Center for Demography and Ecology

University of Wisconsin-Madison

A Re-Examination of the Hispanic Mortality Paradox

Alberto Palloni
Elizabeth Arias

CDE Working Paper No. 2003-01

A RE-EXAMINATION OF THE HISPANIC MORTALITY PARADOX

May 30, 2003

Alberto Palloni
Center for Demography and Ecology

and

Elizabeth Arias
National Center for Health Statistics

A Re-Examination of the Hispanic Mortality Paradox

Abstract

We test three competing explanations of the adult "Hispanic mortality paradox:" data artifact, migration, and cultural or social buffering effects. Based on a series of parametric hazard models estimated on 9 years of mortality follow-up data, our results suggest that the "Hispanic" mortality advantage is a feature found only among foreign-born Mexicans and foreign-born Hispanics other than Cubans or Puerto Ricans. Although data are not available to appropriately test the hypotheses, our analysis suggests that the foreign-born Mexican advantage can be attributed to return migration. However, we were unable to account for the mortality advantaged observed among foreign-born Other Hispanics.

I. INTRODUCTION

Research continues to show that Hispanics in the U.S. experience lower mortality levels in adulthood than do non-Hispanic Whites. It is argued that this phenomenon is paradoxical because Hispanics, generally, have lower socioeconomic status than non-Hispanic Whites. The literature on the relationship between socioeconomic status and health and mortality consistently shows that low socioeconomic status is significantly associated with poor health and mortality outcomes among both non-Hispanic Whites and blacks in the United States (Sorlie et al., 1995). It is therefore deemed "paradoxical" that Hispanics have better health and mortality profiles than non-Hispanic Whites, a population with a more favorable socioeconomic composition.

In this paper we review, critically evaluate, and empirically test alternative conjectures associated with this so-called Hispanic paradox. We begin the paper by reviewing key findings in the literature regarding the existence of a paradox and discuss the mechanisms invoked to explain its existence. The two most important among these mechanisms are "social buffering effects" and "migration effects." Social buffering refers to the existence of integrative social networks that may exert a protective effect on the health of Hispanics. These effects can be direct, when social networks contribute to emotional support and alleviation of stress, or indirect, when membership in a group and insertion in a cultural context encourages individual behaviors with beneficial health effects.

Migration effects refer to two distinct phenomena associated with the migration process. First, there is the well-known "healthy migrant" effect according to which, under very general conditions governing non-forced migration, individuals who migrate tend to be selected on a number of characteristics, including good health. Second, there is the so-called "salmon bias" effect. This theory refers to the possible impact on U.S. statistics of return migration of migrants. Return migration of those who are in poor health artificially depresses measures of prevalence of chronic illnesses and death rates. The "salmon bias" effect is more accurately represented as an example of a process of reverse causality.

We use the National Health Interview Survey (NHIS)-Multiple Cause of Death (MCD) data file to evaluate the validity of the previously mentioned claims. The NHIS-MCD data set is a linked file with nine NHIS surveys (1986-1994) matched, using the National Death Index, to mortality data from 1986 to 1997. We further link the NHIS-MCD data file to 1990 census STF1A data by state and county geocode and append contextual-level measures of residential segregation. We then estimate parametric hazard models for the near decade-long mortality experience of the population aged 35 and above. Our ethnicity measure enables us to distinguish between non-Hispanic Whites and several Hispanic sub-groups including Cubans, Mexicans, Puerto Ricans and Other Hispanics. The parametric models account for demographic, socioeconomic and contextual-level determinants of mortality, for interaction effects of theoretical importance, and for unmeasured individual (fixed) heterogeneity.

II. BACKGROUND

Evidence of the Hispanic Mortality Paradox

Studies based on varied data sources, including national and state vital statistics, local surveys and national linked data files such as the National Longitudinal Mortality Study (NLMS) (a series of linked Current Population Surveys with vital statistics mortality data) and the NHIS-MCD, find general, although varying, support for the Hispanic mortality paradox. Invariably, after controlling for socioeconomic characteristics in multivariate analyses of mortality outcome differentials between non-Hispanic Whites and Hispanics, extant studies report that the Hispanic population as a whole fares better in adult all-cause mortality than non-Hispanic Whites. However, some studies have found that the Hispanic mortality advantage varies by age, sex, nativity and Hispanic sub-group. In addition, while most studies that group the various Hispanic sub-groups find an advantage by selected causes of death, such as cardiovascular disease, cancer and pulmonary diseases, they also find that Hispanics fare worse with respect to such health outcomes as diabetes, infectious diseases and chronic liver disease (Sorlie et al., 1993; Rosenwaike, 1987; Markides and Coreil, 1986, Novello et al., 1991; Becker et al., 1988; Liao et al., 1998; Markides, 1983; Rogers et al., 1996).

In an early study based on Hispanic surnames found in Medicare files in the United States, Spencer (1984) identified a pattern whereby death rates among those with Hispanic surnames were 5 to 30 percent lower than non-Hispanic Whites and that the advantage of Hispanic males was somewhat higher than the advantage found among Hispanic females.

Based on the NLMS, Sorlie et al. (1993) find that among adults ages 25 and older, Hispanics fare better than non-Hispanic Whites in all-cause and selected cause-specific mortality outcomes, controlling for family income in Cox proportional hazard models. For all-cause mortality among men ages 25-44, 45-64, and 65 and older, they report hazard ratios of mortality of Hispanics to mortality of non-Hispanics, controlling for age and family income, of .81, .63, and .66, respectively. For women, comparable hazard ratios were reported as .64, .74 and .83. They report similarly low hazard ratios for cardiovascular and cancer deaths (Sorlie et al., 1993).

Two other studies based on the NLMS, but not exclusively focused on the Hispanic population report similar findings (Singh and Siahpush, 2001; 2002). Findings from fuller multivariate Cox proportional hazard models, estimated separately by sex, for persons ages 25 and older and controlling for age, nativity, race/ethnicity, marital status, rural/urban status, education, occupation and income, show that Hispanic men have an all-cause hazard ratio relative to non-Hispanic White men of .68 and women of .67 (Singh and Siahpush, 2001). In similar analyses of all-cause mortality among those ages 25 and older estimated with Cox proportional hazard models, the same authors found that, controlling for age, marital status, rural/urban status, education, family income and race/ethnicity stratified by nativity, U.S.-born Hispanic men had a hazard ratio relative to U.S.-born non-Hispanic white men of .71; U.S.-born Hispanic women a ratio of .79; foreign-born Hispanic men a ratio of .54 and foreign-born Hispanic women a ratio of .55 (Singh and Siahpush, 2002). They report similar results based on age-specific models, including ages 25-64 and 65 and older.

Studies which disaggregate the Hispanic population by national sub-group have reported varying support for the Hispanic mortality paradox. For example, Abraido-Lanza et al., (1999) also using the NLMS and Cox proportional hazard models estimated separately for men and women ages 25 and older and controlling for age, education and family income, report lower hazard ratios for each of four Hispanic subgroups relative to non-Hispanic Whites. For men the ratios are as follows: Mexican (.57), Puerto Rican (.63), Cuban (.53), and Central and South Americans and Other Hispanics (.61). For women the ratios are: Mexican (.60), Puerto Rican (.45), Cuban (.47), Central and South Americans and Other Hispanics (.56). They further report lower hazard ratios for U.S.-born Hispanics (excluding Cubans and Puerto Ricans) in comparison to U.S.-born non-Hispanic Whites by age group and sex and after adjusting for education and family income. For men they report the following ratios by age: 25-44 (.59), 45-64 (.60), 65+ (.62). For women the equivalent figures by age are .49, .65 and .59, respectively.

Another study looking at an individual Hispanic sub-group, Mexican Americans, but based on an earlier release of the NHIS-MCD (1986-1991) data reports similar findings for Mexican Americans. Based on estimates of Cox proportional hazard models that take into account age, marital status, education and family income-to-needs ratio, minority concentration (% Hispanic and % African American) at the tract level and census tract median income, the mortality hazard ratio for Mexican men ages 18 and older relative to non-Hispanic white men was .73. Similarly, Mexican women exhibited a hazard ratio of .75 relative to non-Hispanic white women (LeClere et al., 1997). Interestingly, the authors found that segregation had what appeared to be a protective effect on Mexican Americans, but, as expected, a negative effect on African Americans. Exploring these mechanisms by age, the authors find that the advantage of Mexican American men remains significant only among those ages 65 and older, although at younger ages they show no difference from their non-Hispanic counterparts.

In contrast to the studies reported above, one study based on 1989-1995 NHIS-MCD data files and on discrete-time hazard models, Hummer et al. (1999), reports that Mexican Americans and Other Hispanics do not differ statistically from non-Hispanic Whites in their mortality outcomes after controlling for age, sex, nativity, education, income and marital status. When adding interactions between ethnic group and nativity, they do find that foreign-born Mexicans and foreign-born Other Hispanics have odds ratios of mortality relative to native-born non-Hispanic Whites of .78 and .74, respectively. However, this foreign-born advantage is found only among the middle-aged (45-64) and elderly groups (65 and older). U.S.-born Mexicans and Other Hispanics evidence no such advantage at any age (Hummer et al., 1999).

Based on 1986-1995 NHIS-MCD data, Hummer et al. (2000) explore mortality differentials comparing each of five Hispanic sub-groups (Mexican, Puerto Rican, Cuban, Central/South Americans and Other Hispanics) with non-Hispanic Whites. Using Cox proportional hazards models controlling for age, sex, nativity, marital status, education, family income, employment status, region of country, and metropolitan residence status, the authors find hazard ratios significantly lower for Mexicans (.84), Central/South Americans (.72) and Other

Hispanics (.85) relative to non-Hispanic Whites. They find no statistical difference between Puerto Ricans and Cubans and non-Hispanic Whites. They also find that the Hispanic advantage is mainly found in the older age group (65+) for both men and women and predominantly among Mexican, Central/South Americans and Other Hispanics (Hummer et al., 2000).

Explanations of the Hispanic Health and Mortality Paradox

There are three standard explanations for the observed advantage that Hispanics experience with regard to mortality risks. The first is based on the idea that there are conditions associated with migration into and out of the U.S. that could favor the health profile of the resident migrant population. The second rests on the argument that the Hispanic population enjoys conditions that lead to “cultural or social buffering effects” (cultural effects for short) that imply different behavioral profiles, secure emotional support from social networks, and enhanced self-efficacy. The third explanation suggests that the Hispanic advantage is an illusion produced by data artifacts. Clearly, the first explanation makes more sense for migrants or non-U.S.-born Hispanics whereas the remaining two may be applicable to all Hispanic groups. However, in all three cases, the main outcome is that, after controlling for measurable determinants, Hispanic mortality will appear to be lower than mortality for individuals who belong to other ethnic groups and who have identical measured characteristics.

Migration Effects

Two hypotheses are part of the explanation based on migration effects: the “healthy migrant effect” and the “salmon bias effect.” If the observed difference between Hispanic and non-Hispanic mortality is a result of migration effects, one cannot conclude that there are characteristics – genetic, socially produced or culturally acquired – conferred upon individuals by virtue of their membership in the group, that translate into health advantages and lower mortality risks. This is so because the observed difference between migrants and non-migrants is net of **measured** characteristics. These measured characteristics almost certainly won’t include all those that are relevant to both migration decision-making and to health and mortality.

The “healthy migrant effect” posits that selection of healthy migrants to the United States accounts for the epidemiological paradox (Sorlie et al., 1993; Abraido-Lanza et al., 1999; Palloni and Morenoff, 2001). Hispanic migrants are selected from the origin population for certain traits, including better physical and psychological health. The population of successful migrants is not a random draw from the health distribution of the origin population. On average, migrants will be healthier than those who do not migrate and possibly healthier than the average individual in the receiving population.

The "salmon bias effect" occurs due to a phenomenon experienced by some (non-U.S.-born) Hispanic sub-groups. This is that they have a propensity to return to the country of origin following a period of temporary unemployment and/or illness (Abraido-Lanza et al., 1999). Return migration will result in artificially lower mortality rates for the Hispanic population for one of two reasons. First, death rates are typically calculated for a period of time using denominators estimated from a baseline population and a numerator that includes all relevant events observed during the period. If return migration occurs, a downward bias in the estimated rates will result, **irrespective of whether return migrants are less healthy than those who stay** (Abraida-Lanza et al., 1999). If there is no significant difference between those staying and those leaving, the rates can be corrected by excluding from the initial exposure counts (denominators) all individuals who are known to have left the U.S. We will refer to this as a Type I "return migrant effect or bias" or Type I for short.

Second, to the extent that returning migrants are more likely to be in poor-health and exposed to higher mortality risks than those who stay, the death rates for a given period will be biased downwardly **even if one were able to adjust denominators by excluding those who left the country.** This bias is shared by all data sources without distinction and cannot typically be corrected **even if one is able to identifies return migrants.** We will refer to this as a Type II "return migrant effect or bias" or Type II for short.

Because it provides an identifying clue, it is important to note that the magnitude of the impact associated with healthy migrant and return migrant effects should, in theory at least, vary by age. The return migrant effect (especially Type II) is likely to have heavier salience at older ages, when increases in morbidity augment the population at risk of experiencing return migration. Return migration effects should lead to mortality rates at older ages that are too low.

On the other hand, the healthy migrant effect should be more visible at younger ages, specifically within the age interval over which the age distribution of recent migrants is largely concentrated. This is so because as both migrant and domestic populations age, their composition by frailty will tend to converge. In the limit, at very old ages, all healthy migrant effects should vanish.

Cultural Effect

The main premise of the "cultural" hypothesis is that Hispanics' mortality advantage is a function of social and cultural characteristics that differentiate this population from the non-Hispanic population, rather than selection (Sorlie et al., 1993; Abraida-Lanza et al., 1999; LeClere et al., 1997). It proposes that culture affects mortality outcomes by influencing individual health and life style behaviors, family structure and social networks.

Culture may influence mortality through two types of effects on individual behaviors. First, culture of origin may shape the behavioral risk profile of individuals. For example, diet is closely tied to cultural practices and to dietary patterns, as are the prevalence of smoking, alcohol consumption, habitual exercise, and the use of preventive medical care.

A second type of effect shapes the nature of the social environment of individuals and operates through norms and beliefs about family relationships and obligations. Such norms and beliefs may influence propensities to live alone or in extended families, increase the density of social networks and the amount of social support exchanged and, finally, enhance the sense of control and self-efficacy (Arias, 1998). Although not conclusively proven, it is suspected that health status and mortality are related to individuals' ability to actively participate in social networks, to establish bonds of reciprocal obligations through which they derive emotional and material support and enhance their sense of control (Mendes de Leon and Glass, 2002). This is a plausible explanation as there are apparently strong physiological benefits in the form of dilution of "allostatic" loads that accrue to social integration (Adler and Ostrove, 1999). One would expect that individuals who are more successful in establishing an identity and forging a social and cultural infrastructure will be exposed, *ceteris paribus*, to conditions more conducive to good health and low mortality risk. By contrast, those who fail to establish social linkages may be left in a disadvantaged position.

It follows that a successful accounting of the Hispanic paradox using the cultural explanation must verify the joint occurrence of the following three regularities: (a) other things being equal, Hispanics who share advantageous mortality and health conditions must also share either beneficial behavioral risk profiles or denser social networks and social, emotional and material support than individuals who do not display the advantage, (b) Hispanics who are not well integrated into social networks and who receive less social support will experience higher exposure to health and mortality risks and will not share the advantage benefitting other members of the same ethnic group and, finally, (c) the advantage should must fade away with increasing assimilation to the receiving country if the latter implies either acquisition of a less healthy behavioral profile or the abandonment of norms and behaviors securing social support.

Data Artifacts

There are three equally salient data problems that may lead to the appearance of a Hispanic advantage. The first two of these are shared by all data sets where mortality (or disease prevalence) rates require ethnic self-identification and self-reporting of ages. The third one is only pertinent to data sets, such as the one we use in this paper, where rates are calculated by matching death and survey records.

Ethnic identification: The first source of error is associated with ethnic identification. Under-reporting of Hispanic origin on U.S. death certificates, which are the prime data collection tool for mortality statistics, is a significant problem. Mortality rates in the U.S. are based on two distinct data sources: the vital statistics system (numerator) and the census population enumeration (denominator). Ethnic identification in the denominator is usually and mainly "self-identification." Individuals answering the census forms choose their own ethnic characterization or a family member in the household does; even in the latter case the person is alive and most likely provides input into what is put down on the form. On the other hand, ethnicity in the numerator is reported by someone other than the decedent; sometimes the reporting person may be a family member or friend, but often the funeral director makes the final assignment by observation. Based on analyses of the NLMS, it has been estimated that about 7% of Hispanic individuals do not get recorded as Hispanic on the death certificate when they die (Rosenberg et al., 1999). Admittedly, this figure is subject to errors associated with the matching algorithm used in the NLMS. The figure is also somewhat crude since it does not reveal to us whether foreign-born Hispanics are more or less prone to be erroneously classified in the death certificate than their counterparts born in the U.S.

