Identifying the Race/Ethnicity of SSI Recipients

Charles Scott, Social Security Administration

The Social Security Administration (SSA) produces data to help the public, Congress, and the research community assess the impacts of its programs on people. Important demographic variables include beneficiary age, race, and sex. People frequently want to know how the beneficiaries of SSA's programs are represented among various demographic groups.

Through the years, the agency has published a wide range of data on recipients of Supplemental Security Income (SSI) to answer many demographic questions. But SSA has not published extensive data on the race of SSI recipients despite collecting information on race since the 1930's. The purposes of this paper are to:

- describe the process for collecting data on the race/ethnicity of SSI recipients,
- explain the problems that limit the publication of consistent data on race,
- suggest ways that the data collection process can be improved, and
- present data on race/ethnicity for SSI recipients.

To assist in the discussion, we have chosen to work with a sample of SSI recipients in November 1998.

How SSA Collects Data on Race

Since the 1930's, SSA has collected data on race/ethnicity from people who apply for Social Security account numbers. Form SS-5, the application for account number, is the source for race/ethnicity data and contains questions about the applicant's name, date and place of birth, mother's maiden name and father's name, and race/ethnic description. Prior to 1980, the choices given on the race/ethnic question were "White," "Negro," or "Other." In the early years, the typical application was filed in order to secure an account number that, in turn, permitted the person to work. The employer then reported wages under that account number, and information from the SS-5 was used many years later to verify eligibility at the time Social Security benefits were sought. SS-5 application forms were stored at the SSA headquarters in Baltimore until a person filed for benefits, and were then returned to the 1,300 field offices across the country to assist in determinations of program eligibility.

In addition to the original application for an account number, SS-5 applications were filed whenever there was a change to any of the information previously submitted. A typical correction was a change in middle name or surname when women married. But, by far, the most common occurrence for an additional application was a request for a replacement card, as people lost their original cards. [1]

Computerizing the SS-5 file

Over time, there were several important changes to both the race/ethnicity codes and the process for reporting them. The first occurred in the mid-1970's when the SS-5 file, housed in the Baltimore headquarters, was converted to a computer file called the Numident (**num**ber **ident**ification). At that time, all existing SS-5's were placed on the new Numident file. Today, that file contains over 700 million records for 400 million account number holders. The Numident was incomplete with respect to race, however, because when the SS-5's were returned to the field after an application for SSA program benefits was filed, a special form was put in its place. This form contained most of the original SS-5 information but lacked the race code. Therefore, race data were missing for many people who were receiving program benefits when the Numident was created. That shortcoming was never corrected, and the Numident still does not have data on race/ethnicity for many people who were receiving benefits on or before 1979.

All was not lost, however. By this time, SSA had

also developed a computer file for the purpose of recording earnings data. This computer file is now called the Master Earnings File (MEF). Records are created on the MEF when an account number is issued and updated with earnings data. The original MEF record includes the race code taken from each SS-5 when it is filed. Therefore, the original race codes were split between two files—the Numident contained the codes for persons who had not yet filed for benefits, and all new SS-5's; and the MEF contained codes for all persons. The MEF, however, had certain limitations of its own with respect to race/ethnicity. While the Numident contains all SS-5 entries, their dates, and corresponding race codes, the MEF contains only a single entry for race, does not update that code, and does not associate a date with the race code. The lack of date for the code would become important in the event the code was changed, and it was in 1980.

Changing the race code

In late 1980, the race question on the SS-5 was changed at the direction of the Office of Management and Budget (OMB). OMB suggested several options. One was to separate racial and ethnic topics into separate questions. SSA decided to continue with a single question by combining the ethnic and racial topics, with permitted responses of "White," "Black," "Hispanic," "Asian or Pacific Islander," or "American Indian or Alaskan Native." This decision was effective in keeping the size of the SS-5 application to a minimum, but also muddied the waters with respect to racial and ethnic distinctions. And, worse, the new code was apparently not compatible with the old one. Account number applicants opting for the "Hispanic" designation after 1980 might well have answered as "White," "Negro," or "Other" under the pre-1980 coding scheme. It was also not clear how Asians or Native Americans would have responded under the older scheme.

Enumeration at birth

In past decades, people typically applied for a Social Security number when they sought their first employment. In recent years, people have needed a Social Security number well before they seek a job. Because the Social Security account number is used for tax purposes and because the number has become the de facto national identifier, many people need the number at birth. Beginning in 1989, SSA entered into agreements with all 50 States to provide "enumeration at birth." When an infant is born, the hospital representative asks the parent if he/she would like the birth certificate data transmitted to SSA so that an account number can be issued. The data are forwarded to the State's vital statistics office, and from there to the Social Security Administration, where a card is issued and a record created on both the Numident and MEF files. The problem with this procedure is that the race/ethnicity information is not included, because it is shown on the birth certificate under "Information for Medical and Health Use Only." This means that SSA gets no race/ethnic data at the point of birth, and receipt of these data is limited to additional applications filed in the ensuing years.

Associating the race codes with SSI recipients

In order to provide data on the racial/ethnic distribution of SSA beneficiary populations, it is necessary to take the appropriate codes from the Numident and MEF source files and place them on the appropriate beneficiary files. Those beneficiary files are the Master Beneficiary Record (MBR) for people receiving Social Security disability or retirement benefits, and the Supplemental Security Record (SSR) for persons receiving means-tested benefits under the Supplemental Security Income (SSI) program. In this paper, we will focus on the SSI population. SSI recipients are particularly useful in illustrating a discussion on race coding because they are more evenly spread among all age groups than in the other two programs.

