The Intergenerational Persistence of Immigrant Mortality Advantage: New Results for U.S. Male Old-Age Mortality

Extended abstract for 2019 PAA proposed paper

Joshua R. Goldstein
UC Berkeley, Dept. of Demography

September 18, 2018

Abstract

The tendency of immigrants to have lower mortality than natives is one of the most widely replicated findings in demography, but there is still no agreed-upon explanation. In this paper, the new CenSoc data set, linking millions of males in 1940 census records to old-age social security deaths, is used to show that this immigrant advantage persists into the second generation. This finding suggest that selective immigration (and return migration) of the first generation is not, as some have hypothesized, the driving force behind immigrant mortality advantages. Instead, health behaviors that can be passed on intergenerationally appear more likely to be the underlying cause. Preliminary results suggest that more than half of the immigrant advantage persists into the second generation.

1 Introduction

The immigrant mortality paradox – that immigrants often live longer than natives despite lower socio-economic status – is one of the most replicated findings in demography. However, there is still no agreed- upon explanation. Some argue that the selective nature of the immigration process is the main force behind immigrant advantage. Others attribute the advantage of immigrants to health behavior and a supportive social environment. Still others argue that much of the effect is due to artifacts of data collection, such as age misreporting, mismatching of numerators and denominators, and loss to follow-up mistaken as survival.

This paper takes advantage of the newly created multi-million record CenSoc data set, which links individuals in the 1940 census records with their social security death records. The analysis of the CenSoc data aims to make two contributions. First, the sample size is large enough to analyze dozens of national origin groups, extending the existing literature on Hispanics to cover European and Asian immigrant groups. Second, the CenSoc data allows the identification of 2nd generation immigrants as well as a rich set of socio-economic and even cultural variables.

The study of the second generation is particularly interesting when addressing the causes of immigrant advantage. The process of immigration – who chooses to come to the United States, who makes it here, and who remains – is known to select for those in good health and against those in poor health. However, this usual explanation of selection for 1st generation immigrant health advantages does not apply directly to the U.S.-born children of immigrants. Instead, the advantages of the second generation – if they exist – are likely due to intergenerationally transmissible characteristics and behaviors.

The proposed paper tries to address the following questions:

1. What is the mortality of first generation immigrants in the United States from a broad range of origin countries? Does the well-documented advantage of Mexican immigrants apply to immigrants from Europe and Asia? Which countries have the largest advantages? Are there origin countries from which immigrants are *disadvantaged*, compared to natives?

2. Do mortality differences between natives and immigrants persist into the 2nd generation? The census items on parental birthplace allow for the first time identification of second generation immigrants in large enough numbers to obtain accurate estimates.

3. What explains the persistence and/or disappearance of immigrant mortality advantage over the generations? The 1940 census allows measurement of language, name giving, and neighborhood ethnic composition, indirect measures of cultural assimilation. Measures of socio-economic status such as income, education, and occupation can also used to explain the immigrant mortality paradox – or make it even more paradoxical.

4. What is the mortality of the 3rd and higher generation? Ethnically distinct last names can be used to create a sample of 3rd and higher generation migrants by country of origin. Using this approach, which can be seen as an extension of the traditional use

of Hispanic surnames classification, we are able to study the persistence of immigrant advantages beyond the 3rd generation.

The paper builds on recent work by Mehta et al (2017), which uses secure Social Security and Medicare data. The CenSoc data set extends their analysis by allowing the addition of covariates observed in the 1940 census, which among other advantages allows identification of the 2nd generation, measurement of education, income, neighborhood characteristics. Another feature of the CenSoc data is that it is based on public records and thus allows replication and extension by other researchers.

2 Data and Methods

The CenSoc data set links transcribed 1940 census records to publicly available death records from the Social Security Administration. The death records cover nearly all deaths in the United States from 1975 to 2005 for individuals over 65. The data set was created at Berkeley by Goldstein and Monica Alexander, based on the full-count transcribed census records created by Ancestry.com and the University of Minnesota. The current version of the data set is based on unique, exact matches by first name, last name, and birth year.

