Data Quality and Adjusted Hispanic Mortality in the United States, 1989–1991

Karen Swallen, MPH, PhD; Abdelhani Guend, MS

Objectives: Hispanics appear to live longer compared to other ethnic groups in the United States. Our main objective was to determine whether data quality biases mortality statistics. We calculated the impact of misclassification of ethnicity on death certificates in order to create adjusted mortality estimates.

Methods: We used the National Mortality Follow-Back Survey of 1993 (NMFS) for our assessment of ethnicity misclassification. We then created misclassification estimates for 10-year age-sex groups, and used these to correct mortality estimates for 1989–1991.

Results: The overall predictive value positive (PV+) and sensitivity were 0.981 and 0.805, respectively, for men; and 0.994 and 0.902, respectively, for women. Age-specific adjustment for misclassification on death rates caused Hispanic male life expectancy to go from a 1.01 year advantage over White non-Hispanics, to a 1.83 year deficit, with a similar pattern found for females.

Conclusions: The apparent Hispanic advantage in life expectancy is influenced by misclassification of ethnicity. Misclassification of ethnicity on death certificates biases Hispanic mortality downward, thereby falsely inflating life expectancy. (*Ethn Dis*. 2003;13:126–133)

Key Words: Bias, Ethnic Groups, Hispanic Americans, Life Expectancy, Life Tables, Mortality, United States, Vital Statistics

From the Center for Demography and Ecology, University of Wisconsin, Madison, Wisconsin.

Address correspondence and reprint requests to Karen Swallen, PhD; Center for Demography and Ecology, Social Sciences; 1180 Observatory Drive; Madison WI 53705; 608-262-2923; 608-262-8400 (fax); kswallen@ssc.wisc.edu

Introduction

As of November 1, 1998, the US Census Bureau estimated that 11.4% of the US population was currently Hispanic.[1] In 1997, 9.7% of the US population was foreign-born.[2] When American residents are divided into 3 groups (native-born citizens, naturalized citizens, and non-citizens), the differences in ethnicity are even more apparent. While only 7.6% of native-born Americans are Hispanic, 27.2% of naturalized citizens, and 53.3% of non-citizen American residents, are Hispanic.[2]

When considering aging populations, nativity and ethnicity are important, yet under-utilized, social variables that may be related to topics as diverse as health, mortality, religious participation, informal care networks, etc. Immigrants, or persons born in countries other than the United States, make up 10% of the age 65+ population in the United States. In some states, including California, New York, and Hawaii, over 20% of the elderly population are foreign-born.[3] This is a sizable fraction of the population; in comparison, Blacks only constitute 8% of the elderly population in the United States. While these populations are an increasingly important component of the elderly population of the United States, very few data sources include sufficient samples of minority populations (other than Blacks) to study minority population aging.[4]

Another body of research reported a Hispanic-American mortality advantage.[5–10] The mortality pattern is not universally better for Hispanics; death rates among Hispanic children and young adults appear to be elevated, relative to non-Hispanic Whites.[10–11] In addition, certain causes of death (eg, homicide) are elevated, while others (eg, cardiovascular disease) decrease.[9,12,13]

While some researchers argue that "Hispanic" is a meaningless term and should be abandoned,[14–16] other researchers have argued that "Hispanic" should be even broader than self-identification, and should include up to third-generation non-self-identified Hispanics in a "Hispanic" category.[17] It seems reasonable to expect that research on health and disease in the United States will continue to utilize race and ethnicity as collected according to the Office of Management and Budget.[18–21] One editorial in a leading medical journal specifically called for more research on ethnic and racial differentials in health and mortality.[22]

The Hispanic advantage in mortality rates, compared to non-Hispanic Whites, has been referred to as an "epidemiological paradox."[23] The paradox is apparent for outcomes other than mortality; for example, Hispanics have been found to have unexpectedly good birth outcomes.[24] Many possible causal explanations of the "paradox" have been proposed, including cultural practices, nutrition, health behaviors, selective migration, and genetic advantages;[5,11,23] however, these hypotheses have not been adequately tested.

If the differential mortality described in the preceding sections is real, then the causal mechanisms should be examined carefully. However, before considering possible explanations for Hispanic longevity, it is necessary to consider the possibility that data quality distorts this finding For example, although many researchers explored causal explanations for the crossover in Black-White mortality, recent findings seem to indicate that age misreporting among Blacks has led to artificially low estimates of

"Many possible causal explanations of the 'paradox' have been proposed, including cultural practices, nutrition, health behaviors, selective migration, and genetic advantages;[5,11,23] however, these hypotheses have not been adequately tested."

old-age Black mortality.[25–26] Therefore, we must exercise caution when considering other anomalous mortality patterns. If data quality is acceptable, or suitable adjustments can be made, then future research may be able to consider causal mechanisms.