Supplemental Security Income is a Federal welfare program for low-income people who are aged, blind, or disabled. In November 1998, there were about 6.5 million recipients of all ages. These eligible persons may apply for benefits at any of the SSA field offices across the country. Once they are found eligible for payments, a record is created on the SSR, the main computer file used in administering the program. At that time, the newly created records are matched to the Numident file to secure the latest information on race/ethnicity. A single code is brought across to the SSR, and no date is attached to it. No further association is made with the

Numident, even if subsequent SS-5's are received in the Numident and even if there was no code available at the point of award.[2] This system was established at the beginning of the SSI program in 1974. It makes two assumptions:

1. There would be only one coding scheme, and

2. There would be a race/ethnic code for almost everyone at the point of award for program benefits.

As it happens, neither of these assumptions was correct. As explained earlier, the coding scheme was changed in 1980. The second assumption was also incorrect because of the enumeration-at-birth policy, and because increasing numbers of people declined to complete the race question on their latest SS-5. Nevertheless, this system has never been changed and presents one of the principal stumbling blocks to presenting better data on race/ethnicity.

Race-coding on the SSR

To explore the system of race/ethnic coding for SSI recipients, we selected a 1-percent sample of recipients from the 6,589,000 recipients in November 1998 from the SSR. Until the very end, this report shows data for these 65,890 sample recipients without adjusting the figures to represent the entire universe of SSI recipients. Table 1 provides an age distribution of the current race/ethnicity coding on the SSR.

By arraying the data by age groups, many of the inadequacies of the SSR code become apparent. The first problem is that the overall percentage of those with some sort of legitimate code is less than 85 percent. It is not immediately clear whether that 15 percent of missing codes represents people who did not answer the question on race/ethnicity, or if the record exists but has not found its way to the SSR. And that 85-percent figure masks larger problems at either end of the age spectrum. Among recipients under the age of nine, the

Table 1. Race Codes for SSI Recipients on the SSR, by Age Group, November 1998

Race code	Total in sample	Age group: Under 9 years	9 to 17 years	18 to 29 years	30 to 39 years	40 to 49 years	50 to 64 years	65 to 74 years	75 or over
Total	**65,890**	**3,454**	**5,521**	**7,153**	**7,744**	**8,589**	**12,912**	**10,776**	**9,741**
Total with codes	55,466	1,443	4,952	6,657	7,095	7,389	10,800	9,170	7,960
Total percent	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
White	53.1	33.3	41.4	58.2	58.9	56.9	56.9	47.9	52.2
Black	29.5	41.8	44.6	30.4	30.1	31.7	27.7	23.3	24.1
Hispanic	8.4	21.3	11.1	6.5	6.3	5.5	8.0	12.5	6.6
Asian	4.7	2.1	1.2	1.8	1.8	2.6	3.6	11.1	8.4
Other	3.5	-	0.4	2.4	2.2	2.3	3.0	4.5	8.5
Indian	0.7	1.5	1.2	0.6	0.8	0.9	0.8	0.7	0.2
Percent coded	84.2	41.8	89.7	93.1	91.6	86.0	83.6	85.1	81.7

completion rate is a dismal 41.8 percent, no doubt the result of the enumeration-at-birth policy. The problems with the oldest group of recipients are likely the result of the inability to capture race data for persons receiving benefits in 1979, as described earlier.

The other big problem with the SSR is that it mixes the two age/ethnicity coding schemes and provides no dates so that they can be separated. The old White, Black, and Other codes issued before 1980 are thrown together with the newer White, Black, Hispanic, Asian or Pacific Islander, and American Indian or Alaskan Native codes obtained since that time.[3]

The newer scheme may not easily collapse into the older scheme. Presumably, people who consider themselves to be White or Black would choose these categories for either coding scheme, but even that assumption can be challenged as people's perceptions of their race/ethnicity change over time. SSA has occasionally published race/ethnic data in the past based on collapsing these two coding schemes. Typically, Hispanic, Asian, and Indian codes have been converted to "Other" in the older scheme. But little empirical work has been done to establish this connection, and even after making that leap of faith, the policymaker is still stuck with the old scheme and its lack of detail.

■ Making Better Use of Existing Codes: the Numident and the MEF

One of the purposes of writing this paper is to explore the possibility that sufficient codes might exist already in the SSA computer system to support better descriptive statistics for either of the two coding schemes. Since the two sources for original codes are the Numident and Master Earnings files, these are the logical places to begin the search.

Going back to the Numident

Because the race code on the SSR is not updated with new Numident entries, it was likely that additional SS-5's on the Numident contain codes where the SSR has none, and new codes where the SSR has old ones. Also, the dates for the race codes on the Numident permit us to separate the old codes from the new codes. The SSI sample was matched to the Numident to select the first old code and the first new code for each person.

We found that for the 65,890 sample recipients, there were 167,393 Numident entries or about 2.5 entries for each recipient. Women tend to have more entries than men because of surname changes due to marriage. Table 2 shows the result of that match.

Overall, nearly 92 percent of the 65,890 recipients had a new scheme or old scheme race/ethnicity code, an improvement over the 85 percent found on the SSR. In total, the percentage of cases with old and new codes was nearly identical—about 64 percent for each group had a legitimate code. By age group, however, the differences between old and new codes are considerable. For those with the new code, the percentage is low for the under-9 category because of enumeration at birth, peaks at 95 percent for the 9-to-17- year group, and declines steadily as recipients get older, reaching 39 percent in the 75-or-older group. There is nothing terribly surprising about this. Generally, you would expect younger people to have the newer race code.

