequent revisions. All references to OMB Directive 15 in this policy for Federal systems of records are to be understood as encompassing all subsequent revisions to Directive 15. For non-Federal systems of records, the minimum standard for civil rights compliance purposes will be the current OMB Directive 15 minimum standard. However, HHS encourages use of subsequent revisions to the Directive in non-Federal systems of records, when feasible.

**C. Groups within the Minimum Standard OMB Directive 15 Racial and Ethnic Categories:** HHS recognizes the diversity of the population within each of these minimum categories and encourages the inclusion of subgroups when such inclusion improves the usefulness of the data.

**D. Administrative Level of Reporting for Program Services Data:** Wherever possible, racial and ethnic data regarding populations served by HHS-funded programs should either be collected and reported at the providing organizational level and program beneficiary level or be available at that level through use of existing data systems (e.g., matching of enrollment and claims data) so as to be

useful in assessing compliance with Title VI of the Civil Rights Act of 1964. Such information would be collected either directly from such entities, or through use and/or matching of existing administrative data sets, including upgrading of such data sets as appropriate to contain information consistent with Directive 15 reporting categories.

**E. Collecting and Reporting Data on Race and Ethnicity:** Data on race and ethnicity must be collected, analyzed, and reported in an objective, accurate, and useful manner. Both the collection and reporting of the data must be sensitive to racial and ethnic communities' concerns about potential misuse or abuse. Such data will not be used by the Department in a way that would stigmatize certain populations or to suggest a biological or genetic connection based on nongenetic studies or when race and ethnicity are actually surrogates for other risk factors. Only those racial or ethnic groups with adequate sample sizes to provide statistically reliable data should be reported. Information on the validity and reliability of the data should be included whenever possible to enable the readers to judge the credibility of the findings.

**F. Data Collection and Reporting Activities Covered by this Policy:** This policy applies to the following types of data collection systems:

1. Statistical data collections (e.g., vital statistics, disease registries, and other research and survey data such as those collected in national health status surveys, longitudinal research surveys, and studies of access, utilization, and financing of health and social services).
2. Administrative records (e.g., those used for research, for general program administrative, contracts, and grants reporting, and for assessing civil rights compliance).
3. Research, evaluation, and other study projects (e.g., intramural research).
4. Applications, grants, and contract proposals submitted to the Department and its agencies or major operating components that collect data from the public.
5. Reporting systems for civil rights compliance under Section 80.6(b) of CAR 45 Part 80, implementing Title VI of the Civil Rights Act of 1964.

### **III. EXEMPTIONS FROM THE POLICY**

A. Exemptions to this inclusion policy for data on race and ethnicity for HHS funded or sponsored data collection activities are as follows:

1. The data collection activities of an HHS Agency, component, or HHS funded program that are directed to one or a limited number of minority racial or ethnic groups are not required to include all the minimum standard categories of OMB Directive 15 but are encouraged to collect and report data on the subgroups within their targeted minority group. An example is the Indian Health Service.
2. A services program may be exempted from the minimum standard categories for civil rights compliance reporting when the program is directed by Federal law to one or a limited number of minority racial or ethnic groups and would include data on only the minority groups or subgroups to which the program was directed.

3. When consultation with the Agency statistician determines that the data on particular racial and ethnic groups are considered statistically unreliable, then such racial and ethnic groups should not be reported separately unless accompanied by the appropriate caveats.
4. Data collection and reporting for activities within HHS that are not health or human services program administrative, research, survey, or services reporting or assessment are not subject to this policy. In some cases, these areas are covered by the policies and regulations of other Departments (e.g., employment activities and information are covered by the rules and regulations of the Office of Personnel Management).
5. Data from activities that have not been sponsored or funded by HHS but which are used by HHS for regulatory, research, or other purposes may be excluded. An example is clinical trial data that the FDA receives in support of product approval.
6. U.S. commonwealths, trusts, and territories and other areas where the Bureau of Census does not use OMB Directive 15 standard categories in either the questions in the decennial census or in reporting the responses are exempt from this policy.
7. Special exemptions may be granted on a case-by-case basis by the HHS Secretary or a designee.

