Why Does the Health of Mexican Immigrants Deteriorate?
New Evidence from Linked Birth Records

Osea Giuntella*

University of Pittsburgh, IZA

January 1, 2017

Abstract

This study uses a unique dataset linking the birth records of two generations of children born in California and Florida (1970–2009) to analyze the mechanisms behind the generational decline observed in birth outcomes of children of Mexican origin. Calibrating a simple model of intergenerational transmission of birth weight, I show that modest positive selection on health at the time of migration can account for the initial advantage in birth outcomes of second-generation Mexicans. Moreover, accounting for the socioeconomic differences between second-generation Mexicans and white natives and the observed intergenerational correlation in birth weight, the model predicts a

*University of Pittsburgh: osea.giuntella@pitt.edu. I am grateful to Daniele Paserman, Randall Ellis, Kevin Lang, and Claudia Olivetti for their comments and advice. I am also thankful to Yasmine Serrano (Florida Department of Health) and Julie Turner (California Department of Public Health), who were extremely helpful with the data collection process. I would like to thank all the participants at the annual meetings of the American Economic Association (2014), the Society of Labor Economics (2013), the Association for the Study of Economics, Culture and Religion (2013), the Northeast Universities Development Consortium Conference (2012), the American Society of Health Economists Conference (2012), the XII Doctoral Workshop in Economic Theory and Econometrics (2012), the EconCon Conference (2012), the European Meeting of the Econometric Society (2012), the XV IZA Summer School (2012), and the European Association of Labor Economists/IZA Conference (2012), as well as all seminar attendees at the University of Connecticut, University of Lugano, Northeastern University, London School of Economics, University of Oxford, Universitat Pompeu Fabra, Syracuse University, University of Queensland, Singapore Management University, the BC-BU Green-Line Labor Meeting (2012), and the Boston University Empirical Micro-Lunch (2012). Any errors are my own. The project was made possible by generous funding from the Boston University Institute for Economic Development.

greater deterioration than that observed in the data. Using a subset of siblings and holding constant grandmother quasi-fixed effects, I show that the persistence of healthier behaviors among second-generation Mexican mothers can explain more than half of the difference between the model prediction and the observed birth outcomes of third-generation Mexicans.

Keywords: Hispanic health paradox, birth outcomes, risky behaviors

JEL Classification Numbers: I10, J15

1 Introduction

With a population of 31.8 million in 2010, Mexican descendants comprise 63% of the US Hispanic population and 10% of the total US population, with births overtaking immigration as the main source of population growth. Immigrants of Mexican origin constitute one the most disadvantaged immigrant groups in terms of socioeconomic status, and the pace of their earnings assimilation is considerably slower than that of other immigrant groups (Duncan and Trejo, 2011b). Mexicans are generally characterized by lower socioeconomic status compared to the natives, and hence, given the poorer average health conditions in Mexico [e.g., life expectancy and incidence of low birth weight (LBW), among others] and evidence of positive socioeconomic gradient in health, they can be expected to be at higher risk of negative health outcomes. However, although the Mexicans are one of the most disadvantaged immigrant groups in the country, a substantial body of research has documented them as healthier than the natives along several dimensions. Furthermore, despite positive socioeconomic assimilation, previous studies have shown that their initial health advantage deteriorates with the time spent in the United States and erodes in the next generation. For these reasons, previous scholars have referred to these stylized facts as the "Hispanic health paradox."

This apparent paradox has been observed in terms of general health status, life expectancy, mortality from cardiovascular diseases, cancer, age of puberty, and infant outcomes (Markides and Coreil, 1986; Antecol and Bedard, 2006; Bates and Teitler, 2008; Elder et al., 2012; Powers, 2013; Giuntella, 2016b), and has captured the attention of the media and policy makers (Tavernise, 2013). The goal of this paper is to shed light on the mechanisms underlying these facts.

Previous studies suggest that selection may explain the first-generation health advantage of immigrants (Jasso et al., 2004; Antecol and Bedard, 2006; Riosmena et al., 2013b). International migrants are not a random sample of their population of origin. Migrants frequently move in search of improved labor market opportunities and health status affects the perceived costs and benefits of migration; crossing the border might be more costly for unhealthy individuals. Moreover, healthy migrants might show higher returns to migration because health may enhance their earning capacities. While most of these studies show health selectivity as one of the main reasons for the Hispanic health paradox, only a few theoretical studies have formally investigated the relationship between health and the migration decision. A handful of empirical studies examining the healthy migrant hypothesis found evidence of a positive but mild selection on health (Crimmins et al., 2005; Barquera et al., 2008; Rubalcava et al., 2008; Ullmann et al., 2011; Riosmena et al., 2013b,a). However, researchers remain puzzled on what causes the subsequent health convergence observed in the second generation.

One possible explanation for these health patterns among the Mexican immigrant population is that health status is only weakly correlated across generations. Because of selection, first-generation immigrants have better health outcomes, but the second generation essentially loses all this initial advantage through a natural process of regression toward the mainstream average (Jasso et al., 2004). Previous studies have shown that negative health selection in return migration, the so-called salmon-bias hypothesis, could also contribute to the Hispanic health paradox (Riosmena et al., 2013a). In particular, Palloni and Arias (2004) suggested that a large part of the lower mortality rates among the Mexican population could be due to selective out-migration. However, Hummer et al. (2007) argue that selective out-migration may not explain the advantages observed in the health outcomes of second-generation children, particularly with regard to first-hour, first-day, or first-week mortality. Finally, several scholars examine the role of behaviors and offer evidence that immigrants initially exhibit fewer risk factors at the time of immigration but engage in riskier behavior once more time is spent in the United States and across generations (Marmot and Syme, 1976; Acevedo-Garcia et al., 2005; Antecol and Bedard, 2006; Fenelon, 2013).

There is a wide set of studies documenting the generational decline in the birth outcomes of Hispanics in the US (Acevedo-Garcia et al., 2004a, 2005). Giuntella (2016b) provides new descriptive evidence on the Hispanic health paradox in birth outcomes documenting important differences in the health trajectories of Hispanics of different origins. In a related paper, Giuntella (2016a) finds evidence of a negative relationship between cultural assimilation and health among second generation Hispanics. However, previous studies did not assess the role of selection and regression to the mean in explaining the unhealthy assimilation in adult second-generation and third-generation children.

The goal of this paper is to explain how much of the generational decline in the birth outcomes of immigrant descendants of Mexican origin can be explained by positive health selection at migration and a subsequent convergence to the average health in the population; and whether behavioral and socioeconomic factors mediate immigrant health trajectories. To this end, I present a simple model of selection and intergenerational health transmission to interpret the health trajectories of Mexican immigrants in the United States. Using the country-level differences in health outcomes, I demonstrate that a modest selection on health can explain why children born to first-generation Mexican women have better birth outcomes than those born to US-born women. This is consistent with the evidence of mild positive selection on health found in studies based on information collected in Mexico prior to migration (Crimmins et al., 2005; Barquera et al., 2008; Rubalcava et al., 2008; Ullmann et al., 2011; Riosmena et al., 2013b) and also with the mild positive selection of women on education (Chiquiar and Hanson, 2005; Moraga, 2011).

This paper exploits a large longitudinal intergenerationally linked data to analyze the birth outcomes of second- and third-generation Mexicans born in California and Florida, two of the top immigrant destination states in the United States, between 1970 and 2009. By linking the birth records of two generations, I am able to investigate the factors affecting the generational decline of birth outcomes among Mexican immigrant descendants. In order to verify whether the erosion of advantage in the third-generation birth outcomes can be explained through a simple process of regression toward the mainstream average, I try to predict the expected incidence of LBW among third-generation birth outcomes based on the observed intergenerational correlation in birth weight and estimates on the intergenerational transmission of health status in the literature. By calibrating the differences in quality of health care based on differences in socioeconomic status, the model not only explains the entire paradox but also overpredicts convergence. Against the non-significant difference observed between third-generation immigrants and white natives, the calibration exercise predicts a fairly large health advantage for natives. In reality, when accounting for the relatively low rate of intergenerational transmission observed in the data and the relatively low socioeconomic conditions they face, third-generation Mexicans are found to exhibit better birth outcomes than that predicted by the model. In other words, accounting for the relatively weak intergenerational correlation in birth outcomes and the socioeconomic conditions of second-generation Mexicans compared with native whites, the calibrated model indicates that the third-generation birth outcomes could be even worse than what the data indicate. Thus, the new puzzle is to ascertain why third-generation birth outcomes do not deteriorate as rapidly as predicted.

From the matched birth records for California and Florida, I find that approximately half of the differences between the model prediction and data can be explained by the persistence of healthier behaviors among the pregnant women of Mexican origin. First-generation Mexican mothers have substantially lower risky behaviors (such as smoking and alcohol consumption) and health risk factors (hypertension) that are known to have serious effects on birth outcomes (Almond et al., 2005; Shireen and Lelia, 2006; Gonzalez, 2011; Kaiser and Allen, 2002; Forman et al., 2009).

Although the risk-factor behavior worsens between the first and second generations, Mexican mothers maintain a sizeable advantage in terms of lower incidence of health risk factors during pregnancy compared to white natives. In particular, second-generation Mexican mothers maintain substantially lower incidence of tobacco use during pregnancy, which is known to be the most modifiable risk factor for LBW (Almond et al., 2005; Currie and Schmieder, 2009). To account for potential endogeneity, I follow the Currie and Moretti (2007) strategy of linking siblings and consider grandmother quasi-fixed effects (QFE)

(Mundlak, 1978). Conditioning on tobacco use during pregnancy and accounting for the persistence in healthy behaviors explains at least 50% of the difference between model predictions and the data.

The paper is organized as follows. Section 2 briefly reviews the literature on immigrant self-selection on health. Section 3 discusses the theoretical framework. Section 4 presents the data. Section 5 discusses the calibration exercise. Concluding remarks are presented in Section 6.

2 Immigrant self-selection

A few papers have proposed a framework to analyze the importance of health selection in explaining the Hispanic paradox. Palloni and Morenoff (2001) propose a simple model of selection on health at migration and show that even a moderate degree of selection at migration can explain the second-generation advantage in birth outcomes. Following this argument, Jasso et al. (2004) suggest that immigrants might self-select on transitory health traits and that their health might naturally revert toward the average health of the population of origin.