The current version of the data set includes about 3.1 million males from birth cohorts from about 1890 to about 1920 who are observed as adults in the 1940 census and whose deaths occur after age 65 in the period from 1975 to 2005. A linked data set for females will be created in the future, based on newly available social security (SS-5) application files.

The CenSoc data has exact dates of birth and death from the social security death records and full 1940 census information, including a rich set of census socio-economic covariates such as years of education, wage income from the pre-census year, occupation, geographic location, race, and place of birth. For a sample of 1-in-20 respondents, the 1940 census also asked mother and father's place of birth.

The CenSoc linked data is most useful for individuals born in the period from 1900 to 1920 and dying after 1975. These individuals are observed age 20 to 40 in the 1940 census and have mortality records for a broad-enough range of ages for reliable estimation.

A notable feature of the CenSoc data is that it is entirely based on publicly-released records: both the 1940 census and the Social Security death records have been made public. The data set is being prepared for public release and will be able to be used by researchers without having to obtain permission to use a secure data environment.

3 Mortality Estimation

The CenSoc data set includes only records of those who have died, and covers only a limited age range, which varies by cohort. Using the assumption that the record linkage process preserves the age-distribution of deaths for each cohort – which we have confirmed on aggregate by comparison with published cohort life tables – it is possible to use a variety of methods to estimate mortality.

Monica Alexander (2018) has developed methods for estimating life tables using doubly-truncated data. The approach is to fit parametric, or semi-parametric mortality schedules to the truncated data. These methods work well when estimated using Bayesian hierarchical models.

For exploratory data analysis, the effect of covariates on mortality can also be estimated by regressing directly on age at death. Fixed-effects for each birth year are introduced to in order to account for the differential truncation by cohort. However, it should be kept in mind that ordinary regression will tend to underestimate the effect of covariates, because it does not consider the unobserved tails of the death distribution.

4 Preliminary Results

In this section, we show the preliminary estimates from the CenSoc data for 1st and 2nd generation male immigrant mortality. The results are descriptive, with further investigation of the explanatory factors for immigrant mortality advantages being the focus of the work to be done for the conference paper. For these preliminary results, regression is to estimate mean ages of death, a method that is conservative in estimating effect sizes, as discussed above.

We begin by estimating the longevity of 1st generation male immigrants by their country of birth. Figure 1 shows life expectancy at age 65 for immigrants from more than 30 countries compared with the longevity of "natives" – defined as the U.S.-born with two U.S.-born parents. Several features emerge from the figure:

- Immigrants from almost all countries have lower old age mortality than natives. Almost all of the the immigrants groups live longer on average than natives, with many groups living 1 or even 2 years longer than natives. This is the same order of magnitude as the advantage associated with attending university vs. grade-school for these cohorts.
- Mexican immigrants have the mortality advantage that has been found in the literature.
- Immigrants from Asia tend to have the greatest advantage,
- Immigrants from some origins – namely those from Ireland, Scotland, and French-speaking Canada – are *disadvantaged* compared to natives.
- The immigrant advantage exists for a wide variety of countries including the Philippines, Turkey, Syria, and Romania.

The variety of mortality advantages and disadvantages shown in Figure 1 raise a number of questions about the causes of the immigration mortality paradox, particularly around the relative roles of migration selectivity and health behavior.

The new approach taken in this research to gain insight into the source of immigrant mortality advantages is to look at the mortality of the immigrant 2nd generation. The kinds of health selection we usually associate with migration (e.g., TB, physical disability,