In general, mortality is calculated by dividing counts of deaths from death certificates by the census population figure. To calculate death rates for women aged 35–40, for example, we divide the number of deaths among women aged 35–40 by the population of women aged 35–40. The problem here is that the independent sources, the death certificates and the census, determine the demographic information separately. Therefore, if the two sources do not agree, numerator-denominator mismatches can occur. For example, if women who die between the ages of 35 and 40 were frequently assigned to an older age group, the death rate would be underestimated since too few deaths would be counted. In the case of the Black-White cross-over mentioned previously, it appears that age for Blacks was more likely to be inaccurately reported on both sources, compared to age for non-Blacks.[25–26] Census reports are collected on the household level, with one member of the household providing data on all members of the household. Death certificate information is collected after the death of the person of interest. Different states authorize different individuals to fill out death certificates; in general, information is completed by a physician, a medical examiner, a coroner, or a funeral director.[27]

The particular numerator-denominator mismatch we examined is the match between Hispanic ethnicity and Hispanic mortality rates. Although some reports find that death certificates and next-of-kin reports usually agree on ethnicity,[29–30] these findings are based on overall agreement, and ignore the fact that for some groups, misclassification may be more severe. For instance, White non-Hispanics are likely to be reported correctly, but a Black Hispanic may be highly likely to have one or both identities misclassified. In general, misclassification of race and ethnicity are fairly well documented problems.[19,31–34] Although several authors have suggested that using Spanish surname matching may be a way to increase the count of Hispanic deaths,[32,34–35] this type of matching may increase the error in ethnic classification, especially for women.[33]

As of 2000, all federally collected racial data have been coded using new categories, making multiple category responses possible.[21] Understanding how racial and ethnic classification schemes may vary between data sources is likely to become even more important in the near future. The census began using new classification schemes for race in 2000. Although the proxy reporting of ethnicity is a limitation of the NMFS, the census also depends on classifications made by one individual per household. Other data sources derive ethnicity from self-report prior to death (such as the National Longitudinal Mortality Study[36] or the NHIS-NDI[37]) and may be preferable for calculating true death rates for ethnic groups, without requiring the adjustment. However, these studies do not provide information to public users on the match of ethnicity between the 2 sources. Using the NMFS, we compare the reporting of ethnicity between 2 sources and then calculate adjusted death rates.

MATERIALS AND METHODS

The National Mortality Follow-Back Study of 1993 (NMFS) was used to assess the level of ethnic misclassification on death certificates. The NMFS is based on a national sample of death certificates of persons aged 15 years and older in 1993. A total of 22,957 death certificates were selected, with persons under age 35, persons over age 99, women, and Blacks, being over-sampled. Next-of-kin and hospital personnel (both groups referred to here as proxy respondents) were then contacted to obtain additional information (such as socioeconomic status or cigarette smoking), and to verify the information on the death certificates. For the purposes of this study, important variables included in the death certificate file include race, age, Hispanic origin, and place of birth. The proxy respondent file also included these variables.[28]

Race and ethnicity were determined separately by both sources. On the death certificate, Hispanic origin is obtained by a yes/no question: "Was decedent of Hispanic descent?" In the NMFS, Hispanic ethnicity is determined by a "yes" response to the question: "Was – – – of Spanish or Hispanic origin or descent?" For both the death certificate and the NMFS data, further coding of Hispanic ethnicity was available, ie, Puerto Rican, Cuban, Mexican, etc. More specific information is often a source of further misclassification,[22] but is not addressed here. Institutional review board approval of this research was obtained prior to any analysis. All data used in these analyses are available

	Hispanic on Death Certificate	Non-Hispanic on Death Certificate
Hispanic by Proxy	Correctly Classified A	Misclassified Hispanic B
Non-Hispanic by Proxy	Misclassified Non-Hispanic C	Correctly Classified D

Fig 1. Possible classifications from death certificates and by proxy report in the NMFS.

publicly and all individuals are anonymous.

The 1986 NMFS was previously used to evaluate the accuracy of racial and ethnic classification;[19] however, only 17 states provided information on Hispanic ethnicity of use to the 1986 NMFS.[29] In this report, we do not simply report the accuracy rates, but also use the misclassification rates calculated from the 1993 NMFS to adjust the national mortality statistics. NMFS[28] provides weights for the proxy respondent section of the survey; weights have been used for all calculations. The final weight we use is a product of the reciprocal of the probability of the selection factor into the survey, a non-response adjustment factor, and a post-stratification adjustment factor.

