AMERICAN SOCIOLOGICAL REVIEW

OFFICIAL JOURNAL OF THE AMERICAN SOCIOLOGICAL ASSOCIATION

ONLINE SUPPLEMENT
to article in

AMERICAN SOCIOLOGICAL REVIEW, 2006, VOL. 71 (FEBRUARY:16–41)

Widowhood and Race

Felix Elwert
Harvard University

Nicholas A. Christakis
Harvard University

SECTION 1

CENSUS VALIDATION

This paper uses a new longitudinal data set derived from national Medicare databases. Here, we ascertain the representativeness of our sample against known population distributions from the 5 percent Public Use Micro Sample (PUMS) of the 1990 Census.[1]

Supplement Table S1 shows summary statistics for all variables common to the PUMS and our data separately for the four couple-race groups included in this analysis. The population for all statistics encompasses married couples in which both spouses are between 67 and 98 years old (on January 1, 1993 in our data; on January 1, 1990 in the PUMS), are of non-Hispanic black or non-Hispanic white race, and lived together in one of the 50 states. We note that our sample is substantially larger than the Census PUMS (410,272 versus 273,306 couples).

Overall agreement between our sample and the Census can be described as good to excellent. In general, males in our sample are older than males in the Census by 1.2 years, and females by 1.1 years on average. The absolute average age difference between spouses in our sample is smaller by .1 years, and age order (wife older) is virtually identical. Despite somewhat different definitions of poverty (state buy-in coverage for dually eligible Medicare and Medicaid recipients as proxy for the federal poverty level versus the actual federal poverty level) the percentages of respondents labeled poor in both sources show close agreement (4.6 versus 5.6 percent). The geographic distribution of couples across Census regions differs by at most 3.1 percentage points per region.

---

[1] The authors thank Andrew Clarkwest for computing the Census tabulations.

Table S1. Comparison Between Medicare-Based Sample (1993) in This Analysis and Census PUMS (1990) by Couple-Race

	All		Couple-Race: Both White		Couple-Race: Both Black	
Variable	Medicare	Census	Medicare	Census	Medicare	Census
N in sample	410,272	273,306	388,794	256,319	17,064	12,183
% in sample	100	100	94.8	93.8	4.2	4.5
% in population (weighted)[a]	100	100	94.8	94.0	4.2	4.8
Age, Husband	76.6	75.4	76.6	75.3	76.6	75.6
Age, Wife	74.2	73.1	74.2	73.1	74.0	73.0
Age Difference[b]	3.3	3.4	3.2	3.4	4.0	4.1
% Wives older than Husband	21.0	19.5	20.8	19.5	24.6	21.7
% Poor[c]	4.6	5.6	3.8	4.9	21.7	19.2
Northeast[d]	18.2	21.3	18.5	21.8	11.3	15.5
Midwest	28.4	25.8	28.9	26.5	18.4	19.3
South	35.2	34.0	33.9	33.2	62.7	57.3
West	18.2	18.9	18.7	18.6	7.6	7.9

Table S2. Continued.

	Couple-Race: Black Husband/ White Wife		Couple-Race: White Husband/ Black Wife	
Variable	Medicare	Census	Medicare	Census
N in sample	2,359	92	2,055	56
% in sample	.57	.03	.50	.02
% in population (weighted)[a]	.10	.03	.09	.02
Age, Husband	76.6	75.4	76.3	76.3
Age, Wife	73.6	73.0	73.2	72.5
Age Difference[b]	4.0	4.2	4.2	4.7
% Wives older than Husband	19.8	18.3	21.4	15.0
% Poor[c]	11.7	9.5	11.6	15.4
Northeast[d]	16.7	23.0	18.6	22.2
Midwest	19.1	21.9	18.8	22.9
South	51.1	26.3	45.5	28.1
West	13.1	28.9	17.0	26.8

Note: Medicare-based sample, authors' calculations. Census calculations by Andrew Clarkwest. This table includes all individual-level and couple-level. N = number of cases.

[a] Sample weighted to represent population proportions. Medicare-based sample weighted to account for oversampling of racially intermarried couples. Census calculations weigh married couples by average of each spouse's person-weight.

[b] Mean absolute age difference between spouses.