Previous studies have found evidence of positive but mild selection on health (Crimmins et al., 2005; Barquera et al., 2008; Rubalcava et al., 2008; Ullmann et al., 2011; Riosmena et al., 2013b; Farré, 2013).[1] In particular, Rubalcava et al. (2008) used longitudinal data from the Mexican Family Life Survey and found weak support for the healthy immigrant hypothesis. More recently, Riosmena et al. (2013a) found evidence consistent with emigration selection in height, hypertension, and self-rated health, while Ro and Fleischer (2014) show evidence that recently migrated women are less likely to be obese than non-migrant women in Mexico. Note that as shown by Palloni and Morenoff (2001), even a modest selection on health can explain the healthier birth outcomes of second-generation immigrants in the United States.

[1] Halliday and Kimmitt (2008) investigate the impact of health on domestic migration in the United States and show that poor health is associated with less geographical mobility.

Thus, these studies provide some empirical support for the selection hypothesis as a plausible explanation for the initial health advantage observed among first-generation immigrants and their children compared with natives. However, these studies do not assess the role of selection and regression toward the mean in explaining the unhealthy assimilation in adult second-generation and third-generation children.[2]

3 Theoretical framework

Building on the above-mentioned research, I develop a simple theoretical framework to analyze the mechanisms driving these health trajectories. Since information available on the birth weight distribution in the country of origin is limited, I cannot directly estimate the original selection using birth records. However, I can calibrate the model from the differences in health outcomes observed between the United States and Mexico to identify the degree of selection among first-generation immigrants. Once the patterns of selection are assessed, I can predict the expected intergenerational trajectories in birth weight by using the observed intergenerational correlation in birth outcomes and estimates of intergenerational correlation in health status and by accounting for the less than full socioeconomic assimilation of second-generation Mexicans in the United States. The intuition behind this exercise is straightforward. Given the differences in distribution of birth weight and the estimated intergenerational correlation in birth weight, one can calibrate the differences from

[2] An extensive literature analyzes the selection of Mexican immigrants based on skills. In a seminal article, Borjas (1987) argued that immigrants from countries with relatively high returns to education and income inequalities tend to be selected from the lower half of the skill distribution in the sending country. However, this result only holds if the migration costs are constant across individuals. Chiquiar and Hanson (2005) show that if the migration costs decrease with education, migrants might be negatively or positively selected depending on the size of costs and shape of skill distribution. Using counterfactual wage densities, Chiquiar and Hanson (2005) provide evidence against the negative-selection hypothesis. Their results suggest that migrants are selected from the middle of Mexican earnings distribution, with evidence of positive selection for Mexican-born women. Similarly, Orrenius and Zavodny (2010), McKenzie and Rapoport (2010), and Kaestner and Malamud (2014) confirm positive or intermediate selection. Yet, several scholars have provided new evidence favoring the negative selection hypothesis (Ibarraran and Lubotsky, 2007; William and Peri, 2012; Moraga, 2011; Reinhold and Thom, 2012). However, since this paper focuses on pregnant mothers, it is noteworthy that all these studies find evidence of intermediate or mild positive selection on education among immigrant women.

one generation to the next. In the following subsection, I first consider a model that purposely ignores the role of behaviors to test whether selection alone can explain the observed intergenerational convergence in birth outcomes.

3.1 Health selectivity and migration

For the sake of simplicity, following Palloni and Morenoff (2001), I assume that, ceteris paribus, health is the only factor affecting migration.[3] From this assumption, I develop a simple model of health selectivity at migration. Let $h_{j,G_1} \sim N(\mu_{j,G_1}, 1)$ be the distribution of first-generation ($G1$) health in country j ($j = \{MX,\ US\}$, respectively, Mexico and U.S.), which is distributed as a random normal variable (μ_{j,G_1}, 1), and μ_{j,G_1} be the average health in country j. Parameter μ_{j,G_1} can be interpreted as the composite effect of genes, quality of health care, technology, socioeconomic environment, environmental conditions (pollution, altitude, etc.), and risk-factor behavior on health.[4]

The migration process can then be defined as follows. An individual from the source country ($j = MX$) will be able to migrate to the destination country ($j = US$) only if his or her health is above a certain threshold t_1. This may be formally represented as

$$Pr(Mig_j) \begin{cases} > 0 & \text{if } h_{j,G_1} \geq t_1 \\ 0 & \text{if } h_{j,G_1} < t_1. \end{cases} \tag{1}$$

Thus, the probability of migration is positive only if the health of the individual is above a certain threshold. The higher the threshold, the more selected is the sample of migrants. These relationships imply that the health of first-generation immigrants in the United States is distributed as a truncated normal random variable, $TN(\mu_{j,G_1}, 1, t_1)$. Thus,

[3] This is a highly stylized model because other factors such as motivation, ability, and education affect the decision to migrate. Yet, this simple model captures the idea that, everything else remaining constant, better health will both reduce the cost of migration and increase the returns to migration.

[4] Because the variance σ_{j,G_1} in birth weight of the two populations differs by less than 0.6%, without loss of generality, I can assume that the distribution of health has identical variance in the two populations considered: σ_{MX,G_1}=σ_{US,G_1}=1.

the birth weight of the second generation (G_2)—the children of first-generation Mexican mothers—is determined as follows:

$$BW_{MX,G_2} = \gamma h_{MX,G_1} + v_{MX,G_2}, \tag{2}$$

where BW_{MX,G_2} is the birth weight of the second generation, h_{MX,G_1} captures maternal health, v_{MX,G_2} is distributed as a random normal variable (0, σ_v^2) reflecting the effect of other unobservable factors on birth weight, and γ captures the effect of maternal health on the child's health. Thus, for second-generation Mexicans, the CDF of BW_{MX,G_2} will be given by the sum of the truncated normal and a random normal variable:

$$BW_{MX,G_2} = H_1 + V_2, \tag{3}$$

where

$$H_1 = \gamma h_{jt} \sim TN(\gamma\mu_{MX,G_1}, \gamma^2, \gamma t_1) \tag{4}$$

$$V_2 = v_{G_2} \sim N(0, \sigma_v^2). \tag{5}$$

Following Turban (2010) and Azzalini (2005), I can show that the birth weight of second-generation Mexicans is distributed according to density as follows:

$$f_{MX}(bw) = \eta e^{-\frac{(bw-\lambda)^2}{2(\sigma_v^2+\gamma^2)}}[\Phi(\frac{bw - t_1 - \alpha}{\beta})], \tag{6}$$

where $\lambda = \gamma\mu_{MX,G_1}$, $\alpha = \frac{\sigma_v^2(bw-\lambda)}{\sigma_v^2+\gamma^2}$, $\beta^2 = \frac{\sigma_v^2+\gamma^2}{\sigma_v^2\gamma^2}$, $\eta = \frac{\sqrt{2\pi}\beta}{2\pi\sigma_v\gamma(\Phi(d))}$, $d = \frac{\lambda - t_1}{\gamma}$, and Φ is the cumulative distribution function (CDF) of a standard normal distribution. For children of US-born women ($j = US$), since no immigration occurs, the CDF of birth weight will be given by the sum of two normal distributions.

Given the distribution of BW_{j,G_2} ($j = MX, US$), the incidence of LBW is determined as follows:

$$LBW_{j,G_2} = p(bw \leq t_2) = \int_{-\infty}^{t_2} f_j(bw)dbw, \tag{7}$$

where t_2 represents the threshold defining LBW.

By setting the key parameters of the model defining the LBW threshold and the process governing the intergenerational transmission of birth weight (t_2, μ_{j,G_1}, and γ), I can straightforwardly compute the expected gap in LBW between second-generation Mexicans and children of US native women for any level of selection (t_1) (see Section 5).

3.2 Unhealthy assimilation

Next, I examine whether the relatively weak intergenerational correlation in birth weight and the less than full socioeconomic assimilation of second-generation Mexicans might explain the convergence observed in the data. For any level of selection at migration, one can predict the expected birth outcomes of the third generation. Third-generation birth outcomes can be described as a function of second-generation health characteristics and other factors. Let the health of second-generation Mexicans be defined as

$$h_{j,G_2} = \rho h_{j,G_1} + u_{G_2}, \tag{8}$$

where h_{j,G_2} is the health of second-generation mothers, u_{G_2} is distributed as a random normal variable (μ_{j,G_2}, $\sigma^2_{u_{G_2}}$) reflecting the effect of other unobservable factors on the health of second-generation mothers, and ρ measures the degree of intergenerational correlation in health between the first and second generations. Note that while μ_{j,G_1} captures the original difference in health distributions between Mexico and the United States, including the differences in contextual factors, quality, and access to care between the two countries, μ_{j,G_2} simply reflects the socioeconomic disparities between second-generation Mexicans and US natives. In other words, if second-generation Mexicans economically assimilate with the US natives, μ_{MX,G_2} will be set equal to μ_{US,G_2}, capturing the assimilation to the mainstream average. If the distribution of health is stable, the following equation holds:[5]

$$\sigma^2_{h_{G_1}} = \sigma^2_{h_{G_2}} = \sigma^2_{u_{G_2}} + \rho^2 = 1. \tag{9}$$

The birth weight of the third generation can be expressed as a function of maternal health:

$$BW_{G_3} = \gamma h_{G_2} + v_{G_3} = \gamma(\rho h_{j,G_1} + u_{G_2}) + v_{G_3}, \tag{10}$$

where BW_{G_3} is the birth weight distribution in the third generation and v_3 is distributed as a random normal variable $(0,\, \sigma^2_v)$ reflecting the effects of other unobservable factors on birth weight.

Assuming that the unobserved random shocks to health and birth weight are not correlated, the covariance between birth weights of the two generations can be rewritten as follows:[6]

$$\text{Cov}(BW_{G_3}, BW_{G_2}) = \text{Cov}(\gamma h_{G_2} + v_{G_3}, \gamma h_{G_1} + v_{G_2}) = \gamma^2 \rho, \tag{11}$$

implying the following:[7]

$$\rho = \frac{\text{Cov}(BW_{G_3}, BW_{G_2})}{\gamma^2} = \frac{\rho_{bw}}{\gamma^2}. \tag{15}$$

[5]Note that the incidence of LBW among natives is stable at approximately 6% throughout the period analyzed.

[6]Allowing for correlation between u_{j,G_2} and v_{G_2} and imposing reasonable restrictions on the model such that $\gamma \geq \rho$, the model generates qualitatively similar results. In particular, for $\rho_{u,v} > 0$, a higher but relatively modest selection on health (with the health threshold percentile ranging between 0.13 and 0.25) can explain the initial advantage of the second generation. The process of regression to the mean predicts greater deterioration in birth outcomes of Mexican descendants than that observed in the data. The distance between model prediction and the data ranges between 0.5 and 1.5 percentage points depending on $\rho_{u,v}$.