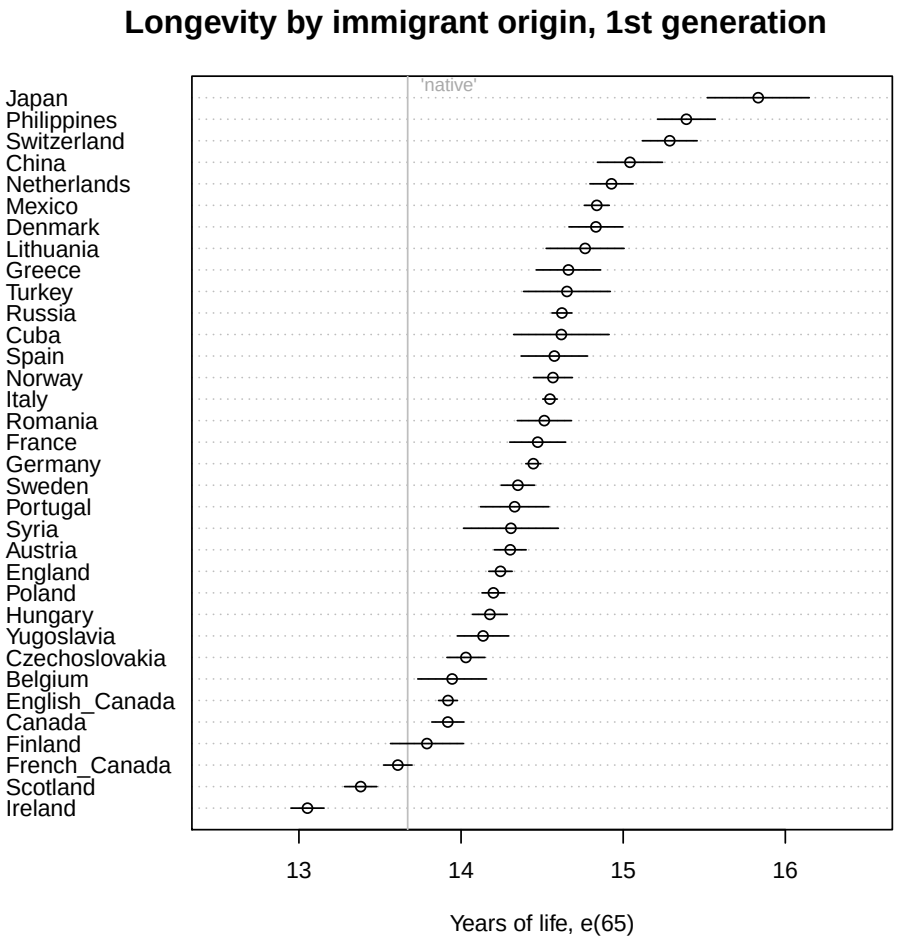

Figure 1: Longevity by immigrant origin, 1st generation males. Regression estimates of life expectancy at age 65 from CenSoc, with ±1 SE bars. "Natives" are defined as the U.S.-born with two U.S.-born parents.

exposure to severe childhood disease) would affect the 2nd generation. Figure 2 shows the results of our preliminary analysis of the 2nd generation men, defined as the native-born with foreign-born fathers. The figure shows results for cohorts of 1st and 2nd generation immigrants born from 1900 to 1920. The immigrant generations are born at the same time, and are not directly related to each other.

Keeping in mind that both the longevity of the 1st and 2nd generations is estimated with some error, some preliminary observations can be made:

- For most immigrant groups the advantage (or disadvantage) in mortality compared to natives persists from the 1st generation to the next. The mean absolute deviation of life expectancy of immigrants compared to natives across all groups is about 0.9 years for the 1st generation and about 0.6 years for the 2nd generation.

- There is some convergence to native mortality. Part of this is statistical regression