The incongruence between classification of Hispanic origin in numerators and denominators leads to artificially low death rates for Hispanics relative to non-Hispanics and to the appearance of an advantage. In the absence of pertinent information we assume that the bias calculated by Rosenberg and colleagues is independent of other social and economic characteristics related to mortality (including nativity). If we take seriously the possibility of this type of error (Rosenberg et al., 1999; Guend and Swallen, 2002), the observed relative risk of Hispanics should be adjusted upward by an amount close to * before one can speak of a Hispanic advantage. This notwithstanding, the data set we use in this paper is not subject to this problem since the ethnic identification that prevails is not that found in the death certificate but the one declared by the individual at the time of the survey. The estimated mortality rates we calculate from our data set will be biased only if there is a relation between the health of an individual and the propensity to provide an erroneous self-identification, a possible but unlikely phenomenon.

<u>Age misreporting:</u> A second source of data artifact is associated with age misreporting. It has been shown that some populations in Latin America (Dechter and Preston, 1991) and some Hispanic subgroups in the U.S. (Rosenwaike and Preston, 1984) manifest a tendency to overstate age, particularly among those who are older (above 55 or 60). Net age overstatement in the population will depress mortality rates at older ages, thus producing the illusion of a more benign mortality pattern. A similar outcome occurs when age overstatement affects the age distribution of deaths. When both take place, namely, when age is overstated in the population and among the deceased, there will be offsetting errors but, typically, the net effect will be to bias downward the death rates. In our data set age at death is derived from the age at the baseline survey and the **date of death**. Thus, age at death will be overstated if and only if age at the survey is also overstated. Even if there is age overstatement in the survey–and this is less likely to be so than in conventional censuses–the absolute value of the downward bias in the mortality rates for Hispanics will be lower than if rates were calculated conventionally from census and vital statistics.

<u>Mismatches</u>: The third source of error that could lead to downward biases in the mortality rates of some ethnic groups only applies to data sets where mortality rates are constructed by matching

deaths that occurred during an interval of time to populations enumerated at the beginning of the interval. The success rate of the matching procedure that links a death record to a population record is variable across individuals and groups. Matching rates are usually associated with missing or erroneous information in some of the key characteristics that make feasible the matching algorithm, such as last names or social security numbers. Although there are no good sources documenting differential matching rates by ethnic groups, it is suspected that they could be lower in populations whose identification via universal identifiers is more difficult to obtain reliably due to legal status. The result of this is to impart a downward bias to mortality rates of ethnic groups that are more heavily composed of individuals whose legal status may be questionable and whose identifiers are likely to be less complete or less reliably recorded.

Although all three sources of error described above could lead to the illusion of a Hispanic advantage because they produce downward biases in mortality rates of some Hispanic groups, only the second source inevitably produces errors that are proportionately higher at older ages. There is no compelling argument or evidence to support the idea that ethnic <u>mis-identification</u> is more likely to occur at older ages or, more generally, that is has a differential impact by age, whatever the pattern may be. If mismatches of records are more likely to occur among undocumented migrants, then it stands to reason that this source of error is more likely to induce biases that are proportionately higher at younger than at older ages. These expected patterns of errors provide a guide, albeit a tenuous one, in the interpretation of the data.

III. DATA SOURCE AND STUDY DESIGN

Data Source

The National Health Interview Survey, undertaken in the United States annually since 1957, is a nationally representative multistage probability sample of the civilian noninstitutionalized American population. The 1985-1994 design obtained information from approximately 49,000 households, including about 132,000 individuals per year, through personal interviews. The annual response rate was over 95 percent. The 1985-1994 NHIS design consisted of two basic parts: (a) a core questionnaire containing basic demographic and health questions; and, (b) one or more modules with questions related to current health topics. The basic or core questionnaire was repeated yearly. Questions included demographic, socioeconomic and health status characteristics such as age, sex, race, education, family income and self-assessed health status. The core also included a set of questions about disability, physician visits, chronic conditions and hospital stays. The special modules with questions on specific health topics changed yearly, and included such topics as alcohol use, smoking, health care and health insurance (NCHS, 1989).

Beginning in 1986 linkage information for NHIS respondents ages 18 and older was collected in order to match the NHIS individual records with other data systems, including the National Death Index (NDI). The NHIS record is linked to the NDI based on a series of combinations of 12 identifiers, including social security number, first and last name, father's surname, and month and year of birth. Once linkage is made and vital status ascertained the records are linked to the national vital statistics multiple cause of death data, resulting in the National Health Interview Survey-Multiple Cause of Death file. Mortality information is available for the NHIS surveys beginning with 1986 and ending with 1994 for deaths through December 31, 1997 (NCHS, 2000).

In addition to vital status, county level information from the 1990 Census was linked to the NHIS for 1986 to 1994. To facilitate our analyses, we use a ten percent random sample of the non-Hispanic white population in combination with the full sample of the Hispanic sub-groups. Since survey years 1986 through 1988 did not include a variable for nativity or duration of residence in the United States, an important variable in our analyses, we use the 1989-1994 NHIS with 1989-1997 mortality follow-up. The population exposed as well as the frequencies of relevant events for selected subgroups are listed in Appendix A. Note that, despite including the entire sample, all Hispanic subgroups but particularly Other Hispanics, are relatively small and the number of events of interest infrequent. This could reduce the stability of some estimates.

Basic model and enhancements

The data set described above enables us to estimate adult mortality over a nine year period for a population aged 35 and older at baseline. Because of the design of the study we can only assess the effects on individual mortality risks of characteristics elicited at the outset, and are not in a position to evaluate changes in those characteristics and the effects of these changes. Our plan of analysis consists of using estimates from standard parametric hazard models to (a) assess the existence and magnitude of a mortality advantage among both foreign and U.S.-born Hispanics, and (b) identify the mechanisms that may produce it. Both of these tasks are complicated because data artifacts and various mechanisms genuinely producing an advantage can lead to similar observable patterns. Throughout we search for identifying signals that help us separate the contribution of artifacts and other mechanisms involved. In the following section we formulate a model to represent mortality patterns, estimate the magnitude of the Hispanic advantage and falsify hypotheses regarding the mechanisms producing it.

We employ a standard parametric hazard model to estimate effects on mortality for the decade-long follow up period for individuals aged 35 and above at the time of the baseline survey.[1] We assume that a Gompertz model represents well the profile of mortality increase for ages (approximately) 35 and above. Throughout, we rescale age to be the difference between age at the onset of the study and 35 so that Gompertz's constant will refer to the mortality rate at age 35. With this modification, the hazard rate t years into the study can be expressed as follows:

$$\mu_i(t \mid X_i; \mathbf{Z}_i) = \mu_o(X_i - 35 + t)\exp(\beta \mathbf{Z}_i) = \alpha \exp(\gamma (X_i - 35 + t)) \exp(\beta \mathbf{Z}_i) \quad (1a)$$

or, alternatively,

$$\mu_i(t \mid X_i; \mathbf{Z}_i) = \alpha \exp(\gamma t) \exp(\gamma (X_i - 35)) \exp(\beta \mathbf{Z}_i) \quad (1b)$$

where $\mu_i(t \mid X_i; \mathbf{Z}_i)$ is the hazard rate t years after the onset of the study for an individual i aged X_i at the outset who is characterized by a vector of attributes $\mathbf{Z}_i$. The expression $\alpha * \exp(\gamma (X_i - 35 + t))$ in (1a) above refers for the standard Gompertz mortality rate evaluated at age $X_i + t$, β is a vector of effect parameters, α is Gompertz's constant scaled to represent the mortality rate at age 35, and γ is Gompertz's ancillary parameter or the slope of the hazard rates above age 35.

[1] Our initial intention was to study the mortality experience at ages above 40. Since individuals who are younger than 40 at baseline will contribute variable amounts of exposure at ages 40 and above during the follow up period, we opted for a compromise solution and included individuals who at baseline were aged 35 and older.

It is clear from expression (1b) that the parameter for age (re-scaled) need not be equal to (. It is only because (1b) is a re-expression of (1a) that this equality is necessary. In fact, one could think of (1b) as the constrained version of the more general model (1c):

$$:_{i} (t \mid X_{i}; \mathbf{Z}_{i}) = \text{"} \exp ((t) \exp(* (X_{i} - 35)) \exp(\$ \mathbf{Z}_{i}) \qquad (1c)$$

But while (1c) is plausible, only (1b) is compatible with the Gompertz model in (1a). This suggests a test to validate (1a): if an estimated model that constrains * to equal (does not fit the data as well as a model that leaves the parameters unconstrained, we have *prima facie* evidence suggesting either that the underlying hazard cannot be reproduced by the Gompertz model in (1a) or that the effects of age and/or duration in the study are biased due to measurement errors. We will return to this issue in the section on estimation.

In addition to verifying that the constrained form of the model (expression 1b) is acceptable, we must ensure that the estimated values of " and (are within the expected range: " must be close to the observed mortality rate in the non-Black population (non-Hispanic Whites and Hispanics) at age 35 and (must fall within the known range for a population such as that of the entire U.S. (between .06 and .12).

To test some of our hypotheses the vector **Z** will include variables reflecting ethnic group, marital status, socioeconomic characteristics, nativity and duration, state of residence and a constructed index of ethnic isolation. In addition, when required, we will include suitable interaction terms. Appendices A through C contain a full description of variables as well as descriptive statistics in the sample.

Finally, investigation of healthy migrant and return migration effects as well as identification of data artifacts will be pursued by generalizing (1a) in three different ways, namely, by assuming that the effects of age are not invariant over the age span, by formulating a model where the slope is a function of covariates and, finally, by posing the existence of unmeasured frailty (unmeasured heterogeneity).

IV. RESULTS

Mortality profiles

Figures 1a through 1d are plots of logs of observed mortality rates among males in 5-year age groups at baseline.[2] These figures are cohort rates cumulated (over the nine year period of follow up), or the ratio of events (deaths over the nine year follow up) to person-years lived by the exposed set (individuals aged x, x+5 at baseline). These rates are similar but **not** necessarily identical to the age-group specific mortality rates observed during the follow up. Figures 1a and 1b consider ethnicity but not nativity whereas Figures 1c and 1d account for both. Overall, Puerto Ricans and Cubans exhibit mortality rates that are slightly worse than the non-Hispanic White population while the rates for Mexicans are virtually identical to those of non-Hispanic Whites. Rates for Other Hispanics are considerably lower, especially at younger ages. Figures 1c and 1d show that disparities between U.S.-born and foreign-born Mexicans and Other Hispanics are significant. The differences between Puerto Ricans[3] and Cubans (U.S.- and non U.S.-born) are unimportant and are not plotted. Foreign-born Mexicans have considerably lower mortality than both U.S.-born Mexicans and non-Hispanic Whites. The differences are particularly salient at older ages, as would be expected if Type II biases (or considerable age overstatement) were of some importance. Both foreign-born and U.S.-born Other Hispanics exhibit lower mortality than non-Hispanic Whites but the differences are especially large among the foreign-born at **any age**.

In summary, on first blush the Hispanic advantage of adult mortality is not uniform across all Hispanic groups. It is especially prominent among Other Hispanics and less so among Mexicans. Furthermore, it is much more salient among Mexicans and Other Hispanics born abroad than among those born in the U.S. This cursory description suggests that Hispanic group of membership and nativity are both important characteristics and need to be considered explicitly in all models.

Although simple to calculate, the rates in Figures 1a-1d are somewhat complex functions

[2] With caveats reviewed below, the figures for females are very similar and are not displayed.

[3] Puerto Ricans are U.S. citizens whether born in the U.S. mainland or on the Island of Puerto Rico. In this study, the term ‘foreign-born’ with respect to Puerto Rico refers to Puerto Ricans born on the Island of Puerto Rico.

of the age-duration specific hazard rates defined by expression (1a). Plots of the latter hazard rates analogous to those in Figures 1a and 1d are difficult to produce efficiently since populations for all five year age groups above age 35 are associated with five different sets of nine-year hazards each (one hazard curve per ethnic group). Furthermore, those rates are hard to represent graphically without reflecting the noise associated with low frequencies of events. For these reasons we only show the log of the hazard rates for both sexes combined and for individuals who, at the outset, were in one of two coarse age groups, 35-64 and 65+ . To avoid cluttering the figures we display the plots for only three groups, namely, non-Hispanic Whites, Mexicans and Other Hispanics. The plots for Puerto Ricans and Cubans did not reveal distinct features worthy of note and are omitted.[4] Figures 2a through 2d display the hazard rates by duration in the study for the combined male-female population, for non-Hispanic Whites as well as for Mexicans and Other Hispanics separately by nativity.

Despite all our concessions to smoothness, the estimated rates are jumpy and with non-trivial standard errors (not shown). These graphs are not a textbook example of what proportionality of hazards should be, but this is in part due to the coarse control for age at the outset. Figures 2a and 2b show that foreign-born Mexicans have a distinct mortality advantage especially at higher durations. They also suggest that, consistent with the existence of Type II biases (and age overstatement), the advantage is larger in the older age group. Figures 2c and 2d show that the absolute magnitude of the advantage is much larger among Other Hispanics than among Mexicans, more a characteristic of the foreign-born than the U.S.-born population and, finally, that the advantage for foreign-born Other Hispanics does not follow a clear age pattern as is the case among foreign-born Mexicans.

Alternative baseline models for males and females

Tables 1a-1b and Table 2 display key statistics for fitted models in the male and female samples respectively. These are all additive models following expressions (1a) and (1b) and covering a broad range of specifications, stretching from the simplest one (null model and

[4] The resulting rates will not reflect differences in mortality intensity only but will obviously include the influence of the age distributions of each group within the two selected age groups. This is the price we must pay to ensure a sufficiently large number of events and exposure for robust estimation. However, the price is modest as these compositional effects should be small and can be safely neglected.

baseline hazard model) to the most complex (including measures of employment). In light of the diagnostics made before from graphic representation, we use refined variables for ethnicity that enable us to discriminate between Hispanics by nativity. We first estimate the most parsimonious model, one where the effects of nativity are zero (Model 2 in column 3, Tables 1a and 1b). Since we use non-Hispanic Whites as the reference category, this model has only 4 free parameters other than those associated with the baseline hazard. The second model (Model 3 in column 4 of Tables 1a and 1b) is one where we allow effects of nativity but force them to be identical across all Hispanic groups. This model has five free parameters. Finally, the least parsimonious model (Model 4 in column 5 of Tables 1a and 1b) is one where each Hispanic-nativity subgroup is assumed to have its own mortality level. This model requires a total of 8 parameters (four Hispanic groups and two nativity categories) in addition to those of the baseline hazard.

Two tests are pertinent. The first compares Models 2 and 3 and the second compares Models 3 and 4. Since these are contrasts involving nested models we rely on conventional chi-square statistics to assess goodness of fit. The first contrast yields a chi-square statistic of 8.8 for males and of 14.8 for females with one degree of freedom (column 4, last row in Tables 1a and 1b). The second test yields a chi-square statistic equal to 13.0 for males and 12.9 for females with 3 degrees of freedom (column 5, last row of Tables 1a and 1b). These values are statistically significant with $p<.001$ and thus justify choosing the most complex model from each contrast. In other words, we have grounds to conclude that (a) the nativity effects are indeed important (first contrast) and (b) that these effects are specific to each Hispanic group (second contrast).

Two additional features of Tables 1a and 1b are worth mentioning. First, according to Model 1 (including only baseline hazard parameters in column 2) mortality rates for all males in the age group 35-39 ought to be close to $\exp(\alpha + 2.5 * \beta)$ or .0017. The estimate for females from Table 1b is 0012. These can be compared with observed values in the national life tables for the White population for 1991 (a year approximately midway through the follow-up period), .0022 and .0010 for males and females respectively. The correspondence is not perfect but quite acceptable for both genders. A second feature is that the slope parameter for males and females (β~.086) falls nicely in the middle of the range for a population like the U.S. Additionally, among males a model that does not constrain the effects of the variable Delta = $(X_i - 35)$ to be identical to (, the Gompertz slope parameter, (Model 5 in column 6 of Tables 1a and 1b) performs just as well as a model that imposes the constraint (Model 4 in column 5). This is an *ad hoc* indication that the functional form imposed on the data captures satisfactorily the age effects on the hazards. The same is not true for the female sample where the fit is slightly better when age effects and slope are not constrained to be identical. All other estimates, including the slope, are insensitive to the specification of the constraint on the parameter for *. As we discuss later, the fact that the constrained and unconstrained model perform equally well not only suggest that the Gompertz formulation is quite reasonable but helps rule out partially the possibility that estimates of ethnicity are contaminated by age overstatement.

The preferred models in Tables 1a and 1b (column 5) lead to four clear-cut inferences already anticipated by our cursory examination of graphic patterns. The first is that the Hispanic advantage is a trait found among Mexicans and Other Hispanics. It is neither a feature of Cubans nor of Puerto Ricans. The second inference is that the advantage is characteristic of foreign-born Mexicans and Other Hispanics, not of U.S.-born individuals. Note that according to the fitted model for males, the overall mortality level among non-U.S.-born Mexicans is approximately equal to exp (-.26) or 77 % as high as the mortality level associated with non-Hispanic Whites, whereas the level for Other Hispanics is even lower, about exp(-.61) or 54% as high. Both these effects are statistically significant. Instead, mortality levels of Puerto Ricans, Cubans and either U.S.- born Mexicans or Other Hispanics are not significantly different from non-Hispanic White mortality rates. The same patterns apply to females but with one notable exception: the advantage of foreign-born Mexican women is of lower magnitude and not at all statistically significant (t statistic is about 1.3). The third and final inference is that, without exceptions, the numerical value of the relative advantage is much larger among Other Hispanics than among Mexicans.