For the old codes, only half of persons in the 18-to-29-year group have such a code, and, of course, none of those under age 18 has an old code, since they were born after the new code was implemented in 1980. The old code is strongest in the 40-to-49-year group and declines with age to 57 percent in the 75-or-over group.

The results from the Numident also gave us a first look at the racial distributions for each code. If the old code is used, Whites make up 63 percent of the recipient population, with Blacks at 30.9 percent and Other at 6.0 percent. If the newer code is used, Whites make up only 45.6 percent of the population, Blacks have 31.9 percent, Hispanics have 14.2 percent, Asians have 7.1 percent, and Native Americans are at 1.2 percent. These differences occur primarily because of the different age distributions of persons with new and old codes.

If statistics are to be published from the two coding schemes, it is important to obtain high completion percentages for each scheme. The 64-percent figures for

Table 2. Race Codes for SSI Recipients from the Numident, by Code Version and Age Group, November 1998

Race code	Total in sample	Age group: Under 9 years	9 to 17 years	18 to 29 years	30 to 39 years	40 to 49 years	50 to 64 years	65 to 74 years	75 or over
Total	**65,890**	**3,454**	**5,521**	**7,153**	**7,744**	**8,589**	**12,912**	**10,776**	**9,741**
Old scheme									
Total with code	42,003	1	-	3,907	6,328	7,709	10,930	7,548	5,580
Total percent	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
White	63.1	-	-	57.4	62.7	63.0	65.8	63.0	62.6
Black	30.9	100.0	-	36.7	32.4	32.8	29.8	29.4	27.2
Other	6.0	-	-	5.9	4.9	4.2	4.5	7.7	10.2
New scheme									
Total with code	41,848	1,988	5,231	5,597	5,817	5,854	7,568	5,967	3,826
Total percent	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
White	45.6	33.6	41.3	55.4	54.6	51.3	48.5	34.5	32.7
Black	31.9	43.1	44.8	32.7	32.6	34.9	30.1	22.3	20.6
Hispanic	14.2	20.1	11.4	8.7	9.1	8.9	14.4	23.8	23.3
Asian	7.1	1.8	1.3	2.5	2.5	3.6	5.5	18.1	22.5
Indian	1.2	1.5	1.1	0.8	1.2	1.3	1.5	1.4	0.9
Percent with either code	91.8	57.6	94.7	97.3	98.5	97.6	96.4	94.1	79.7
Percent with old code	63.7	0.0	0.0	54.6	81.7	89.8	84.6	70.0	57.3
Percent with new code	63.5	57.6	94.7	78.2	75.1	68.2	58.6	55.4	39.3

the two coding schemes were less than exciting results, but were at least a start in the search for more accurate and complete data. The challenge was to fill in some of the missing pieces. Of course, we realized the unlikelihood of finding new codes for many of the older recipients and the impossibility of getting old codes for the youngest recipients. But we hoped to increase our percentages for both young and old recipients, so that it might at least be possible to show some statistics for each group.

Problems with codes for young recipients—enumeration at birth

As mentioned earlier, the policy of enumeration at birth creates a problem in obtaining codes for younger recipients. Since the policy has been in effect since 1989, we created a separate analytical category for this age group so that we could separate out its impact (Table 2). As seen earlier, new race codes were obtained from the Numident for only 58 percent of those recipients age 9 or under. Percentages for older groups were much higher in the 9-to-17 age group and slightly higher in the 18-to-39 age group.

Is this lack of codes long-lasting or temporary; that is, do children receive race codes from additional applications for account numbers filed within a few years of birth, or does the enumeration-at-birth policy permanently impair attempts to gather data on race/ethnicity? To address this question, we looked for the presence of new race codes for individual sample birth cohorts for the last 13 years (Table 3).

Not surprisingly, children born before the policy was implemented in 1989 showed much higher rates for new

Table 3. Percentage of Found Race Codes for SSI Recipients Born Since1984, by Year of Birth and Year of Race Code

Year of birth	All recipients in sample	Year of race code													
		1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
1985	651	51.8	79.0	84.6	87.9	91.1	93.9	94.2	95.1	95.5	95.9	96.0	96.0	96.3	96.6
1986	606		55.1	81.7	88.0	90.1	94.2	95.2	95.5	95.5	95.9	96.0	96.0	96.2	96.2
1987	648			58.0	82.7	88.4	93.4	94.3	95.7	95.8	96.1	96.5	96.6	96.6	96.6
1988	596				58.4	81.5	88.4	90.8	92.1	92.8	93.5	93.8	94.1	94.3	94.3
1989	631					46.6	67.5	72.1	75.6	77.0	78.4	79.4	80.2	81.1	81.5
1990	559						39.9	56.9	60.5	63.9	66.5	68.2	70.1	70.7	71.4
1991	550							37.5	52.0	56.9	59.6	62.0	63.8	66.5	67.5
1992	491								31.6	44.4	49.5	51.9	53.6	54.8	55.8
1993	436									33.5	47.2	51.1	52.5	54.8	56.9
1994	393										35.4	47.6	50.4	52.4	54.2
1995	297											33.7	44.4	46.5	48.1
1996	245												32.2	37.1	39.2
1997	213													30.5	37.6
1998	152														44.1

race codes. The pre-1989 birth cohorts began with well over half of the cases having new codes in the first year, and over 90 percent with codes by age 5. The year 1989 appears to have been a year of transition to the new method of enumeration. After 1989, only about one-third of the cases had a code in the first year. By age 5, a little more than half of the children had picked up a code. It is quite possible that, even with the limitations imposed by the enumeration-at-birth policy, the majority of recipients would have a code by age 18. Moreover, the need by some children to show a Social Security card for working purposes, or name changes due to marriages might produce an upsurge in rates of coding in the late teen years.