[c] Medicare-based sample uses dual eligibility for Medicare and Medicaid as proxy for federal poverty level. Census calculations use 1990 federal poverty level.

[d] This table combines nine geographic Census divisions into four Census regions.

The percentage distributions of endogamously married couples in the population according to each source agree closely with each other (94.8 versus 94.0 percent endogamously married white couples; 4.2 versus 4.8 percent endogamously married black couples).[2] The agreement between our sample and the Census for spouses' ages is even better among endogamously married black couples than among endogamously married white couples, but slightly worse regarding geographic distribution.

---

[2] The race distributions in the two samples do not exactly equal the inferred distributions in the population because of weighting. Our sample uses a stratified probability sample of the elderly married couples identified in the Medicare source files. The reported Census figures weigh each couple by the average of each spouse's person-weight to produce population distributions.

It is more difficult, if not impossible, to assess the representativeness of our sample of intermarried couples from the Census (or any other currently available data set) because the PUMS data contains only a very small number of intermarried couples from the population of interest (n=148). By contrast, our data include 4,414 intermarried couples. Sampling variability in the Census statistics on this group therefore is a real concern. Regardless, our data closely match Census statistics on the age distribution of racially intermarried spouses, and the closeness of the age-match is comparable to the match among endogamously married couples. Similarly, the fractions of intermarried couples that are poor differ only by 2 to 4 percentage points between our sample and the Census. However, the geographic distributions of intermarried couples in the two data sets vary substantially. In relative terms, our data are about twice as efficient in locating intermarried elderly couples in the South, but only about half as good in finding intermarried couples in the West.

Finally, we note that even though the estimated population prevalence of intermarried couples matches to a precision of better than 0.1 percentage points in both data sets, our data nonetheless detect three to four times as many intermarried couples among the elderly as a percentage of the population as does the Census PUMS. There are several possible reasons for this discrepancy. First, and most likely, the Census PUMS might simply be too small to produce accurate estimates for this extremely small population. Second, our data may over-ascertain or the Census may under-ascertain the presence of racially intermarried couples.

Third, the two data sets may categorize borderline cases involving bi- or multi-racial individuals differently. Recall that the Census race variable asks for current racial identity, whereas our race variable was populated from applications for social security cards, which may have been filed decades ago. Racial self-descriptions, particularly of multi-racial individuals, may well have changed as a function of time, but also, crucially, as a function of marriage. In support of this latter possibility, Lieberson and Waters (1993) have found that some multiethnic whites change their self-described ethnic identity over time and also adjust their own ethnic identity after marriage to match that of their spouse.[3] If so, both our sample and the Census may report "accurate" population proportions of elderly, racially intermarried couples, but arrive at different numbers on account of recovering racial identities at different points in time and differing on the classification of bi- and multiracial individuals.

Based on these comparisons between our data and known population distributions, generalization of the findings from our data to the total population of elderly married black and white couples in the United States appears justified.

---

[3] What is more, individuals reporting on the racial identity of their spouse—as a primary respondent would in the Census—may tend to adjust their spouse's race to match their own. By contrast, social security card applications were likely filled out on an individual basis by each spouse separately and thus do not suffer from potential proxy bias in race reporting.

SECTION 2

POSITIVE PREDICTIVE VALUES OF THE BLACK AND WHITE RACE CATEGORIES IN THE VITAL STATUS FILE AND IN THE CENSUS

We explore a central dimension of the accuracy of the race variable used in this study by comparing the positive predictive values (PPV) of the black and white race categories in the Vital Status (VS) file with the corresponding PPVs in the Census. The PPV gives the probability that an individual identified as belonging to a particular race in the focal dataset is identified as belonging to the same race in a reference data set. The comparison with the Census is instructive because the Census is the primary source of recent research on interracial marriage as well as much other research on race.