#### **IV. Racial and Ethnic Data Collection Enhancement Options**

This HHS-wide inclusion policy requires only the inclusion of data on race and ethnicity as defined by OMB Directive 15 or any subsequent revisions. It does not require but does encourage HHS-wide collection and reporting of cultural or other data variables related to race or ethnicity. The Department encourages its Agencies, components, and funded programs to improve the availability of racial and ethnic data to better understand and improve the health and well-being of minority racial and ethnic populations. Therefore:

- A. The Department encourages the collection of data on variables other than race and ethnicity that may be useful in assessing and improving the health and well-being of minority populations and the provision of needed health and social services. Program officials and researchers are encouraged to collect, study, and report on cultural background, socioeconomic status, and other important characteristics and conditions that can assist HHS to fulfill its mission of improved health and well-being for all its constituents.
- B. The Department supports innovative uses of currently available data consistent with the Privacy Act and confidentiality constraints. Where statistically and methodologically appropriate, HHS encourages pooling the data from several years, analyses such as cross-comparisons from different data sets, and specialized studies and linkage of data sets. These techniques can be used with existing data to enhance our understanding of the health status and social services needs of minority racial and ethnic populations.
- C. In cases in which HHS funded or sponsored data activities rely on compiling data collected by States or other entities that do not include racial and ethnic data or do not use the OMB Directive 15 Standard, the Department encourages such primary data collection entities to review their systems and make changes as appropriate.

#### **V. RELATED POLICIES**

Both the National Institutes of Health (NIH) and the Centers for Disease Control and Prevention (CDC) have policies on the inclusion of racial and ethnic minorities in research. These two existing inclusion policies are consistent with this HHS-wide policy and should not require revision regarding statistical reporting.

## REFERENCES

Asian American and Pacific Islander Health Organizational Leaders. *Recommendations for Improving the Health of Asian Americans and Pacific Islanders*. Report by delegates to the First National Health Summit of Asian American and Pacific Islander Health Organizational Leaders held in San Francisco, California. June 21-24, 1995.

American Indian Health Care Association. *Enhancing Health Statistics for American Indian and Alaskan Native Communities: An Agenda for Action*. A report to the National Center for Health Statistics. Hyattsville, Maryland, December 1992.

Centers for Disease Control and Prevention. Use of Race and Ethnicity in Public Health Surveillance. Summary of the CDC/ATSDR Workshop. MMWR 1993; 42(No. RR-10).

Department of Health and Human Services Working Group on Hispanic Issues. *Hispanic Agenda for Action: Improving Services to Hispanic Americans*. A Report to the Secretary. Washington, D.C., July 29, 1996.

Department of Health and Human Services Data Council's Ad Hoc Working Group on Racial and Ethnic Data. *Meeting Current and Future Demands for Racial and Ethnic Data for Health and Human Services*. Rockville, Maryland, July 1996 Draft.

Department of Health and Human Services, National Center for Health Statistics. *Setting a Research Agenda: Challenges for the Minority Health Statistics Grants Program*. U.S. DHHS/CDC/NCHS Working Paper Series No. 40. Hyattsville, Maryland, December 1991.

Heckler, Margaret. *Report of the Secretary's Task Force on Black and Minority Health. Volume I: Executive Summary*. U.S. Department of Health and Human Services. Washington, D.C., 1985.

National Asian Pacific American Families Against Substance Abuse, Inc. Letter to Mr. Peter Edelman. December 17, 1993.

Public Health Service Task Force on Minority Health Data. *Improving Minority Health Statistics*. U.S. Public Health Service. Washington, D.C., 1992.

U.S. Surgeon General's Hispanic/Latino Health Initiative. *One Voice, One Vision -- Recommendations to the Surgeon General to Improve Hispanic/Latino Health*. Washington, D.C., June 1993.

## Appendix II

### Appendix G Policy Statement on Inclusion of Race and Ethnicity in HHS Data Collection Activities

October 24, 1997

**TO:** Heads of Operating Divisions

Heads of Staff Divisions

**SUBJECT:** HHS Policy for Improving Data - ACTION

Agencies and programs of the Department require data on race and ethnicity for a variety of purposes ranging from research, public health surveillance, program administration and civil rights enforcement. While many HHS data collection systems do include data on race and ethnicity, not all do so and no clear policy currently exists. Accordingly, at the recommendation of the HHS Data Council, I am issuing the attached policy on the inclusion of information on race and ethnicity in HHS data collection systems. The policy reaffirms the HHS commitment to the appropriate inclusion of data on minority groups in our research, services and related activities. The policy also will help us monitor HHS programs to determine that funds are being used in a nondiscriminatory manner and to promote the availability of standard race and ethnicity data across agencies when the Department is required to make a coordinated response to major health and human services issues. It is consistent with policies already adopted by the National Institutes of Health and the Centers for Disease Control and Prevention regarding the inclusion of minorities in research.