But Tables 1a and 1b are limiting since the advantage remains ill-defined when it is not associated with Hispanic populations that are **comparable** to the baseline population. It is in models that control for relevant compositional factors where we must search for an advantage, identify the groups affected, and assess its approximate magnitude. The results of these models are in Model 1 (columns 1 and 3) of Table 2 for males and females respectively. This model includes controls for demographic and socio-economic conditions that are known to influence adult mortality. The log likelihood ratio test statistics comparing this model with Model 4 in Tables 1a and 1b (displayed at the bottom of Table 2 in columns 1 and 3) suggests a much better fit for both males and females. Furthermore, the estimates reveal a picture of such remarkable consistency and regularity that one can summarize the main results with four statements:

i. Invariance: Introducing marital status or any other factor designed to control for socioeconomic conditions, such as education, income or employment, does not alter the patterns revealed before. That is, the adult mortality advantage is shared by foreign-born Mexicans and Other Hispanics, but by nobody else. If anything, the magnitude of the advantage is slightly increased in models with extensive controls. In fact, according to Model 4 Mexican males experience mortality levels that are exp(-.43) or 65% as large as non-Hispanic White males (rather than 77% from models in Table 1a), whereas foreign-born Other Hispanic males have mortality levels that are only exp(-.63) or 53% as high (rather than 55% from models in Table 1a). These are slightly larger differences than in the baseline models of Tables 1a and 1b.[5]

ii. Regularity: All factors introduced as controls behave as expected, are properly signed, and exert influences that are statistically significant. Thus, for example, the impact of marital status is powerful and, as expected, indicates that single individuals experience considerably higher mortality than married ones. The same applies to education, income and employment.

iii. Minimal gender contrasts: the pattern of relations for females is not identical to that of males but it is remarkably similar. There are two features that make the female pattern distinct. One is the lack of strength of the advantage among foreign-born Mexican females (at least when compared with males). In fact, although the estimated effect is negative, the corresponding t-statistic is only 1.92, somewhat below the threshold value of 2.0 we use here to allocate statistical relevance. The other feature visible only in the model for females is that there is a positive effect on mortality for U.S.-born Cubans. But, here again, the estimated parameter does not quite reach the significant threshold value. All other observed patterns are concordant with those found among males.

[5] The figures from Tables 1a and 1b are 77% and 55% respectively.

iv. General invariance of constrained and unconstrained models: Tests contrasting models where the effects of age and slope are constrained to be equal to each other with models where the constraint is absent indicate that the constrained model is preferable. Even the exception among models without controls (for females) disappears.[6]

Statistically significant differences constitute important raw materials for testing theories, but may have remarkably little influence in the life of actual individuals. What does the advantage detected in these models really amount to? In order to offer an easily interpretable metric, we translate the differences in mortality rates between foreign-born Mexicans and Other Hispanics, on the one hand, and non-Hispanic Whites, on the other, into differences in life expectancies. Figure 3 displays the predicted life expectancies at ages 45 and above for males who, at the onset of the study, were 45 years old. The apparently innocuous effect of -.62 for foreign-born Other Hispanics (see above) translates into a difference in life expectancy at age 45 of about eight (7.31) years. For foreign-born Mexicans the estimated effect of -.43 translates into an excess of life expectancy of about 6 (4.74) years. For females (figure not shown) the differences are slightly lower. Since male life expectancy at age 45 among Whites in the U.S. is roughly 39.7, the estimated relative risks translate into an advantage in years of life expectancy at age 45 amounting to 18.5 and 11.9 percent for foreign-born Other Hispanics and Mexicans respectively. For females the relative advantage is on the order of 16 and 9 percent respectively.[7]

Are these differences plausible? To place these contrasts in perspective, we also plot in Figure 3 the residual life expectancies in the Coale-Demeny life tables with the highest life expectancy at birth (80 years) in the female life tables. This also happens to be the life expectancy at birth among females in the U.S. during 2000 (Arias, 2002). Note that against this

---

[6] The chi-squared statistics for contrasts associated with models (1) and (2) for males and females in Table 2 are not reported there but they attain values that are small and that do not enable us to reject the null hypotheses that the slope and age effects are identical.

[7] Under a Gompertz function there is an approximate correspondence between relative risks and life expectancy at age 45 that can be expressed as follows: $\ln(E(45)) \sim 3.92 - .99 * RR$, where RR is the relative risk and E(45) is the life expectancy at age 45. When the estimated coefficient of a 0/1 variable is 0, the RR will be 1 and the value of E(45) about 18.7. When the estimated effect of a 0/1 variable is -.60, the associated RR is .55 and E(45) will be roughly 29.23 or about 10 years more than the baseline group. These are all approximations based on a range of values of mortality rates at age 35 not exceeding .0050 but not lower than .0010.

extreme standard the advantage for measured foreign-born Mexicans and Other Hispanics is considerable. The implied differences in life expectancy at age 80, for example, are 7.5 to 9.3 years respectively. These are non-trivial differences, even when considered against the backdrop of a contrast between the non-Hispanic White population and the life table for females (Model West) of about 4.5 years. Are these differences too large and perhaps the product of data artifacts? We now discuss the role that data artifacts may play in our estimates.

Assessing the plausibility of the estimated advantage

The most obvious culprit behind excessively large life expectancies is an across the board downward bias in the estimated mortality rates. As argued before, there are three potential sources for this type of bias. We examine each in turn.

Ethnic misidentification

The very nature of our data set rules out the inconvenience of numerator/denominator biases associated with inconsistent ethnic identification. Ethnic categorization is derived from the baseline NHIS *self-identification* rather than from the death certificate's proxy response, where the bulk of errors and inconsistency are believed to be rooted. However, but very unlikely, when ethnic misidentification is characteristic of the survey a bias will exist if such misclassification is more likely among those who are in worse health or have higher risk of death. Though possible, this scenario is admittedly somewhat farfetched. And yet we have no direct evidence against the existence of the associated flaw. We can say though that we have no knowledge of even a speculation suggesting that ethnic self-reporting is a function of health status and are not aware of any data supporting the speculation (Sandefur et al., 2002).

Mismatching records

A different but related source of error may originate from imperfect, incomplete or impossible matches. The NHIS-NDI matching algorithm has been shown to be highly reliable with an overall 98% of successful matches based on analyses of known decedents from the active follow-up of the National Health And Nutrition Examination Survey (NHANES) I cohort. However, this cohort contained very few Hispanics and there is no way to assess the quality of the algorithm for this population. As a result, the possibility for error due to erroneous matches or mismatches cannot be ignored (NCHS, 2000).

Mismatching death and survey records can indeed lead to downwardly biased death rates. All one needs is to have a sufficiently high number of deaths that cannot be matched to records of live persons. Furthermore, if mismatches occur because of faulty and/or lack of relevant identifiers and this is more likely to occur among populations whose legal status is doubtful, then we could begin to suspect that the estimated advantage of foreign-born Mexicans and Other Hispanics is indeed an illusion. For this to be the explanation of the advantage, though, the matching success rate of death and survey records would have to **be at most** 77 percent among foreign-born Mexicans and at most 54 percent among foreign-born Other Hispanics. These are remarkably poor success rates for a well-tested matching algorithm. Unlikely, but not impossible since with a small population base it would take only a few hundred mismatches to produce such large downward biases. However, as we show later, what makes this explanation questionable is that to account fully for the advantage among foreign-born Mexicans at least, the rate of mismatches should increase with age, a pattern inconsistent with the idea that illegal residence is at the root of faulty matching to begin with. And if not among foreign-born Mexicans, those most likely to count illegal migrants among their ranks, why should the explanation based on mismatches hold for foreign-born Other Hispanics?

Age overstatement

Age overstatement of the baseline population would indeed lead to downward biases of mortality rates and to the mirage of an advantage. However, there are two reasons why this explanation is also unlikely to hold. The first is that the pattern of age overstatement has been found among Hispanic populations **in general** and, in the U.S. at least, among Puerto Ricans in particular (Dechter and Preston, 1991; Rosenwaike and Preston, 1984), not just among foreign-born Hispanics. If age overstatement is indeed the culprit, why is it that we do not find an advantage among other Hispanic groups?

The second reason that undermines this explanation is more technical and relates to the constrained estimation of the coefficients of age and Gompertz's slope. If there were extensive age overstatement, there should be a **downward bias on rates inducing proportionately higher errors at older ages**. This must be reflected in estimated effects of age that do not mirror the passage of time so that the effects on mortality of a one year difference in age is less than the effect of a one year difference in exposure to mortality (controlling for age). Said otherwise, age overstatement leads to an estimate of the coefficient of age that is lower than the estimate of the slope invalidating the use of a Gompertz function. With a singular exception (see Table 1b) the constrained version of the model fits as well as the unconstrained one. The exception is the model for females, an odd finding indeed since age overstatement in the Hispanic populations referred to above is seemingly more serious among males than it is among females!

In sum, while we cannot rule out completely that the estimated advantage for foreign-born Mexicans and other Hispanics is due entirely to artifact, we find all three possible sources of errors empirically unlikely and inconsistent with observable features of the mortality patterns. There is still some explaining to do.

Choosing a parsimonious baseline model

The tenor of our arguments based on examination of Tables 1a, 1b and 2 suggests two additional simplifications may be possible. First, the representation of ethnic-nativity classes within the Hispanic population can be significantly reduced. In light of the fact that what matters most is the advantage of foreign-born Mexicans and Other Hispanics, we could estimate a model that contains five fewer parameters than Model 1 in Table 2 by constraining Puerto Ricans, Cubans, and U.S.-born Mexicans and Other Hispanics to have the same effect on mortality levels. Model 2 (columns 2 and 4) in Table 2 contains the corresponding estimates. Three features of this model are worth mentioning. First, it fits as well as Model 1 but uses five fewer parameters. Second, the effects of the new combined Hispanic group are negative for both males and females but only marginally significant by our standards. Finally, none of the other estimates experience changes worthy of note. In sum, a constrained model such as Model 2 is our best starting point and it is from this model that we calculate the absolute magnitude of the advantage for the two Hispanic groups of interest.

The second simplification is that we have a good basis and justification for using a pooled male-female sample. Indeed, it greatly simplifies our task to focus estimation and discussion on the male-female combined sample rather than moving back and forth between male models and female models. The combined sample will also provide us with additional power to perform tests that must rely on subgroups that, by nature, are very small and experience only a small number of events. Using the combined sample is not just a simplifying shortcut: it is amply justified by our previous findings according to which there are no appreciable male-female differences in the pattern of relations. Further statistical confirmation that this is indeed a wise and thoroughly justified choice will be presented below.[8]

Tracing the roots of the advantage

In what follows we perform a number of tests to identify the mechanisms producing the Hispanic advantage observed in the baseline models. We use a simplified representation of ethnic-nativity groups identical to the one in Model 2 of Table 2, and combine the male and female samples. To account for gender differences in mortality levels we include an additive term for gender.

Rules for choosing among competing models

The strategy we employ below to choose among alternative representations of the data is based on using conventional p-values and log likelihood ratio statistics combined with the Bayesian Information Criterion or BIC (Raftery, 1996). Only when we are reasonably certain about the performance (fit) of a model do we launch into an assessment of pertinent t-statistics and related inferences. Information on estimates, standard errors and associated log likelihood ratio statistics appears in Tables 3a - 3c. Table 4 summarizes all the necessary quantities for the calculation of log likelihood ratio statistics and BIC for all relevant models. The estimates for a few of the models included in Table 4 are not shown in other tables in the paper since they are less important substantively. Pertinent statistics are displayed in Table 4 for the sake of completeness.

Before proceeding with the discussion of our tests, we need to comment on some features of Table 4. First, all likelihood ratio statistics and corresponding BIC values for the models are calculated with reference to the null model, one where mortality is represented by one and only one parameter. Second, a comparison of BIC values for Models 5 and 6 suggests a representation including separate effects for three Hispanic groups only (foreign-born Mexicans,

[8] Since at the outset we suspected that data artifacts as well as potential mechanisms producing advantages worked differently for males and females, all models were estimated separately by sex. But the results are not consistent and the fit of corresponding models not strong enough to justify separate estimation.

foreign-born

Other Hispanics and all Other Hispanics). Third, our choice of a pooled male and female sample is further justified by a comparison of Models 8 and 9 in Table 4. Model 8 includes only the additive effects of gender whereas Model 9 also includes interaction terms between gender and the two variables for foreign-born Mexican and Other Hispanic. If there were important gender differences in relations implicating ethnicity, Model 9 should fit better. As expected from previous discussion, it does not.

In what follows we evaluate a few tests to discard alternative explanations using the combined information in Tables 3 and 4. We first introduce tests designed to detect healthy migrant effects. We follow these with tests to identify the combined occurrence of Type I and II effects. Finally, we evaluate hypotheses regarding cultural effects.

Identifying a healthy migrant effect: the role of duration of stay

Through simulations, Palloni and Morenoff (2001) show that health selection among migrants can have large, potent effects. Mechanisms through which those selection effects can take place are illustrated by Jasso et al. (2002). This means that an apparently sizeable advantage may be completely attributable to initial differences in health status conditions of the populations being compared. If so, alternative explanations invoking cultural and social effects are misguided. Our baseline models suggest that the male and female advantages are far from trivial, as adult mortality rates among foreign-born Mexicans and Other Hispanics are between 35 and 47 percent lower than among non-Hispanic Whites. The corresponding relative risks translate into additional years of life expectancy at age 45 of between approximately five and nine years of life. The work by Palloni and Morenoff suggests that contrasts as large as and even larger than these can be created by healthy migrant effects. But if this is so, it follows that healthy migrant effects must leave at least some observable imprint in mortality data.

We argue below that the first imprint is a necessary convergence of mortality of migrants and non-migrants by duration of stay in the U.S. The second imprint should be a sharp contrast between the mortality experiences of migrants residing in different areas in the U.S.

Convergence of mortality rates for migrants and non-migrants as duration of stay increases will occur as a result of two very different processes.

Assimilation: Assume the existence of an advantage at the lowest duration of stay. This can be produced by migrant health selection or by beneficial cultural endowments. Increasing similarity of mortality rates for migrants and non-migrants may take place because the former group progressively adopts a more adverse profile of risk exposure – shaped by social, cultural and behavioral factors – that resembles that of the non-migrant population. Assimilation implies jettisoning favorable traits and adopting new ones in a tradeoff with negative net health benefits for the migrant population in the following sense: had migrants preserved the original traits, their mortality levels would remain below those of the non-migrant population, except at very old ages. In addition to assimilation, though, conditions associated with the migration experience *per se* such as added stress, poor access to health care, and poor familiarity with sources of assistance, contribute to less favorable health and mortality profiles of the migrant population as duration of stay increases.

Healthy migrant effect: Suppose now that the initial advantage of the migrant population is exclusively a function of health selection. The migrant population is healthier, on average, than the reference domestic population and than the population at origin. To simplify the argument, suppose also that the country of destination has a more favorable mortality regime and that migration does not lead to changes in migrants' risk profiles, that is, that there are no significant negative effects associated with assimilation or with the migration experience *per se*. What should be the patterns of contrast between the migrant and domestic population by duration of stay while age effects are held constant? The answer hinges on a rather subtle feature best illustrated with an example. Assume two migrants identically aged but with different duration of stay in the U.S. The key difference between them is that they migrated at different ages. By assumption, the individual who migrated at a younger age (longer duration) has experienced longer the more benign mortality regime of the country of destination and is, therefore, less selected for health related traits than the one who migrated at an older age. The consequence of this is that the effect of duration of stay should be to attenuate the advantage experienced at lower durations; as in the case of assimilation, the advantage should be diluted as duration of stay increases. Thus, the good news is that a decreasing advantage with duration of stay should be observable if healthy migrant effects prevail. The bad news is that the same pattern is compatible with an explanation that **does not necessarily require** health selection.

All this is moot if there are no detectable effects of duration of stay in the U.S. following the expected pattern. Table 3a provides information to help us make a judgement on this count.[9]

The first column of Table 3a displays the estimated effects associated with foreign-born Mexicans and Other Hispanics in our baseline model, that is, a model including only ethnicity and controls for socioeconomic traits. The second column of Table 3a displays estimates of effects of dummy variables reflecting ethnicity (Mexican and Other Hispanic) and different durations of stay for migrants (< 5 years, 5-9 years, 10-14 years, 15 + years, and Unknown) relative to the non-Hispanic White population. The results of this model are not entirely convincing. To begin with, the model itself does not fit well, at least relative to one that does not include duration in the U.S. as a variable. The chi-square statistic (2.0 with 6 degrees of freedom) is too small and statistically insignificant (see also the BIC statistic and associated information for Model 10 in Table 4). Even if the estimated effects of a few ethnicity-nativity duration dummies are significant, the pattern of effects is inconsistent with either assimilation or health selection effects. In fact, it is at the longest duration of stay that mortality advantage tends to be larger (increasing negative effects). This could reflect attrition of the unhealthier as duration of stay increases, and is certainly consistent with strong return migration effects. However, the fact that the pattern is visible among Other Hispanics but not among Mexicans is reason enough to lend lesser credibility to this alternative interpretation.