[7]Without loss of generality, one can select units of birth weight such that for both generations,

$$\sigma^2_{BW} = \gamma^2_h + \sigma^2_v = 1 \tag{12}$$

and therefore,

$$Cov(BW_{G_3}, BW_{G_2}) = Corr(BW_{G_3}, BW_{G_2}) \tag{13}$$

The foregoing implies that equation ρ can be expressed as a function of ρ_{bw} and γ.

$$\rho = \frac{Corr(BW_{G_3}, BW_{G_2})}{\gamma^2} = \frac{\rho_{bw}}{\gamma^2}\ : \tag{14}$$

Analogous to what is shown for the second generation, the CDF of $BW_{j,G3}$ is given by the sum of truncated normal and a random normal variable:

$$H_2 \sim TN(\lambda, \gamma^2\rho^2, \gamma\rho t_1) \tag{16}$$

$$V_3 \sim N(0, \varepsilon^2), \tag{17}$$

such that the birth weight of third-generation Mexicans is distributed according to the density:

$$f(bw) = \eta e^{-\frac{(bw-\lambda_2)^2}{2(\varepsilon^2+\gamma^2\rho^2)}}[\Phi(\frac{bw - t_1 - \alpha}{\beta})], \tag{18}$$

where $\lambda_2 = \gamma\rho\mu_{j,G_1} + \gamma\mu_{j,G_2}$, $\varepsilon^2 = \gamma^2(1-\rho^2) + \sigma_v^2$ and Φ is the CDF of the standard normal distribution. Given the distribution of BW_{G_3}, the incidence of LBW is determined as follows:

$$LBW_{G_3} = \int_{-\infty}^{t_2} f(bw)dbw, \tag{19}$$

where t_2 represents the LBW threshold. As in the previous section, the predicted gap in LBW for the third generation is easily derived by subtracting the observed incidence of LBW among children whose mothers and grandmothers were US born. Again, the CDF of birth weight of US-born natives will be given by the sum of two normal distributions.[8]

4 Data

To calibrate the model and compare its predictions with actual figures, I use data from three main sources: the California and Florida birth records, the Natality data, and the Current Population Survey.

[8] It is not surprising that substantially identical results are found with simulations with a population of 10,000 potential Mexican migrants who receive a random value for their health at migration drawn from a normal distribution with a mean identical to the mean of their country of origin.

4.1 California and Florida Birth Records

The primary data used in this paper are from the Birth Statistical Master File provided by the Office of Vital Records of the California Department of Health and the Birth Master Dataset provided by the Bureau of Vital Statistics of the Florida Department of Health. Data were obtained from the Florida Department of Health for the years 1970-2009 and from the California Department of Public Health for the years 1970–1981 and 1989–2009.[9]. The data for 1970 on Florida do not include information on the mother's country of origin.

Information on mother's country and state of birth, mother's first and maiden names, and child's full name, date of birth, gender, parity, race, birth weight, hospital of birth, and county of birth are available in both states for the full period considered. However, not all variables are available for each year and each of the two states. For instance, the mother's age is reported for the entire period in California, but only from 1989 in Florida, whereas the mother's education is reported for the entire period in Florida but only from 1989 in California. Information on birth weight is available for the entire period in both states, whereas other important heath measures at birth (Apgar score, gestational length, etc.) are unfortunately available only for the more recent years. While Almond et al. (2005) and Wilcox (2001) have reservations on the notion that birth weight has a causal effect on mortality "in particular" and infant health "in general," there is a general consensus that LBW is an important marker of health at birth and strongly associated with increased mortality and morbidity risk (Currie and Moretti, 2007; Conley and Bennett, 2000).[10] Because this study does not analyze the effects of birth weight and birth weight is the only measure of birth outcomes available for the entire period, I will primarily focus on birth weight and the incidence of low birth weight as indicators of health at birth. However, the deterioration

[9] Due to budget constraints I did not include the years 1982-1988 in California. However, including those years when using cross-sectional data from the Natality Detail Dataset yields similar results.

[10] LBW (<2500 grams or <5.5 pounds) is the single most important factor affecting neonatal mortality and is a determinant of post-neonatal mortality (Source: Center for Disease Control and Prevention (CDC), `http://www.cdc.gov/pednss/what_is/pednss_health_indicators.html`.). Infants weighing less than 2,500 grams are almost 40 times more likely to die during their first four weeks of life compared to infants of normal birthweight (Paneth, 1995).

of birth outcomes observed when analyzing the incidence of LBW is confirmed when using alternative measures (Apgar scores, infant mortality, etc.) of infant health for the years in which such metrics are available.[11] Following the earlier literature using administrative birth records (Fryer and Levitt, 2004; Currie and Moretti, 2007; Royer, 2009), I link the information available at a woman's birth to that of her children if the woman was born in California (Florida) and also gave birth in California (Florida).

In the sample, I define the first generation as all Mexican-born women who delivered in the United States between 1970 and 1985, the second generation as their children born between 1970 and 1985, and the third generation as their grandchildren born between 1989 and 2009. Children are classified according to their mother's country of origin, because information on the father's country of origin is not available prior to 1989 and often missing thereafter. Unfortunately, the data do not contain the year of immigration. The generational status depends only on the birthplace of first-generation mothers: all the children of Mexican-born women are considered second-generation Mexican; all the grandchildren of Mexican-born women are considered third-generation Mexican. Thus, the sample of second-generation children includes children born to first-generation Mexican women and US-born fathers. Second-generation Mexican children are compared to the children born to US-born white women between 1970 and 1985, and third-generation Mexican children are compared to the grandchildren of US-born white women born between 1989 and 2009. Thus, the sample of children born to US-born white women between 1970 and 1985 does not include second-generation Hispanics, but includes third-generation Hispanics. Consequently, the sample of white native children born between 1989 and 2009 does not include third-generation Hispanics, but includes fourth-generation Hispanics.

In the following analysis, I focus on white natives, because the Hispanic health paradox refers to the epidemiological finding that Hispanics tend to be paradoxically as healthy as, or in some cases healthier than, their US-born white counterparts (see for instance Riosmena et

[11] Results are available upon request.

al. (2013b); Fenelon (2013); Palloni and Morenoff (2001), etc.).[12] Since I focus on Mexican descendants, it is noteworthy that about 99.9% of first- and second-generation Mexicans reported a white race. Since I define the immigrants and natives based on their mother's country of origin, the sample of white natives born between 1970 and 1985 includes children of second-generation Hispanic white women, and the sample of white native children born between 1989 and 2009 includes children born to third-generation Hispanic white mothers.

Because of data limitations and since the information on father's ethnicity is limited and not available for all the years, children of mixed couples in the baseline specification are classified according to their mother's country of origin. However, I can identify the father's ethnicity from the father's surname and the latest list of US Census surnames,[13] which tabulates surnames classified by self-reported race/ethnicity. The list is based on 270 million individuals with valid surnames on the 2000 US Census (Word et al., 2008; Elliott et al., 2009).[14] Duncan and Trejo (2007, 2009, 2012, 2011a) point out that selective intermarriage and ethnic attrition can produce significant downward bias in tracking the socioeconomic success of US immigrant groups. Furthermore, in a companion paper (Giuntella, 2016a), I show that intermarriage and cultural assimilation importantly affect the birth outcomes of immigrant descendants. Therefore, as a robustness check, using the information on father's surname, I conduct a separate analysis excluding mixed couples (see Table A.1). In practice, these estimates indicate generational change when ethnic attrition is ignored.

Similarly, by using the information on mothers' surnames, I can identify the mother's ethnicity and define different control groups. Table A.2 excludes children born to Hispanic

[12]Third-generation children of Mexican origin maintain a large advantage with respect to the African-American population (see Table A.3).

[13]Data are available at `http://www.census.gov/topics/population/genealogy.html`. Last accessed on June 20, 2014. Because information on the father's last name is unavailable in California (1971–1980), I proxy it with the child's last name. From the 2000 Census Bureau's surname list, I consider as Hispanic any surname with a posterior probability of being associated with Hispanic ethnicity higher than 0.7, but the results are robust to the use of different thresholds.

[14]Recent studies (Duncan and Trejo, 2007, 2011a) assessed the importance of selective ethnic attrition among Mexican Americans. Spanish surnames, although imperfect, provide a more objective measure of ethnic origin and allow for the identification of individuals of Hispanic origin who did not self-report as Hispanic.

mothers from the white native children group and compares second- and third-generation Mexican children with children born to non-Hispanic whites.[15]

4.1.1 Matching and Selection

For the intergenerational sample, I linked the records of all infants born between 1989 and 2009 whose mothers were born in California or Florida between 1970 and 1985 to the birth records of their mothers. By using information on the mother's first and maiden names and date and state of birth, I could link 87.5% (96.6%) of the children born in California (Florida) of all the children born between 1989 and 2009 and whose mothers were born in California (Florida) between 1970 and 1981 (1971 and 1985) to the information available on the birth records of their mothers. I cannot match the observations of those whose names were misspelled or changed across birth certificates or whose dates of birth were misreported or could not be uniquely identified with the information available.

Table 1 presents the matching rates for the main racial and ethnic groups in the sample. The linked sample is not representative of women (men) born between 1970 and 1985. The final sample includes 1,355,896 (46%) of the 2,952,909 female children born between 1970 and 1985 in California and Florida. This reflects the reality that not all women born in California and Florida between 1970 and 1985 were still living in those states between 1989 and 2009, and that not all these women became mothers before 2009. In particular, the Natality Detail Data, which contain information on the mother's and child's state of birth, indicate that approximately 13.2% of women born in California and Florida between 1970 and 1985 had a child in a different US state before 2004 (the last year for which information on the state of birth of both mother and child is available in this database). Furthermore, according to the American Community Survey (2010), approximately 37% of women born in

[15] The results from using information on mother's and father's ethnicity to exclude children of mixed couples from both Mexican descendants and white natives (see Table A.1) and Hispanics from the white natives (see Table A.2) are not substantially different from the main results presented in Table 2. The sample of higher-order (third and fourth generation) immigrant generations of Hispanics is too small for a similar analysis comparing second- and third-generation Mexican children with higher-order immigrant generations of Hispanic origin.