Looking for better old race codes on the MEF

The second stop in our search for codes was to the Master Earnings file, which, as previously noted, contains additional race/ethnic information for some persons receiving benefits in or before 1979. Since the MEF promised to supply missing codes mostly for older persons, the search was limited to old codes. Because the MEF code does not have a date attached to it, some help was needed in identifying it as an old code. The code was considered to be old if the person had filed for SSI on or before 1980, the year the new coding scheme began, or had begun to work on or before 1980. As a result of the match to the MEF, we were able to capture a few more codes (Table 4).

Overall, the percentage with an old code rose slightly from 64 percent to 70 percent. More importantly, the oldest groups improved dramatically. The 75-or-over group improved from 39 percent to about 80 percent. All of the age groups over 40 showed significant improvement. Overall, the number of age-40-plus recipients with an old code increased from 78 percent to 85 percent. One of the reasons why there are still 15 percent of recipients without a code, even after obtaining all the old codes, is that some SSI aged recipients are noncitizens and do not obtain an account number until they enter the country and file for SSI. Noncitizens filing for SSI after 1980 would have only the new code on their Numident records even if they were very old.

It is also useful to note that the percentages of each race were not changed significantly by the addition of the codes from the MEF. The percentage of Whites (62.6) rose by one-half of a percent, and the percentage of Other recipients (6.6) fell by the same amount. The percentage of Blacks (30.9) stayed exactly the same.

■ Assigning Codes to SSI Recipients

The search of administrative records left us with several problems. The old code is very limited in detail and reasonably complete only for people over 30 years

Table 4. Race Codes for SSI Recipients from the Numident and Using Earnings File Codes, for Old Code Version and Age Group, November 1989

Race code	Total in sample	Age group: Under 9 years	9 to 17 years	18 to 29 years	30 to 39 years	40 to 49 years	50 to 64 years	65 to 74 years	75 or over
Total	**65,889**	**3,454**	**5,521**	**7,153**	**7,744**	**8,588**	**12,912**	**10,776**	**9,741**
Old scheme									
Total with codes	46,227	1	4	3,967	6,430	8,026	11,657	8,274	7,868
Total percent	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
White	62.6	0.0	25.0	56.5	62.8	62.7	65.4	62.1	61.4
Black	30.9	100.0	50.0	36.2	32.0	32.6	29.6	29.4	29.0
Other	6.6	0.0	25.0	7.3	5.2	4.7	4.9	8.5	9.6
Percent with old code	70.2	0.0	0.1	55.5	83.0	93.5	90.3	76.8	80.8

of age. The new code carried much better detail, but is deficient in all groups except those recipients 9 to 17 years old. If anything is to be made of either code, some of the gaps must be filled by creating codes where none currently exists. Assigning race codes is a controversial process for several reasons. First, the race/ethnic classifications are not precise. Second, the applicants for account numbers could ignore the question, and may not care to have a response created for them. We look at two methods. Each of the methods were fairly easy to describe and implement, and each involved assignment of codes based on direct racial data gathered from the administrative records of the recipient or his or her family. The two methods were:

- Assigning codes to recipients based on the codes of their parents, and
- Combining the old and new codes into a single code

Assigning codes to recipients based on the race of their parents

Assigning codes to recipients based on their parents' codes fills the need to get better codes for young SSI recipients and is based on race information specific to that family. Because information about parents who live in the household with SSI child recipients is often required for program purposes, it is possible to identify the account numbers of the parents from SSI administrative records.[4] Those account numbers can then be matched to the Numident file to secure race codes.[5] We divided the study sample into four groups:

1. Those recipients who had a new code and whose mother or father had a new code (8,463).
2. Those recipients who did not have a new code, but whose mother or father had a new code (2,202).
3. Those recipients who had a new code, but whose mother and father did not (33,385).
4. Those recipients who did not have a new code, and whose mother and father also did not have a new code (21,840).

We then use the recipient and parental race information from the first group (8,463) to assign a code to recipients in the second group (2,202). The other groups

were not affected, because the third group already had new codes, and the codes for the fourth group could not be improved because there was no information about the parents.

Table 5 shows the combination of parental codes for the first group. Each cell in Table 5 shows three things for each parent race combination—the number of cases with each parental combination, the most frequent recipient code for that combination, and the percentage of cases with the most frequent outcome. For example, of the 8,463 cases in group 1, there were 1,203 recipients with two White parents. Of that subgroup, 97 percent of the recipients were listed as White. The White code was then assigned to anyone in the second group (2,202) where both parents were shown as White. The process was then repeated for all combinations of parents from group 1.

For most parental combinations of any size, the recipient outcome was very clear cut. Ninety-seven percent of recipients with two White parents identified themselves as White. Ninety-nine percent of recipients with two Black parents identified themselves as black. The results were similar for Asians (98 percent) and Hispanics (97 percent).