Figure 1 shows the possible states of classification used. Results include reports on predictive value positive [A/(A+C)], predictive value negative [D/(B+D)], sensitivity [A/(A+B)] and specificity [D/(C+D)]. Predictive value positive (PV+) is the correctly classified proportion among those classified as Hispanic on the death certificate. Sensitivity is the proportion correctly classified as Hispanic on the death certificate among the true Hispanics. Predictive value negative (PV−) and specificity are similar, but relate to non-Hispanics rather than Hispanics. Relative bias is the amount by which the percentage classified as Hispanic on the death certificate differed from the proxy-identified Hispanic group, as a percentage of the proxy-identified group. From Figure 1, this is calculated as [(A+C)/(A+B)]−1. (It can also be calculated as (sensitivity/PV+)−1.)

We calculated all our life tables using the US population distribution by age and sex, as estimated by the 1990 US Census, and the 3-year average of the deaths recorded for 1989–1991. We extracted the population data from the Public Use Microdata Sample. The counts of population are re-weighted to derive the complete count for each life table calculated. We re-coded the variables of interest to match the standards life table. Age was re-coded from 0 to 90 years and over, as 0 years, 1 to 4 years, and in 5-year intervals for ages 5 to 90 years. The open age group 90 and over is the highest age group provided by the census data. Although the deaths were tabulated to over 100 years, we had to set the age limit for the life tables at 90 years, because census data are not tabulated beyond this age. We tabulated the census data by race, detailed race, nativity, and Hispanic origin, according to the same standard age groups, and for both males and females. A further cross tabulation of race and nativity, by Hispanic origin, was performed in order to calculate the corresponding life tables. The states of Louisiana, New Hampshire, and Oklahoma did not provide Hispanic origin as a category on the death certificates; therefore, we excluded these 3 states from the census data for the calculation of life tables where "Hispanic origin" is involved. The death statistics included an unknown age group, as well as unknown "Hispanic origin." Although the unknown group was not large enough to have a significant impact on the final result, we distributed it proportionally to the weight of each group as follows: the deaths with unknown place of birth (unknown nativity) were distributed among the age group/sex according to the weight of the age group to the total within-sex category; the deaths with unknown age were also distributed among age groups according to the weight of each age group within the total deaths, separately for males and females. For consistency in matching the deaths to the statistics of the population, we defined US-born as deaths of persons born in the 50 states and DC. This definition excludes some populations of US territories for the benefit of a strict matching of deaths and population statistics, given that these territories are not coded strictly in the same way in the census and the death data.

The adjustment method is described below. Considering classification by ethnicity alone, the following formula adjusts for counts of death by ethnicity[39]:

$$EH = (PH * IH) + [(1 - PH) * INH]$$

where

- EH = estimate proportion of "true" Hispanics (or A + B from Figure 1)
- PH = proportion classified as Hispanic on death certificate ([A + C]/all deaths)
- IH = proportion self-identified as Hispanic among those classified as Hispanic on death certificates (A/[A+C] or PV+)
- INH = proportion self-identified as Hispanic among those classified as non-Hispanic by the death certificate (B/[B+D] or 1-PV−)

We adjusted the life tables using estimates of PV+ and PV− calculated from the data of the NMFS for groups aged 15 years and older, and using a published rate[19] for age groups less than 15 years. PV+ and PV− for the 4 age groups (0, 1–4, 5–9, and 10–15) are calculated in Table 2 as reported by Hahn.[19(p262)] The above adjustment has been used to adjust for cancer incidence,[40] but not to adjust for all-cause mortality among Hispanics. We used

PV+ and PV− calculated for 15-year age groups to smooth the adjustment.

In addition to life expectancy, an overall measure of mortality in a given population, we present some intermediary results, such as: the raw count of deaths by race-sex-ethnicity group, both before and after adjustment for misclassification; the magnitude of the effect of the adjustment on the death counts, shown by presenting data on the change in the deaths after adjustment, as a percentage of the unadjusted counts; and the age-specific standardized mortality ratio (SMR) for the total Hispanic population, both before and after adjustment. The SMR is calculated by dividing the age-sex-ethnicity-specific death rate by the standard mortality rate, in this case, the rate for non-Hispanics.

RESULTS

Results from the NMFS indicate that while Hispanic ethnicity is not perfectly classified, the errors are not as large as might have been predicted from results of some previous studies. For the entire sample, PV+ was 0.9872, while sensitivity was 0.848. Results varied by factors such as age of decedent, respondent's relationship to decedent, and race. The age pattern was unclear; PV+ increased with age, but sensitivity exhibited no linear pattern. There was more error in the classification of Hispanic ethnicity for Blacks and Asian/Pacific Islanders, than for Whites or "other race." In general, PV+ was higher than sensitivity, indicating that the relative bias is toward under-reporting Hispanic ethnicity, as expected. A high PV+ indicates a high probability that a person reported on the death certificate as Hispanic really was Hispanic. A relatively lower sensitivity indicates that persons who were reported as Hispanic on the survey were often not classified as Hispanic on the death certificate. For example, when the proxy reporter was an institutional staff person as opposed to a relative, every decedent reported as Hispanic on the death certificate was also reported as Hispanic by the staff person. However, death certificates captured only 67% of those the staff person reported as Hispanic. The relative bias in this case was −33%. Only in one instance was the relative bias positive, indicating that Hispanics were overidentified; this occurred when a neighbor provided the proxy report. Complete results are reported in Table 1.