California and Florida between 1970 and 1985 did not have a child by 2009. Data problems such as misspelled or missing information for women (men) born between 1970 and 1985 account for the rest of the attrition.

While the matching rate for children of Mexican origin is 58%, it is 41% for children born to US-born white women. The matching rate also depends on the socioeconomic background, which is clearly associated with infant health, mobility, and age of mother at first birth. Children of first-generation mothers residing in poor zip codes (in the lowest income quartile) are more likely to be linked to the records of their offspring than children of first-generation mothers living in wealthier zip codes (in the highest income quartile).

To address the selection bias concern arising from the matching process selecting women who both were born and have given birth in either California and Florida, I verify the external validity of the results using data from the Natality Detail Data, which contain detailed data on all births in the United States and mother's birth place and state of residence. However, note that this analysis considers only the descendants of first-generation mothers who arrived in the United States before 1985 and does not generalize to more recent immigrant cohorts.[16]

4.2 Natality Data

Natality data are drawn from the National Vital Statistics System of the National Center for Health Statistics and provide demographic and health data for births occurring during the calendar year. The microdata are based on information abstracted from birth certificates filed in the vital statistics offices of each state and district of Columbia. Demographic data include

[16] Before the Immigration Reform and Control Act (IRCA) of 1986, which required employers to attest to their employees' immigration status and made it illegal to hire or recruit illegal immigrants knowingly, a large proportion of Mexican-born women were unauthorized immigrants. As previous studies have shown, gaining legal status through IRCA had substantial effects on the labor market and some evidence of health effects (Phillips and Massey, 1999; Baker, 2010). Therefore, one may expect some variation in generational change in health among groups with different levels of exposure to authorization. Furthermore, studies have shown differences in the origin, skills, and "quality" of cohorts of Mexican immigrants since the mid-1980s (Borjas, 1985, 1995, 2015). Giuntella and Stella (2016) examining BMI find evidence of important cohort differences in both the extent of the immigrant health advantage upon arrival and the pace of convergence. Unfortunately, the data used in this paper do not allow for testing how these cohort differences affected the intergenerational transmission of infant health, although this is definitely a question that deserves further research.

variables such as date of birth, age, and educational attainment of parents; marital status; live-birth order; race; sex; and geographic area. Health data include items such as birth weight, gestation, prenatal care, attendant at birth, and Apgar score. The public version of the Natality Detail Data does not allow for cross-generational record linking, because the dataset does not release information on the names of child and mother. Geographic data include state, county, city, standard metropolitan statistical area (SMSA, 1980 onward), and metropolitan and non-metropolitan counties. From 2005 onward, the data do not include any geographic variables, such as state, county, or SMSA.

Using the Natality data, I can analyze cross-sectionally the entire United States for the 1970–2004 period and address the concern that the results obtained from the California and Florida data could suffer from selection bias due to attrition in the matching process.

4.3 Current Population Survey

In the calibration exercise (see Section 5.3), I account for the socioeconomic disparities between second-generation Mexicans and white natives using information from the Current Population Survey (CPS). The CPS is a monthly US household survey conducted jointly by the US Census Bureau and the Bureau of Labor Statistics. The CPS provides detailed information on sociodemographic characteristics and, since 1994, collects information on the country of birth of each respondent and his or her parents. Using this information, I define all foreign-born Mexicans arriving at age 16 or later as first generation, all US-born individuals with at least one parent born in Mexico as second-generation Mexicans, and all individuals with both parents born in the United States as US-born natives.

5 Calibration exercise

5.1 Parameterization

The model presented in Section 3.1 can be solved for different levels of selection t_1 by defining t_2 (the LBW threshold), μ_{j,G_1} (the mean of unobservable factors affecting the health of the first generation), and ρ (the estimated intergenerational correlation in health).

Table A.4 illustrates the parameterization of the model. Native health is used as a benchmark; thus, I set μ_{US,G_1}—the mean of unobservable factors affecting the health of US-born mothers—equal to 0. The LBW threshold, t_2, is set to -1.555 to match the average incidence of LBW observed in the data (0.06) over the entire period studied (1970–2009) throughout the US population (excluding African-Americans). However, choosing different low-birth weight thresholds (e.g., considering the average incidence of LBW in the entire population) yields similar results. While I consider different values of γ (see Section 5), the discussion of the model focuses on the predictions based on baseline specification (ρ=0.35), which implies that $\gamma = .76$. The mean of the unobservable factors affecting the health of first-generation Mexican mothers, μ_{MX,G_1}, is set to reflect the incidence of LBW in Mexico (10.6%, see Buekens et al. (2012)); γ is determined from eq. 15. As mentioned above, μ_{MX,G_1} captures the differences in socioeconomic, behavioral, and environmental factors affecting birth outcomes in Mexico and the United States.

Although I find no intergenerational correlation in health status (ρ) or effect of maternal health on birth outcomes (γ), I can directly estimate the degree of intergenerational correlation in birth weight (ρ_{bw}). Furthermore, from the existing estimates of intergenerational correlation in adult health outcomes, I can infer a plausible value range for ρ. Previous studies have estimated the intergenerational correlation in longevity and mortality as ranging between 0.2 and 0.3 (Ahlburg, 1998), the intergenerational correlation in Body Mass Index (BMI) as approximately 0.35 (Classen, 2010), and the intergenerational correlation in height as ranging between 0.4 and 0.5 (Bjorklund and Jantti, 2012). The intergenerational correlation between the birth weights of the mother and child is estimated at approximately 0.2 by Currie and Moretti (2007).[17]

Despite the different outcomes considered and methodologies employed to measure intergenerational correlations of health, most previous estimates of the intergenerational transmission of health range between 0.2 and 0.5. Therefore, I focus on the range of values of $\rho \in [0.2, 0.5]$. In the model, given equation 15, these restrictions on ρ indicate that γ must be in the range $\in [0.6, 1]$. Next, I discuss a model with baseline ρ equal to 0.35 (see Table 2); this corresponds to the middle of the plausible range for ρ. However, Table 3 illustrates the predictions of the model for different values of ρ and γ. Finally, I consider the mother-child intergenerational correlation in birth weight, ρ_{bw}, to be equal to 0.2, as estimated by Currie and Moretti (2007) and verified in my data.[18] [19]

5.2 Predicting second-generation ($G2$) birth outcomes

Using this parametrization and under the assumption of health having identical effects on birth weight in the two populations considered, the model can be solved semi-analytically for different levels of selection on health at migration (t_1). Figure 1 depicts the predicted differences in incidence of LBW between children of first-generation Mexican immigrants ($LBW_{MX,G2}$) and children of white natives ($LBW_{US,G2}$) by the extent of selectivity at migration. The x-axis describes the percentiles of the first-generation Mexican health distribu-

[17] Analyzing the relationship between parental and child outcomes, previous studies have estimated the intergenerational correlation in years of schooling as approximately 0.4 (Hertz et al., 2007) and in earnings as approximately 0.3 (Bjorklund and Jantti, 2009).

[18] As a caveat, note that intergenerational correlation may understate the true intergenerational transmission because it accounts for only one dimension of socioeconomic status (Bjorklund and Jantti, 2012). When analyzing sibling similarity rather than the correlation between parental and child outcomes, intergenerational correlations are somewhat higher (Bjorklund and Jantti, 2012; Mazumder, 2011; Halliday and Mazumder, 2014). For instance, Halliday and Mazumder (2014) account for the fact that health is a latent variable and estimate a sibling correlation of 0.5. If the sibling correlation is a lower bound of the lack of mobility (parameter ρ in the model), as argued by Bjorklund and Jantti (2012), then parameter ρ may be higher than 0.5.

[19] Note that the coefficient of intergenerational correlation is substantially similar across the two ethnic groups considered (0.19356 for white natives, and 0.19354 for Mexican descendants). As a robustness check, I consider a range of values for intergenerational correlation in birth weight ($\rho_{bw} \in [0.15, 0.5]$). The main implications of the model remain substantially unchanged.

tion corresponding to different values of the selection threshold (t_1). For any given level of threshold, the probability of migration will be positive only if an individual's health is above a given percentile of the population health distribution.

The predictions of the model are compared with the observed differences in the birth outcomes of Mexican descendants and white natives presented in Table 2, Panel A. Column 1 (2) presents the difference in the incidence of low birth weight between children of first-generation (second-generation) Mexicans and white natives for our main sample. In columns, 3-6 the sample is restricted to Florida. Columns 3 and 4 replicate the analysis presented in columns 1 and 2 in Florida. Column 5 includes a control for tobacco use during pregnancy. Column 6 includes further controls for socioeconomic status and family background (see Section 5.5). Panel B, adds gestational hypertension and alcohol use to the behaviors controlled for in column 5 and 6. Panel C documents the observed differences in maternal behaviors and Panel D illustrates the differences in socioeconomic status which are also used in the calibration exercise (see Section 5.3-5.4).

The dashed line marks the raw difference (-0.008) in LBW in the data between second-generation Mexicans and white natives (column 1, Table 2, Panel A). These differences control for child's state and year of birth fixed effects.[20] The figure suggests that the initial advantage can be entirely explained by relatively moderate selection. If Mexicans with health below the 13th percentile do not migrate to the United States, positive selection explains the lower incidence of LBW observed among second-generation Mexicans.

Unfortunately, no method exists to directly estimate the extent of selection on birth weight because, to the best of my knowledge, no dataset contains information on the conditions just prior to migration, while allowing for comparing the birth outcomes of migrants to the United States and the non-migrants who gave birth in Mexico. However, the modest selection implied by the calibration exercise is consistent with the evidence on both mild

[20]If sociodemographic controls are included, the coefficient increases, indicating a larger advantage in health at birth with respect to children of white US-born women with similar socioeconomic background (see Table A.8).

positive selection on education among women (Chiquiar and Hanson, 2005; Moraga, 2011), and mild positive selection on health (Riosmena et al., 2013a; Rubalcava et al., 2008).[21]

5.3 Accounting for socioeconomic assimilation

To predict third-generation birth outcomes, I first define $\mu_{MX,G2}$ as the mean of unobservable factors (u_{MX,G_2}) affecting second-generation health (h_{MX,G_2}). As mentioned above, μ_{US,G_2} is used as a benchmark and set equal to 0. I set parameter μ_{MX,G_2} to reflect the lower socioeconomic status of second-generation Mexicans with respect to white natives. To identify the effects of socioeconomic assimilation and account for the socioeconomic gradient in health, I rely on previous estimates of causal income effects on birth weight. Cramer (1995) finds that a 1% change in income-to-poverty ratio increases birth weight by approximately 1.05 grams. More recently, Almond et al. (2009) find similar marginal effects by analyzing the effect of food stamps on birth outcomes. Using the CPS data (1994–2009) for California and Florida, I estimate the average family income-to-poverty ratio among Mexicans as 49% lower than that among US natives (see Table 2, Panel D, column 2).[22] For comparability with the analysis based on vital statistics, I restrict the sample to white women between ages 15 and 44. Using the nativity status and country of origin information, which is available in the CPS for 1994 onward, I consider as first generation all women born abroad but reporting Mexican origin, and as second-generation immigrants, all individuals reporting Mexican origin but with at least one foreign-born parent. In the sample of US-born white women, I include all women with two US-born parents.[23]

To compute the income-to-poverty ratio, I consider the total family income. All estimates use household weight and include state and year fixed effects. Columns 3 and 4 report the

[21]The evidence of mild positive selection on height and other health outcomes has been documented by Crimmins et al. (2005), Barquera et al. (2008), Ullmann et al. (2011), and Riosmena et al. (2013b) using other anthropometric information for Mexicans living in the United States and Mexico.