In many situations, there was only a mother listed on SSI records as living in the household. In this situation, 92 percent of recipients with a White mother identified themselves as White, 98 percent of those recipients with a Black mother identified themselves as Black,

Table 5. Parent New Race Code Combinations, by Most Frequent Race Code of Recipient, Number of Recipients with the Parent Race Combination, and Percent of Recipients with Outcome

	Father's race					
Mother's race	White	Black	Hispanic	Asian	Indian	Unknown
White	W N=1203 97%	B N=38 58%	H N=58 53%	W N=4 50%	I N=12 75%	W N=2022 92%
Black	B N=11 64%	B N=578 99%	B N=6 67%	B N=3 67%	B N=1 100%	B N=2825 98%
Hispanic	W N=37 54%	B N=9 67%	H N=299 97%	W N=1 100%	I N=2 100%	H N=594 87%
Asian	W N=9 56%	B N=5 60%	A N=1 100%	A N=51 98%	U N=0	A N=54 69%
Indian	W N=10 70%	B N=1 100%	H N=2 100%	U N=0	I N=25 100%	I N=40 78%
Unknown	W N=300 94%	B N=151 95%	H N=92 85%	A N=9 67%	I N=9 67%	U N=0

and 87 percent of those with Hispanic mothers identified themselves as Hispanic. Some situations were inconclusive, such as Black and White parental combinations, but the sizes of those categories were very small. Overall, selecting the category of the recipient based on the most frequent parental race combination would have yielded the recipient result 94 percent of the time.

The process was repeated for recipients who had old codes, and whose parents had old codes. Since SSA captures the parents' numbers only for SSI recipients under age 18, and since the program is only 25 years old, it would be unlikely that we would find parents for anyone much over the age of 40. Therefore, like the previous exercise, this exercise was also concentrated on people in the younger age groups. Unlike the previous exercise, we realized that since recipients under age 18 could not have an old code, assignment would have to be based on the experience of recipients age 18 or over. The same four groups were organized:

1. Those recipients who had an old code and whose mother or father had an old code (3,258).

2. Those recipients who did not have an old code, but whose mother or father had an old code (8,349).

3. Those recipients who had an old code, but whose mother and father did not (38,745).

4. Those recipients who did not have an old code, and whose mother and father also did not have an old code (15,538).

Again, the procedure was to use the recipient and parental information from the first group (3,258) to assign a code to recipients in the second group (8,349). A difference, however, was that the smaller group was now being used to assign codes to a much larger group.

As with the new codes, old parent code combinations produced fairly clear-cut results. In most cases, where both parents were of the same race or the mother was alone, the recipient was overwhelmingly of the same race. Again, racial combinations produced mixed results, but, overall, selecting the category of the recipient based on the most frequent parental race combination would have yielded the correct result 94 percent of the time. The results of assigning parent codes to recipients is contained in Table 7.

Table 6. Parent Old Race Code Combinations, by Most Frequent Race Code of Recipient, Number of Recipient with the Parent Race Combination, and Percent of Recipients with the Most Frequent Outcome

Mother's race	Father's race			
	White	Black	Other	Unknown
White	W N=980 97%	B N=9 44%	W N=20 70%	W N=809 92%
Black	W N=8 38%	B N=268 99%	U N=0	B N=930 97%
Other	W N=33 64%	O N=1 100%	O N=40 78%	O N=56 59%
Unknown	W N=80 95%	B N=20 95%	O N=4 100%	U N=0

The impact of new code assignment is readily apparent. The youngest group of recipients jumped from a 58-percent completion rate to a 94-percent completion rate. All groups under age 40 showed some improvement. Overall, the completion rate for the under-40 group rose from 78 percent to 87 percent.

As a result of the assignments of new codes, it appears that little was done to the racial/ethnic distribution. Percentages by racial group were very similar to those produced from using the Numident alone.

The results of assigning old codes were just as dramatic. There was a substantial increase in codes for the

Table 7. Race Codes for SSI Recipients from the Numident and Using Parent and Earnings File Codes, by Code Version and Age Group, November 1998

Race code	Total	Age group							
		Under 9 years	9 to 17 years	18 to 29 years	30 to 39 years	40 to 49 years	50 to 64 years	65 to 74 years	75 or over
Total	**65,890**	**3,454**	**5,521**	**7,153**	**7,744**	**8,589**	**12,912**	**10,776**	**9,741**
Old scheme									
Total with code	54,363	2,267	4,356	5,375	6,537	8,029	11,657	8,274	7,868
Total percent	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
White	61.5	53.7	53.0	59.3	62.9	62.7	65.4	62.1	61.4
Black	32.2	40.5	42.7	34.4	31.9	32.6	29.6	29.4	29.0
Other	6.3	5.7	4.3	6.3	5.2	4.7	4.9	8.5	9.6
New scheme									
Total with code	44,050	3,231	5,452	6,105	6,026	5,875	7,568	5,967	3,826
Total percent	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
White	45.8	40.2	41.6	54.7	54.6	51.3	48.5	34.5	32.7
Black	32.0	40.2	44.0	32.6	32.5	34.9	30.1	22.3	20.6
Hispanic	14.1	16.3	11.8	9.5	9.3	8.9	14.4	23.8	23.3
Asian	6.8	1.8	1.4	2.5	2.5	3.6	5.5	18.1	22.5
Indian	1.2	1.4	1.1	0.8	1.2	1.3	1.5	1.4	0.9
Percent with either code	98.6	97.0	99.5	99.0	99.2	98.9	98.8	98.1	97.8
Percent with old code	82.5	65.6	78.9	75.1	84.4	93.5	90.3	76.8	80.8
Percent with new code	66.9	93.5	98.8	85.3	77.8	68.4	58.6	55.4	39.3

three youngest age categories. Overall, the percentage of those with old codes rose from 44 percent for recipients under age 40 to 78 percent. This result was much improved from that obtained just from the recipient, but was still well short of the results for the new code.