In general, the policy described in the attached material requires the inclusion of information on race and ethnicity in all HHS-sponsored data collections systems, with certain exceptions. The policy also requires that the minimum standards specified by the Office of Management and Budget (OMB) for race and ethnicity data collection and reporting be employed, including any subsequent revisions to the OMB standards. The policy will go into effect as of November 1, 1997. I am directing Heads of OPDIVs and STAFFDIVs to implement the policy within your organizations in accordance with normal agency data planning, OMB clearance, data collection and analysis cycles.

Please make every effort to assure the successful implementation of this policy. Questions about the policy should be directed to your agency's representative on the HHS Data Council (attached).

/s/

Donna E. Shalala

Attachments

cc:

Co-chairpersons

HHS Data Council

## **Appendix III**

THE SECRETARY OF HEALTH AND HUMAN SERVICES  
WASHINGTON, D.C. 20201

### **CHARTER**

#### **NATIONAL COMMITTEE ON VITAL AND HEALTH STATISTICS**

##### **PURPOSE**

Collection, analysis and dissemination of health and health-related information is a crucial aspect of the responsibilities of the Department of Health and Human Services. The Department also plays a national leadership role in health data standards and health information privacy policy, and is charged with the responsibility for implementation of the Administrative Simplification provisions of the Health Insurance Portability and Accountability Act of 1996. In addition, the Department engages in cooperative efforts with other countries and the international community to foster health data standards, comparability and cross-national research.

The National Committee on Vital and Health Statistics is the Department's statutory public advisory body on health data, statistics and national health information policy. This Committee shall serve as a national forum on health data and information systems. It is intended to serve as a forum for the collaboration of interested parties to accelerate the evolution of public and private health information systems toward more uniform, shared data standards, operating within a framework protecting privacy and security. The Committee shall encourage the evolution of a shared, public/private national health information infrastructure that will promote the availability of valid, credible, timely and comparable health data. With sensitivity to policy considerations and priorities, the Committee will provide scientific-technical advice and guidance regarding the design and operation of health statistics and information systems and services and on coordination of health data requirements. The Committee also shall assist and advise the Department in the implementation of the Administrative Simplification provisions of the Health Insurance Portability and Accountability Act, and shall inform decision making about data policy by HHS, states, local governments and the private sector.

##### **AUTHORITY**

42 U.S.C. 242k(k), Section 306(k) of the Public Health Service Act, as amended. The Committee is governed by provisions of Public Law 92-463, as amended, (5 U.S.C. App. 2), which sets forth standards for the formation and use of advisory committees.

##### **FUNCTION**

It shall be the function of the Committee to assist and advise the Secretary through the

Department of Health and Human Services Data Council, on health data, statistics, privacy, national health information policy, and the Department's strategy to best address those issues.

Specifically, the Committee shall advise the Department in the following matters:

(A) Monitor the nation's health data needs and current approaches to meeting those needs; identify emerging health data issues, including methodologies and technologies of information systems, databases, and networking that could improve the ability to meet those needs.

2

(B) Identify strategies and opportunities to achieve long-term consensus on common health data standards that will promote (i) the availability of valid, credible, and timely health information, and (ii) multiple uses of data collected once; recommend actions the federal government can take to promote such a consensus.

(C) Make recommendations regarding health terminology, definitions, classifications, and guidelines.

(D) Study and identify privacy, security, and access measures to protect individually identifiable health information in an environment of electronic networking and multiple uses of data.

(E) Identify strategies and opportunities for evolution from single-purpose, narrowly focused, categorical health data collection strategies to more multi-purpose, integrated, shared data collection strategies.

(F) Identify statistical, information system and network design issues bearing on health and health services data which are of national or international interest; identify strategies and opportunities to facilitate interoperability and networking.

(G) Advise the Department on health data collection needs and strategies; review and monitor the Department's data and information systems to identify needs, opportunities, and problems; consider the likely effects of emerging health information technologies on the Department's data and systems, and impact of the Department's information policies and systems on the development of emerging technologies.

