MEASURING RACE: AN EMPIRICAL INVESTIGATION

Hee-Choon Shin, NORC
55 E. Monroe St., Suite 4800, Chicago, IL 60603

Key Words: Race, Measurement Errors

I. Introduction

In studying individual behavior or a social phenomenon of a multiracial society like U.S.A., race of an individual or racial composition of a community is one of the most important explanatory factors or independent variables along with other demographics such as sex and age (Smith, 1995; 1997). In most branches of social sciences (e.g., economics, sociology, demography, political science, and education) and in some biological sciences (e.g., population genetics, epidemiology, eugenics and medicine), race is a very significant explanatory factor or variable.

Population counts or estimation by racial group is an essential component of the decennial U.S. census. Indeed, in U.S.A., this population counts by racial group are the *de facto* criteria for power and resource allocation among racial groups. Each member of this society is influenced by this statistics. Heated controversy surrounding Census methods is an indication of the effects of this count on the individual welfare.

In most large-scale social surveys, race is a sampling variable to minimize the sampling variance on which stratification is based. A relative homogeneity of a certain characteristic is assumed within a racial group. Frequently, over-sampling in certain racial minority groups is arranged to get a decent sample size of minority members.

II. Definition of Race

The concept of race has been controversial and will be so for a long time. The controversial nature of the concept was well described in *Suicide*, a hundred-year-old classic written by Durkheim. Even Durkheim gave us a stern warning on the use of race in scientific research: "the sociologist must be very careful in searching for the influence of races on any social phenomenon." As Durkheim implicated, race might be vaguely understood to mean an aggregate of individuals with clearly common traits. Implying the impossibility of a clear definition of race, Durkheim used the following Morselli's classification for his study of suicide in Europe: 1) the Germanic type; 2) the Celto-roman type; 3) the Slav type; the Ural-Altaic type. This race is not equivalent to ethnicity nor to nationality. The fundamental question is this: Is race determined by civilization or by blood? According to the above Morselli's classification, the answer should be *neither,* or *both.*

The 5-category race is used in the U.S. census, i.e., white, black (or African-American), Asian, American-Indian, and other. Some of major national surveys (e.g., General Social Survey; National Longitudinal Survey of Youth) utilize a 3-category race: white, black, and other. We see that, in U.S.A. context, the race is mainly defined by blood rather than by civilization. The skin color is determined by blood. Therefore, in U.S.A. context, race might be defined as an aggregate of individuals with clearly common skin color. However, we should note that the 5-category Census definition reflects other (other than skin color) criteria as well as the skin color.

III. Statements of the Problem and Data

Once we decide to use skin color as a measurement, race might be measured as a continuous variable by measuring the melanin content. In practice, race is measured as a categorical or discrete variable. Since the categorization is being done by a human being (respondent or interviewer), it is quite a subjective phenomenon. The main objective of this paper is to investigate the degree of uncertainty in measuring race. We will also examine the effects of different categorization on one key dependent variable, i.e., income. Data are from the first 14 panels (1979-1993) of the National Longitudinal Survey of Youth[1]. In this survey, race has been independently measured by an interviewer annually. We are using the race measured by observation rather than self-identified race (Smith, 1997). Since a minority is, if she or he is ever, discriminated by others (not herself or himself), the race by observation is preferable. Ideally, the racial category of an individual should be a constant across all the different survey years. Yet the existence of ambiguous or mixed racial groups is a reality. Therefore, we expect to see a certain degree of uncertainty in measuring race.

IV. Findings

Table 1 shows some transition matrices of race from the base year (1979) to a survey year. Assuming no measurement errors in measuring race, all the diagonal entries should be one and all the off-diagonal entries be zero. To the contrary, all three matrices in Table 1 show less than one diagonal elements and non-zero off-diagonal elements. Among those youths who had been identified as

[1]NLSY was not fielded in 1987.

blacks in 1979, as shown in the top panel in Table 1, 1.75% were classified as white in 1980. Meanwhile, 0.30% of whites were identified as black in 1980. In particular, we see a higher degree of measurement errors between whites and others. Among those youths who had been identified as other in 1979, 50.41% were classified as white in 1980. The same pattern is observed in the other two transition matrices. The presence of off-diagonal elements in the transition matrices indicates the degree of uncertainty in measuring race. Table 2 shows the proportions (%) of off-diagonal elements in some transition matrices. As we see in Table 2, race of about 2.5% - 3.0% of youths have been differently identified from 1979 to a survey year.