[22]The earliest year for which information on the birthplaces of fathers and mothers is available in the CPS is 1994.

[23]In Table A.9, I exclude from the sample native women who had US-born parents but reported Hispanic ethnicity. The coefficients are only marginally different from those reported in Table 2, yielding substantially identical results in the calibration exercise.

same estimates only for Florida.[24] The Cramer (1995) estimate can be used to predict the expected differences in birth outcomes. Because of socioeconomic differences, ceteris paribus, the birth weight of Mexicans is on average 51 grams lower than that of natives. From these calculations, μ_{MX,G_2} is set equal to -0.1 (the native benchmark μ_{US,G_2} is equal to 0) to reflect the expected differences due to lower socioeconomic status of second-generation Mexicans with respect to the US-born white benchmark. I can thus derive the difference between the health distribution of second-generation Mexicans and US natives, assuming full assimilation with white natives on other unobservable characteristics affecting health (including behavioral risk factors), which is equivalent to assessing the implications of a regression toward the average health of Americans with similar socioeconomic background.

5.3.1 Predicting third-generation ($G3$) birth outcomes

With this parameterization, one can solve the model semi-analytically and obtain the incidence of LBW for the third generation. I can thus predict the expected differences in LBW between the children of second-generation Mexican immigrants ($LBW_{MX,G3}$) and white natives of the same cohort ($LBW_{US,G3}$) for any given level of initial selection.

Table 2 Panel D suggests that second-generation Mexicans assimilate with the first-generation but are not likely to be exposed to an identical quality of care or environmental and socioeconomic characteristics as those experienced by the average white native (Duncan and Trejo, 2011b). As mentioned above, the baseline model considers ρ equal to 0.35. To test for sensitivity of the analysis, Table 3 illustrates the model predictions for different values of ρ. By accounting for the socioeconomic gradient in health and the positive, but less than full, socioeconomic assimilation observed among second-generation Mexicans (see columns 1-4, Table 2, Panel D), the model not only explains the paradox but third-generation birth outcomes are predicted to be worse than they actually are. The model now predicts that third-generation Mexican should have an incidence of low birth weight that is approximately

[24] The results are identical when I include men or extend the analysis to all individuals in the working age (15–65); see Table A.10.

1.4 percentage points higher than natives (see Table 3). More generally, for values of $\rho \in [0.2, 0.5]$, the model consistently overpredicts convergence, and the distance between the model's prediction and data ranges between 1 ($\rho = .5$) and 1.9 percentage points ($\rho = .2$).[25] The results tend toward the same direction when the entire Hispanic group is considered or the socioeconomic information at the zip-code level is utilized.

5.4 Accounting for maternal risky behaviors

Thus far, I have not considered the role of risky behaviors. Abundant literature shows that risky behaviors affect health and birth outcomes. Administrative records provide only limited information on health behavior during pregnancy, and that too only for more recent years. Therefore, I cannot directly verify how intergenerational changes in significant risk factors, such as smoking during pregnancy, affect the intergenerational transmission of health at birth. However, I can provide cross-sectional evidence of differences between second-generation US-born Mexican immigrants and first-generation Mexicans. Information on adult behaviors and health conditions is limited in California, whereas the data on Florida report tobacco use, alcohol consumption, and weight gain during pregnancy from 1989 onward and gestational hypertension from 2004. Therefore, to analyze the role of behavioral assimilation, I focus on the Florida sample. In addition, I validate this analysis using the information on behaviors and risk factors given in the Natality Detail Data for the entire United States (see Table 4).

Table 2 Panel C illustrates the mean differences in incidence of these risk factors between first-generation Mexicans and white natives (column 3) and between second-generation Mex-

[25]Note that assuming no socioeconomic assimilation ($\mu_{MX,G_2} = \mu_{MX,G_1}$), the model would predict a much more rapid deterioration; the incidence of LBW among third-generation Mexican immigrants would thus rapidly converge to the levels observed in the non-migrant Mexican population. Conversely, assuming full socioeconomic assimilation and accounting for the persistent differences observed in behaviors ($\mu_{MX,G_2} = 0.1 >= \mu_{US,G_2}$), third-generation Mexican children would be expected to exhibit better statistics than natives for LBW, but they do not. In particular, the model predicts that if Mexicans were to maintain their lower incidence of smoking and fully assimilate socioeconomically, they would retain more than half (approximately 60%) of their initial advantage. This calculation is based on the expected advantage in birth outcomes due to lower incidence of tobacco use during pregnancy among first-generation Mexican women.

icans and white natives (column 4). All estimates include state and year fixed effects.

While first-generation immigrants show a substantially lower incidence of risk factors compared to white natives, second-generation immigrants exhibit some convergence toward less healthy behaviors and higher incidence of risk factors of natives, but retain a fairly sizeable advantage over the natives. When sociodemographic controls are included, the differences in risky behaviors persist. Table A.5 reports summary statistics by ethnic group.[26]

Overall, these differences remain similar when analyzing the Natality Detail Data (see Table 4, Panel C, and Table A.6). However, note that I cannot distinguish the second-generation from the higher-order generation immigrants in the Natality Detail Data, which likely explains the more marked deterioration in behaviors observed in column 2 of Table 4.

When controlling for tobacco use during pregnancy (Panel A), which, as mentioned earlier, is considered the most modifiable risk factor for LBW (Almond et al., 2005; Currie and Schmieder, 2009), the difference in LBW incidence of the third generation relative to the native benchmark ($LBW_{MX,G3}$-$LBW_{US,G3}$) increases from -0.001 to 0.011 (see column 5 of Table 2, Panel A). When alcohol use and gestational hypertension (Panel B) are included, the coefficient remains substantially unchanged.[27] In other words, accounting for tobacco use during pregnancy explains 80% (see Table 3, row 4, column 7) of the difference between the model prediction after accounting for socioeconomic differences (see Table 3, row 4, column 5) and the data (-0.001, column 2, Table 2).[28] Again, the results remain very similar when considering the Natality Detail Data. Controlling for risky behaviors, the difference in LBW incidence of the third generation ($LBW_{MX,G3}$-$LBW_{US,G3}$) increases from 0.001 to 0.008 (see

[26]When individual controls are included, the differences in risky behaviors for both first-generation and second-generation Mexicans substantially remain unchanged, but they become slightly smaller than those reported in Panel C of Table 2 when zip code fixed effects are included. Results are available upon request.

[27]When controlling for only alcohol use and gestational hypertension in separate regressions (as in Panel A for tobacco use), the difference in the incidence of LBW between third-generation Mexicans and white natives becomes marginally larger (0.004), but the coefficient is not substantially different from the unconditional difference reported in column 4 (0.003).

[28]More specifically, depending on whether the LBW differences in the United States or California and Florida samples are considered, controlling for behavior and health conditions helps us explain between 66% and 83% of the difference between model predictions and the data. Despite these differences, these results demonstrate that the model fits the pattern observed in the data fairly well once both the persistence in healthy behaviors and less-than-complete socioeconomic assimilation are accounted for.

column 3, Table 4, Panel A).

5.5 Within-family analysis

The extent to which health is transmitted from one generation to the next one depends on genetic factors and background characteristics that are not directly observable. To isolate the role of immigrant selectivity and partially account for potential unobservable genetic and family background characteristics, I include grandmother QFE (see Mundlak (1978)) and exploit the differences between siblings (within a family) in the covariates under analysis. As in any sibling study, there are both genetic and family background characteristics which may differ across siblings and confound the role of immigrant health selectivity, behavior and socio-economic assimilation (Currie and Moretti, 2007). I identify the siblings born between 1970 and 1985 using information on the maternal grandmother (mother's mother). To match grandmothers ($G1$) across the different birth certificates of their children ($G2$), I use information on the grandmother's name, child's last name, mother's race, and mother's state of birth; thus, children born to the same mother but of different fathers would not be considered in my sample of siblings. I drop individuals for whom the matching variables are missing.[29] As the conditional difference in the incidence of LBW is time-invariant across individuals sharing the same grandmother, to account for grandmother fixed effects, I follow Mundlak's approach (Mundlak, 1978) and use QFE. In practice, I use random-effects regression and include the group mean of independent variables in the model. Formally, I estimate the following model:

$$LBW_{3,t} = \beta_1 MX_{1,t} + \beta_2 LBW_{2,t} + \gamma X_{2,t} + \hat{X}_{GM} + \epsilon_{3,t}, \tag{20}$$

[29] With regard to the matching of mothers to grandmothers in California, I matched only one daughter in 84% of the cases, two daughters in 12% of the cases, and three or more daughters to each grandmother in 4% of the cases. In Florida, I matched only one daughter in 80% of the cases, two daughters in 17% of the cases, and three or more daughters to each grandmother in approximately 3% of the cases. Over the entire sample, the average number of children matched to each mother is 1.91, and the average number of grandchildren linked to each grandmother is 2.50; this figure would be 4.20 if it were conditioned on linking at least two second-generation sisters to their offspring.

where $LBW_{3,t}$ is an indicator of whether the third-generation child was born with a birth weight below 2,500 grams. $MX_{1,t}$ is a dummy equal to 1 when the first-generation grandmother was born in Mexico. I include controls for adequacy of prenatal care, gender, a quadratic in mother's and father's age; an indicator for whether the mother was born with LBW; an indicator for reported marital status; dummies for number of children; four dummies for education (high-school drop-out, high-school graduate, college, college and more); an indicator for plural birth, parity[30], and mother's and child's year of birth fixed effects; $\gamma X_{2,t}$, an indicator for mother born with LBW ($LBW_{2,t}$); and group means of the second-generation characteristics varying within the grandmother group ($\hat{X}_{GM}$).[31] Controlling for the birth weight of the second-generation mother ($LBW_{2,t}$) and including grandmother QFE ($\hat{X}_{GM}$) allows me to control for genetic and environmental factors that are constant within the family. In particular, by controlling for the common characteristics of mothers (sisters) and comparing the birth outcomes of third-generation cousins, I can partially capture the effect of characteristics related to migrant selection that might also affect the intergenerational transmission of healthy behaviors.