Table 1. Sensitivity, predictive value positive, and relative bias for Hispanic ethnicity, NMFS, 1993

	Sensitivity	Predictive Value Positive	Relative Bias*
Age			
15–29	0.866	0.974	−11
30–44	0.874	0.975	−10
44–59	0.716	0.990	−28
60–74	0.850	0.988	−14
75–89	0.836	0.993	−16
90+	0.990	0.996	−1
Race			
White	0.864	0.993	−13
Black	0.540	0.811	−33
Am Ind/Eskimo/Aleut	0.528	1.0	−47
Asian/Pacific Islander	0.143	0.653	−78
Reporter			
Spouse	0.762	0.984	−23
Parent	0.890	0.997	−11
Step-parent	0.931	1.0	−7
Child	0.766	0.978	−22
Sibling	0.745	0.968	−23
In-law	0.983	0.990	−1
Other relative	0.966	0.996	−3
Friend	0.874	0.936	−7
Neighbor	0.967	0.961	1
Institutional staff	0.673	1.0	−33
Person	0.716	0.987	−27
Other			
Gender			
Male	0.805	0.981	−18
Female	0.902	0.994	−9

* Relative bias is calculated as [(Sensitivity/PV+) − 1] × 100.

Using the NMFS estimates of misclassification, period life tables were created for 1989–1991 US populations adjusting for misclassification by sex, age, and race. Table 2 shows the unadjusted and adjusted counts of deaths by gender and race-ethnicity. Table 2 also demonstrates that some race/ethnic groups were much more affected by misclassification than were others. The overall range of impact is from −0.3% (White non-Hispanic females) to +89% (Black Hispanic females). Leaving race aside, Hispanic male deaths were increased by 17% and Hispanic female deaths by 11%.

Table 3 shows how these adjusted deaths alter life expectancy at birth for the various groups. Life expectancy is a more detailed measure than the death counts in Table 2, since life expectancy takes the age pattern into account. As Table 3 shows, adjustment for misclassification was more important for some groups than for others. For example, White non-Hispanic male life expectancy at birth rises by 0.16 years while White Hispanic male life expectancy

Table 2. Unadjusted and adjusted counts of deaths for Hispanics and non-Hispanics in the United States. Change in count of death made by adjustment, as a percentage of the unadjusted count of deaths. Calculated using the 1989–1991 complete mortality statistics of the United States and the National Mortality Follow-Back Survey of 1993 for estimates of misclassification

	Males			Females		
	Unadj.	Adjust.	Change*	Unadj.	Adjust.	Change*
White non-Hispanic	876410	869058	−0.8%	845999	843174	−0.3%
White Hispanic	46910	54330	16%	31153	33978	9%
Black non-Hispanic	137670	136675	−0.7%	113976	113284	−0.6%
Black Hispanic	1173	2168	85%	777	1469	89%
Am Indian non-Hispanic	4247	4215	−0.8%	3029	2974	2%
Non-White non-Hispanic	154106	153003	−0.7%	125846	125089	−0.6%
Non-White Hispanic	1628	2731	68%	1055	1812	72%
Hispanic (regardless of race)	47897	56232	17%	32120	35729	11%

* Change is calculated as (adjusted count − unadjusted count)/unadjusted count.

falls by 2.5 years. Ignoring race, the life expectancy for Hispanic males falls by 2.68 years; life expectancy for Hispanic females falls by 1.62 years. While Table 1 indicates significant variation in misclassification depending on who reports the ethnicity on the survey, we cannot take advantage of this in our adjustment since we have no knowledge about who provided the report of ethnicity for the death certificates.

Figure 2 shows more detail for one group—Hispanic males. The SMRs demonstrate that although the overall age pattern of Hispanic male mortality remains the same, the level is altered. The change is especially pronounced at older ages, where the Hispanic advantage is significantly muted after adjustment for misclassification.

DISCUSSION

Our results indicate that previous reports of a longer average lifespan for Hispanics is not supported by the most recent American data, even before adjusting for possible misclassification of ethnicity on death certificates, when we take race into account as well. However, if we compare the White, non-Hispanic group to the Hispanic group as a whole, Hispanics do appear to live longer on average; approximately 1 year longer for men, and 2 years longer for women. After adjusting these 2 populations for misclassification of Hispanic ethnicity, the life expectancy gap is eliminated for both men and women. After adjustment, Hispanic men appear to live approximately 2 years less than White non-Hispanics. For women, life expectancy is extremely close for the 2 groups: 79.33 years for White non-Hispanic women, and 79.49 years for Hispanic women.