All in all, this first test is inconclusive since it suggests that the data do not reveal the patterns one would expect if either selection effects or assimilation operated in these populations.

Identifying a healthy migrant effect: the role of region of residence

Migrant populations are not homogeneous and neither is the degree of health selection within each of them. Differences in health selection across migrant populations should in part be reflected in the destination place: those living in areas that demand largely unskilled labor and

[9] All models in Table 3a constrain the effects of Puerto Rican, Cuban, U.S.-born Mexican and U.S.-born Other Hispanic to be identical. They all include controls for marital status, education, family income and employment status (see Table 2). To avoid cluttering, this table only displays estimates associated with the ethnic-nativity groups of interest and with the variable or variables introduced in the discussion. All other estimates are omitted.

that offer higher and perhaps easier accessibility to points of entry (lowering the cost of migration), should be regions where health selection is less rigorous.[10] The implication is that foreign-born Mexicans living in or nearby the border areas ought to be less selected than those residing elsewhere. The same argument applies to Puerto Ricans but since Puerto Ricans show no advantage to begin with, we test the implication only with adult Mexicans. To do this we create a variable for state of residence for the entire sample. It attains a value of 0 if the individual lives in Texas or California and 1 otherwise. To test the implication that the effects of being a foreign-born Mexican are larger for those residing in states other than Texas and California, we create an interaction term using the dummy variable for state of residence and the dummy variable for foreign-born Mexican. Estimates for the models including the new variables are displayed in columns 3 and 4 of Table 3a.

Our expectation is that the interaction term should be negative and significant, pointing toward a higher advantage among foreign-born Mexicans residing in non-neighboring states. The expectation is borne out with ample room to spare. Regardless of ethnic group, the effect of residing outside Texas or California is to reduce adult mortality by about (1-exp(.-.086)) or close to 8 percent. While this contrast is of some interest by itself, it is the **additional advantage** for foreign-born Mexicans of residing in other states that concerns us here. This is captured by the estimated coefficient of the interaction term which is properly signed (negative), large and statistically significant (-.545 with t =~4.1). The implications of this estimate are quite interesting. A foreign-born Mexican living in either Texas or California has a mortality level that is exp(-.34), or about 71 percent as high as a non-Hispanic White who resides in the same states. But the contrast between foreign-born Mexicans and non-Hispanic Whites living in states **other** than Texas and California is much larger, namely, exp(-.34-.55), approximately 41 percent lower mortality rates. Foreign-born Mexicans living in states other than Texas and California have lower mortality than foreign-born Mexicans living in those two states not just by virtue of universal effects that apply to everyone who resides elsewhere (exp(-.085), about 8 percent) but

[10] Although this inference is *ad hoc*, it can be retrieved from models of migration where the risk of migration is made a function of both contrasts in the price of labor between sending and destination regions and of the average migration costs (See Jasso et al., 2002).

because of an extra advantage that characterizes foreign-born Mexicans settled in other states (exp(-.55), about 58 percent).

Although the interaction effect is statistically significant with an exceedingly low significance margin ($p<.0001$) and is also of very large magnitude, we think this test provides only weak support for the conjecture regarding selection effects. First, while the conventional goodness of fit statistic suggests that the addition of the two variables for state of residence improves the fit of the model (last row of column 3 in Table 3a), the BIC criterion associated with the model is somewhat unsatisfactory relative to a model that excludes those variables (compare BIC values for Models 12 and 8 in Table 4). Second, note that the absolute magnitude and statistical significance of the foreign-born Mexican advantage is as strong as it was to begin with. Thus, although the test is suggestive, the results we obtain do not offer strong support for the idea that health selection effects reflected in state of residence account for the observed Hispanic advantage among foreign-born Mexicans and Other Hispanics. This can occur for one of two reasons: either there are no health selection effects at all or, if they exist, they are not reflected well in state of residence as assumed by the test.

Identifying the return migrant effect

We now turn to an assessment of Type I and Type II selection effects. Since our data set does not enable us to empirically distinguish between these, we will refer to them as the 'return migrant effect' and treat them as a single bundle. The tests we discuss below involve contrasts of mortality levels and slopes across ethnic groups that should hold under both type of errors.

<u>The importance of age effects:</u> The first question we need to answer is whether or not the estimates discussed before hold for the entire age span. If the model estimated above is true, then the answer to this question is obviously affirmative. But since this is not known for sure, we should estimate an alternative model allowing the effect of being (foreign-born) Mexican and Other Hispanic to be **different at ages above and below some threshold, say age 65**. The justification for this argument is as follows: if return migrant effects are influential, we would expect the advantage to be proportionately larger at older ages. Furthermore, because the magnitude of these effects is a function of return migration rates, the advantage is more likely to occur among foreign-born Mexicans than among foreign-born Other Hispanics whose countries

of origin are less easily reachable. Thus, we expect the age differences in the advantage to be trivial for Other Hispanics but significant for Mexicans.[11]

To test this conjecture we define a new dummy variable, 'breage', as 0 if age at onset of the study is younger than age 65 and set it equal to 1 if age at onset is 65 and older. We then estimate two models, one where the effect of an interaction term between the variables for ethnicity-nativity (foreign-born Mexican and Other Hispanic) and the dummy for age group are identical for both Mexicans and Other Hispanics, and a second model where the effects of the interaction term are unconstrained. The results are displayed in the last two columns of Table 3a. The constrained model (column 4) yields a negative and significant effect of the interaction term between the dummy for older age and the dummies for foreign-born Mexican and Other Hispanic. This means that, as expected if there is a return migrant effect, the advantage is larger for those who were aged 65 and above at the beginning of the study. But the main effects for foreign-born Mexican and Other Hispanic still persist. It is in the unconstrained model (last column of Table 3a) where we find three features expected by our conjecture: (a) the effect of older age applies to foreign-born Mexicans only (not to foreign-born Other Hispanics); (b) the advantage among foreign-born Mexicans vanishes and is replaced by pure age effects; and, finally, (c) the advantage enjoyed by foreign-born Other Hispanics remains unchanged. These three regularities are exactly as one would conjecture if there are significant return migrant effects and if these are more likely among Mexicans than among Other Hispanics.

Can these effects be interpretable as data artifacts? The only source of data artifact that could remotely be brought to bear is age overstatement. Yet since all models were estimated constraining age and slope effects and in all cases the constraint holds, there is no reason to suspect large effects of age overstatement. Finally, only very peculiar patterns of flawed self-reported ethnicity and of record mismatching can produce the set of effects reported above.

Yet, despite its apparent success, the model has three drawbacks. First, while the tests of

[11] Behind this reasoning may hide an important simplification. Journalistic reports suggest that return migration for some non-Mexican Hispanic groups, such as Salvadoreans and Guatemalans, has been fairly large during the period 1995-2000. In part the flows have been fueled by massive deportations but at least some of them originate in voluntary repatriation, so it should follow that age differences could also be important among Other Hispanics.

significance suggest the existence of age effects, the goodness of fit of the preferred model is only marginally better, if at all, than the constrained model. Thus, the estimate of the interaction effect with foreign-born Mexican (-.372) is large, negative, and statistically significant, and the difference between it and the interaction term for foreign-born Other Hispanic is of the opposite sign, large (.489) and also statistically significant (t~3.5). But the likelihood ratio test for the contrast between Models 4 and 5 in Table 3a is only 3.6. With 1 degree of freedom this is only marginally significant (last row of last column, Table 3a). Similarly, Table 4 shows that the unconstrained model (Model 13 in Table 4) exhibits a BIC statistic suggesting an inferior value relative to the constrained model (Model 12 in Table 14). These results indicate that we should opt for the constrained model, that is, one where the effects of the age variable are the same across both groups. But the main problem with this decision is that the fit of the constrained model is not as good as that of the baseline model (Model 8).

Second, the results of the test reflect a shift in mortality levels for older cohorts. While this is an expected consequence of return migration effects, it can also be the outcome of cohort changes in mortality. Although unlikely because the shift would require a **deterioration** of mortality levels at older ages for those who were members of **younger** cohorts, it is an unsettling possibility. In fact, even if the goodness-of-fit statistics (chi-squared and BIC) of the preferred model were stronger than what they are, the interpretation of the observed pattern would not be unique and Model 5 could be a good representation of empirical relations **without implying the existence of any return migration effects at all.**

The third and final problem is that the age we chose as a cut point (65) is arbitrary and there is no reason to select it over many other potential candidates which may yield completely different results.

Mortality slopes: There is still an unexplored possibility. Rather than resulting in a shift of rates after a certain age, return migration effects should exert a gradual influence, spread out over a large range of duration; this should be reflected in a **reduced slope of the mortality pattern**. The gradual depletion of individuals in bad health leads, via return migration of the unhealthy spread out through the duration of the study, to a mortality pattern with a downwardly biased slope.[12] Our argument is that we need to make room for effects operating not just on the level of mortality but also on the rate of change of mortality risks with duration in the study. If this is so, Gompertz's slope must vary across ethnic groups, even after constraining age effects to be identical to what is common in the slope effect across ethnic groups. More formally, in equation (1a) we replace (by (= ($_o$ + . Z_e where ($_o$ is a constant, Z_e is a dummy variable reflecting membership in ethnic group e, and . is an effect on the slope. We then constrain *, the effect of age (re-scaled) at the beginning of the follow-up period, to be equal to ($_o$. Thus, we let the slope change but not the constant slope and the age effect. The slopes should be lowest in groups most affected by return migration. Unlike the age-related shift in the advantage documented above, a slope effect is more difficult to account for by discrete cohort effects, it does not require an arbitrary choice of age as a cut point and, finally and not trivially, it cannot be confused with the impact of age overstatement since it only reflects duration effects. Letting the slope be a function of covariates is tantamount to saying that the underlying hazard model is not proportional at all, and that the effects of covariates cannot be completely captured by constant shifts of the log of hazards.

To test this possibility we estimate two models where the slope of the Gompertz curve is a function of ethnicity. The first model seeks to determine whether or not the slope is lower for foreign-born Mexicans than it is for everybody else, as it should be if a return migrant effect is stronger among Mexican migrants than among all other groups. The second model tests the same implication but for both foreign-born Mexicans and foreign-born Other Hispanics.

The first two columns of Table 3b contain the results for these two models. The estimated effect of being Mexican on the slope is, as expected, negative and statistically significant (t~3.0). Furthermore, all the information for this model in Table 4 (Model 14, Table 4) indicates that it fits the data well, marginally better than a model where the slope is constrained to be a constant (compared with Model 8 in Table 4).[13] Finally, and more importantly, the effects of being

[12] This will be the case if and only if Type II effects are proportionally greater at older ages, as they should be under conditions of unrestricted return migration.

[13] Note that in Table 4, Model 7 is NOT nested in Model 14 and that conventional likelihood ratio tests do not apply. This is the reason for the missing chi-square statistic in column 2 in Table 3b. The same applies to Model 15.

foreign-born Mexican vanish while those associated with being Other Hispanic remain strong.

Is this pattern also evident for Other Hispanics? If this were the case, our explanation would lose credibility since such a pattern is expected if and only if the observed mortality advantage is accounted for by return migration. To the extent that return migration among Other Hispanics is much less likely, we should not expect an effect on the slope similar to that just estimated among foreign-born Mexicans. In fact, this is the case. The third column of Table 3b displays estimates of a model in which we allow effects of being either foreign-born Mexican or foreign-born Other Hispanic on the slopes of the mortality curve. But only the former are in the expected direction and significant. The effect of being Other Hispanic on the slope is positive and insignificant. Furthermore, the estimated advantage increases rather than becoming weaker. Finally, the fit of the model (see Table 4, Model 15) is poor when compared to its simpler version. These are all signs that return migration effects cannot possibly explain the mortality advantage among Other Hispanics.

These findings lead to the following two propositions. First, the advantage of foreign-born Mexicans is completely accounted for by the smaller slope of their mortality curve, a telltale sign of return migrant effects. Neither duration of stay nor residential patterns account for much of the observed advantage, suggesting that health selection effects may be less important than return migration effects. Second, the advantage for Other Hispanic is remarkably robust. It is not related to return migration (as reflected in slope or age effects) and it does not weaken when controlling for duration effects. While this is inconsistent with the existence of health selection effects, we cannot rule out completely its influence in observed patterns. Thus, robust as it may be, the advantage enjoyed by Other Hispanics may still reflect heavy selection effects.

Bringing cultural and social effects back in

As stated at the outset of this paper, the data set available to us contains no information on individual connections and social networks, integration into a community or the like. Thus, it is impossible to design a rigorous test of the cultural hypothesis, according to which some of the Hispanic advantage that remains to be explained is associated with cultural conditions. To bridge the gap we propose two decidedly humbler tests. The first relies on the argument that the primary line of social protection is derived from an individual's family ties. If so, the universally protective effects of marriage ought to be even stronger. This suggests that the health benefits among married migrants must exceed those of the general population. The model we estimate (results not shown) includes an interaction term for marital status and being either foreign-born Mexican or foreign-born Other Hispanic. But although the effects are negative as expected (protective effects of marriage are higher among these groups) they are statistically insignificant and the model's fit is mediocre relative to other models.

The second test seeks to control for the type of community in which migrants live. If the advantage is associated with community ties, it must be the case that communities that are ill-prepared to offer an abundant supply of ties will not be good for migrants. Communities that do not score high on an index of segregation or isolation of migrant groups are not fertile grounds in which to cultivate social ties and to establish social networks. Further, these communities may have a negative impact by magnifying immigrants' sense of isolation, loneliness, or discrimination. If the hypothesis of cultural and social effects has any relevance, it should be the case that the mortality advantage vanishes once we control for the type of community in which migrants live.

Although we just showed that the advantage associated with being foreign-born Mexican disappears when we account for slope effects, we will estimate a model including isolation indicators with no slope effects to assess whether or not it can account for observed patterns as does the one with variable slopes. The last column of Table 3b shows pertinent estimates of the models. Interaction terms between foreign-born Mexican and Other Hispanic and the index of isolation of the community within which each individual lives were also estimated (results not shown) (see Appendix C for a definition of the variables). Once again, although the effects are in the expected direction, neither is large enough to rise beyond very modest levels of statistical significance. And, worst of all, neither the advantage associated with Mexicans nor with Other Hispanics changes much. Indeed, they are as strong as they were in our baseline models.

Modest as they might be, neither of these tests supports the validity of the cultural explanation. Surely, our failure to detect effects should not be interpreted to mean that social and cultural factors are immaterial. It may just be that they are not properly reflected by marital status nor by measures of ethnic group isolation.

Unmeasured heterogeneity

We end this section by asking whether or not the set of estimates obtained from the data in some of the most acceptable models are robust to model specification, namely, whether we need to explicitly address matters of unmeasured heterogeneity. Although there are no model-based solutions to the problem, the question arises: are the estimates for a hazard model contaminated by the effects of unmeasured conditions? Since unmeasured heterogeneity can have large effects on the curvature of a mortality pattern, this question is of crucial importance as contamination may invalidate the interpretation given to slope effects. Unmeasured heterogeneity is a larger concern in our case since center stage is occupied by the set of unmeasured conditions that **jointly affect migration status and mortality**. Because we only observe mortality patterns among Hispanic migrants we cannot hope to identify what those conditions are. We can, however, estimate a model posing the existence of some type of unmeasured condition so that we can assess if the estimates obtained before are sensitive to such specification. If they are, we need to reconsider the validity of our inferences. If they are not, we cannot conclude that heterogeneity is unimportant, only that the tenor of our conclusions, including all the uncertainty inherent in them, can be maintained.

To simplify matters, we estimate the baseline model (Model 1 in Table 3a) with a gamma-distributed unmeasured variable.[14] The estimation procedure yields a set of estimates of the various parameters and an estimate of the variance of the unmeasured characteristic. Two tests are necessary. The first determines whether estimates of parameters change significantly. The second determines if the variance of the unmeasured conditions is different from zero. The first two columns of Table 3c contain the relevant information. First, changes in the estimates of our parameters relative to the model without the gamma-distributed variable are trivial, and we can confidently conclude that these are robust to the specified type of unmeasured conditions (gamma-distributed). The second is that the variance of the unmeasured condition is not different from zero and that, therefore, its effects are irrelevant.

The result of this test enables us to end with a weak but important inference: there is no

[14] It is well-known that misidentification of the parametric distribution is consequential for estimation of some parameters included in the model.

evidence of gamma-distributed, fixed frailty in the data. If present at all, it is not enough to affect estimates of the advantage nor the curvature of the mortality pattern. In other words, the slope effects found before cannot reflect gamma distributed, fixed frailty. If so, our interpretation of these slope effects is robust to this potential source of bias.