As a result of the assignments of old codes, it again appears that little was done to the racial/ethnic distribution as compared with results from the Numident and the MEF.

The percentage of Blacks rose by slightly more than 1 percent, and that of Whites fell by the same amount.

Assigning codes by combining the old and new codes into a single code

Another approach is to combine the two codes into one. Although this approach has some pitfalls, it still relies on information about race supplied by the recipient's household.

To combine the two codes, it is necessary to somehow squeeze the new W, B, H, A, and I code values into the old W, B, O code values. To assist in this process, we identified four groups:

1. 33,455 recipients who had both old and new codes resulting from all previous steps in the analysis,
2. 10,595 recipients who had only a new code,
3. 20,908 recipients who had only an old code, and
4. 932 recipients who had neither code.

The process then was to assign codes to the second group using the code combinations from the first group. This first group represented over half of the original sample (65,890) and provided a substantial base for making these assignments. The intersection of the old and new codes is shown for persons in the first group in Table 8.

Unsurprisingly, the great majority (98 percent) of persons claiming to be White or Black in the new coding scheme had previously chosen the same category in the old. Most Hispanics, however, had previously listed themselves as White (75 percent), most Asians had listed themselves as Other (74 percent), and most Native Americans had previously chosen Other (67 percent). If these old codes are assigned based on the most frequent code category, the result would be correct 95 percent of the time. The biggest flaws were that a substantial minority of Hispanics (21 percent) chose "Other," and substantial minorities of Asians (21 percent) and Native Americans (26 percent) chose "White."

These old codes were then assigned to the recipients who had only a new code and added to the old code from previous steps. The results are shown in Table 9.

Combining codes produced a result that is not too different from that obtained entirely by using only old codes. The percentage of Whites remained the same, while the percentage of Blacks dropped by over two

Table 8. Converting New Codes to Old Codes

Old code	New code					
	Total	White	Black	Hispanic	Asian	Indian
Total	**33,455**	**16,068**	**12,277**	**4,116**	**518**	**476**
White	19,314	15,836	141	3,103	108	126
Black	12,323	98	12,001	166	29	29
Other	1,818	134	135	847	381	321

Table 9. Combining Race Codes from New and Old Coding Schemes

Old scheme race code	Total	Age group							
		Under 9 years	9 to 17 years	18 to 29 years	30 to 39 years	40 to 49 years	50 to 64 years	65 to 74 years	75 or over
Total	**65,890**	**3,454**	**5,521**	**7,153**	**7,744**	**8,589**	**12,912**	**10,776**	**9,741**
Not combined									
Total with codes	54,363	2,267	4,356	5,375	6,537	8,029	11,657	8,274	7,868
Total percent	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
White	61.5	53.7	53.0	59.3	62.9	62.7	65.4	62.1	61.4
Black	32.2	40.5	42.7	34.4	31.9	32.6	29.6	29.4	29.0
Other	6.3	5.7	4.3	6.3	5.2	4.7	4.9	8.5	9.6
Combined									
Total with codes	64,958	3,351	5,492	7,081	7,679	8,497	12,760	10,573	9,525
Total percent	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
White	61.1	54.9	52.0	63.2	63.9	62.1	65.0	60.2	59.7
Black	29.7	40.0	43.4	30.4	30.3	31.6	27.7	23.9	24.7
Other	9.1	5.0	4.6	6.4	5.8	6.3	7.3	15.8	15.6

percentage points, and the percentage of Others increased by the same amount. The increase in the percentage of Others can be mainly attributed to a large number of elderly noncitizens who come from Asian countries and who have only the new code on their Numident records. In fact, combining the codes probably produces a result that is more accurate because it includes this group.

■ SSI Distributions by Racial/Ethnic Status

The foregoing analysis provides two ways of classifying SSI recipients by race and ethnicity:

- ❒ Race/ethnic distributions for SSI recipients under age 40 based on the new code, and
- ❒ Racial distributions for all SSI recipients based on the old code.

Table 10 provides a set of distributions for SSI recipients under age 40 using the new SSA codes. Counts have been inflated to the universe of recipients.

Some interesting patterns emerge. Hispanics and Blacks are concentrated in the younger recipient groups, and Whites are a majority of recipients between the ages of 18 and 40. Black males and Hispanic females are particularly apparent among child recipients. Asians are concentrated in the over-18-to-40 age groups, and Native Americans are represented evenly and sparsely throughout the under-40 age range.

In looking at SSA geographic regions, Blacks are the largest racial/ethnic group in the Atlanta and Dallas regions.[6] Whites are the largest group in all other regions. Hispanics are well represented in the San Francisco and New York regions, Asians in the San Francisco region, and Native Americans in the Denver region.

Noncitizens show a much different racial/ethnic makeup than citizens. Nearly 38 percent of noncitizens are Hispanic, and 35 percent are Asians. This compares to only 10 percent of citizens who are Hispanic and 1 percent who are Asians.

Distribution by year of application shows that all members of racial/ethnic groups who remain on the rolls in 1998 have come onto the rolls at about the same rate throughout the program's history.

Table 11 includes the same statistics using the combined codes. It contains data for the entire SSI age range and is inflated to the universe.

One immediately striking pattern is the large number of Other races among recipients over age 65. Nearly 16 percent of all recipients in that age range are shown as Other, no doubt the result of the steady influx of noncitizens from Asia and Latin America. Among male recipients over age 75, the percentage is 23.9. Percentages of Black recipients gradually decrease with age.