Based on the NMFS, misclassification of ethnicity is a problem on death certificates. While the misclassification may appear to be relatively minor (overall PV+ was 98%), using simple adjustment measures demonstrates that these low levels of misclassification result in an overstatement of life expectancy for White Hispanic males of approximately 2 years. Non-White Hispanics appear to be subject to much higher rates of misclassification. For example, life expectancy for Black Hispanics falls by over 10 years for both men and women when ethnic misclassification is taken into account. This may indicate a bias against reporting more than one minority classification on death certificates. Further study of the magnitude of this effect is important, especially given the recent change in race reporting on Census 2000. Individuals are now able to choose a combination of race and ethnicity, and, potentially, multiple races as well.

The NMFS is the best source for potential misclassification we were able to examine. The biggest potential limitations of using NMFS data for this study include the loss of follow-up due to non-response or inability to find next-of kin, and the possibility that next-of-kin reports do not agree with self-report. Our major concern about using the NMFS data is the procedure of asking a proxy respondent a series of questions after the decedent's death, therefore not actually sampling the decedent's ethnicity, since that would require speaking directly to the individuals. This further means that we cannot compare the de-

Table 3. Life table calculations of average life expectancy (e_0) for Hispanics and non-Hispanics in the United States, adjusted for misclassification of Hispanic ethnicity. Calculated using the 1989–1991 complete mortality statistics of the United States, the Census population for 1990, and the National Mortality Follow-Back Survey of 1993 for estimates of misclassification.

	Males		Females	
	Unadjusted	Adjusted	Unadjusted	Adjusted
White non-Hispanic	72.63	72.79	79.28	78.33
White Hispanic	65.65	63.15	75.59	74.48
Black non-Hispanic	63.67	63.84	73.09	73.23
Black Hispanic	77.28	65.01	89.15	74.47
Am Indian non-Hispanic	71.67	71.79	82.65	79.34
Non-White non-Hispanic	66.24	66.33	74.97	74.88
Non-White Hispanic	92.58	89.88	93.61	91.58
Hispanic (regardless of race)	73.64	70.96	81.11	79.49

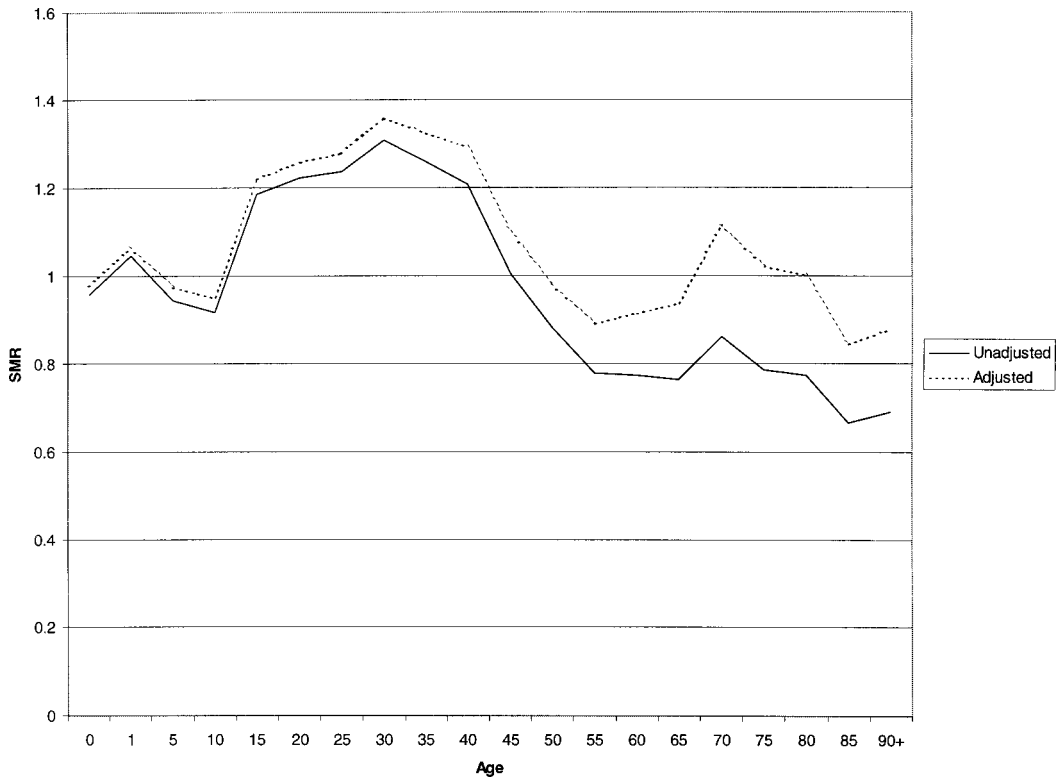

Fig 2. Standardized mortality ratios (SMRs), comparing age-specific mortality rates for Hispanic males, unadjusted and adjusted, to unadjusted non-Hispanic male mortality rates. Calculated using the 1989–1991 complete mortality statistics of the United States, the Census population for 1990, and the National Mortality Follow-Back Survey of 1993 for estimates of misclassification