Forcing a choice of model

We believe that a model with a variable slope is the most appropriate for the data. It does account for the advantage of foreign-born Mexicans though without simultaneously reducing the advantage of foreign-born Other Hispanics. It is an appropriate formulation that captures the presence of return migration effects and is much less amenable to be interpreted as a result of data artifacts. Our test for unmeasured frailty suggests that an admittedly narrow class of frailty effects did not change our results at all, thus increasing our confidence that what we observe is less likely to be produced by unobserved factors. A more serious threat to the interpretation we offer is associated with the possibility that a similar pattern of results would be observable due to age overstatement (Dechter and Preston, 1991; Rosenwaike and Preston, 1984). But this is unlikely for three reasons. First, age declaration among the living population is better in NHIS than in censuses and, by the very nature of the data set, age overstatement of the deceased is less of a problem. Second, the tendency to overstate ages has been attributed to Hispanics in general, not to particular groups. It is hard, then, to reconcile this interpretation with the observed pattern whereby negative slope effects are only prevalent among the non-U.S.-born Mexican population but not among Other Hispanics. Third and more importantly, age overstatement should not lead to slope effects at all. The effects of age overstatement should be reflected in estimates of age effects that are lower than estimates of the slope.

Although the variable slope model is theoretically sound, is resistant to artifacts and appears to be statistically preferable to others, how much better is it than its competitors? So far, we have relied on conventional p-values associated with chi square statistics and on the less liberal BIC values. Neither of these provides a concrete metric of fit, one with a straightforward interpretation and with practical implications. One solution is to compare the predicted number of deaths during the period of observation derived from the models we consider as top choices. Table 5 provides the observed and expected counts of deaths for each of four different ethnic groups computed from each of four competing models. A cursory examination of the tables reveals the key problem: they are all so good at predicting the count of deaths by ethnic groups that it is impossible to tell them apart.

Thus, our main conclusion must be necessarily subdued: while we have a theoretical preference for the more elegant model with variable slope, it does not perform significantly better than competing models in accounting for observed counts. However, additional data reviewed below provides additional justification for our choice.

Reconsidering the 'salmon bias': comparing migrants

One of the main conclusions of this paper is that the bulk of the foreign-born Mexican advantage is related to return migration of those who are in poor-health. We reach this conclusion using indirect rather than direct evidence. In order to reinforce it, we now briefly describe the results of evaluation of data that can bring us very close to a direct type of evidence.

The ideal test for the return migrant hypothesis is to compare mortality of recent return migrants to mortality of migrants who remain in the country of destination. Such a comparison is difficult since there is no follow-up of return migrants. However, we do have available information on the heath status of adults 50 and over interviewed in Mexico during the year 2000.[15] This data set (MHAS) also provides information on the migrant status of individuals residing in Mexico, and a limited migration history. With this data set we can compare self reported health status of return migrants with self-reported status of individuals in the NHIS-MCD sample at various points during the follow up. If return migration effects are strong, we expect self-reported health status among return migrants to be worse than among migrants who stayed in the U.S. To carry out the comparison we select the sub-sample of adults 50 and over interviewed by MHAS who resided in the U.S. and who had returned to Mexico in the ten years before the survey.[16] This group will be compared to three different NHIS-MCD samples. The

[15] The data source is MHAS (Mexican Health and Aging Study) conducted with a nationally representative sample of adults 50 and over and their surviving spouses. See Soldo et al., 2002.

[16] It is obviously better to choose a sub-sample of more recent returnees. While this is feasible the frequencies involved are too small to make meaningful comparisons. Choosing an interval of ten years or less was the best compromise we could find to resolve the tension between sample size and recency of return migration.

most inclusive (sample 1) among them consists of all Mexicans in the NHIS-MCD baseline study. The second sample (sample 2) consists of all Mexicans who were in the initial NHIS-MCD sample and who were not matched to a death record for a death that occurred before the midpoint of the follow up period. The third sample (sample 3) consists of all those enumerated in the baseline survey who were not matched to a death certificate at all during the entire follow up period. Because of the progressive removal of those who die during the follow up, the health status composition of the first sample should be the worst and that of the last one should be best. Though all three samples include return migrants, they are most heavily represented in the last sample since all individuals who died during the follow up period are removed. Neither of these three samples corresponds to a true sample of stayers but the first one should provide a lower bound for the health status of stayers whereas the last one provides an upper bound.

The comparison should be suggestive and not decisive for a number of reasons. First, we do not know much about the biases inherent in self reports nor about how they change with ethnicity, with place of residence, and with duration of stay in the U.S. Second, the samples we compare are not consistent in terms of age nor in terms of timing of migration. Third, the NHIS-based self-reports include individuals who, in due course, become return migrants and it is thus not a sample of true stayers. Fourth, return migrants who were in the worst health are probably not represented at all among surviving return migrants in Mexico as they may have died before they could become part of the sampling frame of MHAS. Most of these difficulties, however, play in our favor since the differences in self-reported health status between the MHAS sample and the NHIS-MCD sample will **underestimate the differences between a true sample of very recent returnees and a true sample of stayers.**

Table 6 displays the fraction of individuals self-reporting in bad health and in fair or bad health in three age groups in all four samples previously considered. As expected, the health distribution of sample 1 is the worst and that of sample 3 is the best. This simply reflects the fact that self-reported health is a moderately good predictor of mortality. Also as expected, the health status of those at older ages deteriorates in all samples.

The comparison with the MHAS sample is a bit difficult since the latter is based on a very small number of cases. Note, however, that, as would be expected if return migrants were selected among those in bad health, MHAS respondent health status is worse than any of the NHIS-MCD samples. Though consistent with the return migration conjecture, this finding is fragile for two reasons. First, the sample frequencies are too small to justify more than modest enthusiasm. Second, a comparison between the MHAS sample and the sub-sample of return migrants shows that health status among the latter is marginally better than among the entire MHAS sample. While this pattern is **not inconsistent at all** with return migration effects – indeed, they could be found when initial health selection in the immigration flow to the U.S. is combined with return migration effects – the pattern begs for an explanation. Until this explanation is offered, the evidence just produced must be taken as circumstantial. Matters may be more complicated than what we are assuming in this simple test.

IV. SUMMARY AND CONCLUSION

After a fairly dense battery of tests, we can draw the following conclusions:

a. The Hispanic advantage in adult mortality is not "Hispanic." The adult mortality advantage is a feature first and foremost of foreign-born Other Hispanics and of foreign-born Mexicans. It is not a characteristic of all other Hispanics, including Puerto Ricans and Cubans, whether born in the U.S. or abroad.

b. The foreign-born Mexican and Other Hispanic advantage in adult mortality is not trivial. It amounts to experiencing mortality rates that are between 30 and 50 percent lower than those experienced by non-Hispanic Whites. In turn, these differences translate into approximately six to nine years of additional life expectancy at age 45.

c. The behavior of mortality slopes produces strong signs of return migrant effects for foreign-born Mexicans but not for Other Hispanics. While the model we use to confirm this pattern fits the data as well as or only marginally better than competing models, our conclusion is robust to a class of unmeasured heterogeneity and receives additional support from a comparison of several NHIS-MCD samples and the MHAS sample. Indications for the presence of health selection effects are reduced and circumscribed to effects of state of residence.

d. The observed advantage favoring Other Hispanics persists even after accounting for indirect consequences of healthy migrant effects (duration of stay, state of residence) and it is resistant to age and slope effects as well as to unmeasured heterogeneity. This does not mean that healthy migrant effects are absent but that, if they exist, they are not reflected strongly enough in the mechanisms we were able to identify (duration or residence effects).

e. The cultural hypotheses received no support. We uncovered effects suggesting that those who live in ethnically more cohesive communities have lower mortality, as one would expect from the cultural hypotheses. But these conditions do not account for the Hispanic advantage nor do they alter the effects of membership in a group. It is not because foreign-born Mexicans or Other Hispanics have higher propensities to live in cohesive communities that they experience lower mortality than non-Hispanics. And it is not because there are extra gains accruing from residence in those communities among some Hispanics that there are mortality advantages.

This has been a partially successful exploration of the problem. We were able to justify a model that accounts for part of the advantage and attribute it to return migration effects. However, the preferred model which spawns this interpretation neither rests on robust, uncontested grounds nor is it complete, as part of the advantage – the part associated with Other Hispanics – remains thoroughly unexplained.

References

Abraido-Lanza, Ana F., Bruce P. Dohrenwend, Daisy S. Ng-Mak and J. Blake Turner. 1999. "The Latino Mortality Paradox: A Test of the "Salmon Bias" and Healthy Migrant Hypotheses." *American Journal of Public Health*, Vol 89(10).

Adler, N.E. and J.M. Ostrove. 1999. "Socioeconomic status and health: what we know and what we don't." In N.E. Adler, M. Marmot, B.S. McEwen, and J. Stewart (eds.) 1999. *Socioeconomic Status and Health in Industrial Nations: Social, Psychological and Biological Pathways*. Annals of the New York Academy of Sciences, Vol 896, pp. 3-16.

Arias, Elizabeth. 1998. The Demography of Assimilation: The Case of Cubans in the United States. Doctoral Dissertation. Department of Sociology, The University of Wisconsin-Madison.

_________. 2002. United States Life Tables, 2000. *National Vital Statistics Reports*, Volume 51, Number 3. Hyattsville, Maryland: National Center for Health Statistics.

Becker, T.M., C. Wiggins, C.R. Key and J.M. Samet. 1988. "Ischemic Heart Disease Mortality in Hispanics, American Indians, and non-Hispanic Whites in New Mexico, 1958-1982. *Circulation*, Vol 78.

Dechter, Aimée and Samuel H. Preston. 1991. "Age Misreporting and its Effects on Adult Mortality Estimates in Latin America." *Population Bulletin of the United Nations*, Nos. 31/32: 1-16.

Guend, Abdelhani and Karen C. Swallen. 2002. "Abridged Life Tables by Sex, Race, and Hispanic Ethnicity; US Population 1989-1991." CDE Working Paper 2002-03, Center for Demography and Ecology, University of Wisconsin-Madison.

Hummer, Robert A., Richard G. Rogers, Charles B. Nam and Felicia B. LeClere. 1999. "Race/Ethnicity, Nativity and U.S. Adult Mortality." *Social Science Quarterly*, Vol 80(1).

Hummer, Robert A., Richard G. Rogers, Sarit H. Amir, Douglas Forbes and W. Parker Frisbie. 2000. "Adult Mortality Differentials Among Hispanic Subgroups and Non-Hispanic Whites." *Social Science Quarterly*, Vol 81(1).

LeClere, Felicia B., Richard G. Rogers and Kimberley D. Peters. 1997. "Ethnicity and Mortality in the United States: Individual and Community Correlates." *Social Forces*, Vol 76(1).

Jasso, Guillermina, Douglas Massey, Mark Rosenzweig and James Smith. 2002. "Immigrant Health: Selectivity and Acculturation." National Academy of Science Conference on Racial and Ethnic Disparities in Health. March 2002.

Liao, Youlian, Richard S. Cooper, Guichan Cao, Ramon Durazo-Arvizu, Jay S. Kaufman, Amy Luke and Daniel L McGee. 1998. "Mortality Patterns among Adult Hispanics: Findings from the NHIS 1986 to 1990." *American Journal of Public Health*, Vol 88(2).

Markides, Kyriakos S. 1983. "Mortality Among Minority Populations: a Review of Recent Patterns and Trends." *Public Health Reports*, Vol 98(3).

Markides, Kyriakos S. and Jeannine Coreil. 1986. "The Health of Hispanics in the Southwestern United States: an Epidemiologic Paradox." *Public Health Reports*, Vol 101(3).

Massey, Douglas S. and Nancy A. Denton. 1988. "The Dimensions of Residential Segregation." *Social Forces*, Vol 67(2).

Mendes de Leon, C. and T. Glass, 2002. "The Role of Social Networks and Personal; Resources in Ethnic Disparities in Aging Health" National Academy of Science Conference on Racial and Ethnic Disparities in Health. March 2002.

National Center for Health Statistics. 1989. "Design and Estimation for the National Health Interview Survey, 1985-94." *Vital and Health Statistics, Data Evaluation and Methods Research*, Series 2 (110). Hyattsville, MD: National Center for Health Statistics.

National Center for Health Statistics. 2000. "National Health Interview Survey Multiple Cause of Death Files - 2000 Linkage." Hyattsville, MD: National Center for Health Statistics.

Novello, A.C., P.H. Wise, and D.V. Kleinman. 1991. "Hispanic health: time for data, time for action." *Journal of American Medical Association*, Vol 265.

Palloni, Alberto and Jeff Morenoff. 2001. "Interpreting the paradoxical in the 'Hispanic Paradox': Demographic and Epidemiological Approaches." in M. Weinstein, Al Hermalin and Mike Stoto (eds). *Population Health and Aging*, New York Academy of Sciences.

Raftery, Adrian. 1996. "Bayesian Model Selection in Social Research" (with discussion):11-195

In P.V. Mardsen (ed). *Sociological Methodology*, Cambridge, MA: Blackwell.

Rogers, Richard G., Robert A. Hummer, Charles B. Nam and Kimberley Peters. 1996. "Demographic, Socioeconomic, and Behavioral Factors Affecting Ethnic Mortality by Cause." *Social Forces*, Vol 74(4).

Rosenberg, Harry M., Jeffrey D. Maurer, Paul D. Sorlie, Norman J. Johnson, Marian F. MacDorman, Donna L. Hoyert, James F. Spitler and Chester Scott. 1999. "Quality of Death Rates by Race and Hispanic Origin: A Summary of Current Research, 1999." *Vital and Health Statistics*, Series 2 (128). Hyattsville, MD: National Center for Health Statistics.

Rosenwaike, Ira. 1987. "Mortality Differentials among Persons born in Cuba, Mexico, and Puerto Rico Residing in the United States, 1979-81." *American Journal of Public Health*, Vol 77(5).

Rosenwaike, I. and S.H. Preston. 1984. "Age Overstatement and Puerto Rican Longevity." *Human Biology*, Vol 56(3).

Sandefur, G., M.E. Campbell and J. Eggerling-Boeck. 2002. "Racial and Ethnic Identity, Official Classifications, and Health Disparities." National Academy of Science Conference on Racial and Ethnic Disparities in Health. March 2002.

Singh, Gopal K. and Mohammad Siahpush. 2001. "All-Cause and Cause-Specific Mortality of Immigrants and Native Born in the United States." *American Journal of Public Health,* Vol 91(3).

_________. 2002. "Ethnic-Immigrant Differentials in Health Behaviors, Morbidity, and Cause-Specific Mortality in the United States: An Analysis of Two National Data Bases." *Human Biology*, Vol 74(1).

Soldo, B., R. Wong and A. Palloni. 2002. "Migrant health selection: evidence from Mexico and the U.S." Paper presented at the Population Association of America Meeting, Atlanta, GA, May 9-11, 2002.

Sorlie, Paul D., Eric Backlund, Norman J. Johnson and Eugene Rogot. 1993. "Mortality by Hispanic Status in the United States." *Journal of American Medical Association*, Vol 270(20).

Sorlie, Paul D., Eric Backlund, and Jacob Keller. 1995. "U.S. Mortality by Economic, Demographic, and Social Characteristics: The National Longitudinal Mortality Study." *American Journal of Public Health*, Vol 85:949-56.

Spencer, G. 1984. "Mortality among the Elderly Spanish Surnamed Population in the Medicare Files: 1968 to 1979." Paper presented at the Meetings of the Population Association of America, Minneapolis, 3-5 May.