(H) Stimulate the study of health data and information systems issues by other organizations and agencies, whenever possible.

(I) Review and comment on findings and proposals developed by other organizations and agencies with respect to health data and information systems and make recommendations for their adoption or implementation.

(J) Assist and advise the Secretary in complying with the requirements imposed under Part C of Title XI of the Social Security Act;

(K) Study the issues related to the adoption of uniform data standards for patient medical record information and the electronic interchange of such information, and report to the Secretary not later than August 21 2000 recommendations and legislative proposals for such standards and electronic exchange;

(L) Advise the Secretary and the Congress on the status of the implementation of Part C of Title XI of the Social Security Act;



(M) Submit to the Congress and make public, not later than one year after the enactment of the Health Insurance Portability and Accountability Act, and annually thereafter, a report regarding the implementation of Part C of Title XI of the Social Security Act. Such report shall address the following subjects, to the extent that the Committee determines appropriate:

3

- The extent to which persons required to comply with Part C of the Act are cooperating in implementing the standards adopted under such part;
- The extent to which such entities are meeting the security standards adopted under such part and the types of penalties assessed for noncompliance with such standards.
- Whether the federal and State Governments are receiving information of sufficient quality to meet their responsibilities under such part.
- Any problems that exist with respect to implementation of such part.
- The extent to which timetables under such part are being met.

(N) Assist and advise the Secretary in the development of such reports as the Secretary or Congress may require.

In these matters, the Committee shall consult with all components of the Department, other federal entities, and non-federal organizations, as appropriate.

#### STRUCTURE

The Committee shall consist of 18 members, including the Chair. The members of the Committee shall be appointed from among persons who have distinguished themselves in the fields of health statistics, electronic interchange of health care information, privacy and security of electronic information, population-based public health, purchasing or financing health care services, integrated computerized health information systems, health services research, consumer interests in health information, health data standards, epidemiology, and the provision of health services. Members of the Committee shall be appointed for terms of up to four years. The Secretary shall appoint one of the members to serve a two year, renewable term as the Chair.

Of the members of the Committee, one shall be appointed by the Speaker of the House of Representatives after consultation with the minority leader of the House of Representatives; one shall be appointed by the President pro tempore of the Senate after consultation with the minority leader of the Senate, and 16 shall be appointed by the Secretary.

Membership terms of more than two years are contingent upon the renewal of the Committee by appropriate action prior to its termination. Any member appointed to fill a vacancy occurring prior to the expiration of the term for which his or her predecessor was appointed shall be appointed only for the remainder of such term. A member may serve 180 days after the expiration of that member's term if a successor has not taken office.

Standing and ad hoc subcommittees, composed solely of members of the parent Committee, may be established to address specific issues and to provide the Committee with background study and proposals for consideration and action. The Chair shall appoint members from the parent Committee to the subcommittees and designate a Chair for each subcommittee. The subcommittees shall make their

recommendations to the parent Committee. Timely notification of the subcommittees, including charges and membership, shall be made in writing to the Department Committee Management Officer by the Executive Secretary of the Committee. The HHS Data Council, through the Assistant Secretary for Planning and Evaluation, shall oversee and coordinate the overall management and staffing of the Committee. Professional, scientific, and technical staff support shall be provided by all components of the Department. The National Center for Health Statistics shall provide executive secretariat and logistical support services to the Committee.

#### MEETINGS

Meetings shall be held not less than annually at the call of the Chair, with the advance approval of a Government official, who shall also approve the agenda. A Government official shall be present at all meetings.

Meetings of the subcommittees shall be held at the call of the Chair, with the advance approval of a Government official, who shall also approve the agenda. A Government official shall be present at all subcommittee meetings. All subcommittees shall report their

findings to the Committee. Meetings shall be open to the public except as determined otherwise by the Secretary; notice of all meetings shall be given to the public. Meetings shall be conducted, and records of the proceedings kept, as required by the applicable laws

and departmental regulations.

#### COMPENSATION

Members who are not full-time Federal employees shall be paid at a rate not to exceed the daily equivalent of the rate in effect for an Executive Level IV of the Executive Schedule for each day they are engaged in the performance of their duties as members of the Committee.

All members, while so serving away from their homes or regular places of business, may be allowed travel expenses, including per diem in lieu of subsistence, in the same manner as such expenses are authorized by Section 5703, Title 5, U.S. Code, for employees serving intermittently.