cedent's stated ethnicity with the death certificate, but rather must compare the proxy report of ethnicity. Although we have treated the proxy report as if it matched that of the respondent precisely, it is not, in fact, a self-report of ethnicity. Since the proxy may not be aware of the actual self-reported ethnicity, we believe that true misclassification may be even higher. However, we note that census responses are often by proxy as well, with one household member reporting on the entire household. An even better source might be a study such as the National Longitudinal Mortality Study (NLMS), which matches Current Population Surveys with death records,[36] or the National Health Interview Survey matched with the National Death Index.[37] Since neither of these surveys releases ethnicity from the death certificate, matching between self-report prior to death and on the death certificate cannot be performed.

". . . if we compare the White, non-Hispanic group to the Hispanic group as a whole, Hispanics do appear to live longer on average . . ."

We also are concerned that the match between self-reported ethnicity and NMFS ethnicity may vary by the identity of the proxy. In particular, we had expected that proxies who are non-family members might be especially prone to mis-report ethnicity. Table 1 demonstrates that sensitivity was highest when the proxy reporter was an in-law or neighbor. Unexpectedly, sensitivity was low when the reporter was a child, spouse, or sibling of the decedent. Likewise, PV+ is highest (at 1.0, indicating perfect agreement) when the reporter was a step-parent or institutional staff person. Reports by spouses and children are much less likely to agree with the death certificate. The aim of the NMFS was to use the best-informed proxy possible. Overall, 74% of respondents to the NMFS were relatives. The most common relationships the proxies had to the decedents were filial (31%) or spousal (29%). Only 8% of the respondents were not related to the decedent in any way. Nonetheless, the reporter bias could be a design flaw in the NMFS. When no close relative was available, neighbors or institutions may have been asked questions about the decedent. Those non-relatives might have had the same biases as the person who filled out the death certificate. Therefore, an elderly man who would call

himself Hispanic might be categorized as non-Hispanic on the death certificate, and non-Hispanic by a nursing home staff member. On the NMFS, however, his ethnicity would have appeared to be correctly classified, since the two sources would be in agreement. Only by obtaining ethnicity prior to death can this bias be eliminated.

Another potential problem with the NMFS is the fact that it is not a random sample of all deaths, nor does it contain an oversample of Hispanic deaths. Other data sources, such as the National Longitudinal Mortality Survey (NLMS), also have assessed mortality of Hispanics.[9,36] While the NLMS has a larger sample of deaths, the results do not lead to an assessment of misclassification. Examining the results reported for the NLMS, we find that while Hispanic death rates are low, rates are reported for White-Hispanics rather than for all Hispanics.[36(pp154–155)] In the 2000 US Census, of the 35.5 million people who claimed Hispanic ethnicity, 47.9% also said they were White, 42.2% said they were of "some other race," 6.3% reported 2 or more races, 2% were Black, 1.2% were American Indian (or Alaskan Native), 0.3% were Asian, and 0.1% were Native Hawaiian or Pacific Islander.[41] On examining the results reported in the NLMS for respondents who said they were of "other" race, we found very high death rates, especially at young ages.[36(pp44–45)] Therefore, while the NLMS may appear to indicate that mortality rates for Hispanics are low, the racial classification of White Hispanics may be influencing these conclusions.

In addition, we were able to compare reporting from two sources, and then calculate adjusted death rates; a method unavailable to the NLMS or the NHIS-NDI. Many researchers examining Hispanic death rates may be relying on life tables derived from NCHS data; therefore, the adjustment methodology might be particularly valuable. Clearly, those researchers with access to self-reported ethnicity in their own studies should rely on that information, rather than using a nationally representative sample, such as we propose here. In addition, users who want to adjust for misclassification of ethnicity in their own data may be able to apply this methodology and create more accurate death rates for a diverse set of uses.

Our future plans include assessing the San Francisco-Oakland Cancer Registry study of Hispanic misclassification to reassess the issue of misclassification of Hispanic ethnicity.[33,38] We also would like to assess misclassification of Hispanic ethnicity on some larger data sources that are collected prior to death and then matched to death certificate data (such as the NLMS or the NHIS). Neither the NLMS nor the NHIS reports the ethnicity ascribed to the decedent on the death certificate in their public data files.