Column 6 in Table 2 indicates that when partially controlling for family unobserved heterogeneity and other sociodemographic characteristics, accounting for the tobacco use during pregnancy can explain approximately 64% of the difference between the model and the data compared to the 80% found without controlling for grandmother QFE (see Table 3, row 4, column 9). Again, when including alcohol use and gestational hypertension (Panel B), the coefficient remains substantially unchanged.

Table 3, columns 8 and 9 indicate that when considering values of $\rho \in [0.2, 0.5]$, risky

[30] The number of times a woman has given birth.

[31] I include group means (at the grand-mother level) for second-generation birth weight, tobacco use, alcohol use, gestational hypertension, adequacy of prenatal care, marital status, age, parity, and education (four groups).

behaviors explain at least 50% of the difference between model predictions and the data.[32][33]

Notably, I cannot account for a limited set of behaviors for which data are available. Dietary practices have been shown to be significant determinants of birth outcomes. In particular, fruit and vegetable consumption has been shown to be important, and their intake has been shown to differ significantly across ethnic groups (Guendelman and Abrams, 1995). Therefore, the unexplained difference between model predictions and the data is likely related to other types of behavior, such as dietary habits, for which I do not have data but which are known to significantly affect birth outcomes.

Overall, the model suggests that a combination of selection and persistence in lower incidence of health risk factors can explain the Hispanic paradox in LBW fairly well. A modest selection on health at migration can account for the second generation's advantage in birth outcomes. As a natural process of regression toward the mainstream average and less than complete socioeconomic assimilation, third-generation birth outcomes would be expected to deteriorate to a greater extent than observed in the data. The persistence of healthy behaviors during pregnancy can explain a substantial portion of the difference between model predictions and the data.[34]

[32] As a robustness check, I considered a broad range of values for the intergenerational correlation in birth weight ($\rho_{bw} \in [0.15, 0.5]$). The qualitative implications of the model are substantially unchanged and the magnitudes of predictions are only slightly different. A relatively modest selection on health can explain the initial advantage in the LBW of Mexican children ($LBW_{G_2,MX} - LBW_{G_2,US}$). The model overpredicts convergence for third-generation Mexicans ($LBW_{G_3,MX} - LBW_{G_3,US}$), with the difference between model predictions and the data ranging between .8 and 2 percentage points. Controlling for risky behaviors explains more than half of the difference between model prediction and the data, and at least 40% of it when including grandmother QFE.

[33] Table A.7 restricts the analysis to the linked sample presenting the birth outcomes of second generation only for the mothers I could link to the records of their own children. Results go in the same direction.

[34] The importance of socioeconomic factors and risky behaviors is confirmed in a companion paper (Giuntella, 2016a), which analyzes the differences in health convergence among second-generation Hispanics and shows that the birth outcomes of third-generation Hispanics in the United States importantly correlate with socioeconomic factors and behavioral assimilation.

6 Conclusion

This paper confirms that although second-generation Mexicans have lower incidence of LBW than the children of US-born white native mothers, this advantage declines substantially in the third generation. I then demonstrate that a modest selection on health can explain the better birth outcomes of second-generation Mexican children compared to the children of US white native women, and that, given the relatively weak intergenerational correlation in health status and birth weight, third-generation birth outcomes are predicted to deteriorate even more than what is observed in the data. By accounting for the socioeconomic differences between second-generation Mexicans and natives, the model not only explains the apparent paradox but also predicts that third-generation birth outcomes should be worse than what the data indicate. I show that approximately half of the difference between the model predictions and actual data is explained by the differences in tobacco use during pregnancy. Although evidence shows a generational deterioration with respect to engaging in risky types of behavior, second-generation pregnant women maintain a significantly lower level of risk-factor incidence than white natives. Between the first and second generations, behaviors do worsen, but to a limited extent. This holds even after accounting for potentially confounding factors by controlling for grandmother fixed effects.

This study revises the interpretation of the apparent Hispanic paradox in birth outcomes. Overall, the findings presented in the paper indicate that the health trajectories observed among Mexican descendants cannot be entirely explained by statistical selection and regression to the mean. While this paper shows that the birth outcomes of immigrant descendants would mechanically deteriorate as a consequence of selection and the natural convergence to the population average health, socioeconomic and behavioral factors mediate the transmission of health across generations.

The model presented in the paper could be used to analyze the health trajectories of immigrants in other contexts where it is not possible to directly observe selection on health at migration. Furthermore, as suggested by previous scholars (Marmot and Syme, 1976;

Fernandez, 2011), immigrant health trajectories provide the opportunity to study the effects of health transitions associated with a change in the environment that is similar to the one observed over a longer time in countries undergoing important epidemiological transitions and are therefore informative of the possible health risks associated with these transitions (Delavari et al., 2013).

Policies designed to reduce disparities in the access to health care and to maintain healthy behaviors may positively affect immigrant health trajectories mediating the natural process of convergence. In particular, the relevance of tobacco use during pregnancy suggests that policies aimed at counteracting the tobacco industry's increasing efforts to market cigarettes to Hispanic immigrants in the U.S. (Osypuk and Acevedo-Garcia, 2010; Acevedo-Garcia et al., 2004b) may contribute to reduce the risk of low birth weight among Mexican descendants. Because second-generation births are overtaking migration as the main source of growth in the American population, such policies could have important effects in promoting the health of future generations.

References

Acevedo-Garcia, Dolores, Elizabeth Barbeau, Jennifer Anne Bishop, Jocelyin Pan, and Karen Emmons (2004a) 'Undoing an epidemiological paradox: The tobacco industry's targeting of US industry.' *American Journal of Public Health* 14(12), 2188–2193

Acevedo-Garcia, Dolores, Elizabeth Barbeau, Jennifer Anne Bishop, Jocelyn Pan, and Karen M Emmons (2004b) 'Undoing an epidemiological paradox: the tobacco industry's targeting of US immigrants.' *American Journal of Public Health* 94(12), 2188–2193

Acevedo-Garcia, Dolores, Jocelyn Pan, Hee-Jin Jun, Theresa L. Osypuk, and Karen M. Emmons (2005) 'The effect of immigrant generation on smoking.' *Social Science and Medicine* 61, 1223–1242

Ahlburg, Dennis (1998) 'Intergenerational transmission of health.' *American Economic Review, Paper and Proceeding* 88(2), 265–270

Almond, Douglas, Hilary W. Hoynes, and Diane Whitmore Schanzenbach (2009) 'Inside the war on poverty: The impact of food stamps on birth outcomes.' *Review of Economics and Statistics* XCIII(2), 387–403

Almond, Douglas, Kennet Y. Chay, and David S. Lee (2005) 'The cost of low birth weight.' *Quarterly Journal of Economics* 120(3), 1031–1083

Antecol, Heather, and Kelly Bedard (2006) 'Unhealthy assimilation: Why do immigrants converge to american health status levels?' *Demography* 43(2), 337–360

Azzalini, Adelchi (2005) 'The skew-normal distribution and related multivariate families.' *Scandinavian Journal of Statistics* 32(2), 159188

Baker, Scott (2010) 'Effects of legal status and health service availability on mortality.' *Discussion Papers, Stanford Institute for Economic Policy Research* pp. 09–018

Barquera, S., R.A. Durazu Arvizu, A. Luke, G. Cao, and R.S. Cooper (2008) 'Hypertension in Mexico and among Mexican Americans: Prevalence and treatment patterns.' *Journal of Human Hypertension* 22, 617–626

Bates, Lisa M., and Julien O. Teitler (2008) 'Immigration and low birth weight in the US: The role of time and timing.' Working Papers 1085, Princeton University, Woodrow Wilson School of Public and International Affairs, Center for Research on Child Wellbeing., Jul. `http://ideas.repec.org/p/pri/crcwel/1085.html`. Accessed Nov. 2012

Bjorklund, Anders, and Markus Jantti (2009) 'Intergenerational income mobility and the role of family background, oxford handbook of economic inequality.' In *Oxford Handbook of Economic Inequality,* ed. W. Salverda, B. Molan, and T.M. Smeding (Chicago, IL: The Oxford University Press) pp. 491–591

— (2012) 'How important is family background for labor-economic outcomes?' *Labour Economics* 19, 464–474

Borjas, George J (1985) 'Assimilation, changes in cohort quality, and the earnings of immigrants.' *Journal of Labor Economics* pp. 463–489

Borjas, George J. (1987) 'Self-selection and the earnings of immigrants.' *The American Economic Review* 77(4), 531–553

Borjas, George J (2015) 'The slowdown in the economic assimilation of immigrants: Aging and cohort effects revisited again.' *Journal of Human Capital*

Borjas, GJ (1995) 'Assimilation and changes in cohort quality revisited: what happened to immigrant earnings in the 1980s?' *Journal of Labor Economics* 13(2), 201–245

Buekens, Pierre, Caitlin Canfield, Nicolas Padilla, Elia Lara Lona, and Rafael Lozano (2012) 'Low Birthweight in Mexico: A systematic review.' *Maternal and Child Health Journal* pp. 1–7

Chiquiar, Daniel, and Gordon Hanson (2005) 'International migration, self-selection, and the distribution of wages: Evidence from Mexico and the United States.' *Journal of Political Economy* 113(221), 239–281

Classen, Timothy J. (2010) 'Measures of the intergenerational transmission of body mass index between mothers and their children in the United States, 1981-2004.' *Economics and Human Biology* 8(4), 1074–3529

Conley, Dalton, and Neil G. Bennett (2000) 'Is biology destiny? Birth weight and life chances.' *American Sociological Review* 65(3), 458–476

Cramer, James (1995) 'Racial and ethnic differences in birth weight: The role of income and financial assistance.' *Demography* 32(2), 231–247