When all recipients are counted, Whites have a majority of SSI recipents in all regions of the country. Blacks have substantial minorities in Atlanta (43.5 percent), Chicago (35.6 percent), Philadelphia (34.3 percent), and Dallas (33.0 percent). San Francisco had the largest percentage of Other races.

A substantial number of SSI recipients were noncitizens. They showed a much different racial makeup than citizens. Noncitizens were listed as Other (35.6 percent) and White (58.9 percent), while citizens were less likely to be Other (6.0 percent) and equally likely to be White (61.6 percent). Noncitizens were much less likely to be Black (5.5 percent) than were citizens (32.4 percent).

■ Conclusions About the Future of Race Coding for SSI Recipients

SSA would like to provide comprehensive data on racial/ethnic status for people who use its programs. In this paper, we have discussed several barriers to producing better statistics on race/ethnicity for SSI recipients.

- ❒ First, SSA needs to improve the methods it uses for capturing and sharing these codes among the various administrative files.

Table 10. SSI Recipients Under Age 40 (Using New Codes), by Race, Sex, Age Group, and Region, November 1998

Age, sex, region	Total	With codes	Total percent	White	Black	Hispanic	Asian	Native American
All	**2,387,100**	**2,081,300**	**100.0**	**49.0**	**36.8**	**11.1**	**2.1**	**1.1**
Under 9 years	345,400	323,100	100.0	40.2	40.2	16.3	1.8	1.4
9 to 17 years	552,100	545,200	100.0	41.6	44.0	11.8	1.4	1.1
18 to 29 years	715,300	610,500	100.0	54.7	32.6	9.5	2.5	0.8
30 to 39 years	774,300	602,500	100.0	54.6	32.5	9.3	2.5	1.2
Male	-							
Under 9 years	214,900	201,600	100.0	40.1	40.6	16.0	1.8	1.5
9 to 17 years	355,800	351,500	100.0	41.9	43.9	11.6	1.4	1.2
18 to 29 years	387,900	326,800	100.0	54.3	32.2	10.3	2.4	0.9
30 to 39 years	377,500	279,000	100.0	53.5	32.4	10.4	2.4	1.3
Female	-							
Under 9 years	130,500	121,500	100.0	40.4	39.7	16.8	1.8	1.3
9 to 17 years	196,300	193,700	100.0	41.2	44.2	12.2	1.4	1.0
18 to 29 years	327,400	283,700	100.0	55.1	33.0	8.6	2.6	0.7
30 to 39 years	396,800	323,500	100.0	55.5	32.6	8.3	2.5	1.0
Region	-							
Boston	103,000	85,300	100.0	63.4	14.4	19.1	2.7	0.4
New York	243,600	206,200	100.0	39.0	33.7	24.7	1.9	0.7
Philadelphia	234,800	200,600	100.0	55.1	38.5	5.0	1.2	0.1
Atlanta	558,000	499,300	100.0	46.2	49.0	3.9	0.5	0.4
Chicago	451,000	395,200	100.0	51.0	42.5	4.1	1.4	1.0
Dallas	289,800	258,500	100.0	39.1	41.2	17.7	0.8	1.2
Kansas City	96,200	81,500	100.0	68.0	28.0	2.6	0.7	0.7
Denver	51,700	45,300	100.0	70.0	5.7	15.0	1.3	7.9
San Francisco	293,400	252,600	100.0	43.0	22.6	23.7	8.5	2.3
Seattle	65,700	56,900	100.0	79.4	7.9	6.2	3.5	3.0
Citizenship status:								
Citizen	50,600	42,400	100.0	16.5	10.6	37.5	35.4	-
Noncitizen	2,336,600	2,039,000	100.0	49.7	37.3	10.6	1.4	1.1
Year of application:	-							
1974-1979	135,100	82,600	100.0	46.2	39.0	12.7	0.7	1.3
1980-1984	190,400	149,500	100.0	55.0	32.8	9.2	2.1	0.9
1985-1989	349,800	298,500	100.0	52.7	33.4	10.2	2.4	1.3
1990-1994	993,800	902,600	100.0	46.6	39.4	11.0	2.1	1.0
1995-1998	718,100	648,200	100.0	49.7	35.3	11.9	2.1	1.0

Table 11. SSI Recipients, by Race (Using Old Code), Sex, Age Group, Citizenship Status, Region, and Year of Application, November 1998