Future work must also consider other reporting errors on death certificates. While this project focuses on misclassification of ethnicity, other sources of data error include age, the mis-reporting of which has been thought to be relatively unimportant among the oldest-old Whites including both Hispanics and non-Hispanics.[42] However, since age mis-reporting generally biases mortality estimates downward,[43] age matching also should be considered wherever possible, as an additional factor potentially introducing error to mortality estimates.

These results indicate that the Hispanic paradox cannot be wholly explained by faulty data. Comparing the White non-Hispanic group with the Hispanic group as a whole, indicates that much of the Hispanic advantage in life expectancy is erased by adjusting for misclassification of ethnicity, at least for men (see Table 3). However, the Hispanic pattern is still anomalous. Since life expectancy calculations are dominated by early life, our measure of average life expectancy at birth may understate the Hispanic advantage at older ages. The NLMS found Hispanic male mortality to be especially high at young ages.[36(p154)] Causal mechanisms should be investigated, in addition to data quality issues.

ACKNOWLEDGMENTS

Support for this research was provided to Dr. Swallen and Mr. Guend by the National Institute on Aging (1 R03 AG18190-01). Additional support was provided for Dr. Swallen by the Wisconsin Alumni Research Fund. Lisa Wade also provided valuable research assistance to Dr. Swallen. The University of Wisconsin institutional review board approved this research.

REFERENCES

1. US Bureau of the Census, Population Division. *United States Population Estimates, by Age, Sex, Race, and Hispanic Origin, 1990 to 1997.* Washington, DC: US Census Bureau; 1999. Release PPL-91.
2. Schmidley D, Alvarado HA. The foreign-born population in the United States: March 1997 (update). *Current Population Reports, Population Characteristics.* Washington, DC: US Census Bureau; 1998. P20-507.
3. Treas J. Older Americans in the 1990s and beyond. *Popul Bull.* 1996;50(2):1–46.
4. LaVeist TA. Data sources for aging research on racial and ethnic groups. *Gerontologist.* 1995;35(3):328–339.
5. Shai D, Rosenwaike I. Mortality among Hispanics in metropolitan Chicago: an examination based on vital statistics data. *J Chronic Dis.* 1987;40(5):445–451.
6. Polednak AP. Mortality in the Hispanic population of Suffolk County, New York. *N Y State J Med.* 1990;90(9):442–446.
7. National Center for Health Statistics. *Health, United States, 1990.* Hyattsville, Md: US Dept of Health and Human Services; 1991. Publication (PHS) 91-1232.
8. Council on Scientific Affairs, American Medical Association. Hispanic health in the US. *JAMA.* 1991;265:248–252.
9. Sorlie PD, Backlund E, Johnson NJ, Rogot E. Mortality by Hispanic status in the United States. *JAMA.* 1993;270:2464–2468.
10. Liao Y, Cooper RS, Cao G, et al. Mortality patterns among adult Hispanics: findings from the NHIS, 1986 to 1990. *Am J Public Health.* 1998;88(2):227–232.
11. Swallen KC. *Morbidity, Mortality, and Nativity: The Influence of Early-Life Effects and Health Selection on Health at Old Age* [dissertation]. Berkeley, Calif: Department of Demography, University of California, Berkeley; 1996.
12. Smith JC, Mercy JA, Rosenberg ML. Suicide and homicide among Hispanics in the Southwest. *Public Health Rep.* 1986;101:265–270.