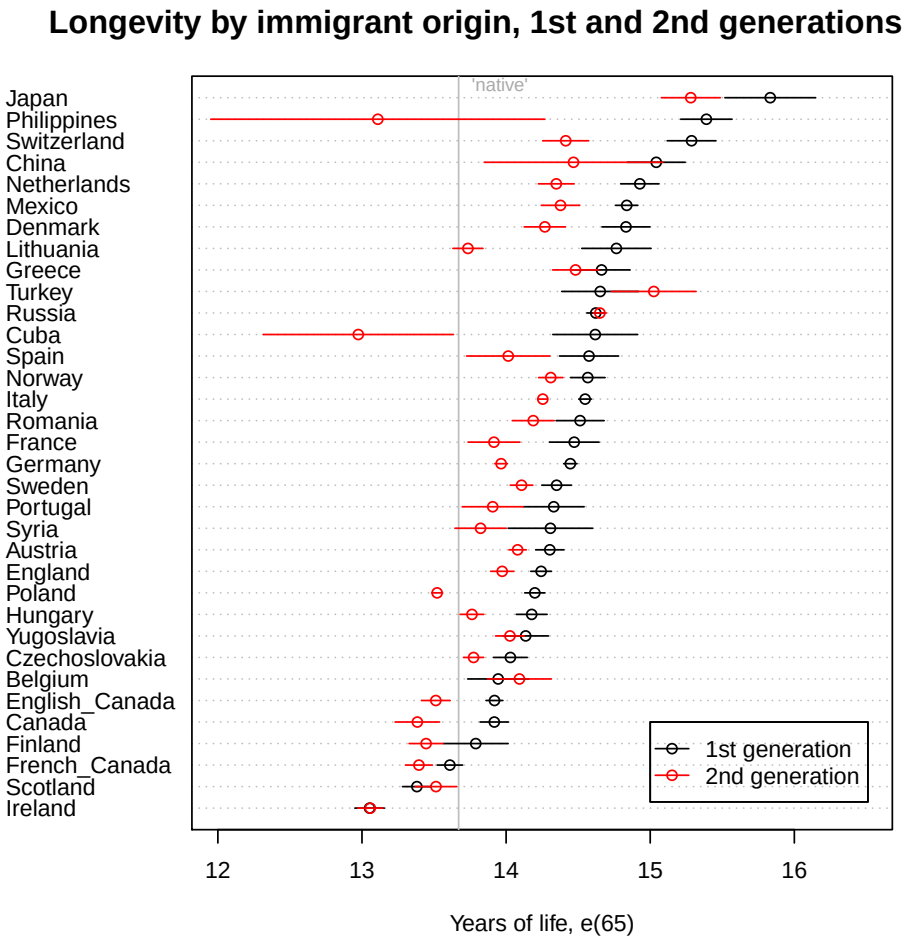

Figure 2: Longevity by immigrant origin, 1st and 2nd generation males. Regression estimates of life expectancy at age 65 from CenSoc, with ±1 SE bars. "Natives" are defined as the U.S.-born with two U.S.-born parents. 2nd generation are U.S.-born males with foreign-born fathers.

to the mean, particularly for groups like the Cubans or the Syrians whose 1st generation estimates were fairly uncertain because of small sample sizes. But part of this convergence also appears to be a real phenomenon across most groups.

- Disadvantaged groups, notably the Irish, see no or little change between the 1st and 2nd generations.

All of these results point to the potential importance of transmissability of health behaviors. Figure 2 raises questions about what health advantages are being passed down from Japanese immigrants to their children and similarly what disadvantages are being passed down from Irish immigrants to their children.

Figure 3 plots the estimates of longevity for the 1st and 2nd generations against one another, allowing us to see the strength of the correlation: $r = 0.60$ for all of the countries