Table 1a: Estimates of basic hazard Models, Males aged 35+[(a)]

	Null Model	Model 1	Model 2	Model 3	Model 4	Model 5
Parameter			Estimates and S.E. (in parentheses)			
Baseline Hazard						
Constant	-4.016	-6.62	-6.56	-6.58	-6.57	-6.57
	(.021)	(.066)	(.086)	(.083)	(.084)	(.124)
Delta	-	.086	.085	.086	.086	.083
	(.0014)	(.0016)	(.0016)	(.0017)	(.0015)	
Gamma[(b)]	-	-	-	-	-	.086
		-	-	-	-	(.0015)
Ethnicity/Nativity						
Non-HW	-	-	-	-	-	-
U.S.-born PR	-	-	.101	.344*	-.225	-.225
			(.105)	(.156)	(.272)	(.272)
Fg-born PR	-	-	-	.059	.145	.145
			-	(.103)	(.113)	(.110)
U.S.-born CU	-	-	.026	.218	-.244	-.245
			(.136)	(.120)	(.513)	(.515)
Fg-born CU	-	-	-	-.067	.004	.003
			-	(.147)	(.141)	(.141)
U.S.-born MX	-	-	-.071	.048	.062	.063
			(.062)	(.087)	(.100)	(.097)
Fg-born MX	-	-	-	-.237**	-.260**	-.260**
			-	(.044)	(.040)	(.040)
U.S.-born OH	-	-	-.354**	-.180	-.058	-.057
			(.106)	(.130)	(.147)	(.147)
Fg-born OH	-	-	-	-.466**	-.606**	-.606**
			-	(.087)	(.084)	(.083)
Sample Size			17,825			
Events			1,632			
LL	-6729.0	-5652.9	-5646.3	-5641.9	-5635.4	-5635.2
Diff D.F.[(c)]	-	1	4	1	3	1
Chi-Sq St[(d)]	-	2,133	13.2	8.8[(e)]	13.0	.4
Upper Tail[(f)]	-	<.000	<.010	<.003	<.005	<.53

Footnotes to Table 1a:

(a) Fg.-Born = Foreign-born; HW= Hispanic White; PR= Puerto Rican; CU= Cuban; MX= Mexican; OH= Other Hispanic; * = p<.0114 or absolute value of t-statistic is larger than 2 but smaller than 3.5; ** = p<.000116 or absolute value of t-statistic is larger than 3.5
(b) When not shown the parameter Gamma is constrained to be equal to the parameter for Delta
(c) Diff D.F is the difference in the degrees of freedom between model j and model (j-1) displayed in the preceding column. In case of Model 1, the contrast is with the null model (constrained model)
(d) LL is the log Likelihood of a model; Chi-Sq St is the chi square statistic or -2*(LLc-LLu) where LLc and LLu are the log-likelihood functions of the constrained and unconstrained models. The contrast is always between models in column j and in column (j-1), except when so noted
(e) Model 3 constrains the differences of effects for foreign-born and U.S.-born to be invariant across ethnic groups.
(f) Upper tail is is the cumulated probability above the observed value of the Chi-Square Statistic

Table 1b: Estimates of basic hazard models, Females aged 35+[(a)]

	Null Model	Model 1	Model 2	Model 3	Model 4	Model 5
Parameter	Estimates and S.E. (in parentheses)					
Baseline Hazard						
Constant	-4.340	-7.06	-6.99	-7.01	-6.99	-7.14
	(.015)	(.039)	(.051)	(.054)	(.058)	(.090)
Delta	-	.086	.085	.086	.086	.085
		(.0014)	(.0014)	(.0015)	(.0017)	(.0015)
Gamma[(b)]	-	-	-	-	-	.132
		-	-	-	-	(.0014)
Ethnicity/Nativity						
Non- HW	-	-	-	-	-	-
U.S.-born PR	-	-	.096	.427**	-.200	-.213
			(.089)	(.118)	(.334)	(.337)
Fg-born PR	-	-	-	.034	.080	.082
			-	(.082)	(.086)	(.086)
U.S.-born CU	-	-	.038	.382	.822	.840
			(.110)	(.125)	(.440)	(.431)
Fg-born CU	-	-	-	-.011	-.110	-.098
			-	(.114)	(.080)	(.089)
U.S.-born MX	-	-	-.092	.065	-.054	-.061
			(.060)	(.062)	(.059)	(.058)
Fg-born MX	-	-	-	-.327*	-.140	-.098
			-	(.102)	(.104)	(.089)
U.S.-born OH	-	-	-.315**	-.100	.058	.055
			(.060)	(.210)	(.133)	(.133)
Fg-born OH	-	-	-	-.498**	-.691**	-.698**
			-	(.095)	(.188)	(.188)
Sample Size	21,189					
Events	1,590					
LL	-6897.6	-5713.6	-5706.9	-5699.5	-5693.1	-5685.2
Diff D.F.[(c)]	-	1	4	1	3	1
Chi-Sq St[(d)]	-	2,367	13.4	14.8[(e)]	12.9	15.7
Upper Tail[(f)]	-	<.000	<.009	<.000	<.005	<.000

Footnotes to Table 1b:

(a) Fg. Born = Foreign-born; HW= Hispanic White; PR= Puerto Rican; CU= Cuban; MX= Mexican; OH= Other Hispanic; * = p<.0114 or absolute value of t-statistic is larger than 2 but smaller than 3.5; ** = p<.000116 or absolute value of t-statistic is larger than 3.5

(b) When not shown the parameter Gamma is constrained to be equal to the parameter for Delta

(c) Diff D.F is the difference in the degrees of freedom between model j and model (j-1) displayed in the preceding column. In case of Model 1, the contrast is with the null model

(d) LL is the log Likelihood of a model; Chi-Sq St is the chi square statistic or -2*(LLc-LLu) where LLc and LLu are the log-likelihood functions of the constrained and unconstrained models. The contrast is always between models in column j and in column (j-1), except when so noted

(e) Model 3 constrains the differences of effects for foreign-born and U.S.-born to be invariant across ethnic groups

(f) The upper tail probability is the cumulated probability above the observed value of the Chi-Square Statistic

Table 2. Models with controls for demographic and socioeconomic conditions, Males and Females age 35+[a]

	MALES		FEMALES	
	Model 1	Model 2[b]	Model 1	Model 2[b]
Parameter	Estimates and S.E (in parentheses)			
Baseline Hazard				
Constant	-6.07(.044)	-6.07(.039)	-6.88(.087)	-6.905(.085)
Delta	.067(.003)	.067(.008)	.073(.0025)	.073(.0022)
Gamma[c]	-	-	-	-
Ethnicity/Nativity				
Non- HW	-	-	-	-
U.S.-born PR	-.267(.243)	-.098(.039)*	.184(.348)	-.027(.047)
Fg-born PR	-.106(.104)	-		-.058(.114)
	-			
U.S.-born CU	-.365(.508)	-	.828(.430)	-
Fg-born CU	-.049(.117)	-	-.127(.075)	-
U.S.-born MX	-.082(.074)	-	-.122(.084)	-
Fg-born MX	-.432(.034)**	-.433(.036)**	-.242(.126)*	-.232(.127)*
U.S.-born OH	-.125(.153)	-	.050(.139)	-
Fg-born OH	-.629(.105)**	-.629(.104)**	-.730(.200)**	-.724(.148)**
Demographic /Economic				
Married	-	-	-	-
Unmarried	.219(.032)**	.216(.033)**	.171(.062)*	.180(.056)**
<High School	-	-	-	-
High school	-.273(.069)**	-.275(.072)**	-.037(.080)	-.019(.083)
>High school	-.247(.080)*	-.249(.083)*	-.091(.070)	-.064(.078)
Family income, 1st Quartile	-	-	-	-
Family income 2nd Quartile	-.074(.039)*	-.073(.038)*	-.058(.090)	-.059(.092)
Family income 3rd Quartile	-.132(.039)**	-.133(.037)**	–.089(.093)	-.090(.088)
Family income 4th Quartile	-.255(.092)*	-.255(.035)**	-.270(.077)**	-.273(.076)**
Employed	-	-	-	-
Unemployed	.359(.293)	.362(.301)	.046(.440)	.038(.464)
Out of labor force	.565(.101)**	.563(.099)**	.468(.053)**	.461(.047)**
Sample Size	17,825		21,189	
Events	1,632		1,590	
LL	-5564.4	5565.0	5661.9	5668.6
Diff D.F.[d]	8	5	8	5
Chi-Sq St[e]	142.0	1.2	62.4	13.4
Upper Tail[f]	<.000	<.94	<.000	<.03

Footnotes to Table 2:

(a) Fg. Born = Foreign-born; HW= Hispanic White; PR= Puerto Rican; CU= Cuban; MX= Mexican; OH= Other Hispanic; * = p<.0114 or absolute value of t-statistic is larger than 2 but smaller than 3.5; ** = p<.000116 or absolute value of t-statistic is larger than 3.5

(b) Model 2 constrains the effects of Puerto Rican, Cuban, U.S.-born Mexican and U.S.-born Other Hispanic to be the same.
(c) When not shown the parameter Gamma is constrained to be equal to the parameter for Delta
(d) Diff D.F is the difference in the degrees of freedom between model j and model (j-1) displayed in the preceding column. In case of Model 1, the contrast is with the Model 4 in Table 1a (Table 1b for females). In case of Model 2, the contrast is with Model 1 and the difference in DF refers to the number of parameters of model 1 that are constrained in model 2
(e) LL is the log Likelihood of a model; Chi-Sq St is the chi square statistic or -2*(LLc-LLu) where LLc and LLu are the log-likelihood of the constrained and unconstrained models. The contrast for Model 1 uses Model 4 in Table 1a (Table 1b for females) as the constrained model. The contrast for Model 2 uses Model 1 as the unconstrained model
(f) The upper tail probability is the cumulated probability above the observed value of the Chi-Square Statistic

Table 3a: Models including duration if stay, state of residence and age-at-onset effects[a]

	Model 1[b]	Model 2	Model 3	Model 4[c]	Model 5
Parameters	Estimates and S.E. (in parentheses)				
Ethnicity/Nativity					
Fg Born MX	-.340**	-	-.342**	-.222**	-.131
	(.055)	-	(.065)	(.071)	(.091)
Fg Born OH	-.690**	-	-.711**	-.585**	-.760**
	(.080)	-	(.084)	(.105)	(.126)
Duration					
Fg Born MX 1-4	-	-.332**	-	-	-
	-	(.066)	-	-	-
Fg Born MX 5-9	-	-.465	-	-	-
				-	(.253)
	-	-	-		
Fg Born MX 10-14	-	-.414	-	-	-
				-	(.233)
	-	-	-		
Fg Born Mx 15+	-	-.327**	-	-	-
	-	(.055)	-	-	-
Fg Born OH 1-4	-	-1.030*	-	-	-
	-	(.352)	-	-	-
Fg Born OH 5-9	-	-.507	-	-	-
	-	(.365)	-	-	-
Fg Born OH 10-14	-	-.543**	-	-	-
	-	(.109)	-	-	-
Fg Born OH 15+	-	-.721**	-	-	-
	-	(.095)	-	-	-
Residence					
Non CA or TX	-	-	-.086*	-	-
	-	-	(.033)	-	-
Interaction *Fg Born MX	-	-	-.545**	-	-
	-	-	(.133)	-	-
Age at onset					
Age >=65	-	-	-	-.164*	-.165*
	-	-	-	(.064)	(.061)
Interaction*Fg Born MX	-	-	-	-.205**	-.372**
	-	-	-	(.042)	(.061)
Interaction*Fg Born OH	-	-	-	-.205**	.117
	-	-	-	(.042)	(.140)
Sample Size		39,013			
Events		3,253			
LL	-11250	-11249	-11244.8	11245.2	11243.4
Diff D.F.[d]	-	6	2	2	1
Chi-Sq St [e]	-	2.0	10.4	9.6	3.6
Upper Tail [f]	-	<.92	<.006	<.008	<.06

Footnotes to Table 3a:

(a) Fg. Born = Foreign-born; HW= Hispanic White; PR= Puerto Rican; CU= Cuban; MX= Mexican; OH= Other Hispanic; * =

p<.0114 or absolute value of t-statistic is larger than 2 but smaller than 3.5; ** = p<.000116 or absolute value of t-statistic is larger than 3.5

(b) All models estimated on the pooled male and female sample. All models include dummies for ethnicity-nativity and controls for demographic and economic conditions as defined in Table 2 (Model 2)

(c) In Model 4 the estimates of effects for the two interaction effects with age at onset are constrained to be the same.

(d) Diff D.F is the difference in the degrees of freedom between model j and model (j-1) displayed in the preceding column.

(e) LL is the log Likelihood of a model; Chi-Sq St is the chi square statistic or -2*(LLc-LLu) where LLc and LLu are the log-likelihood of the constrained and unconstrained models. Model 1 is the baseline against which Models 2 through 4 are contrasted. Model 5 is contrasted against Model 4

(f) The upper tail probability is the cumulated probability above the observed value of the Chi-Square Statistic

Table 3b: Models including duration tests for slopes and cultural effects[(a)]

	Model 1[(b)]	Model 2[(c)]	Model3[(c)]	Model 4[(d)]
Parameters	Estimates and S.E. (in parentheses)			
Baseline				
Constant	-6.219	-6.250	-6.234	-6.194
	(.042)	(.045)	(.049)	(.044)
Slope				
Constant	.070	.071	.070	.070
	(.0013)	(.0011)	(.0014)	(.0013)
Fg-Born MX	-	-.0070	-.0067**	-
	-	(.0024)	(.0025)	-
Fg-Born OH	-	-	.0072	-
	-	-	(.0047)	-
Ethnicity/Nativity				
Fg-Born MX	-.340**	-.089	-.102	-.358**
	(.055)	(.136)	(.132)	(.078)
Fg-Born OH	-.690**	-.686**	-.943**	-.700**
	(.080)	(.082)	(.183)	(.087)
Isolation Index				
1st Quartile	-	-	-	-
2nd Quartile	-	-	-	-.078**
	-	-	-	(.017)
3rd Quartile	-	-	-	-.024
	-	-	-	(.036)
4th Quartile	-	-	-	.004
	-	-	-	(.045)
Sample Size		39,013		
Events		3,253		
LL	-11250	-11249	-11245	-11248
Diff D.F.[(e)]	-	-	-	3
Chi-Sq St [(f)]	-	-	-	5.4
Upper Tail [(g)]	-	-	-	<.15

Footnotes to Table 3b:

(a) Fg. Born = Foreign-born; HW= Hispanic White; PR= Puerto Rican; CU= Cuban; MX= Mexican; OH= Other Hispanic; * = p<.0114 or absolute value of t-statistic is larger than 2 but smaller than 3.5; ** = p<.000116 or absolute value of t-statistic is larger than 3.5

(b) All models estimated on the pooled male and female sample. All models include dummies for ethnicity-nativity and controls for demographic and economic conditions as defined in Table 2 (Model 2)

(c) Model 1, our baseline model, is not nested in Models 2 or 3 and conventional chi square statistics do not apply.

(d) Model 1 is nested in Model 4 and Model 3 and conventional test apply.

(e) Diff D.F is the difference in the degrees of freedom between model j and model (j-1) displayed in the preceding column.

(f) LL is the log Likelihood of a model; Chi-Sq St is the chi square statistic or -2*(LLc-LLu) where LLc and LLu are the log-likelihood of the constrained and unconstrained models.

(f) The upper tail probability is the cumulated probability above the observed value of Chi-Square Statistic

Table 3c: Models with fixed Gamma-distributed unmeasured traits[(a)]

	Model 1[(b)]	Model 2[(b)]
Parameters	Estimates and S.E. (in parentheses)	
Baseline		
Constant	-6.219	-6.218
	(.042)	(.041)
Slope	.070	.070
	(.0013)	(.0013)
Ethnicity/Nativity		
Fg Born MX	-.340**	-.340**
	(.055)	(.055)
Fg Born OH	-.690**	-.690**
	(.080)	(.080)
Gamma Heterogeneity		
theta [(c)]	-	.0001(.0003)
Sample Size	39,013	
Events	3,253	
LL	-11250	-11249.8
Diff D.F. [(e)]	-	-
Chi-Sq St [(f)]	-	-
Upper Tail [(g)]	-	-

Footnotes to Table 3c:

(a) Fg. Born = Foreign-born; HW= Hispanic White; PR= Puerto Rican; CU= Cuban; MX= Mexican; OH= Other Hispanic; * = p<.0114 or absolute value of t-statistic is larger than 2 but smaller than 3.5; ** = p<.000116 or absolute value of t-statistic is larger than 3.5
(b) Models estimated on the pooled male and female sample. All models include dummies for ethnicity-nativity and controls for demographic and economic conditions as defined in Table 2 (Model 2)
(c) Theta is an estimate of the variance of the gamma distribution
(d) Diff D.F is the difference in the degrees of freedom between model j and model (j-1) displayed in the preceding column.
(e) LL is the log Likelihood of a model; Chi-Sq St is the chi square statistic or -2*(LLc-LLu) where LLc and LLu are the log-likelihood of the constrained and unconstrained models. Models 1 and 2 are not nested
(f) The upper tail probability is the cumulated probability above the observed value of the Chi-Square Statistic

Table 4: Quantities for calculation of BIC statistics in all models estimated with combined samples

Model J	Table Location	Remarks about model J	Kj	Pj	-LLj	LRTj	-BICj
0	not displayed	Null Model (one constant)	1	-	13651	-	-
1	not displayed	Two baseline (Gompertz) parameters	2	1	11447	4408	4400
2	not displayed	As model 1 plus gender	3	2	11377	4548	4531
3	not displayed	As model 2 + 4 parameters for ethnicity	7	6	11364	4574	4525
4	not displayed	As model 3 + 1 parameter for nativity	8	7	11347	4608	4551
5	not displayed	As model 2 + 8 parameters for ethnicity/nativity	11	10	11344	4614	4533
6	not displayed	As model 5 + 9 parameters for SES controls	20	19	11248	5806	4652
7	not displayed	As model 2 + 3 parameters for ethnicity/nativity	6	5	11350	4602	4561
8	Table 3a	As model 7 + 9 parameters for SES controls	15	14	11250	4802	4689
9	not displayed	As model 8 + 4 parameters for gender interactions with MX and OH foreign-born	19	18	11346	4610	4464
10	Table 3a	As model 8 + 6 parameters for duration in the U.S.	21	20	11249	4804	4642
11	Table 3a	As model 8 + 2 parameters for residence and interaction with foreign-born MX	17	16	11245	4812	4683
12	Table 3a	As model 8 + 2 parameters for age>=65 at onset and ONE interaction for foreign-born MX and OH	17	16	11245	4812	4683
13	Table 3a	As model 12 but with TWO interaction terms	18	17	11244	4814	4677
14	Table 3b	As model 8 + one extra parameters for slope (Effect of foreign-born MX)	16	15	11249	4804	4683
15	Table 3b	As model 8 + two extra parameter for slope (Effect of foreign-born MX and OH)	17	16	11248	4806	4677
16	Table 3b	As model 8 + three parameters for isolation of community of residence	18	17	11248	4806	4669
17	not displayed	As model 16 + three interaction terms of Isolation and foreign-born MX and OH	21	20	11246	4810	4648