13. Becker TM, Wiggins C, Key CR, Samet JM. Ischemic heart disease mortality in Hispanics, American Indians, and non-Hispanic Whites in New Mexico, 1958–1982. *Circulation.* 1988;78:302–309.
14. Gimenez ME. Latino/"Hispanic"—who needs a name? The case against a standardized terminology. *Int J Health Serv.* 1989;19(3): 557–571.
15. Culebras A. Hispanic. An epidemiologically meaningless term. *Arch Neurol.* 1995;52(5): 533–534.
16. Sacco RL. Hispanic. An epidemiologic meaningful term needed for public health planning. *Arch Neurol.* 1995;52(5):531–532.
17. Zimmerman RS, Vega WA, Gil AG, et al. Who is Hispanic? Definitions and their consequences. *Am J Public Health.* 1994;84(12): 1985–1987.
18. Crews DE, Bindon J. Ethnicity as a taxonomic tool in biomedical and biosocial research. *Ethn Dis.* 1991;1:42–49.
19. Hahn RA, Mulinare J, Teutsch SM. Inconsistencies in coding of race and ethnicity between birth and death in US infants. A new look at infant mortality, 1983 through 1985. *JAMA.* 1992;267(2):259–263.
20. Osborne NG, Feit MD. The use of race in medical research. *JAMA.* 1992;267:275–279.
21. Office of Management and Budget. Revisions to the standards for the classification of federal data on race and ethnicity. *Federal Register.* 1997.
22. Nickens HW. A compelling research agenda. *Ann Intern Med.* 1996;125(3):237–239.
23. Markides KS, Coreil J. The health of Hispanics in the southwestern United States: an epidemiologic paradox. *Public Health Rep.* 1986;101(3):253–265.
24. Collins JW, Shay DK. Prevalence of low birth weight among Hispanic infants with US-born and foreign-born mothers: the effect of urban poverty. *Am J Epidemiol.* 1994;139:184–192.
25. Coale AJ, Kisker EE. Mortality crossovers: reality or bad data? *Popul Stud.* 1986;40:389–401.
26. Preston SH, Elo IT, Rosenwaike I, Hill M. African-American mortality at older ages: results of a matching study. *Demography.* 1996; 33(2):193–209.
27. National Center for Health Statistics. *Funeral Directors' Handbook on Death Registration and Fetal Death Reporting.* Hyattsville, Md: US Department of Health and Human Services; 1987.
28. National Center for Health Statistics, Centers for Disease Control and Prevention. *National Mortality Followback Survey—Provisional Data, 1993* [data file]. Hyattsville, Md: National Center for Health Statistics, Centers for Disease Control and Prevention; 1998.
29. Poe GS, Powell-Griner E, McLaughlin JK, et al. Comparability of the death certificate and the 1986 National Mortality Follow-Back Survey. National Center for Health Statistics. *Vital Health Stat.* 1993;2(118).
30. Sorlie PD, Rogot E, Johnson NJ. Validity of demographic characteristics on the death certificate. *Epidemiology.* 1992;3(2):181–184.
31. Blustein J. The reliability of racial classifications in hospital discharge abstract data. *Am J Public Health.* 1994;84(6):1018–1021.
32. Polednak AP. Estimating mortality in the Hispanic population of Connecticut, 1990 to 1991. *Am J Public Health.* 1995;85(7):998–1001.
33. Swallen KC, West DW, Stewart SL, et al. Predictors of misclassification of Hispanic ethnicity in a population-based cancer registry. *Ann Epidemiol.* 1997;7(3):200–206.
34. Swallen KC, Glaser SL, Stewart SL, et al. Accuracy of racial classification of Vietnamese patients in a population-based cancer registry. *Ethn Dis.* 1998;8(2):218–227.
35. Rosenwaike I, Hempstead K, Rogers RG. Using surname data in US Puerto Rican mortality analysis. *Demography.* 1991;28(1):175–180.
36. Rogot E, Sorlie PD, Johnson NJ, Schmitt C. *A Mortality Study of 1.3 Million Persons by Demographic, Social, and Economic Factors: 1979–1985 Follow-Up.* Bethesda, Md: National Institute of Health; 1992. Publication No. 92-3297.
37. US Department of Health and Human Services, National Center for Health Statistics. *National Health Interview Survey, 1989: Multiple Cause of Death, Dates of Death, 1989–1991* [computer file]. Hyattsville, Md: US Dept of Health and Human Services, National Center for Health Statistics; 1994.
38. Stewart SL, Swallen KC, Glaser SL, et al. Comparison of methods to classify Hispanic ethnicity in a population-based cancer registry. *Am J Epidemiol.* 1999;149(11):1063–1071.
39. Stewart SL, Swallen KC, Glaser SL, et al. Adjustment of cancer incidence rates for ethnic misclassification. *Biometrics.* 1998;54:774–781.
40. Krieger N, Quesenberry C, Peng T, et al. Race/ethnicity, social class, and incidence of breast, cervix, colon, lung, and prostate cancer. *Cancer Causes Control.* 1999;10(6):525–537.
41. Grieco EM, Cassidy RC. *Overview of Race and Hispanic Origin: Census 2000 Brief.* Washington, DC: US Census Bureau; 2001. Publication No. C2KBR/01-1.
42. Hill ME, Preston SH, Rosenwaike I. Age reporting among White Americans aged 85+: results of a record linkage study. *Demography.* 2000;37(2):175–186.
43. Preston SH, Elo IT, Stewart Q. *Effects of Age Misreporting on Mortality Estimates at Older Ages.* Philadelphia, Pa: Population Aging Research Center, University of Pennsylvania; 1998. Working Paper No. 98-01.

AUTHOR CONTRIBUTIONS

Design and concept of study: Swallen, Guend
Acquisition of data: Guend
Data analysis and interpretation: Swallen, Guend
Manuscript draft: Swallen, Guend
Statistical expertise: Swallen, Guend
Acquisition of funding: Swallen
Administrative, technical, or material assistance: Swallen, Guend
Supervision: Swallen