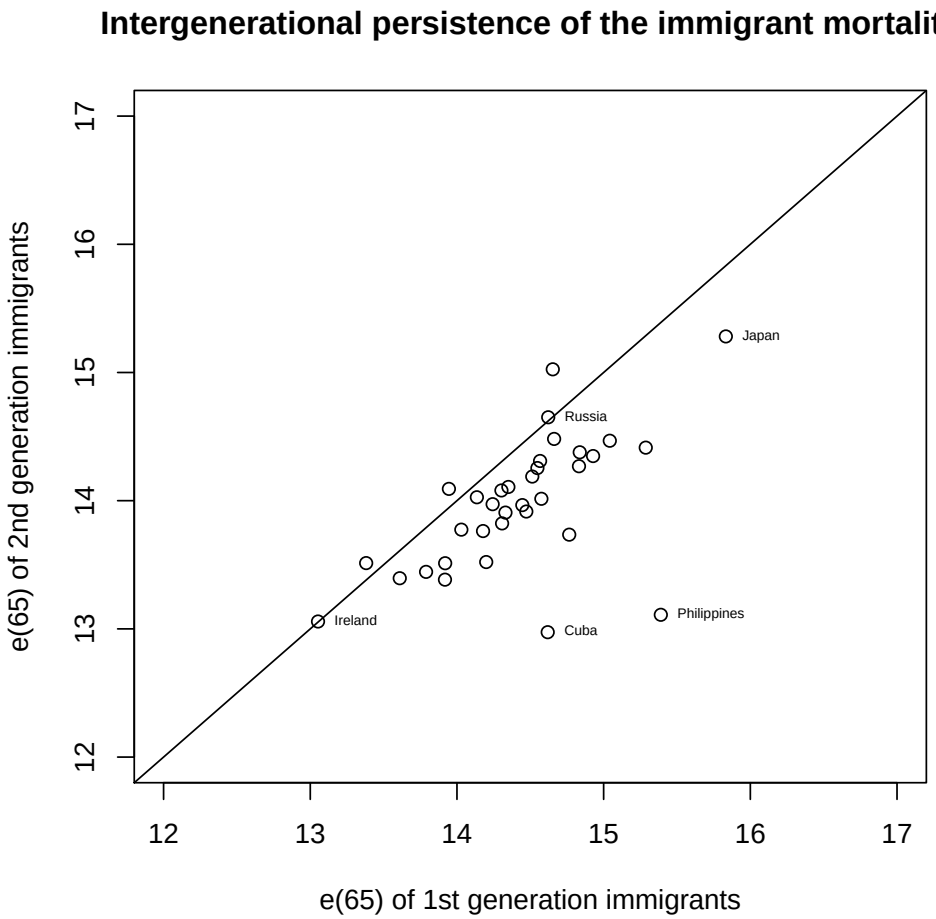

Figure 3: Intergenerational persistence of the immigrant mortality paradox. Regression estimates of life expectancy at age 65 by immigrant origin, 1st and 2nd generation males. See figure 2 for full set of labels and standard errors. The observed correlation is 0.60, which rises to 0.84 without Cuba and the Philippines.

and $r = 0.84$ without the highly-uncertain estimates from the Philippines and Cuba.

Finally, Figure 4 shows a preliminary analysis of the possible effect of socio-economic status on explaining the immigrant mortality paradox. The scale of these figures is given in terms of differences from "natives" in order to enhance comparability. The estimates control for years of education (dummy-coded) and urban/rural estimates, both of which have a powerful effect on expected mortality.

We see in the top panel that the immigrant advantage tends to grow larger when controls for socio-economic status are introduced, strengthening the paradox rather than explaining it. For most of the immigrant groups, the estimates with controls are larger in magnitude than without. This is particularly the case for Mexican immigrants. Interesting exceptions include the Japanese and the Irish for whom education and urban/rural residence are a (small) part of the explanation for their respective advantages and disadvantages.