Highlighted entries, Models 5 through 9, correspond to the set of models out of which the main baseline is selected (Model 8) . Models with lower (more negative) BIC values should be preferred

Kj the number of free parameters in model j
Pj the number of degrees of freedom associated with the test comparison of model j and null model
-LLj the (-1)* Log Likelihood of model j
LRTj the log likelihood ratio test statistic for the contrast between model j and null model or -2*(LL0-LLj)
BICj the quantity (-LRTj+ Pj*lnE) where E is the number of events
Sample Size = N=39,014
Number of Events=E=3,253

Table 5: Comparison of expected and observed death counts from competing models

Ethnic Group	Exposed Individuals	Observed Deaths	Expected Deaths			
			Model 1	Model 2	Model 3	Model 4
Non-Hispanic Whites	22554	2250	2218	2214	2218	2216
Foreign-born MX	3706	196	190	192	190	192
Foreign-born OH	3022	97	94	97	94	94
Remaining Hispanics	9745	701	693	693	692	693

Model 1: refers to the baseline or Model 8 in Table 4 (ethnicity and controls for SES)
Model 2: is Model 1 with added variables for dummy for age at onset of study (>65)
Model 3: is Model 1 with added variables for state of residence
Model 4: is Model 1 with slope a function of one covariate (foreign-born Mexican)

Table 6: Percentage self reported in poor (P) and in poor and fair health (P+F) for three different NHIS-MCD samples and for a sub sample of U.S. return migrants living in Mexico

	NHIS-MCD Samples [1]					
	Sample 1		Sample 2		Sample 3	
Age Group	P	P+F	P	P+F	P	P+F
50-59	8.5	28.0	7.7	27.4	7.6	26.2
60-74	11.2	35.4	10.1	33.2	9.7	32.1
75+	16.2	40.7	11.8	35.0	11.4	34.5

	MHAS Sub-Sample		
Age Group	N [2]	P	P+F
50-59	112	11.6	50.8
60-74	63	15.9	63.4
75+	10	20.1	60.1

(1) See text for definition of samples 1, 2 and 3 and for the definition of the MHAS Sub-sample
(2) Refers to number of cases (unweighted) in each age group