Crimmins, Eileen M., Beth J. Soldo, Jung Ki Kim, and dawn E. Alley (2005) 'Using anthropometric indicators for Mexicans in the United States and Mexico to understand the selection of migrants and the "Hispanic" paradox.' *Biodemography and Social Biology* 52(3-4), 164–177

Currie, Janet, and Enrico Moretti (2007) 'Biology as destiny? Short- and long-run determinants of intergenerational transmission of birth weight.' *Journal of Labor Economics* 25(2), 231–263

Currie, Janet, and Johannes Schmieder (2009) 'Fetal exposures to toxic releases and infant health.' *American Economic Review, American Economic Association* 99(2), 117–83

Delavari, Maryam, Anders Larrabee Sønderlund, Boyd Swinburn, David Mellor, and Andre Renzaho (2013) 'Acculturation and obesity among migrant populations in high income countries–a systematic review.' *BMC Public Health* 13(458), 2–11

Duncan, Brian, and Stephen J Trejo (2007) 'Ethnic identification, intermarriage, and unmeasured progress by mexican americans.' In 'Mexican immigration to the United States' University of Chicago Press pp. 229–268

_ (2009) 'Ancestry versus ethnicity: The complexity and selectivity of Mexican identification in the united states.' *Research in Labor Economics* 29, 31–66

_ (2011a) 'Intermarriage and the intergenerational transmission of ethnic identity and human capital for mexican americans.' *Journal of Labor Economics* 29(2), 195

_ (2011b) 'Low-skilled immigrants and the U.S. labor market.' *IZA Discussion Paper No. 5964.* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1929664. Accessed Nov. 2012

_ (2012) 'The complexity of immigrant generations: Implications for assessing the socioeconomic integration of Hispanics and Asians.' *IZA Discussion Paper No. 6276*

Elder, Todd E., John H. Goddeeris, and Steven J. Haider (2012) ‘Disparate disparities: Understanding differences in infant mortality across ratial and ethnic groups.’ https://www.msu.edu/~telder/DD_Current.pdf. Accessed Nov. 2012

Elliott, Marc N, Peter A Morrison, Allen Fremont, Daniel F McCaffrey, Philip Pantoja, and Nicole Lurie (2009) ‘Using the census bureaus surname list to improve estimates of race/ethnicity and associated disparities.’ *Health Services and Outcomes Research Methodology* 9(2), 69–83

Farré, Lídia (2013) ‘New evidence on the healthy immigrant effect.’ *IZA Discussion Paper*

Fenelon, Andrew (2013) ‘Revisiting the Hispanic mortality advantage in the United States: The role of smoking.’ *Social Science & Medicine* 82, 1–9

Fernandez, Raquel (2011) ‘Does culture matter?’ In *Handbook of Social Economics,* ed. Jess Benhabib, Matthew O. Jackson, and Alberto Bisin, vol. 1A (The Nederlands: North-Holland) pp. 481–510

Forman, J.P., M.J. Stampfer, and G.C. Curhan (2009) ‘Diet and lifestyle risk factors associated with incident hypertension in women.’ *Journal of the American Medical Association* 302(4), 401–411

Fryer, Roland G., and Steven D. Levitt (2004) ‘The causes and consequences of distinctively black names.’ *Quarterly Journal of Economics* 119(3), 767–805

Giuntella, Osea (2016a) ‘Assimilation and health. evidence from linked birth records of second and third-generation Hispanics.’ *Demography*

Giuntella, Osea (2016b) ‘The hispanic health paradox: New evidence from longitudinal data on second and third-generation birth outcomes.’ *SSM-Population Health* 2, 84–89

Giuntella, Osea, and Luca Stella (2016) ‘The acceleration of immigrant unhealthy assimilation.’ *Health economics*

Gonzalez, Carmen (2011) 'Diet for pregnancy and hypertension.' `http://www.livestrong.com/article/370938-diet-for-pregnancy-hypertension`. Accessed Nov. 2012

Guendelman, Sylvia, and Barbara Abrams (1995) 'Dietary intake among mexican-american women: generational differences and a comparison with white non-hispanic women.' *American Journal of Public Health* 85(1), 20–25

Halliday, Timothy, and Bhashkar Mazumder (2014) 'An analysis of sibling correlations in health using latent variable models'

Halliday, Timothy J., and Michael C. Kimmitt (2008) 'Selective migration and health in the USA.' *Population Studies* 62(3), 321–334

Hertz, Tom, Tamara Jayasudera, and Patrizio Piraino (2007) 'The inheritance of educational inequality: International comparison and fifty-year trends.' *The B.E. Journal of Economic Analysis & Policy*

Hummer, Robert A., Daniel A. Powers, Starling G. Pullum, Ginger L. Gossman, and W. Parker Frisbie (2007) 'Paradox found (again): Infant mortality among the Mexican-origin population in the United States.' *Demography* 44(3), 441–457

Ibarraran, Pablo, and Darren Lubotsky (2007) 'Mexican Immigration and Self Selection: New Evidence from the 2000 Mexican Census.' In *Mexican Immigration to the United States,* ed. George J. Borjas (Chicago, IL: University of Chicago Press)

Jasso, Guillermina, Douglas S. Massey, Mark R. Rosenzweig, and James P. Smith (2004) 'Immigrant health: Selectivity and acculturation.' *The Institute for Fiscal Studies Working Paper* 4(23), 1–46

Kaestner, Robert, and Ofer Malamud (2014) 'Self-selection and international migration: New evidence from mexico.' *Review of Economics and Statistics* 96(1), 78–91

Kaiser, Lucia Lynn, and Lindsay Allen (2002) 'Position of the American Dietetic Association: Nutrition and lifestyle for a healthy pregnancy outcome.' *Journal of the American Dietetic Association* 102(10), 1479 – 1490

Markides, K S, and J Coreil (1986) 'The health of Hispanics in the Southwestern United States: an epidemiologic paradox.' *Public Health Reports* 101(3), 253–265

Marmot, Michael G., and S.Leonard Syme (1976) 'Acculturation and coronary heart disease in Japanese-Americans.' *American Journal of Epidemiology* 104(3), 225–247

Mazumder, Bashkar (2011) 'Family and community influences on health and socioeconomic status: Sibling correlations over the life course.' *The B.E. Journal of Economic Analysis and Policy*

McKenzie, David, and Hillel Rapoport (2010) 'Self selection patterns in Mexico-U.S. migration: The role of migration networks.' *The Review of Economics and Statistics* 92, 811–821

Moraga, Jesus Fernandez-Huertas (2011) 'New evidence on emigrant selection.' *The Review of Economics and Statistics* 93(1), 72–96

Mundlak, Yair (1978) 'On the pooling of time series and cross-section data.' *Econometrica* 46, 69–85

Orrenius, Pia M., and Madeline Zavodny (2010) 'Self-selection among undocumented immigrants from Mexico.' *Journal of Development Economics* 78, 215–250

Osypuk, Theresa L, and Dolores Acevedo-Garcia (2010) 'Support for smoke-free policies: A nationwide analysis of immigrants, US-born, and other demographic groups, 1995–2002.' *American Journal of Public Health* 100(1), 171

Palloni, Alberto, and Elizabeth Arias (2004) 'Paradox lost: Explaining the Hispanic adult mortality advantage.' *Demography* 41(3), 385–415

Palloni, Alberto, and Jeffrey D. Morenoff (2001) 'Paradox lost: Explaining the Hispanic adult mortality advantage.' *Annals of The New York Academy of Sciences* 954, 140–174

Paneth, Nigel S (1995) 'The problem of low birth weight.' *The future of children* pp. 19–34

Phillips, Julie A, and Douglas S Massey (1999) 'The new labor market: Immigrants and wages after irca.' *Demography* 36(2), 233–246

Powers, Daniel A. (2013) 'Paradox revisited: A further investigation of racial/ethnic differences in infant mortality by maternal age.' *Demography* 50, 495–520

Reinhold, Steffen, and Kevin Thom (2012) 'Migration experience and earnings in the Mexican.' `https://files.nyu.edu/kt44/public/MigExperienceAndEarnings.pdf`. Accessed Nov. 2012

Riosmena, Fernando, Rebeca Wong, and Alberto Palloni (2013a) 'Migration selection, protection, and acculturation in health: a binational perspective on older adults.' *Demography* 50(3), 1039–1064

Riosmena, Fernando, Rebeca Wong, and Alberto Palloni (2013b) 'Paradox lost: Explaining the Hispanic adult mortality advantage.' *Demography* 50, 1039–1064

Ro, Annie, and Nancy Fleischer (2014) 'Changes in health selection of obesity among mexican immigrants: A binational examination.' *Social Science & Medicine* 123, 114–124

Royer, Heather (2009) 'Separated at girth: US twin estimates of the effects of birth weight.' *American Economic Journal: Applied Economics* 2(2), 49–85

Rubalcava, Luis N., Graciela M.Teruel, Duncan Thomas, and Noreen Goldman (2008) 'The healthy migrant effect: New findings from the Mexican family life survey.' *American Journal of Public Health* 98(1), 78–84

Shireen, Meher, and Duley Lelia (2006) 'Exercise or other physical activity for preventing pre-eclampsia and its complications.' *Cochrane Database of Systematic Reviews* (3), 1233–1241

Tavernise, Sabrina (2013) 'The health toll of immigration.' *The New York Times*

Turban, Sebastien (2010) 'Convolution of a truncated normal and a centered normal variable.' `http://www.columbia.edu/~st2511/notes/Convolution%20of%20truncated%20normal%20and%20normal.pdf`. Accessed Nov. 2012