Table 3 shows the relative magnitude of racial identification of each race for those who have provided more than two valid responses. The distribution of each column is based on the total sample. Therefore, each entry shows the degree of column characteristic. For example, 1.5 of the youth were identified as 70%-79% white, implying a certain degree of mix. 77.2% of the youth were consistently identified as white in every survey, while 12.7% of the youth were identified as nonwhite (black or other). 12.3% of black was consistently identified as black. Only 0.2% were consistently identified as other.

Based on Table 3, a new classification was attempted. Table 4 shows a 7-category classification. 10.3% of the youth were classified as a member of mixed race. 7.0% were identified as white and other, 2.4% as white and black, 0.2% as white and other. 0.6% were identified as a mixture of three different races. Table 4 shows the degree of uncertainty in measuring race. For 10.3% of youths, the membership of a specific racial category is quite arbitrary and subjective.

Finally, Table 5 shows average family incomes for different racial identification. For example, the average family income of blacks is $28,337.94 if the race of 1979 is used. If the race of 1993 is used, the average family income is $27,799.12. When the combined race is used, the average family is $26,793.97. For example, the income difference between whites and blacks might be larger than we observed. The same logic can be carried to other research areas such as biology and medicine. Here is a question to think over: *Which race should we use?*

Table 1. Some Transition Matrices of Racial Identification from the Base Year (1979) to a Follow-Up Year.

Racial Identification in 1979 (n)	Racial Identification in a Follow-up Year		
	White	Black	Other
	1980		
White (8,253)	.9840	.0030	.0130
Black (2,988)	.0175	.9766	.0059
Other (655)	.5041	.0294	.4665
	1988		
White (7,065)	.9792	.0031	.0176
Black (2,693)	.0256	.9711	.0033
Other (551)	.5719	.0289	.3993
	1993		
White (5,720)	.9792	.0029	.0179
Black (2,687)	.0187	.9736	.0078
Other (528)	.6135	.0270	.3595

Note: The sampling weight of the base year (1979) is used for the the analyses.

Table 2. Proportions (%) of Off-Diagonal Elements in Some Transitional Matrices.

Transition	Percent	n
79-80	3.1	11,896
79-81	2.8	11,562
79-82	2.8	11,628
79-83	2.5	11,710
79-84	2.3	11,745
79-85	2.6	10,583
79-86	2.5	10,352
79-88	3.0	9,956
79-89	2.5	10,068
79-90	2.7	10,113
79-91	2.6	8,564
79-92	2.7	8,552
79-93	2.7	8,635

Note: The sampling weight of the base year (1979) is used for the analyses.

Table 3. Distribution of Relative Racial Identification Given the Number of Valid Identifications for Respondents with more than two Identifications (n=12,567).

Percent	Intensity		
	White	Black	Other
0	12.7	84.5	91.9
1-9	1.2	1.2	1.8
10-19	0.4	0.3	1.2
20-29	0.5	0.0	1.6
30-31	0.5	0.0	0.9
40-49	0.4	0.0	0.6
50-59	0.9	0.1	0.8
60-69	0.9	0.0	0.4
70-79	1.5	0.1	0.3
80-89	1.2	0.2	0.2
90-99	2.6	1.2	0.1
100	77.2	12.3	0.2

Note: The sampling weight of the base year (1979) is used for the analyses.

Table 4. A Classification of Racial Identification Using All the 14 Racial Identifications for Respondents with more than two Identifications (n=12,567).

Racial Identification	%
Consistently White	77.2
Consistently Black	12.3
Consistently Other	0.2
Identified as White and Black	***2.4***
Identified as White and Other	***7.0***
Identified as Black and Other	***0.2***
Identified as White, Black, and Other	***0.6***

Note: The sampling weight of the base year (1979) is used for the analyses.

Table 5. Family Income by Race, 1992

Racial Iden -tification	Race of 79	Race of 93	Combined: 79-93
White (W)	45,706.56	45,628.57	46,190.33
Black (B)	28,337.94	27,799.12	26,793.97
Other (O)	37,624.92	36,962.55	---
W, B			41,360.35
W, O			37,486.01
B, O			36,071.58
W, B, O			34,940.73

Note: The sampling weight of the base year (1979) is used for the analyses.

References

Durkheim, Emile. (Tr. John A. Spaulding and George Simpson). 1897/1951. *Suicide: A Study in Sociology*. New York: Free Press.

Smith, Tom W. 1995. *A Review of Ethno-Racial Measures on the General Social Survey. GSS Methodological Report.* No. 85. Chicago: NORC.

Smith, Tom W. 1997. *Measuring Race by Observation and Self-Identification. GSS Methodological Report.* No. 89. Chicago: NORC.