Characteristic	Total	With codes	Total percent	White	Black	Other
All	6,589,000	6,495,800	100.0	61.1	29.7	9.1
Under 9 years	345,400	335,100	100.0	54.9	40.0	5.0
9 to 17 years	552,100	549,200	100.0	52.0	43.4	4.6
18 to 29 years	715,300	708,100	100.0	63.2	30.4	6.4
30 to 39 years	774,400	767,900	100.0	63.9	30.3	5.8
40 to 49 years	858,900	849,700	100.0	62.1	31.6	6.3
50 to 64 years	1,291,200	1,276,000	100.0	65.0	27.7	7.3
65 to 74 years	1,077,600	1,057,300	100.0	60.2	23.9	15.8
75 or over	974,100	952,500	100.0	59.7	24.7	15.6
Male						
Under 9 years	214,900	208,800	100.0	54.3	40.2	5.5
9 to 17 years	355,800	353,800	100.0	51.8	43.5	4.7
18 to 29 years	388,000	383,800	100.0	63.2	30.3	6.5
30 to 39 years	377,600	373,900	100.0	65.6	28.7	5.7
40 to 49 years	362,500	358,500	100.0	62.5	31.2	6.2
50 to 64 years	461,000	454,300	100.0	64.3	27.8	7.9
65 to 74 years	341,100	333,800	100.0	62.4	19.8	17.9
75 or over	220,000	214,000	100.0	57.0	19.2	23.9
Female						
Under 9 years	130,500	126,300	100.0	56.0	39.7	4.4
9 to 17 years	196,300	195,400	100.0	52.4	43.3	4.3
18 to 29 years	327,300	324,300	100.0	63.2	30.5	6.3
30 to 39 years	396,800	394,000	100.0	62.4	31.8	5.8
40 to 49 years	496,400	491,200	100.0	61.7	31.9	6.4
50 to 64 years	830,200	821,700	100.0	65.4	27.7	6.9
65 to 74 years	736,500	723,500	100.0	59.3	25.9	14.9
75 or over	754,100	738,500	100.0	60.5	26.4	13.2
Region						
Boston	293,600	287,100	100.0	78.2	12.5	9.3
New York	744,300	722,800	100.0	60.0	28.1	11.8
Philadelphia	601,700	595,400	100.0	61.3	34.4	4.3
Atlanta	1,511,900	1,499,700	100.0	54.6	43.5	2.0
Chicago	977,900	967,600	100.0	59.8	35.4	4.8
Dallas	808,400	802,000	100.0	60.9	33.6	5.5
Kansas City	210,500	208,200	100.0	74.9	21.7	3.4
Denver	121,000	119,600	100.0	80.9	4.7	14.5
San Francisco	1,150,000	1,127,300	100.0	59.9	14.6	25.5
Seattle	169,700	166,100	100.0	80.6	5.2	14.2
Citizenship status:						
Noncitizen	660,000	625,300	100.0	58.9	5.5	35.6
Citizen	5,929,000	5,870,500	100.0	61.4	32.3	6.3
Year of application:						
1974-1979	847,600	836,100	100.0	62.2	32.7	5.1
1980-1984	563,200	556,700	100.0	60.8	30.6	8.6
1985-1989	1,118,900	1,108,400	100.0	60.7	28.4	10.9
1990-1994	2,448,900	2,416,100	100.0	60.1	30.0	9.9
1995-1998	1,610,400	1,578,500	100.0	62.5	28.4	9.1

- Second, the policy of enumeration at birth deprives the agency of race/ethnic data for a period of time. As discussed, this may be a temporary problem, and assignment of the parents' codes may be a useful way to make up for this lack of data for young SSI recipients.

- Finally, and most importantly, changes to the coding scheme itself are the biggest barriers to producing consistent race/ethnic data. Administrative data on race currently utilize two different coding schemes. This paper describes how the two codes can be collapsed into one.

The techniques described in this paper can be used to provide better data on race/ethnicity for SSI recipients. Where race data are needed for the entire caseload, it is possible to find and create enough codes to support tables based on the old coding scheme—Black, White, and Other. If the area of analytic interest is confined to younger recipients, it is possible to find and create enough codes to support tables based on the new coding scheme—Black, White, Hispanic, Asian, and Native American.

Over the years, there will be fewer and fewer old codes to support the process of combining old and new codes. Therefore, the errors introduced by both techniques will increase, and, at some point, it will make sense to discontinue any distribution employing the old codes. For the new code, the future promises an increase in the age range covered, but that will be decades away.

In October 1997, the Office of Management and Budget announced that revisions would be made to the standards for classification of Federal data on race and ethnicity. The new race code calls for 5 categories—American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, and White. There will be a separate question for ethnicity with two categories—Hispanic or Latino, and Not Hispanic or Latino. There will also be an opportunity for people to choose more than one category. The census for year 2000 will include these changes in coding. Other Federal agencies have until January 2003 to comply with the new guidelines.

Until that time, SSA should be able to cope with its statistical problems by improving its internal processing and applying some fairly conservative techniques for inferring race or ethnicity.

■ Footnotes

[1] SSA has always screened these subsequent applications to ensure that multiple account number issuance was kept to a minimum.

[2] There are very few instances where there is no Numident record, but a race code would not be available if the SS-5 applicant did not answer the question about race/ethnicity.

[3] The actual race code designations on the SSR, MEF, and Numident vary from alpha to numeric. The codes are referred to as alpha throughout the report for the sake of consistency.

[4] The main purpose for this information is to permit income to be "deemed" from the parents to the child. The result is a lower SSI payment to the child where the parents have income. Since some children are in institutions and not households, parent data, including account number may not exist.

[5] Beginning in 1998, the account numbers of parents have been collected by SSA using the SS-5 form. Although the data are not currently included on the Numident, it may be possible to use this source in the future, rather than retouching to the SSI administrative records.

[6] The Boston region includes Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont. The New York region includes New Jersey and New York. The Philadelphia region includes Delaware, the District of Columbia, Maryland, Pennsylvania, Virginia, and West Virginia. The Atlanta region includes Alabama, Florida, Georgia, Kentucky, Mississippi, North

Carolina, and Tennessee. The Chicago region includes Illinois, Indiana, Michigan, Minnesota, Ohio, and Wisconsin. The Dallas region includes Arkansas, Louisiana, New Mexico, Oklahoma, and Texas. The Denver region includes Colorado, Montana, North Dakota, South Dakota, Utah, and Wyoming. The San Francisco region includes Arizona, California, Hawaii, Nevada, and the Northern Mariana Islands. The Seattle region includes Alaska, Idaho, Oregon, and Washington.