HHS Public Access

Differential Record Linkage by Hispanic Ethnicity and Age in Linked Mortality Studies: Implications for the Epidemiologic Paradox

Joseph T. Lariscy, MA

University of Texas at Austin

Abstract

Objectives—This study examines how the linkage of surveys to death records differs for Hispanics and non-Hispanic whites and how such differences impact estimates of ethnic differences in U.S. adult mortality.

Methods—I use data from the 1989–2000 National Health Interview Survey (NHIS) linked to the National Death Index (NDI) through 2002. Analyses assess how match score and match class vary by ethnicity, nativity, and age and whether mortality hazard ratios are sensitive to shifts in match criteria.

Results—Linkage quality is lower for Hispanic and foreign-born adults than for non-Hispanic white and U.S.-born adults. Modification of the linkage criteria determine whether or not the Hispanic mortality advantage is observed among middle-aged adults.

Discussion—The accuracy of adult mortality estimates depends on the quality of the linkage between surveys and death records.

Keywords

mortality; Hispanics; demography

Introduction

Health and demographic studies of the U.S. adult population consistently find that the mortality risk of Hispanics is lower than or comparable to that of non-Hispanic whites (Arias, 2010; Elo, Turra, Kestenbaum, & Ferguson, 2004; Hummer, Rogers, Amir, Forbes, & Frisbie, 2000; Markides & Eschbach, 2011; Sorlie, Backlund, Johnson, & Rogot, 1993). Yet, the reasons for this Hispanic mortality advantage are not fully determined. While immigration patterns, health behaviors, and community and familial ties are offered as explanations for the low death rates of Hispanics in the U.S. (Markides & Eschbach, 2005),

Address: Department of Sociology and Population Research Center, 1 University Station G1800, University of Texas, Austin, TX 78712, jlariscy@prc.utexas.edu, Telephone: (512) 471-9698, Fax: (512) 471-4886.

the advantage may be artifactual; Hispanics' apparent low mortality could be due to or exaggerated by data quality issues.

Mortality estimates of U.S. racial/ethnic groups are calculated using data from a variety of sources (Hummer, Rogers, Masters, & Saint Onge, 2009). Calculations of official U.S. mortality rates (Xu, Kochanek, Murphy, & Tejada-Vera, 2010) use data from two different sources: death counts from the U.S. National Vital Statistics System (NVSS) as the numerator and population estimates developed by the U.S. Census Bureau as the denominator. Since reports of ethnicity in the two sources may not agree, official mortality rates contain biases that result in artificially low Hispanic adult mortality rates unless corrections are made for ethnic misclassification (Arias, Eschbach, Schauman, Backlund, & Sorlie, 2010; Elo et al., 2004; Rosenberg et al., 1999). Nationally-representative surveys linked to the NDI reduce these biases since, most importantly, race and ethnicity are self-reported or reported by a proxy in the household. However, linked data are limited by uncertainty in the quality of record linkage. The linkage between survey and death record for respondents whose files do not contain all of the information used to link the files may be inaccurate and the surveys of Hispanics and the foreign-born may be linked to death records at different rates than those of non-Hispanic whites and the native-born population (Arias et al., 2010; Elo & Preston, 1997; Hummer et al., 2000; Markides & Eschbach, 2005; Palloni & Arias, 2004; Turra & Goldman, 2007). Linkage may be less accurate for foreign-born Hispanics if they are undocumented or reside in the U.S. temporarily and are therefore less likely to provide accurate social security numbers. Also, errors may occur when matching records for Hispanics by first and last name and middle initial since naming practices among Hispanics often differ from the U.S. convention.

This article examines potential ethnic, nativity, and age differences in record linkage in the National Health Interview Survey Linked Mortality Files (NHIS-LMF). I address the following research questions: Is the linkage of NHIS surveys to NDI death records less accurate for Hispanics than for non-Hispanic whites? How do nativity and age influence the quality of record linkage? How does Hispanic mortality risk relative to that of non-Hispanic whites change with modification of the criteria developed by the National Center for Health Statistics (NCHS) to link death records to survey respondents?

Ethnic Differences in Adult Mortality

Death rates among Hispanics (particularly foreign-born and older adults) are comparable to, and perhaps lower than, death rates of non-Hispanic whites in the U.S. (Arias, 2010; Elo et al., 2004; Hummer et al., 2000; Markides & Eschbach, 2011; Sorlie et al., 1993). In the first official set of life tables for U.S. Hispanics, Arias (2010) estimated life expectancy at birth to be 80.6 years for Hispanics and 78.1 years for non-Hispanic whites in 2006. The term "epidemiologic paradox" describes this phenomenon since Hispanics resemble non-Hispanic blacks in socioeconomic status yet exhibit substantially lower mortality risk (Markides & Coreil, 1986). Three explanations are given for why data consistently yield low death rates among Hispanics relative to other racial/ethnic groups: cultural effects, migration effects, and data quality issues.

Hispanics' health behaviors, community networks, family structure, and other cultural practices may be partially responsible for their longevity (Bond Huie, Hummer, & Rogers, 2002; Franzini, Ribble, & Keddie, 2001; Markides & Eschbach, 2011). Foreign-born Hispanics in the U.S. have relatively low rates of smoking and alcohol abuse (Lopez-Gonzalez, Aravena, & Hummer, 2005). Familial and community connections among foreign-born Mexican Americans may reduce the impact of lower socioeconomic status on their health (Angel & Angel, 1992). Yet, neighborhood concentration of Hispanics—which likely reinforces ethnic cultural patterns—has a weak influence on mortality risk and does little to explain the Hispanic mortality advantage (Palloni & Arias, 2004).