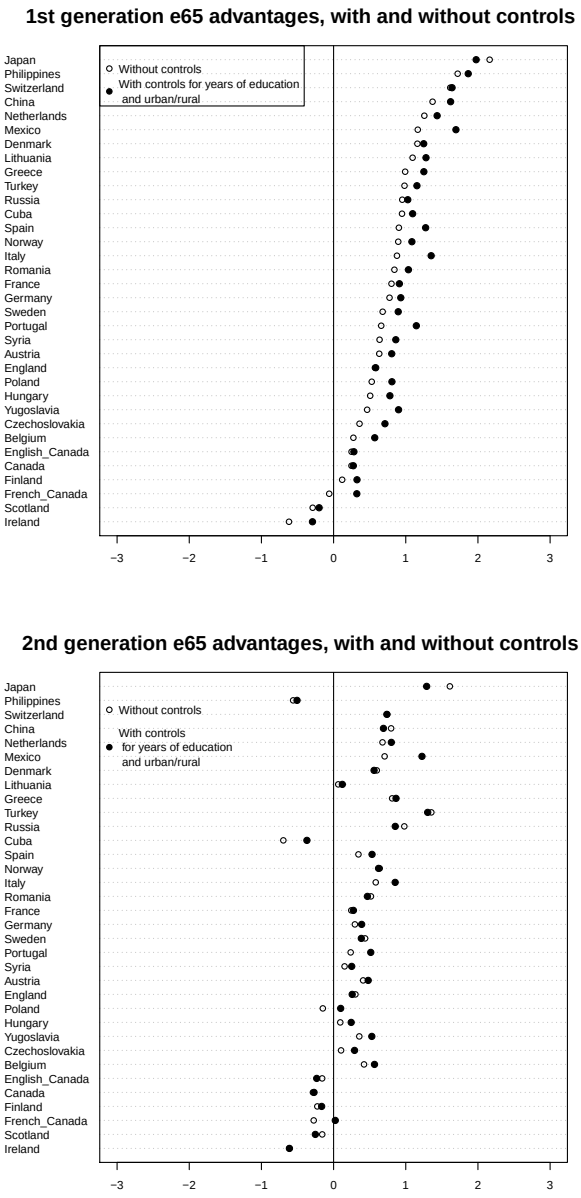

Figure 4: 1st and 2nd generation immigrant mortality, with and without controls. Note that the immigrant advantage increases in the 1st generation once controls are added but that there is little consistent effect in the 2nd generation.

The lower panel on the second generation tends to tell a less clear story. Much of this is due to the large degree of uncertainty in the estimates. However, the overall similarity of estimates with and without these controls tends to suggest that observable characteristics like education, which for some groups increased greatly from the 1st to the 2nd generation, does not appear to be an explanation of the persistence of immigrant group mortality advantages.

5 Brief Discussion of Preliminary Results

In our preliminary results we found a tendency across almost all groups for strong immigrant advantages in mortality, advantages that persisted from the 1st to the 2nd generation, and which appear to be difficult to explain using simple socio-economic variables.

The explanation that this persistent advantage suggests is that immigrant selection, at least selection based on the somatic health of the immigrant, can only be small part of the story behind immigrant advantage. Instead, there is some characteristic of immigrants that is passed intergenerationally from parents to children. Genetic inheritance is one possibility, but an unlikely one that does not appear in the literature. A more common explanation given for first generation advantage is that of "immigrant culture" and its effects on health behaviors – culture that could be passed down to subsequent generations. One of the interesting questions to investigate is whether, within immigrant groups, those who acculturate more lose their mortality advantages. This is a topic that can be investigated further with the CenSoc data.

6 Future Plans

For the PAA conference, the analysis will be extended in several directions.

First, links will be created from the 1940 to the 1930 census. This will allow a roughly 20-fold increase in the sample size of the 2nd generation, because in 1930 all respondents, not just a 1-in-20 sample, were asked to provide mother and father's birthplace. The 1930 census also includes year of immigration for the 1st generation and will allow us to look at immigrant advantage by age-at-entry into the United States.

Second, models will include a richer set of socio-economic controls in order to obtain more precise measurement of the differences between immigrant and native mortality among otherwise like individuals. Income and occupation as well as state and county of residence are additional control variables that can be included.

Third, I plan to explore the effect of cultural assimilation on immigrant mortality. The 1940 census allows measurement of whether immigrants live in the same neighborhood as other people from their country – forming "enclaves" which are thought to be protective. First name (and last names) also can provide insights into the cultural assimilation of immigrants and their children (Goldstein and Stecklov, 2016).

Fourth, it may be possible to extend the analysis of the 2nd generation to take into account both mothers' and fathers' birthplace. For the children of mixed couples, this allows the potential identification of some of the factors that may influence health behavior. For example, a stronger effect of mother's origin might suggest that diet plays a role, whereas a stronger effect of father's origin might suggest that other factors such as smoking might play a role.

Finally, it may be possible to analyse the mortality advantage of third-and-higher generations by identifying "natives" with ethnically distinctive last names. Some last names such as "O'Donnell" (Ireland), "Rossi" (Italy), "Tanaka" (Japan), "Carlson" (Sweden), "Pap-

pas" (Greece) are likely to be limited to those whose paternal lines descended from specific immigrant groups. The specificity of last names can be measured using the names of 1st generation immigrants by their place of birth. The name data in CenSoc may provide a means to look beyond the 2nd generation.

The documentation of mortality differences by country of origin and immigrant generation promises to give us insights into the nature of the immigrant mortality paradox and its possible causes.