We estimate the PPVs of the Census from the 2001 Census Quality Survey (CQS) (U.S. Dept. of Commerce, 2004). The CQS sampled about 55,000 households in a split panel design where each primary respondent was asked twice over two waves to report the race of all household members using in turn the race questions of the 1990 and the 2000 Census. The responses from the CQS were then matched to the same households' responses from the official 2000 Census. We performed de novo analyses where we computed the PPVs of the black and white categories of the Census race variable by comparing the answers of married couples across waves in the CQS, and between the CQS and the 2000 Census. We excluded imputed values and restrict the comparison to households where the same spouse was the primary respondent across waves (best case scenario). Results were weighted to adjust for stratification and clustering. Multiple race responses differing by at least one race were treated as different answers. Depending on whether we compared between consecutive responses to the same race question on the Census and the CQS, or between the different questions asked in the two waves of the CQS, and whether we treat "Hispanic" as a category distinct from black and white, the PPV of the black race code in the Census ranges from 95.6 percent to 98.8 percent, and the PPV for the white race code ranges from 95.6 percent to 97.5 percent. These values are comparable to the PPVs for the Vital Status file – the file that lies at the core of the race variable in the present research – reported by Arday et al. (2000) as 96.1 percent for blacks and 98.4 percent for whites.

Given that CQS and the 2000 Census were administered within months of each other, whereas Arday et al.'s study compared responses from the 1997 Medicare Current Beneficiary Survey to VS race records generated by different measurement protocols and dating back decades in most cases, the PPVs for the white and black categories in the VS file appear especially high, instilling considerable confidence in their accuracy. Furthermore, the race variable in the VS file records each spouse's racial self-classification (via individual social security applications), whereas the Census race variable relies on proxy reporting for one spouse of each married couple. The identification of endogamously and interracially married couples in this study would thus appear to be no more problematic than the identification of endogamously and interracially married couples in the Census.

REFERENCES

Arday Susan L., David R. Arday, Stephanie Monroe, and Jianyi Zhang. 2000. "HCFA's Racial and Ethnic Data: Current Accuracy and Recent Improvements." Health Care Financing Review 21(4):107-116.

Lieberson Stanley, and Mary C. Waters. 1993. "The Ethnic Responses of Whites: What Causes Their Instability, Simplification and Inconsistency?" Social Forces 72:421-450.

U.S. Department of Commerce. 2004. Census Quality Survey Public Use Data File, 2001 [computer file, ICPSR release]. Washington, DC: U.S. Dept. of Commerce, Bureau of the Census [producer]. Ann Arbor, MI: Inter-University Consortium for Political and Social Research [distributor].

(Table S2 on next pages.)

Table S2. Cox Models for Proband's Hazard of Death as a Function of Widowhood, Couple-Race, and Control Variables

Variable	Men	Women
Widowhood (main effect = both white)		
Widowed(t)	1.184**	1.162**
	(1.169 – 1.199)	(1.146 – 1.177)
Widowhood * Race interactions		
Widowed(t) * im_black_hus[a]	1.027	.877
	(.882 – 1.195)	(.751 – 1.025)
Widowed(t) * im_black_wif[b]	.797**	.938
	(.671 – .946)	(.799 – 1.102)
Widowed(t) * endog_black[c]	.852**	.864**
	(.808 – .898)	(.819 – .911)
Race (main effects; ref. cat. = both white)		
Black husband, white wife	.982	.951
	(.924 – 1.044)	(.868 – 1.042)
White husband, black wife	1.01	1.009
	(.946 – 1.079)	(.916 – 1.111)
Both black	.987	1.047*
	(.961 – 1.014)	(1.010 – 1.085)
Other Control Variables		
Age, husband	1.133**	1.055**
	(1.112 – 1.156)	(1.030 – 1.081)
Age, wife	1.030**	1.013
	(1.008 – 1.052)	(.987 – 1.039)
Age squared, husband	1.000**	1.000**
	(1.000 – 1.000)	(1.000 – 1.000)
Age squared, wife	1.000**	1.001**
	(1.000 – 1.000)	(1.000 – 1.001)
Wife older	1.058**	1.046**
	(1.043 – 1.073)	(1.028 – 1.064)
Poverty, couple	1.368**	1.471**
	(1.343 – 1.394)	(1.439 – 1.503)
Moderate Charlson, 1991 husband	1.444**	.963**
	(1.419 – 1.470)	(.940 – .988)
Severe Charlson, 1991 husband	1.831**	.993
	(1.800 – 1.864)	(.968 – 1.019)
Moderate Charlson, 1992 husband	1.491**	.992
	(1.465 – 1.517)	(.968 – 1.016)
Severe Charlson, 1992 husband	2.094**	.974*
	(2.063 – 2.127)	(.952 – .996)
Moderate Charlson, 1991 wife	.998	1.748**
	(.976 – 1.021)	(1.708 – 1.788)
Severe Charlson, 1991 wife	.983	2.234**
	(.958 – 1.009)	(2.180 – 2.289)
Moderate Charlson, 1992 wife	1.008	1.838**
	(.987 – 1.030)	(1.798 – 1.879)
Severe Charlson, 1992 wife	1.003	2.791**
	(.980 – 1.027)	(2.731 – 2.851)
Weeks in hospital, 1991 husband	1.020**	.997
	(1.017 – 1.023)	(.993 – 1.002)
Weeks in hospital, 1991 wife	1.001	1.020**
	(.997 – 1.005)	(1.017 – 1.024)
Weeks in hospital, 1992 husband	1.037**	.995**
	(1.034 – 1.039)	(.991 – .998)
Weeks in hospital, 1992 wife	.994**	1.030**
	(.991 – .998)	(1.027 – 1.033)
Middle Atlantic	1.048**	1.042**
	(1.023 – 1.075)	(1.010 – 1.075)
East North Central	1.043**	1.031