Immigrants to the U.S. tend to be healthier than the individuals who remain in the country of origin (Akresh & Frank, 2008; Rubalcava, Teruel, Thomas, & Goldman, 2008). Thus, the immigration process selects immigrants at least partially on the basis of good health. Yet, over time and across generations, as immigrants and their children adopt U.S. cultural practices, this health advantage decreases (Hummer, Benjamins, & Rogers, 2004; Singh & Siahpush, 2001). While immigration improves a group's overall health by drawing healthy individuals from abroad, emigration may remove the oldest and least healthy members from the U.S. The return migration of elderly Hispanics from the U.S. to their country of origin, known as "salmon bias", could result in artificially low mortality rates among Hispanics if they are included in a survey but excluded from vital status follow-up (Abraído-Lanza, Dohrenwend, Ng-Mak, & Turner, 1999). While salmon bias may exist, a Hispanic mortality advantage is observed even among populations who are extremely unlikely to leave the country or whose deaths are recorded even after emigrating. Hummer et al. (2007) found a Hispanic mortality advantage among a population unlikely to leave the country and thus fail to have their deaths reported in U.S. records: infants recently born to Mexican origin mothers. In the first hours and days following a birth, when the mother's return to Mexico with her child is nearly impossible, mortality rates among infants born to Mexican origin women are roughly 10% lower than those of infants born to U.S.-born non-Hispanic white women. Turra and Elo (2008) performed the first direct investigation of the bias introduced by outmigration using social security/NUMIDENT data, which record the deaths of primary beneficiaries who die abroad. They confirmed the existence of salmon bias for foreign-born Hispanics, but concluded that its effect is not strong enough to explain the Hispanic mortality advantage relative to non-Hispanic whites. While NDI-linked mortality studies miss some deaths due to migration out of the U.S. vital registration area (regardless of whether emigrants are selected on poor health), many of these deaths are recovered since NCHS also link surveys to deaths reported in files maintained by the Social Security Administration and the Centers for Medicare and Medicaid Services (National Center for Health Statistics, 2009a).

Numerous data sources are used to estimate ethnic mortality differences, each with limitations. Hispanic death rates calculated using death records from the NVSS and population estimates from the U.S. Census are affected by errors in reporting, classification, and coverage (Rosenberg et al., 1999). Ethnicity is self-reported on the Census but reported by a funeral director on the death certificate. If Hispanics are not identified as such on the death certificate, the number of Hispanic decedents in vital statistics data will be understated and Hispanic death rates will be underestimated. However, Arias and colleagues (2010)

showed that Hispanic ethnicity reporting on death certificates is reasonably good: roughly 5% lower than reporting on surveys. Arias (2010) and Elo et al. (2004) corrected for ethnic misclassification in vital statistics by applying ethnic classification ratios derived from the National Longitudinal Mortality Study (NLMS) (Arias et al., 2010; Rosenberg et al., 1999). While adjusted Hispanic mortality rates are higher than conventional mortality rates, they are still lower than those of non-Hispanic white adults. Additionally, Elo and colleagues (2004) found that mortality rate ratios of Hispanic adults to non-Hispanic white adults in adjusted vital statistics are similar to rate ratios calculated with social security/NUMIDENT data, which contain more accurate age reports than Census and NVSS data.

Surveys linked to NDI death records, such as the NLMS and the NHIS-LMF, avoid the ethnic misclassification problem because ethnicity is derived from the respondents' self-report on the survey rather than a funeral director's report on the death certificate. While this strategy removes the risk of ethnic misclassification, failure to correctly match surveys and death records can bias mortality estimates. Although linked mortality studies have some limitations, the continuously expanding NHIS-LMF is crucial to addressing the lingering questions surrounding the Hispanic mortality paradox since it contains a large Hispanic sample, is regularly revised to include questions to assess contemporary health issues, and includes socioeconomic and behavioral covariates not available in vital statistics or social security data.

Differential Record Linkage among Subpopulations

The NHIS relies on probabilistic matching to the NDI (passive follow-up) rather than contacting respondents or next of kin (active follow-up) to determine vital status of respondents. NDI record linkage relies on correspondence between twelve items listed in both data sources: social security number, first name, middle initial, last name, birth month, birth day, birth year, state of birth, state of residence, sex, race, and marital status. NHIS records must contain one of three combinations of personal identifiers to be eligible for linkage. The matching of surveys to death records is a two step process (see National Center for Health Statistics, 2005: 9 for a flowchart of the matching process described below). First, the NDI identifies potential matches based on seven criteria involving combinations of social security number, first and last name, middle initial, and date of birth. Second, an algorithm assigns probabilistic scores to each potential match to distinguish true matches from false matches. Scores are the sum of weights assigned to each of the items on which records are matched. If the items agree for the potential match, the weight is positive; disagreement yields a negative weight.

Potential matches are then placed into one of five mutually exclusive classes, based on the items on which they agree. Class 1 matches are considered true matches (i.e., deceased); records agree on at least eight digits of the social security number as well as first and last name, middle initial, birth year (with an allowance of ± three years), birth month, sex, and state of birth. Class 2 matches agree on fewer items than Class 1 matches: at least seven social security number digits and at least five additional items. Class 3 matches take two forms; either social security number is unknown but last name and at least seven items agree or social security number matches on six or fewer digits and at least eight other items agree.

Potential matches classified as Class 4 have unknown social security number on one of the records and fewer than eight of the items required for Class 3 agree. Class 5 matches agree on fewer than seven social security number digits or agree on at least seven social security number digits but agree on fewer than five other items. All Class 5 observations are considered survivors. The NDI linkage process may initially identify multiple potential death record matches for a survey, but only the single best match is included in the NHIS-LMF.

For Classes 2, 3, and 4, different match score cut-points are used to ascertain vital status. The cut-points are 47 for Class 2, 45 for Class 3, and 40 for Class 4; that is, respondents who receive a score of 47 or greater and are placed into Class 2 are considered deceased while Class 2 respondents with a score of 46 or less are considered alive. These match score cut-points were chosen during the NHANES I Epidemiologic Follow-Up Study (NHEFS) to maximize correct identification of true matches while simultaneously minimizing false matches (National Center for Health Statistics, 2005). However, such cut-points may not be most suitable among subpopulations within the dataset. Since the study included few minority participants, the NDI's accuracy in ascertaining vital status for Hispanics is uncertain (Herrera, Stern, Goff, & Villagomez, 1994). Linkage may be less accurate for foreign-born Hispanics if they are undocumented or reside in the U.S. temporarily and are therefore less likely to provide accurate social security numbers. Also, errors may occur when matching records for Hispanics by first and last name since their naming practices often differ from the U.S. convention.