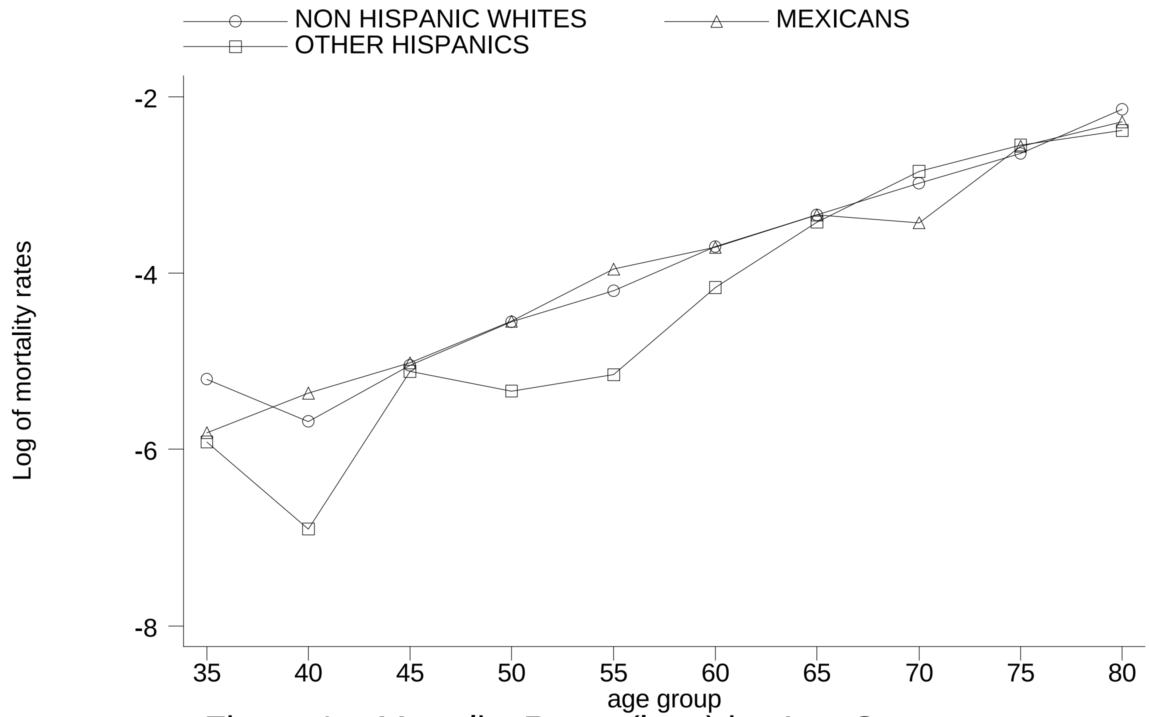

Figure 1a: Mortality Rates (logs) by Age Groups

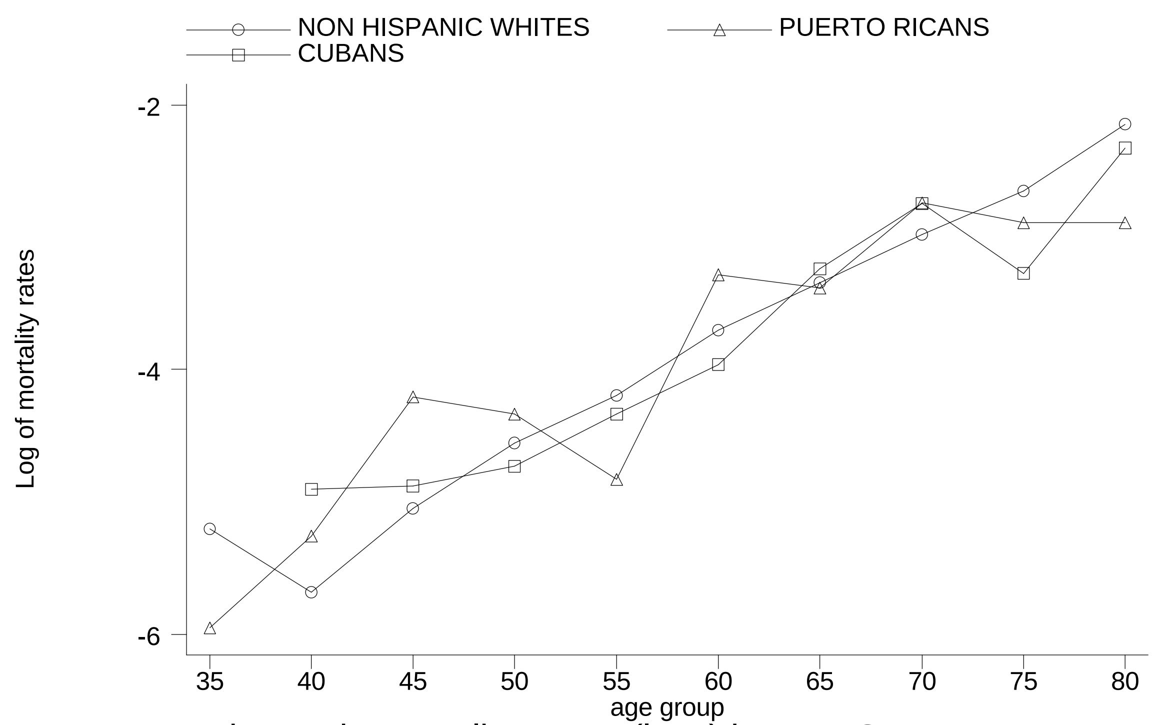

Figure 1b: Mortality Rates (logs) by Age Groups

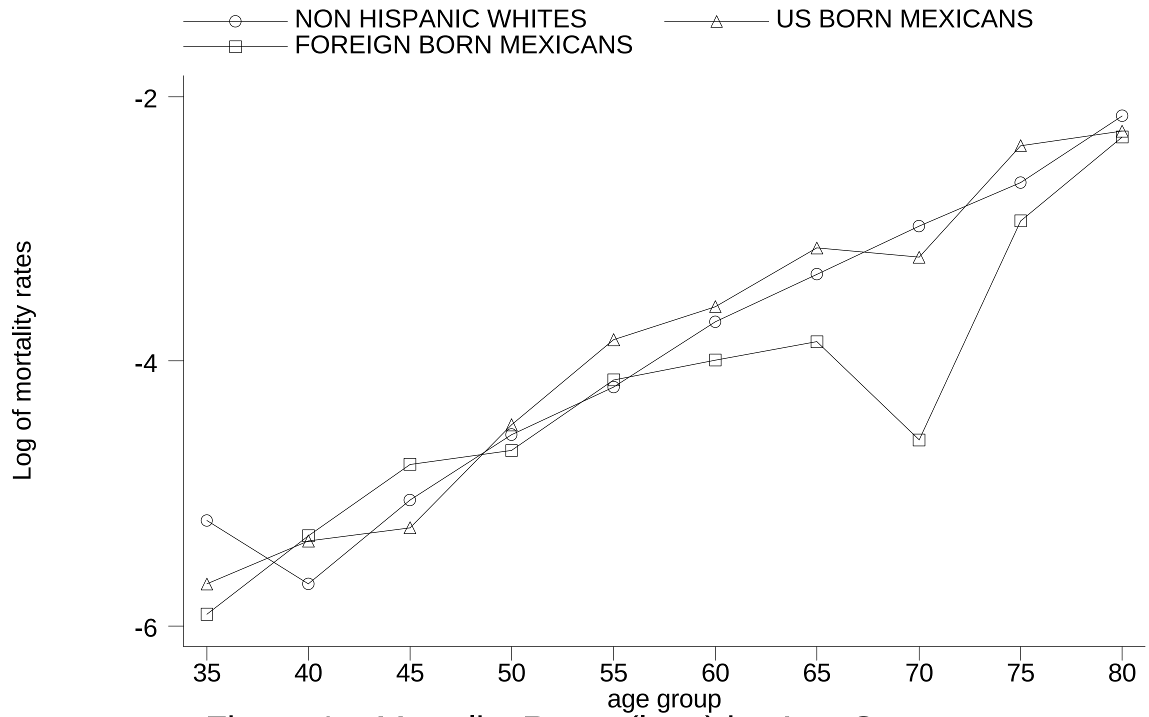

Figure 1c: Mortality Rates (logs) by Age Groups

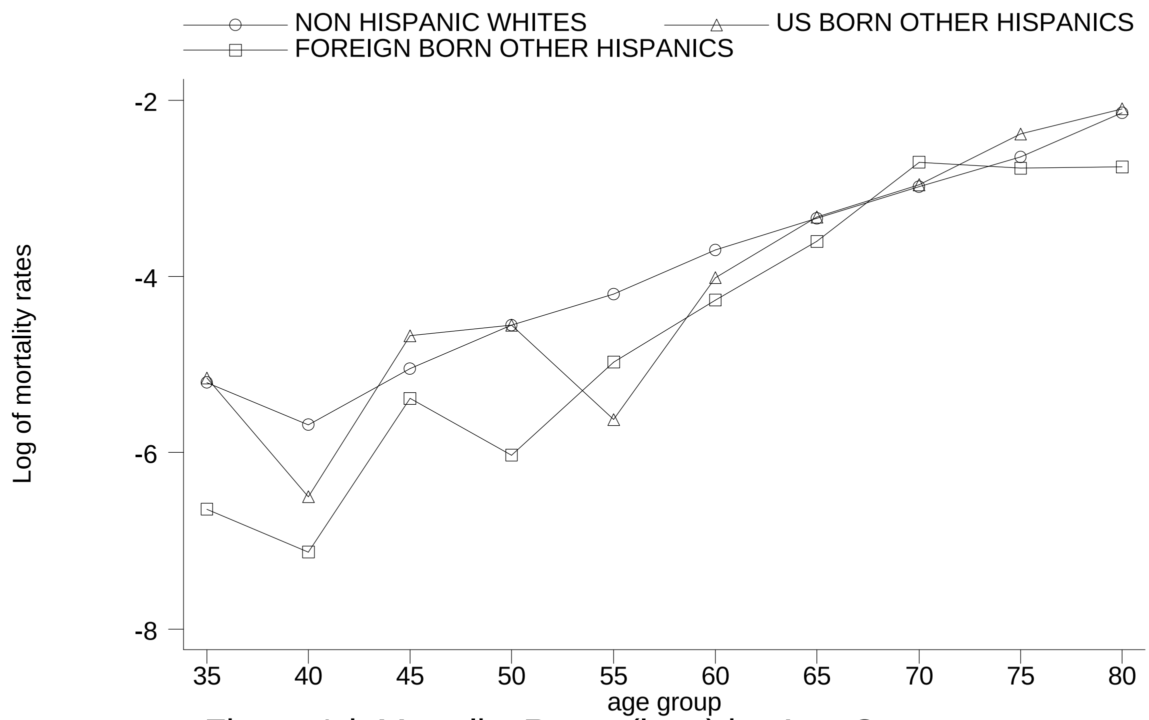

Figure 1d: Mortality Rates (logs) by Age Groups

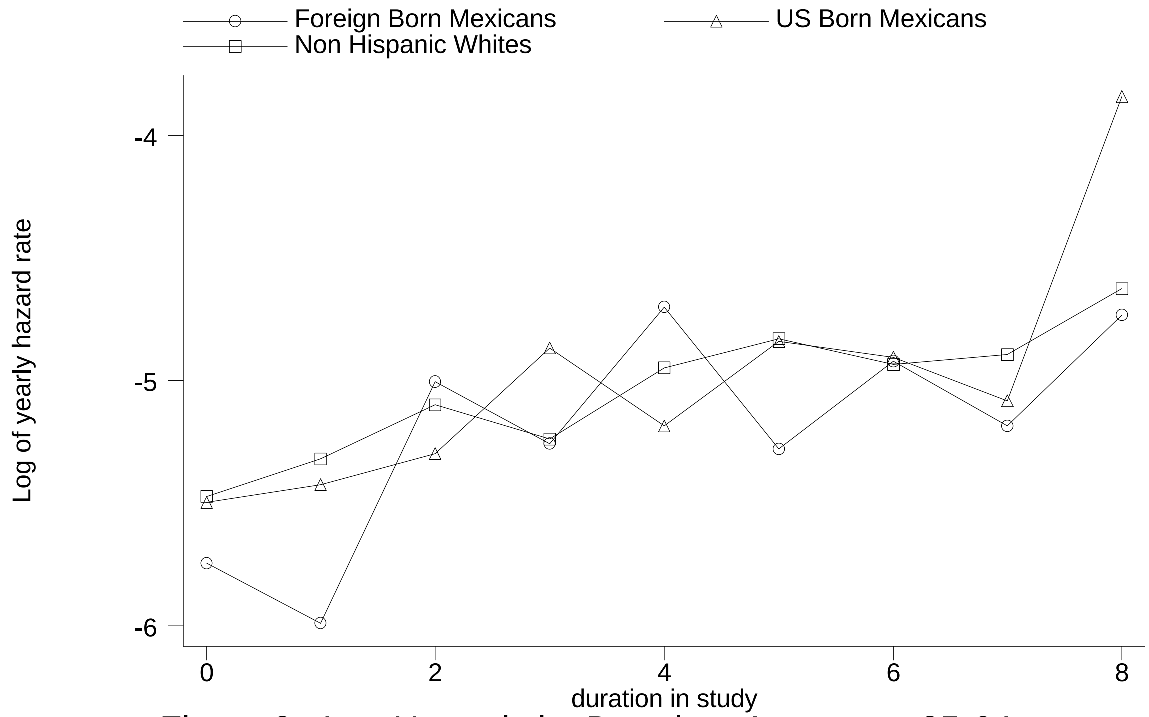

Figure 2a:Log Hazards by Duration, Age group 35-64

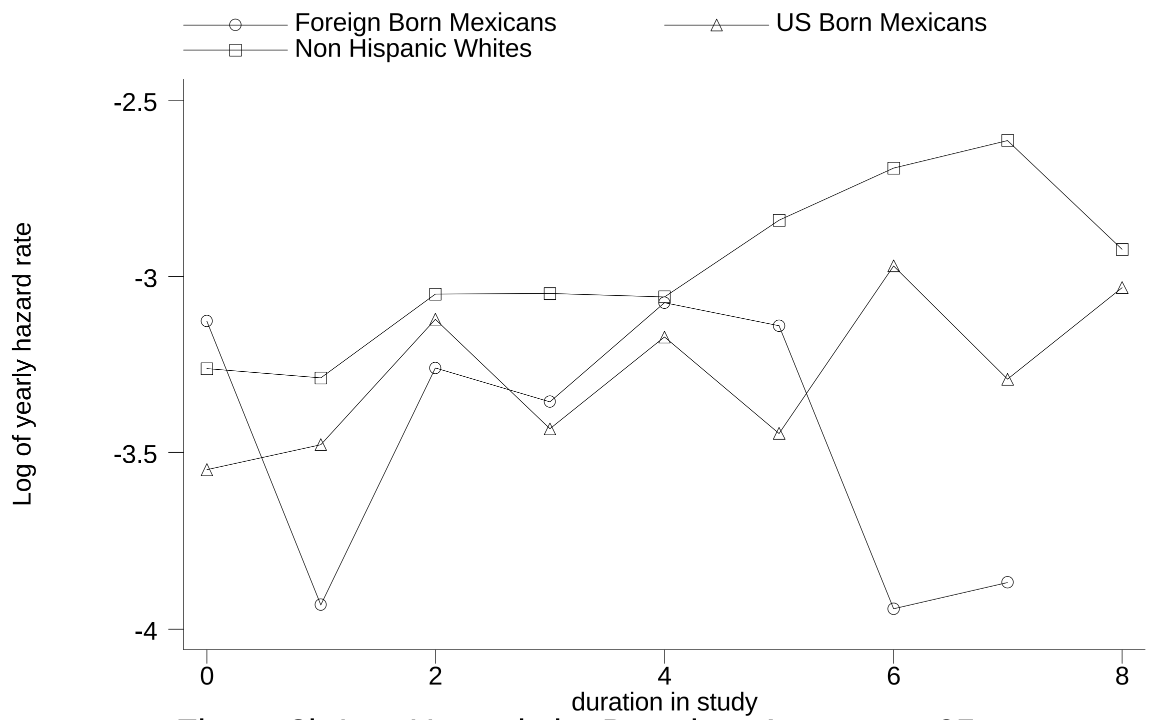

Figure 2b:Log Hazards by Duration, Age group 65+

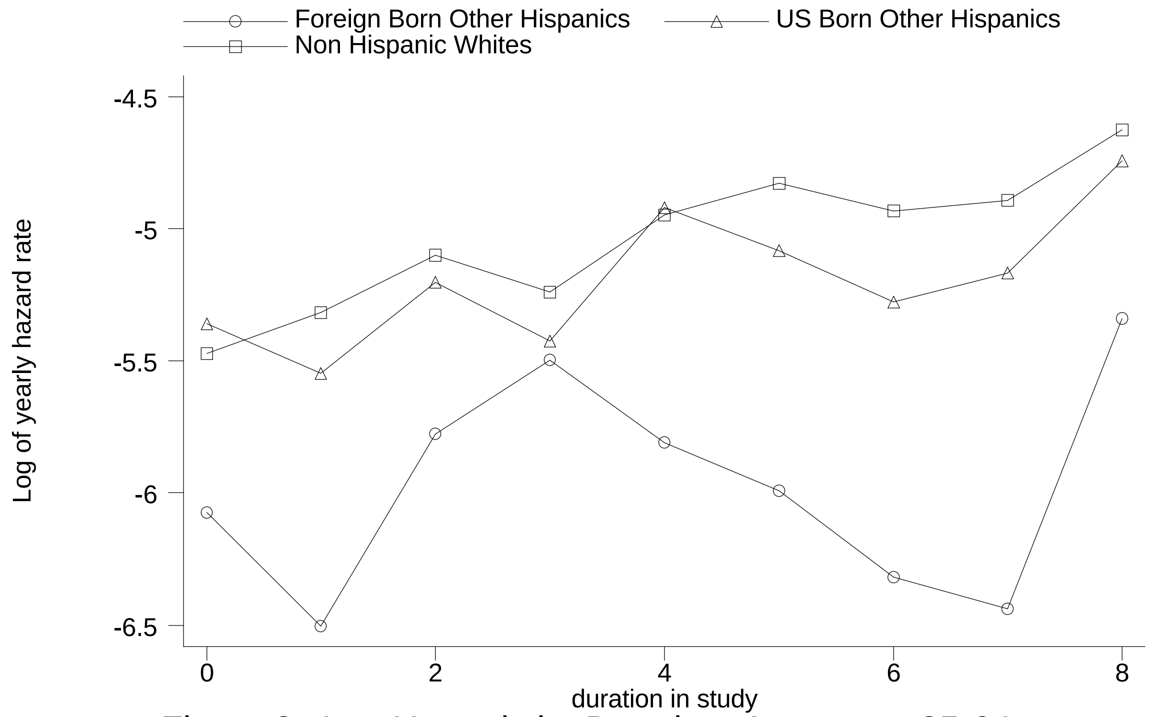

Figure 2c:Log Hazards by Duration, Age group 35-64

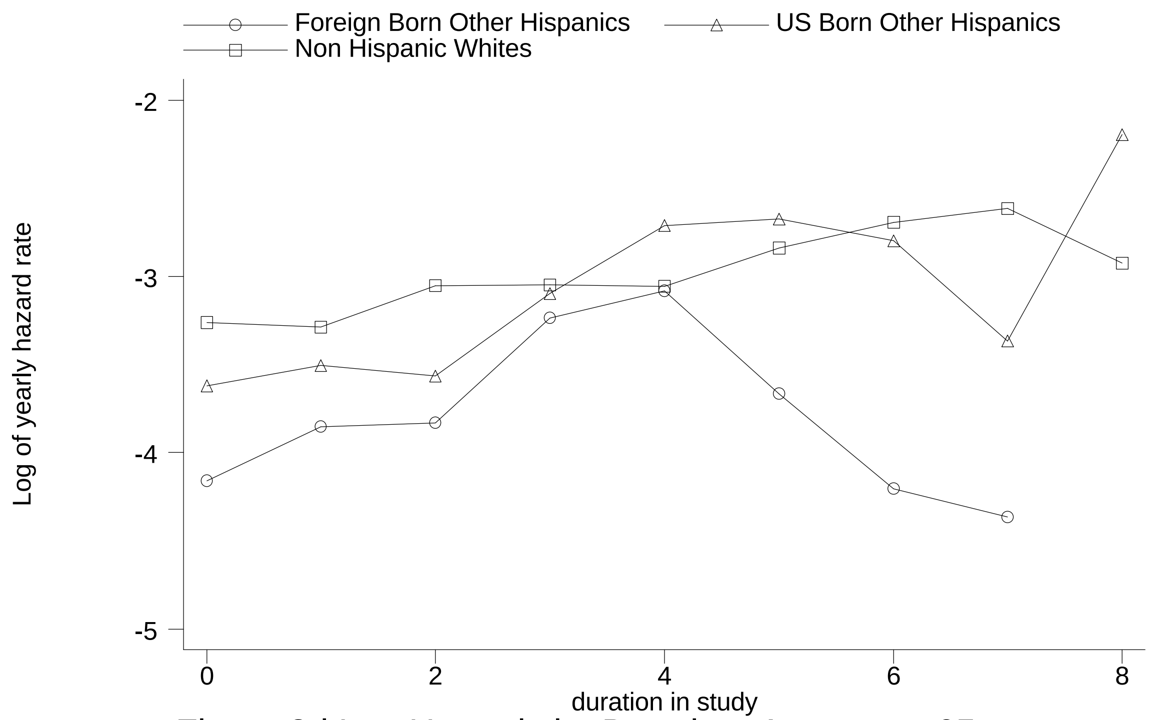

Figure 2d:Log Hazards by Duration, Age group 65+

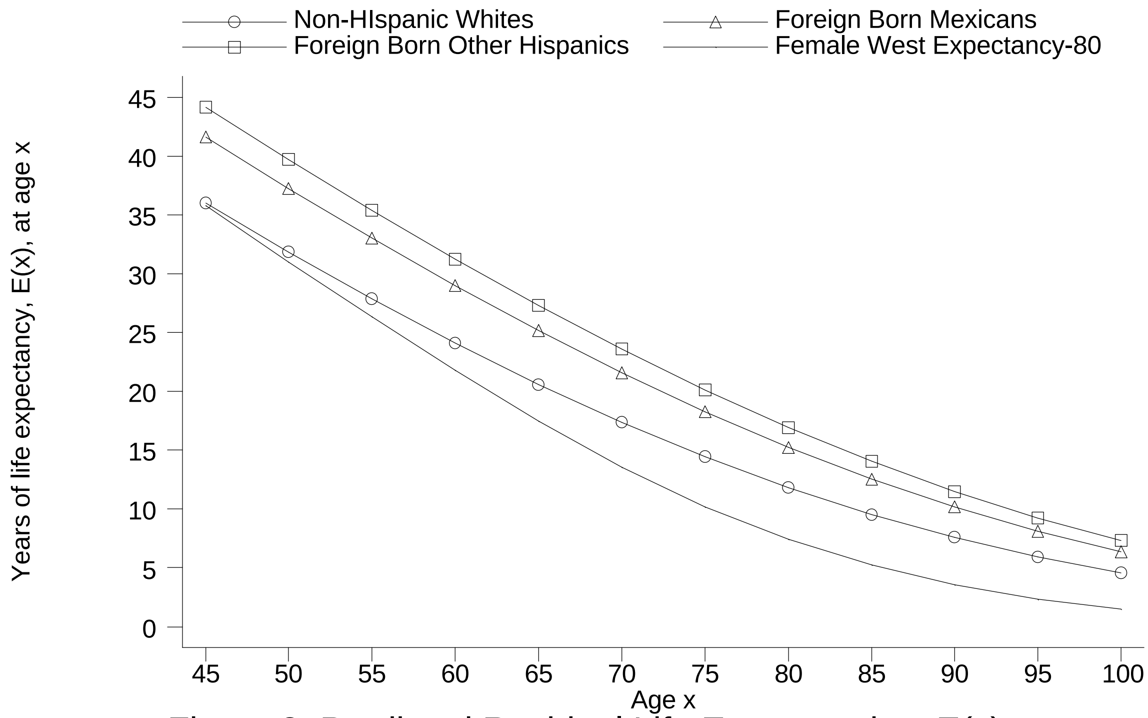

Figure 3: Predicted Residual Life Expectancies, E(x)

Appendix A. Sample Size and Mortality Outcome Characteristics, Males and Females

Ages 35 and Older

Ethnic Group

	non-Hispanic White		Cuban		Mexican		Puerto Rican		Other Hispanic	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Sample Size	10389	12176	664	800	4047	4463	860	1123	1990	2792
Alive	9269	11045	588	724	3768	4243	795	1059	1879	2668
Deaths	1120	1131	76	76	279	220	65	64	111	124

Ages 35-59

Ethnic Group

	non-Hispanic White		Cuban		Mexican		Puerto Rican		Other Hispanic	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Sample Size	6503	7007	394	447	3150	3295	661	841	1527	2051
Alive	6324	6886	382	439	3067	3238	639	815	1503	2030
Deaths	179	121	12	8	83	57	22	26	24	21

Ages 60 and Older

Ethnic Group

	non-Hispanic White		Cuban		Mexican		Puerto Rican		Other Hispanic	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Sample Size	3886	5169	270	353	897	1168	199	282	463	282
Alive	2989	4187	208	290	727	1019	160	247	378	645
Deaths	897	982	62	63	170	149	39	35	85	96

Appendix B. Description of Variables Used in Analyses

Variable Definitions:

Ethnicity includes non-Hispanic White, Cuban, Mexican, Puerto Rican and Other Hispanics. Ethnicity is self-identified at the time of interview by persons ages 18 and older.

Age is defined as age stated in years at time of interview.

Delta is age measured in number of years elapsed since the lower bound for which the hazard is applicable. For example, for the sample of those ages 35 and older Delta=Age-35; for those ages 60 and older it is Age-60.

Marital Status includes three categories: Married, Not Married and Unknown. Not married includes never married, separated, divorced and widowed persons. This variable refers to the marital status of the individual at baseline.

Education includes four categories: less than High School, High School Graduate, More than High School, and Unknown. This variable refers to the educational attainment of the individual at baseline.

Family Income includes four empirical quartiles of family income distribution. The 1^{st} Quartile is the lowest 25^{th} percentile of the income distribution and the 4^{th} Quartile is the highest 25^{th} percentile of the family income distribution. This variable refers to the family income reported by respondent at time of interview.

Employment status is made up of three categories: Currently employed, Currently unemployed and Not in the Labor Force. This variable refers to the employment status of the individual at time of survey interview.

Nativity/Duration of residence is made up of 6 categories: Foreign-born in U.S. for less than 5 years; Foreign-born in U.S. between 5 and 9 years; Foreign-born in U.S. between 10 and 14 years; Foreign-born in U.S. 15 or more years; U.S.-born; and Unknown. Finer categories of duration of residence were not possible as these are how the categories are reported in the NHIS. This variable refers to the Nativity/Duration status of the individual at time of interview.

Isolation Index was generated from Census STFA1 1990 file and appended to the NHIS-MCD data set by matching the data sets by FIPS State and County Codes. It is a physical segregation or measure of exposure. This measure can be interpreted as the probability that a selected minority member is exposed to only other members of his/her minority group. It is estimated as follows:
${}_xP_x = \Sigma_i [x_i / X] * [x_i / t_i]$, where x_i and t_i are the numbers of minority members and the total population in unit i, respectively. The unit of analyses in this case is the county tract. X represents the total number of x minority members in the county. The index ranges from 0 to 1 and may be interpreted as the "probability that a randomly drawn X member shares a unit with another X member." (Massey and Denton, 1988)

For the purposes of this study X refers to the total Hispanic population in unit i (county tract), and the index is summed to the county level. The distribution of the Index is broken down into four empirical quartiles where 1^{st} Quartile refers to the lowest 25^{th} percentile of the Index's distribution (or least amount of segregation) and the 4^{th} Quartile to the highest 25^{th} percentile of the Index's distribution (or the greatest segregation).

State of Residence is a two category variable reflecting residence in 1) California/Texas and 2) Residence in any other States. It was constructed to refer to the state residential status of Mexican Americans at time of survey interview. California and Texas reflecting border states with Mexico.

Appendix C. Descriptive Statistics of Sample of Males and Females Ages 35 and Older

	Ethnic Group									
	non-Hispanic White		Cuban		Mexican		Puerto Rican		Other Hispanic	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Age										
35-59	63.0	57.5	57.8	53.3	78.2	74.3	76.9	74.5	76.5	73.5
60 +	37.0	42.6	42.2	46.7	21.8	25.7	23.1	25.5	23.5	26.5
Marital Status										
Married	81.8	65.0	78.8	56.6	83.5	65.8	77.6	51.8	80.9	58.7
Not Married	18.1	34.8	21.1	43.5	16.4	34.0	22.3	48.0	19.0	41.1
Unknown	0.1	0.2	0.1	0.0	0.1	0.2	0.1	0.2	0.1	0.3
Education										
<H.S.	20.5	22.0	38.5	42.6	55.8	59.1	47.9	51.7	30.8	38.0
H.S.	33.7	41.2	23.9	29.1	22.0	24.7	28.5	26.8	27.6	30.0
> H.S.	45.3	36.2	37.1	27.5	20.6	15.2	22.4	20.6	40.9	31.3
Unknown	0.5	0.6	0.5	0.9	1.6	1.0	1.3	0.9	0.7	0.7
Family Income Distribution										
1st Quartile	13.9	22.1	23.8	29.7	28.3	36.7	30.0	42.3	18.9	29.3
2nd Quartile	23.5	25.3	32.6	31.6	30.0	27.0	26.2	24.2	26.4	24.7
3rd Quartile	29.0	27.0	24.0	20.3	24.3	22.1	23.1	18.4	27.3	24.4
4th Quartile	33.6	25.6	19.7	18.5	17.4	14.3	20.7	15.1	27.5	21.7
Employment Status										
Employed	66.4	47.0	63.2	43.0	70.7	46.2	62.8	38.2	73.2	53.0
Unemployed	2.0	1.6	3.1	2.2	4.8	3.2	2.4	2.3	3.5	2.9
Not in Labor Force	31.6	51.4	3.4	54.8	24.5	50.7	34.8	59.5	23.4	44.0
Nativity-Duration of Residence in U.S.										
< 5 years	0.4	0.3	5.1	4.9	3.1	2.9	4.5	5.8	6.2	6.7
5-9 years	0.2	0.2	6.2	4.2	4.6	3.2	4.7	4.3	11.4	9.4
10-14 years	0.4	0.3	14.7	8.2	6.6	5.7	4.4	5.3	10.0	9.6
15+years	3.5	4.6	59.6	67.5	31.1	29.5	61.9	61.0	34.9	38.1
U.S.-Born	95.4	94.6	13.7	13.5	53.3	57.4	21.5	20.1	36.2	34.8
Unknown	0.0	0.1	0.8	1.8	1.3	1.4	2.9	3.5	1.4	1.4
Isolation Index Distribution (0-1)										
1st Quartile	39.6	40.3	6.0	5.0	2.1	1.8	4.2	5.3	7.5	7.0
2nd Quartile	33.1	33.3	9.6	11.3	10.6	9.7	20.3	16.5	20.3	21.9
3rd Quartile	19.2	19.5	14.6	11.1	32.9	36.1	42.3	41.3	34.4	34.8
4th Quartile	8.2	6.9	69.8	72.6	54.4	52.4	33.1	37.0	37.8	36.3

Center for Demography and Ecology
University of Wisconsin
1180 Observatory Drive Rm. 4412
Madison, WI 53706-1393
U.S.A.
608/262-2182
FAX 608/262-8400
comments to: palloni@ssc.wisc.edu
requests to: cdepubs@ssc.wisc.edu