(Continued on next page.)

Table S2. *(continued)*

Variable	Men	Women
West North Central	.991	.949**
	(.964 – 1.019)	(.917 – .983)
South Atlantic	1.077**	1.049**
	(1.050 – 1.105)	(1.016 – 1.083)
East South Central	1.085**	1.045*
	(1.054 – 1.117)	(1.007 – 1.083)
West South Central	1.062**	1.054**
	(1.033 – 1.091)	(1.018 – 1.091)
Mountain	1.03	1.03
	(.999 – 1.062)	(.991 – 1.070)
Pacific	1.035*	1.043*
	(1.008 – 1.064)	(1.009 – 1.079)
Population density, county	1	1
	(1.000 – 1.000)	(1.000 – 1.000)
Violent crime, county	1.001**	1.003**
	(1.001 – 1.002)	(1.002 – 1.004)
MDs/1000, county	1.004	1.007**
	(1.000 – 1.008)	(1.002 – 1.012)
Hospital beds/1000, county	.997**	.996**
	(.995 – .998)	(.994 – .998)
Urbanization index, zip	1.000**	1.001**
	(1.000 – 1.000)	(1.000 – 1.001)
Black, %, zip	1	1
	(1.000 – 1.000)	(1.000 – 1.000)
Other race, %, zip	1.002**	1.001
	(1.001 – 1.003)	(1.000 – 1.002)
Aged 65+, %, zip	.994**	.995**
	(.993 – .995)	(.994 – .997)
Men widowed, %, zip	1.003	1.008**
	(.999 – 1.007)	(1.003 – 1.014)
Women widowed, %, zip	1.005**	1
	(1.003 – 1.007)	(.998 – 1.002)
Foreign born, %, zip	.999	1.001
	(.998 – 1.000)	(1.000 – 1.003)
Linguistically isolated, %, zip	.996**	.993**
	(.994 – .999)	(.990 – .997)
Median Home Value, $, zip	1.000**	1.000**
	(1.000 – 1.000)	(1.000 – 1.000)
Male unemployment, %, zip	1.005**	1.008**
	(1.002 – 1.007)	(1.005 – 1.011)
Log median income, $, zip.	.996	.976
	(.970 – 1.023)	(.943 – 1.009)
High school median, %, zip	.984	.985
	(.955 – 1.014)	(.949 – 1.023)
Some college median, % zip	.923**	.942**
	(.892 – .954)	(.903 – .983)
College + median, %, zip	.864**	.888**
	(.827 – .903)	(.840 – .938)
Log likelihood	–2577349.8	–1618839

Note: Authors' calculations. Hazard ratios and 95 percent confidence intervals (in parentheses) from separate Cox models for male and female probands. The results in this table form the basis of Figure 2.

[a] im_black_hus = intermarried, black husband and white wife.

[b] im_black_wif = intermarried, white husband and black wife.

[c] endog_black = endogamously married, both spouses black.

*$p < .05$, **$p < .01$ (two-tailed tests).