Few previous studies have examined the effect of record linkage on racial/ethnic mortality differentials among U.S. adults. Liao and colleagues (1998) used an earlier linkage of NHIS and NDI files to calculate mortality rate ratios for Hispanics and non-Hispanic blacks relative to non-Hispanic whites based on three different matching criteria: using the NCHS-recommended criteria; identifying all Class 1 and 2 matches as deaths; and identifying all Class 1, 2, and 3 matches and Class 4 matches above the match score cut-point as deaths. They found consistent black-white rate ratios regardless of the criteria used but changes in the Hispanic-white rate ratio when different criteria were used to ascertain mortality. The dataset used by Liao and colleagues included only five years of the NHIS with six years of mortality follow-up, which precluded a thorough examination of the impact of nativity, in addition to ethnicity, on match quality. I utilize a version of the NHIS linked to the NDI that is more recent, contains a larger Hispanic sample, and allows for measurement of the effect of nativity. Hummer et al. (2000), using 1986–1994 NHIS data linked to the NDI through 1995, found that using a more liberal matching standard than that recommended by NCHS resulted in higher mortality risk among Hispanics vis-à-vis non-Hispanic whites than when the NCHS cut-points were used. Palloni and Arias (2004) discussed how differential record linkage and other data artifacts could bias Hispanic mortality rates downward. They argued that while matching rates would have to be extraordinarily poor to fully account for lower mortality among foreign-born Hispanics relative to non-Hispanic whites, the implications of differential linkage on the Hispanic mortality advantage are inadequately documented.

Objectives of Study

This article examines how the linkage of surveys and death records differ by Hispanic ethnicity. Specifically, I test whether match quality is lower for Hispanics than for non-Hispanic whites, while monitoring the effect of nativity, age, and sex on the association between Hispanic ethnicity and match quality. I also assess whether shifts in the match score cut-points for the determination of survivor/decedent status result in substantial changes in ethnic differentials in adult mortality. This article is of importance to demographic and epidemiologic researchers because the NHIS-LMF is one of the principal sources of health and mortality information for U.S. adults (Preston & Taubman, 1994; Rogers, Hummer, & Nam, 2000). It will also contribute to the ongoing debate over the existence and causes of the epidemiologic paradox.

Methods

Sample

This study utilizes the NHIS-LMF: 1986–2000 data from the NHIS with NDI mortality follow-up through December 31, 2002 (National Center for Health Statistics, 2006). The NHIS is a nationally-representative probability survey of the U.S. non-institutionalized civilian population and interviews roughly 100,000 individuals each year. For years 1986 through 2002, the NCHS determined vital status of NHIS adult respondents aged 18 and over by linking surveys to the NDI mortality database, a computerized registry of death certificates filed in the U.S. since 1979. More than 99 percent of deaths occurring in all 50 states and the District of Columbia are registered. The NHIS-LMF offers many advantages over other datasets used in the study of ethnic and nativity differentials in adult mortality: current data; critical measures of ethnicity, nativity, and other sociodemographic characteristics collected at baseline; a high response rate (about 95% of eligible households for most years); and large Hispanic and foreign-born samples. I use NHIS data from years 1989 forward since respondents did not report nativity in prior years. Only participants 18 years old and older with sufficient identifying information listed on both data sources are eligible for linkage. "Ineligibles" are excluded from analyses (National Center for Health Statistics, 2009a). In the 1989–2002 NHIS-LMF, 3.3% of respondents 18 years old and older are ineligible for NDI linkage. However, ethnic groups are not equally likely to be excluded from matching due to insufficient data: 3.0% and 4.2% of non-Hispanic whites and Hispanics, respectively, are ineligible. While ineligibles are excluded from all analyses in this article, researchers should be aware that Hispanics' higher percentage of ineligibility indicates an additional concern about ethnic differences in record linkage. The analytic sample contains 635,655 respondents, 64,737 of whom were identified by NCHS as deceased by the end of the 14-year follow-up period. In the sensitivity analyses, relaxing the match score cut-points increases the number of respondents identified as deceased and reduces the number of survivors while tightening the cut-points decreases deaths and increases survivors.

Variables and Measurement

The special request file of the NHIS-LMF used here contains all the variables of the public use data but also includes match score and match class. Match score ranges from 0 to 116 in

the analytic sample. Higher scores indicate more items matched between the NHIS and NDI. Respondents are also placed into one of five mutually exclusive classes depending on which identifiers match between the two data sources. Match class is not included in multivariate analyses given its high correlation with match score (r = −.976).

NHIS respondents self-report their race and whether or not they are of Hispanic origin. Ethnicity is measured here as Hispanic and non-Hispanic white. Hispanics in the NHIS are predominantly Mexican American; small sample sizes preclude analysis by Hispanic subgroup. NHIS respondents who report a race or ethnicity other than Hispanic and non-Hispanic white are excluded from analyses. While the data quality issues affecting mortality estimates of Hispanics could also bias estimates for other racial/ethnic groups, the immigration and mortality profiles of other groups are beyond the scope of this article; moreover, their sample sizes are much smaller. For years 1989 through 2000, foreign-born respondents are asked how many years they have resided in the U.S. Since time in the U.S. is not a focus of this article, responses are collapsed into two categories: foreign-born and U.S.-born. Nativity is considered in all analyses since foreign-born individuals generally have a lower risk of mortality than their U.S.-born counterparts (Singh & Siahpush, 2001). Failure to control for nativity when studying populations with a large proportion of foreign-born individuals, such as Hispanics in the U.S., would likely bias mortality estimates downward (Hummer, Rogers, Nam, & LeClere, 1999). Also, immigrants may have lower match scores due to missing or incorrect social security numbers in the NHIS or NDI. I adjust for or stratify by age and sex in multivariate analyses. NHIS top-codes age at 99 years in 1989–1995, 90 years in 1996, and 85 years from 1997 forward. I limit age to 25–89 years due to small numbers of deaths at the youngest ages and survivors at the oldest ages.