#### ANNUAL COST ESTIMATE

Estimated annual cost for operating the Committee, including compensation and travel expenses for members but excluding staff support, is \$639,200. Estimated annual personyears

of staff support required is 5.6, at an estimated annual cost of \$478,460.

#### REPORTS

In the event a portion of a meeting is closed to the public, a report shall be prepared which

shall contain, as a minimum, a list of members and their business addresses, the Committee's

functions, dates and places of meetings, and a summary of Committee activities and recommendations made during the fiscal year. A copy of the report shall be provided to the

Department Committee Management Officer.

**TERMINATION DATE**

Unless renewed by appropriate action prior to its expiration, the charter for the National Committee on Vital and Health Statistics will expire on January 16, 2006.

**APPROVED:**

JAN 13 2004 /s/

4

Date:

Secretary of Health and Human Services

## Appendix IV

September 26, 2003

The Honorable Tommy G. Thompson  
Secretary  
Department of Health and Human Services  
200 Independence Avenue, SW  
Washington, D.C. 20201

Dear Secretary Thompson:

The National Committee on Vital and Health Statistics (NCVHS) commends your initiatives to eliminate racial and ethnic disparities in health care. Disparities in access to and delivery of health care to racial and ethnic minorities are well documented by the research community as well as by federal agencies. The need for the collection of adequate and comparable data for racial and ethnic populations is also well documented. Your initiatives confirm the need to collect information of the race and ethnicity of individuals in order to manage, monitor and evaluate programs to prevent disease and promote better health outcomes.

One of the nation's important health goals is to eliminate racial and ethnic disparities in our health care system to ensure that all Americans receive quality health care. We commend HHS for taking the lead in promoting the collection of racial and ethnic data in the private sector, such as the issuance of the Food and Drug Administration (FDA) *Guidance for Industry on the Collection of Race and Ethnicity Data in Clinical Trials for FDA Regulated Products*. This is an important step toward obtaining accurate health-related data.

Without the collection of standardized racial and ethnic data in health plans, progress toward achieving the national goal of eliminating racial and ethnic disparities cannot be monitored. Medical service provider administrative data are a critical source of information on the race and ethnicity of individuals. However, a uniform data collection infrastructure does not exist. Thus, health plans use a variety of strategies to collect data on race and ethnicity (e.g., administrative data, electronic medical records, enrollee surveys, federal and state enrollment files for Medicare and Medicaid beneficiaries, and data linkages). Most of these efforts have been limited to members representing small subsets of assorted health plans (e.g., new enrollees, patients with particular health conditions, or a random sample of enrollees). Through testimony gathered from public and private sector health plans, large employers and business coalitions, quality oversight organizations, measurement experts, state and federal health data agencies and other interested stakeholders, the NCVHS Quality Workgroup has also identified significant data gaps.

The NCVHS recommends that HHS strongly encourage and provide support as outlined below to public and private sector health plans to collect accurate and complete racial and

ethnic data in accordance with the revised Office of Management and Budget (OMB) standard categories. To accomplish this, HHS should:

- Promote racial, ethnic, and primary language data collection and reporting by public and private health plans and provide information and expertise to assist in the accomplishment of this goal.
- Inform insurers, health plans, employers, providers, entities, and the general public that data collection and reporting by race, ethnicity, and primary language are legal and often required by law.
- Raise awareness that data collection is needed to achieve Healthy People 2010 goals, and to comply with Title VI nondiscrimination requirements.
- Support research on the best practices for collection and reporting of data by race, ethnicity, and primary language.
- Facilitate the collection of racial and ethnic data using the OMB's revised standard categories and the collection of primary language in appropriate administrative transactions mandated under HIPAA.

To do so would increase health plan capacity to:

- Provide data for identifying and correcting disparities in health care delivery.
- Become consistent with emerging health information standards that are part of the National Health Information Infrastructure and the Consolidated Healthcare Informatics Initiative.
- Study racial and ethnic differences in access to health care, health status and health care delivery.
- Facilitate development of culturally appropriate outreach, prevention and intervention programs.

Thank you for your consideration of these recommendations.

Sincerely,

/s/

John R. Lumpkin, M.D., M.P.H.  
Chair, National Committee on Vital and Health Statistics  
cc: HHS Data Council Co-Chairs