Analysis

The analysis involves three stages: description of match class and match score distribution by ethnicity and nativity, ordinary least squares (OLS) regression analysis of match score on ethnicity and nativity, and sensitivity analyses that examine how ethnic mortality differentials change with modification of the NCHS-recommended criteria for death ascertainment. I use SUDAAN 10.0 to account for the complex sampling design of the NHIS (Research Triangle Institute 2008). All analyses are weighted using NCHS-provided weights, which account for ineligibility, so that results represent the non-institutionalized civilian U.S. adult population.

The first stage of the analysis presents cross-tabulations for class by ethnicity and nativity separately for decedents and survivors. Mean match class and mean match score are also shown. This analysis shows how match score and match class vary by ethnicity and nativity prior to adjustment for covariates believed to affect record linkage. Mean scores for survivors should be interpreted with caution since their match score values are truncated at zero (83.5% of survivors have a score of zero).

For the second set of analyses, I regress match score on Hispanic ethnicity, nativity, and covariates for decedents. Negative coefficients indicate a decrease in linkage quality between NHIS and NDI records while positive coefficients indicate an improvement. I present three models to show how adjustment for nativity affects the ethnicity-match score

association. Model 1 shows the association between Hispanic ethnicity and match score net of age and sex. Model 2 includes nativity additively and Model 3 includes a Hispanic × Foreign-born interaction term. The aim of this analysis is to demonstrate whether and in which direction the average match score for Hispanics differs from that of non-Hispanic whites, net of covariates. I adjust for sex (reference group = females) and age in five-year groups (reference group = age 60–64). The choice of adults aged 60–64 years as the reference group is based on preliminary analyses showing that this group has the highest average match score among NHIS respondents who died rather than theoretical expectations about quality of record linkage by age. Due to the large proportion of survivors assigned a score of zero, a meaningful comparison is not performed for survivors.

For the third set of analyses, I use Cox proportional hazard models to investigate whether hazard ratios change when match score cut-points are modified. The outcome variable is all-cause mortality. Duration is calculated using quarter-year of interview and quarter-year of death (median = 8.125 years, standard deviation = 3.6). Respondents are assumed to be exposed to the risk of death from the mid-point of the quarter of interview to the mid-point of the quarter of death. Those who survive throughout the follow-up period are right-censored at December 31, 2002. For respondents whose matches are assigned to Classes 2, 3, and 4, hazard ratios are estimated relative to U.S.-born non-Hispanic whites using NCHS-recommended cut-points, controlling for age and sex. Then, the class-specific cut-points are tightened (increased by five points) and relaxed (decreased by five points) to see whether the hazard ratios increase, decrease, or remain constant. A five point band is more subtle than the shifts by Liao et al. (1998) and is comparable to the four point band used by NCHS staff in a sensitivity demonstration using the entire NHIS-LMF sample (National Center for Health Statistics, 2005). I obtain similar results using two points or ten points rather than five points to relax and tighten criteria (not shown, available by request). Tightening and relaxing can be considered conservative and liberal standards, respectively, by which vital status is ascertained. The match score cut-points are 47 for Class 2, 45 for Class 3, and 40 for Class 4; respondents with a score greater than or equal to their class's cut-point are considered deceased (National Center for Health Statistics, 2005). Vital status of Class 1 and Class 5 respondents does not change in the sensitivity analyses since their classes do not have cut-points; all Class 1 matches are considered deaths while all Class 5 matches are considered survivors.

Results

Table 1 shows class distribution, mean class, and mean score by ethnicity and nativity separately for deaths and survivors identified by NCHS. The class distributions show that, for both vital status groups, non-Hispanic whites have higher percentages in the groups with the most certainty in vital status ascertainment than do Hispanics (i.e., Class 1 for deaths and Class 5 for survivors). For instance, among survivors, about 85% of non-Hispanic whites are placed into Class 5 while about 75% of Hispanics are in this class. The mean class for Hispanic respondents identified as deaths are higher than those for non-Hispanic white respondents identified as deaths, suggesting less certainty in record linkage among Hispanics. Match quality is lower for the foreign-born, particularly among Hispanics. The mean score for U.S.-born Hispanics identified as deceased is about 10 points greater than

that of foreign-born Hispanic decedents while the nativity difference among non-Hispanic white decedents is about one point.

A multivariate analysis is necessary to determine how covariates influence the association between ethnicity and match score. Thus, Table 2 shows the results from the OLS regression of match score on ethnicity, nativity, sex, and age for decedents. Model 1 shows the basic relationship between Hispanic ethnicity and match score, net of sex and age. The coefficient for Hispanics indicates that the average match score for Hispanic decedents is 7.8 points lower than the average match score for non-Hispanic whites. The coefficient for Hispanic ethnicity is reduced but remains significant when nativity is added in Model 2, indicating that a portion of Hispanic decedents' lower match score is due to the high proportion of foreign-born individuals among deceased Hispanic respondents. Ascertainment as deceased is less certain for the foreign-born compared to the U.S.-born; the average match score among foreign-born respondents is nearly three points lower than that of the U.S.-born, net of Hispanic ethnicity, sex, and age. The positive coefficient for males indicates that record linkage is more certain for males than for females. Vital status ascertainment in linked mortality studies is generally less certain for females than for males since some females change their last name at marriage, thus leading to the possibility of different last names listed on the survey and death record. Model 3 allows for interaction between ethnicity and nativity. The average match score for foreign-born Hispanics who died during follow-up is about twelve points lower than the average score of U.S.-born non-Hispanic whites. Adults aged 50–69 years have the highest average match scores. Compared to the average score of 60–64 year old adults who died during follow-up, the average score of the youngest decedents (ages 25–29 years) is nearly ten points lower and the average score of the oldest decedents (ages 85–89 years) is more than two and a half points lower. Thus, record linkage for adults who are interviewed when they are middle-aged and die during follow-up is more certain than for decedents who were interviewed at younger or older adult ages.

Since nativity influences the relationship between ethnicity and vital status ascertainment (particularly for Hispanics), ethnic groups are stratified by nativity status in the sensitivity analyses. That is, respondents are placed into one of four mutually exclusive ethnicity-nativity categories: foreign-born Hispanics, U.S.-born Hispanics, foreign-born non-Hispanic whites, and U.S.-born non-Hispanic whites (reference group). Table 3 shows how hazard ratios change when three different criteria are used to ascertain vital status: the NCHS-recommended cut-points; relaxed (lowered by five points) match score cut-points for Classes 2, 3, and 4; and tightened (increased by five points) match score cut-points for Classes 2, 3, and 4. A hazard ratio greater than 1.0 indicates higher risk of death during follow-up relative to the reference group while a value less than 1.0 indicates lower mortality risk. For foreign-born Hispanics, relaxing and tightening the NCHS-recommended cut-point results in substantial changes to mortality risk relative to U.S.-born non-Hispanic whites. When the NCHS-recommended criteria are applied, the mortality risk at follow-up for foreign-born Hispanics is not statistically different from that of U.S.-born whites. The tightened criteria lead to a mortality advantage (significantly lower mortality risk relative to native-born non-Hispanic whites) while the relaxed criteria lead to a 24% greater risk of death during follow-up for foreign-born Hispanics compared to U.S.-born non-Hispanic whites. The mortality risk during follow-up for U.S.-born Hispanics is significantly greater

relative to U.S.-born non-Hispanic whites when NCHS-recommended cut-points are applied and remains greater when applying more liberal or conservative linkage standards. Regardless of linkage standard used, foreign-born non-Hispanic whites have a roughly 20% lower risk of dying during follow-up than U.S.-born non-Hispanic whites.

Since age and sex are important determinants of mortality risk, Table 4 presents results of the sensitivity analysis in 6 separate sex- and age-group panels. I control for age in each of the six panels. Among younger adults (25–44 years), both foreign-born and U.S.-born Hispanics have a significantly greater risk of mortality during follow-up relative to U.S.-born non-Hispanic whites. The risk of death is generally higher for foreign-born Hispanic young adults than for U.S.-born Hispanic young adults. For instance, using the NCHS linkage standard, foreign-born Hispanic young men are 87% more likely to die during follow-up relative to U.S.-born whites while the risk among U.S.-born Hispanic young men is only 49% greater than that of U.S.-born whites. Among middle-aged adults (45–64 years), relaxing and tightening the linkage criteria determine whether or not the Hispanic mortality advantage is observed. The risk of death during follow-up for foreign-born Hispanic women in this age group is significantly greater than that of the reference group using the relaxed criteria, not statistically different using the NCHS-recommended cut-points, and significantly lower using the tightened criteria. Among the oldest adults, foreign-born Hispanics are less likely to die during follow-up than U.S.-born non-Hispanic whites regardless of which matching standard is applied. Mortality risk among the oldest U.S.-born Hispanics is not statistically different from that of U.S.-born non-Hispanic whites. For all age and sex groups, the mortality risk for foreign-born non-Hispanic whites is lower than or equal to the risk of their U.S.-born counterparts and changes little when alternative matching criteria are used. Middle-aged foreign-born white adults have particularly low relative mortality risk; using the NCHS standard, foreign-born white females are 45% less likely to die during follow-up than U.S.-born white females while foreign-born white males are 33% less likely to die than their native-born counterparts. In general, the shift from relaxed to tightened match standards reduces mortality risk relative to U.S.-born non-Hispanic whites.

Discussion

Researchers examining the association between race/ethnicity and adult mortality in the U.S. generally find that Hispanics have a lower risk of mortality than non-Hispanic whites. This association is observed in both official reports and surveys linked to the NDI; for various periods of time; and at community-, regional-, and national-levels. While several explanations for this paradoxical finding are suggested, data quality issues receive inadequate attention as a possible explanation for the Hispanic mortality advantage. This article contributes to this research by examining differences by ethnicity, nativity, and age in two measures of linkage quality, match class and match score, as well as how these linkage differences affect mortality estimates. Results suggest that match quality is lower for Hispanic and foreign-born adults than for non-Hispanic white and U.S.-born adults. Mortality differences between foreign-born Hispanic and U.S.-born non-Hispanic white adults under age 65 years change substantially when the criteria used to determine vital status are modified.

The sensitivity analyses yield findings similar to those of Liao et al. (1998): shifts in the vital status ascertainment criteria of the NHIS-LMF impact ethnic mortality differentials. Unlike the data used by Liao and colleagues, the 1989–2000 NHIS data linked to the NDI through 2002 used in this article contain sufficient numbers of foreign-born Hispanics and non-Hispanic whites to test the influence of nativity on the association between ethnicity and record linkage. I demonstrate that a portion of the change in mortality risk that occurs when ascertainment criteria are relaxed or tightened is a result of immigrants' lower data quality in the NHIS and NDI and Hispanics' higher proportion of foreign-born individuals. The hazard ratios shown using the NCHS-recommended criteria resemble those calculated by other researchers. Elo and colleagues (2004), using NVSS and Census data corrected for ethnic misclassification, found similar results for older Mexican origin males but less favorable mortality among older Mexican origin females, compared to the results reported in the current study. For instance, among adults ages 65–89, they found that the death rate ratio of Hispanics born in Mexico to non-Hispanic whites ranges from 0.72 to 0.84 for males and from 0.89 to 1.08 for females. Results support Palloni and Arias' (2004) suggestion that record linkage differentials are not large enough to fully account for the Hispanic mortality advantage at all ages, but they reduce the accuracy of mortality risk estimation of U.S. Hispanics.

Adjustment for age reveals how match quality influences ethnic differentials in record linkage throughout the life course. While adults interviewed in middle age and who subsequently die have higher rates of linkage than decedents who were interviewed at the youngest or oldest ages, ethnic and nativity differences in mortality are most sensitive to relaxing or tightening of matching criteria among young and middle-aged adults. This pattern contributes to the mixed findings regarding ethnic mortality differences at younger adult ages. While all data sources used to study ethnic differences in adult mortality contain limitations, they generally come to the same conclusions of lower or comparable Hispanic mortality risk relative to non-Hispanic whites. One uncertainty that remains is whether the Hispanic mortality advantage exists among younger adults. Some studies find that mortality is higher among younger Hispanics relative to non-Hispanic whites and that a cross-over occurs (Arias, 2010; Hummer et al., 2000). Others find that the Hispanic mortality exists at all ages beyond childhood (Xu et al., 2010) or that, compared to non-Hispanic white young adults (ages 15–44) in California and Texas, mortality rates for most causes of death are lower among foreign-born Hispanic young adults (Eschbach, Stimpson, Kuo, & Goodwin, 2007). Thus, studies using a variety of data sources find a Hispanic mortality advantage at older adult ages (where most deaths occur) but mixed results regarding ethnic mortality differences at younger adult ages (where mortality hazard ratios are most sensitive to modification of match score cut-points).

While this article shows that mortality differences vary substantially when adjustments are made to vital status ascertainment criteria, it is unable to show whether alternative match score cut-points produce more accurate estimates of mortality differences. Therefore, the NCHS-recommended criteria should continue to be used to determine survivor/decedent status in the NHIS-LMF until a more suitable linkage process is developed for Hispanics and the foreign-born. However, researchers must be cautious when applying these criteria to subpopulations since they have not yet been validated for minority populations and may

produce biased mortality estimates for Hispanics and the foreign-born (Herrera et al., 1994). The linkage process matches NHIS and NDI records on more items or on items that are given greater weight for non-Hispanic whites than for Hispanics. This special request file of the NHIS-LMF includes match score and match class but not the personal identifiers reported on NHIS surveys and death records in order to maximize data access while protecting the confidentiality of NHIS respondents (Lochner, Hummer, Bartee, Wheatcroft, & Cox, 2008). Therefore, I do not examine the exact items responsible for yielding differential match score and match class distributions. NHIS participants' refusal to report social security number is largely responsible for reduced match quality in the NHIS-LMF. For instance, the percentage of missing social security numbers for adult males in the NHIS increased from 17.5% in 1986 to 56.4% in 2000 (National Center for Health Statistics, 2009b). One solution is to collect fewer digits of respondents' social security numbers (i.e., the last four digits rather than all nine) (Sayer & Cox, 2003). Another strategy to improve record linkage between the NHIS and NDI among subpopulations is to actively follow-up potential matches with scores closest to their class's cut-point. Since these are the individuals most at risk of being incorrectly matched by the algorithm, confirmation of vital status by contacting the respondent or next of kin would greatly improve the quality of record linkage.

Mortality estimates of other subpopulations, including the elderly, individuals with low levels of educational attainment, and other racial/ethnic groups, may also be biased by differential record linkage. The quality of record linkage among the oldest old could not be examined with this NHIS release due to top-coding of age at 99 years in 1989–1995, 90 years in 1996, and 85 years in 1997 forward. Further investigation of record linkage among the oldest adults is needed since age is a primary item used in the linkage process and elderly respondents are at a greater risk of misreporting age on surveys (Preston, Elo, & Stewart, 1999).

Interest in the influence of differential record linkage and other data quality issues on accurate measurement of Hispanics' mortality risk has increased recently as researchers continue to debate the existence and causes of the epidemiologic paradox. Only in the last few decades have administrative records and social science surveys used to study health recorded Hispanic ethnicity. Strategies must be developed to reduce the racial/ethnic, nativity, and age differences in data quality uncovered by this article before accurate estimates of mortality differentials can be produced for the racial/ethnic minority and foreign-born U.S. adult population.

Acknowledgements

References

Abraído-Lanza AF, Dohrenwend BP, Ng-Mak DS, Turner JB. The Latino mortality paradox: A test of the 'salmon bias' and healthy migrant hypotheses. American Journal of Public Health. 1999; 89(10): 1543–1548. [PubMed: 10511837]

Akresh IR, Frank R. Health selection among new immigrants. American Journal of Public Health. 2008; 98(11):2058–2064. [PubMed: 18309141]

Angel JL, Angel RJ. Age at migration, social connections, and well-being among elderly Hispanics. Journal of Aging and Health. 1992; 4(4):480–499. [PubMed: 10125149]

Arias E. United States life tables by Hispanic origin. National Center for Health Statistics. Vital and Health Statistics. 2010; 2(152):1–33. [PubMed: 21313788]

Arias E, Eschbach K, Schauman WS, Backlund EL, Sorlie PD. The Hispanic mortality advantage and ethnic misclassification on U.S. death certificates. American Journal of Public Health. 2010; 100(Supplement):S171–S177. [PubMed: 19762677]

Bond Huie SA, Hummer RA, Rogers RG. Individual and contextual risks of death among race and ethnic groups in the United States. Journal of Health and Social Behavior. 2002; 43(3):359–381. [PubMed: 12467258]

Elo, IT.; Preston, SH. Racial and ethnic differences in mortality at older ages. In: Martin, LG.; Soldo, BJ., editors. Racial and Ethnic Differences in the Health of Older Americans. Washington, DC: National Academy Press; 1997.

Elo IT, Turra CM, Kestenbaum B, Ferguson BR. Mortality among elderly Hispanics in the United States: Past evidence and new results. Demography. 2004; 41(1):109–128. [PubMed: 15074127]

Eschbach K, Stimpson JP, Kuo Y-F, Goodwin JS. Mortality of foreign-born and U.S.-born Hispanic adults at younger ages: A reexamination of recent patterns. American Journal of Public Health. 2007; 97(7):1297–1304. [PubMed: 17538049]

Franzini L, Ribble JC, Keddie AM. Understanding the Hispanic paradox. Ethnicity and Disease. 2001; 11(3):496–518. [PubMed: 11572416]

Herrera CR, Stern MP, Goff D, Villagomez E. Mortality among Hispanics. Journal of the American Medical Association. 1994; 271(16):1237. [PubMed: 8151895]

Hummer, RA.; Benjamins, MR.; Rogers, RG. Racial and ethnic disparities in health and mortality among the U.S. elderly population. In: Anderson, NB.; Bulatao, RA.; Cohen, B., editors. Critical Perspectives on Racial and Ethnic Differences in Health in Later Life. Washington, DC: National Academies Press; 2004.

Hummer RA, Powers DA, Pullum SG, Gossman GL, Frisbie WP. Paradox found (again): Infant mortality among the Mexican-origin population in the United States. Demography. 2007; 44(3): 441–457. [PubMed: 17913005]

Hummer RA, Rogers RG, Amir SH, Forbes D, Frisbie WP. Adult mortality differentials among Hispanic subgroups and non-Hispanic whites. Social Science Quarterly. 2000; 81(1):459–476. [PubMed: 17879490]

Hummer, RA.; Rogers, RG.; Masters, RK.; Saint Onge, JM. Mortality patterns in late life. In: Uhlenberg, P., editor. International Handbook of Population Aging. New York: Springer; 2009.

Hummer RA, Rogers RG, Nam CB, LeClere FB. Race/ethnicity, nativity, and U.S. adult mortality. Social Science Quarterly. 1999; 80(1):136–153.

Liao Y, Cooper RS, Cao G, Durazo-Arvizu R, Kaufman JS, Luke A, et al. Mortality patterns among adult Hispanics: Findings from the NHIS, 1986 to 1990. American Journal of Public Health. 1998; 88(2):227–232. [PubMed: 9491012]

Lochner K, Hummer RA, Bartee S, Wheatcroft G, Cox C. The public-use National Health Interview Survey Linked Mortality Files: Methods of reidentification risk avoidance and comparative analysis. American Journal of Epidemiology. 2008; 168(3):336–344. [PubMed: 18503037]

Lopez-Gonzalez L, Aravena VC, Hummer RA. Immigrant acculturation, gender and health behavior: A research note. Social Forces. 2005; 84(1):581–593.

Markides KS, Coreil J. The health of Hispanics in the southwestern United States: An epidemiologic paradox. Public Health Reports. 1986; 101(3):253–265. [PubMed: 3086917]

Markides KS, Eschbach K. Aging, migration, and mortality: Current status of research on the Hispanic paradox. Journals of Gerontology: Social Sciences. 2005; 60B:68–75. (Special Issue II).

Markides, KS.; Eschbach, K. Hispanic paradox in adult mortality in the United States. In: Rogers, RG.; Crimmins, EM., editors. International Handbook of Adult Mortality. New York: Springer; 2011.

National Center for Health Statistics. The 1986–2000 National Health Interview Survey Linked Mortality Files: Matching methodology. Hyattsville, MD: 2005. <ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/linked_mortality/archived_files/mort_2004_release/matching_methodology_nhis_final.pdf>.

National Center for Health Statistics. NHIS computer file and documentation. Hyattsville, MD: 2006.

National Center for Health Statistics. NHIS 1986–2004 Linked Mortality Files: Analytic guidelines. Hyattsville, MD: 2009a.

National Center for Health Statistics. NHIS 1986–2004 Linked Mortality Files: NHIS mortality linkage eligibility. Hyattsville, MD: 2009b.

Palloni A, Arias E. Paradox lost: Explaining the Hispanic adult mortality advantage. Demography. 2004; 41(3):385–415. [PubMed: 15461007]

Preston SH, Elo IT, Stewart Q. Effects of age misreporting on mortality estimates at older ages. Population Studies. 1999; 53(2):165–177.

Preston, SH.; Taubman, P. Socioeconomic differences in adult mortality and health status. In: Martin, LG.; Preston, SH., editors. Demography of Aging. Washington, DC: National Academy Press; 1994.

Research Triangle Institute. SUDAAN language manual, release 10.0. Research Triangle Park, NC: Research Triangle Institute; 2008.

Rogers, RG.; Hummer, RA.; Nam, CB. Living and dying in the USA: Behavioral, health, and social differentials of adult mortality. San Diego, CA: Academic Press; 2000.

Rosenberg HM, Maurer JD, Sorlie PD, Johnson NJ, MacDorman MF, Hoyert DL, et al. Quality of death rates by race and Hispanic origin: A summary of current research, 1999. National Center for Health Statistics. Vital and Health Statistics. 1999; 2(128):1–13. [PubMed: 10611854]

Rubalcava LN, Teruel GM, Thomas D, Goldman N. The healthy migrant effect: New findings from the Mexican Family Life Survey. American Journal of Public Health. 2008; 98(1):78–84. [PubMed: 18048791]

Sayer B, Cox CS. How many digits in a handshake? National Death Index matching with less than nine digits of the social security number. Joint Statistical Meetings: Section on Survey Research Methods. 2003:3666–3672.

Singh GK, Siahpush M. All-cause and cause-specific mortality of immigrants and native born in the United States. American Journal of Public Health. 2001; 91(3):392–399. [PubMed: 11236403]

Sorlie PD, Backlund E, Johnson NJ, Rogot E. Mortality by Hispanic status in the United States. Journal of the American Medical Association. 1993; 270(20):2464–2468. [PubMed: 8031341]

Turra CM, Elo IT. The impact of salmon bias on the Hispanic mortality advantage: New evidence from Social Security data. Population Research and Policy Review. 2008; 27(5):515–530. [PubMed: 19122882]

Turra CM, Goldman N. Socioeconomic differences in mortality among U.S. adults: Insights into the Hispanic paradox. Journal of Gerontology: Social Sciences. 2007; 62B(3):184–192.

Xu J, Kochanek KD, Murphy SL, Tejada-Vera B. Deaths: Final data for 2007. National Center for Health Statistics. National Vital Statistics Reports. 2010; 58(19):1–136. [PubMed: 25075874]

Table 1

Class Distribution and Mean Score by Hispanic Ethnicity and Nativity, 1989–2002

NCHS-identified deaths

	Hispanic		Non-Hispanic white		
Class[a]	Foreign-born	US-born[b]	Foreign-born[b]	US-born	Total[b]
1	670 (32.5%)	819 (50.0%)	1,651 (57.4%)	36,408 (62.6%)	39,548 (61.1%)
2	311 (15.1%)	332 (20.3%)	500 (17.4%)	8,745 (15.0%)	9,888 (15.3%)
3	610 (29.5%)	335 (20.5%)	629 (21.9%)	11,083 (19.1%)	12,657 (19.6%)
4	473 (22.9%)	153 (9.3%)	96 (3.4%)	1,921 (3.3%)	2,644 (4.1%)
Mean class (SD[c])	2.43 (1.00)	1.89 (0.91)	1.71 (0.94)	1.63 (0.91)	1.67 (0.93)
Mean score (SD[c])	67.48 (17.8)	77.08 (15.7)	79.05 (16.5)	80.22 (15.8)	79.68 (16.1)

NCHS-identified survivors

	Hispanic		Non-Hispanic white		
Class[a]	Foreign-born	US-born[b]	Foreign-born	US-born	Total
2	26 (0.1%)	15 (0.1%)	34 (0.1%)	412 (0.1%)	487 (0.1%)
3	256 (0.8%)	72 (0.3%)	85 (0.4%)	1,893 (0.4%)	2306 (0.4%)
4	8,759 (27.0%)	5,444 (22.8%)	2,884 (12.3%)	74,306 (15.1%)	91,393 (16.0%)
5	23,409 (72.1%)	18,370 (76.9%)	20,491 (87.2%)	414,464 (84.4%)	476,733 (83.5%)
Mean class (SD[c])	4.71 (0.39)	4.76 (0.37)	4.87 (0.38)	4.84 (0.40)	4.83 (0.40)
Mean score (SD[c])	5.63 (8.7)	4.17 (7.5)	2.64 (8.0)	2.86 (7.9)	3.07 (8.0)
Deaths	2,065	1,640	2,875	58,157	64,737
Survivors	32,450	23,901	23,493	491,074	570,918
N	34,515	25,541	26,368	549,231	635,655

[a]No deaths are in Class 5 and no survivors are in Class 1.

[b]Percentages do not sum to 100.0% due to rounding.

[c]SD = standard deviation

Table 2

OLS Regression of Match Score on Hispanic Ethnicity and Covariates for Deceased Respondents, 1989–2002

	Model 1	Model 2	Model 3
Ethnicity			
Hispanic	−7.80***	−6.29***	−2.85***
Non-Hispanic white	ref	ref	ref
Nativity			
Foreign-born		−2.95***	−0.92**
US-born		ref	ref
Hispanic × Foreign-born			−8.09***
Male	1.55***	1.55***	1.54***
Age (five-year groups)			
25–29	−9.52***	−9.43***	−9.26***
30–34	−6.13***	−6.04***	−5.84***
35–39	−4.37***	−4.33***	−4.20***
40–44	−2.49***	−2.45***	−2.37***
45–49	−0.91*	−0.88*	−0.80
50–54	−0.33	−0.33	−0.30
55–59	−0.37	−0.40	−0.37
60–64	ref	ref	ref
65–69	−0.36	−0.33	−0.35
70–74	−0.80**	−0.78**	−0.77**
75–79	−1.40***	−1.37***	−1.37***
80–84	−2.38***	−2.27***	−2.30***
85–89	−2.78***	−2.54***	−2.62***
Intercept	80.73***	80.81***	80.72***
R-squared	0.0261	0.0281	0.0308
Deaths	64,737	64,737	64,737

***
p<.001,

**
p<.05,

*
p<.10

ref = reference group

Table 3

Hazard Ratios for the Association between Ethnicity and Mortality Using Three Linkage Criteria, 1989–2002

	Relaxed	NCHS cut-points	Tightened
Ethnicity and nativity[a]			
FB Hispanic	1.24[***]	0.97	0.78[***]
US Hispanic	1.14[***]	1.10[**]	1.06[*]
FB NH white	0.81[***]	0.78[***]	0.77[***]
US NH white	ref	ref	ref
Male	1.52[***]	1.55[***]	1.58[***]
Age (25–89, continuous)	1.09[***]	1.09[***]	1.10[***]
Deaths	69,272	64,737	61,400
Survivors	566,383	570,918	574,255
−2 [*] LL	541,614.2	507,710.2	482,380.2

[***] p<.001,

[**] p<.05,

[*] p<.10

[a] FB = Foreign-born; US = US-born; NH = Non-Hispanic

ref = reference group

Table 4

Hazard Ratios for the Association between Ethnicity and Mortality, Stratified by Age and Sex, 1989–2002

	Females, age 25–44			Males, age 25–44		
	Relaxed	NCHS cut-points	Tightened	Relaxed	NCHS cut-points	Tightened
Ethnicity and nativity[a]						
FB Hispanic	3.12***	2.35***	1.43***	2.66***	1.87***	1.20*
US Hispanic	1.56***	1.52***	1.39**	1.59***	1.49***	1.37***
FB NH white	0.86	0.88	0.89	0.89	0.75**	0.72**
US NH white	ref	ref	ref	ref	ref	ref
Deaths	2,937	2,173	1,786	4,306	3,341	2,808
Survivors	153,643	154,407	154,794	150,922	151,887	152,420

	Females, age 45–64			Males, age 45–64		
	Relaxed	NCHS cut-points	Tightened	Relaxed	NCHS cut-points	Tightened
Ethnicity and nativity[a]						
FB Hispanic	1.22**	0.99	0.80**	1.14**	0.88**	0.73***
US Hispanic	1.14*	1.10	1.05	1.19**	1.14**	1.10
FB NH white	0.61***	0.55***	0.56***	0.70***	0.67***	0.62***
US NH white	ref	ref	ref	ref	ref	ref
Deaths	7,586	6,804	6,355	10,310	9,418	8,893
Survivors	96,226	97,007	97,457	88,077	88,969	89,494

	Females, age 65–89			Males, age 65–89		
	Relaxed	NCHS cut-points	Tightened	Relaxed	NCHS cut-points	Tightened
Ethnicity and nativity[a]						
FB Hispanic	0.74***	0.70***	0.66***	0.78***	0.72***	0.69***
US Hispanic	1.03	1.00	0.99	0.94	0.94	0.93
FB NH white	0.85***	0.83***	0.83***	0.81***	0.81***	0.80***
US NH white	ref	ref	ref	ref	ref	ref

Deaths	22,861	22,184	21,291	21,272	20,816	20,267
Survivors	46,906	47,583	48,477	30,609	31,065	31,613

***
p<.001,

**
p<.05,

*
p<.10

^{a}FB = Foreign-born; US = US-born; NH = Non-Hispanic

ref = reference group