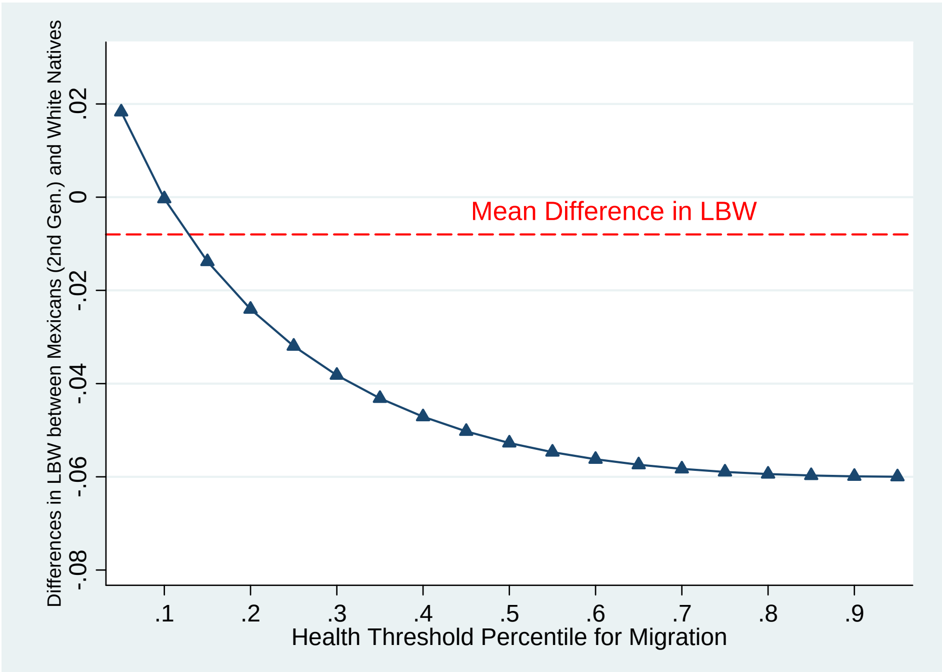
Ullmann, S.Heidi, Noreen Goldman, and Douglas Massey (2011) 'Healthier before they migrate, less healthy when they return?' *Social Science and Medicine* (73), 421–428

Wilcox, J. Allen (2001) 'On the importance and the unimportance of birth weight.' *International Journal of Epidemiology* 30(6), 1233–1241

William, Ambrosini J., and Giovanni Peri (2012) 'The determinants and selection of Mexico-US migrants.' *The World Economy* (35), 11–151

Word, David L, Charles D Coleman, Robert Nunziata, and Robert Kominski (2008) 'Demographic aspects of surnames from census 2000.' *Unpublished manuscript, Retrieved from http://citeseerx. ist. psu. edu/viewdoc/download*

Figure 1: Selection on health at migration and differences in the incidence of LBW between second-generation Mexicans and white natives ($\gamma = 0.75, \rho = 0.35$)



Notes - The plotted curve reports the predicted LBW differences between second-generation Mexicans and white natives for each level of selection on health at migration, assuming that the intergenerational correlation in health ρ is equal to 0.35 and the effect of maternal health on birth weight, γ, is equal to 0.75 (baseline). $\mu_{MX_{G_1}}$ is set equal to -0.42 to reflect the difference between the incidence of LBW in Mexico and in the United States (see Section 5.1). The dashed line describes the observed raw difference in incidence of LBW between second-generation Mexicans and white natives born between 1970 and 1985 in California and Florida (see Table 2, col. 1).

Table 1: Matching quality. Second-generation women born in California and Florida, 1970–1985 (G2)

	Observations			**Birth Weight (grams)**			**Low Birth Weight (below 2500 grams)**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Sample:	Overall	Linked	Matching rate	Overall	Linked	Nonlinked	Overall	Linked	Nonlinked
Overall	2,952,909	1,355,896	0.46	3,274	3,275	3,272	0.072	0.067	0.076
Us born whites	2,082,743	859,326	0.41	3,300	3,318	3,286	0.067	0.056	0.074
Mexicans	283,822	163,812	0.58	3,332	3,347	3,312	0.050	0.044	0.060
Zip code level income:									
1st income quartile	471,251	236,068	0.50	3,252	3,255	3,248	0.076	0.071	0.082
2nd income quartile	542,832	267,325	0.49	3,251	3,253	3,249	0.079	0.074	0.084
3rd income quartile	796,457	360,497	0.45	3,273	3,276	3,271	0.072	0.067	0.076
4th income quartile	700,271	296,500	0.42	3,299	3,300	3,298	0.064	0.059	0.068

Notes - Data are drawn from the California and Florida Vital Statistics, (1970−1985, 1989−2009). The linked sample is composed of all the women born between 1970 and 1985 for whom I was able to link the information available at their birth to the birth records of their children born in California and Florida between 1989 and 2009.

Table 2: Differences between 1st, 2nd generation Mexicans and white natives, California and Florida

	(1)	(2)	(3)	(4)	(5)	(6)
	CA-FL		**Florida**			
	$MX_1 - N$	$MX_2 - N$	$MX_1 - N$	$MX_2 - N$	$MX_2 - N$	$MX_2 - N$
Panel A: CA-FL Vital Statistics	G2	G3	G2	G3	G3	G3
Low birth weight	-0.008***	-0.001***	-0.005**	0.003	0.011***	0.009**
	(0.000)	(0.000)	(0.002)	(0.004)	(0.004)	(0.004)
Control for tobacco use	NO	NO	NO	NO	YES	YES
Grandmother QFE	NO	NO	NO	NO	NO	YES
Socio-demographic characteristics	NO	NO	NO	NO	NO	YES
Observations	4,613,690	2,023,959	1,309,282	384,373	384,373	384,373
	$MX_1 - N$	$MX_2 - N$	$MX_1 - N$	$MX_2 - N$	$MX_2 - N$	$MX_2 - N$
Panel B: CA-FL Vital Statistics	G2	G3	G2	G3	G3	G3
Low birth weight	-0.008***	-0.001***	-0.005**	0.003	0.011***	0.009**
	(0.000)	(0.000)	(0.002)	(0.004)	(0.004)	(0.004)
Control for risk factors	NO	NO	NO	NO	YES	YES
Grandmother QFE	NO	NO	NO	NO	NO	YES
Socio-demographic characteristics	NO	NO	NO	NO	NO	YES
Observations	4,613,690	2,023,959	1,309,282	384,302	384,302	384,302
	$MX_1 - N$	$MX_2 - N$	$MX_1 - N$	$MX_2 - N$	$MX_2 - N$	$MX_2 - N$
Panel C: FL Vital Statistics			G1	G2		
Tobacco consumption			-0.167***	-0.157***		
			(0.001)	(0.006)		
Observations			518,248	409,107		
Alcohol consumption			-0.003***	-0.004***		
			(0.000)	(0.001)		
Observations			519,590	411,172		
Gestational hypertension			-0.026***	-0.006		
			(0.001)	(0.005)		
Observations			228,883	174,331		
	$MX_1 - N$	$MX_2 - N$	$MX_1 - N$	$MX_2 - N$	$MX_2 - N$	$MX_2 - N$
Panel D: CA-FL CPS, 1994-2009	G1	G2	G1	G2		
Socioeconomic status	-0.761***	-0.489***	-0.702***	-0.568***		
log(family income/poverty)	(0.012)	(0.015)	(0.047)	(0.069)		
Observations	36,622	31,634	10,182	9,840		

Notes - Panel A and B document the observed differences in the incidence of low birth weight between Mexican descendants and white natives. Column 1 (2) presents the difference in the incidence of low birth weight between children of first-generation (second-generation) Mexicans and white natives for our main sample. In columns, 3-6 the sample is restricted to Florida. Columns 3 and 4 replicate the analysis presented in columns 1 and 2 in Florida. Column 5 includes a control for tobacco use during pregnancy. Column 6 includes further controls for socioeconomic status and family background (see Section 5.5). Panel B, adds gestational hypertension and alcohol use to the behaviors controlled for in column 5 and 6. Panel C documents the observed differences in maternal behaviors (see Section 5.4) and Panel D illustrates the differences in socioeconomic status (see Section 5.3). All estimates in Panel A-C include fixed effects for child's state and year of birth. Data used in Panel A and B are drawn from the California and Florida Birth Records (1970−1985, 1989−2009). Data on risk factors -tobacco use during pregnancy, alcohol use during pregnancy, and gestational hypertension- (rows 2−5) are drawn from Florida Birth Records (1989−2009). Information on gestational hypertension (row 5) is available from 2004 onwards. Note that data drawn from the California Vital Statistics do not contain information on these risk factors for the period under analysis. Data on socioeconomic assimilation are drawn from the Current Population Survey (CPS, 1994−2009). All estimates in Panel D include state and and year fixed effects. Information on parental birth place is available in the CPS only since 1994.

Table 3: Model parameters and predictions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Model Parameters		Second Generation (G2)	Third Generation (G3)	Third Generation (G3)				
			Selection Matching Initial Advantage	Predicted Gap	Distance between Model and the Observed Data (Unconditional)	(Conditional on risky behaviors)			
	ρ	γ_{bw}	$F(h_{MX,G1} < t_1^*)$	$LBW_{MX} - LBW_{US}$	β=-0.001	OLS $\beta = 0.011$	% Explained	QFE β=0.009	% Explained
	0.20	1.00	0.082	0.018	0.019	0.007	63%	0.010	50%
	0.25	0.89	0.095	0.017	0.018	0.006	67%	0.009	53%
	0.30	0.82	0.112	0.015	0.016	0.004	75%	0.007	60%
Baseline	*0.35*	*0.76*	*0.126*	*0.014*	*0.015*	*0.003*	*80%*	*0.006*	*64%*
	0.40	0.71	0.138	0.012	0.013	0.001	92%	0.004	75%
	0.45	0.67	0.151	0.011	0.012	0.000	103%	0.003	84%
	0.50	0.63	0.162	0.009	0.010	-0.002	120%	0.001	100%

Notes - For this calibration exercise the intergenerational correlation in birth weight ρ_{bw} is set equal to 0.2 (see Section 3.4.1). $\mu_{US,G1}$ and μ_{US,G_2} are set equal to 0 as a benchmark. $\mu_{MX,G1}$ is set equal to -0.42 to be such that the incidence of low birth weight in Mexico (see Section 5.1). μ_{MX,G_2} is set equal to -0.1 to reflect the lower socio-economic conditions of second-generation Mexicans (see Section 5.3). Column 1 reports different values of the intergenerational correlation in health $\rho \in [0.2, 0.5]$. The effect of maternal health on birth weight (γ_{bw}=$\frac{\rho_{bw}}{\rho})^{1/2}$) is reported in column 2. Column 3 reports the health threshold percentile (t_1^*) for which the model matches the observed data second generation children. Column 4 presents the difference in low birth weight between third-generation Mexicans ($LBW_{MX,G3}$) and natives predicted by the model when using the parameters of columns 1-3. In column 5, I report the difference between the model prediction and the unconditional mean difference presented in column 2 of Table 2. Columns 6 and 8 illustrate the distance between the model prediction and the data once we condition on observable behaviors and include grandmother quasi-fixed effects. Column 7 and 9 show how much of the gap reported in column 5 is explained by conditioning on risky behaviors and including grandmother quasi-fixed effects.

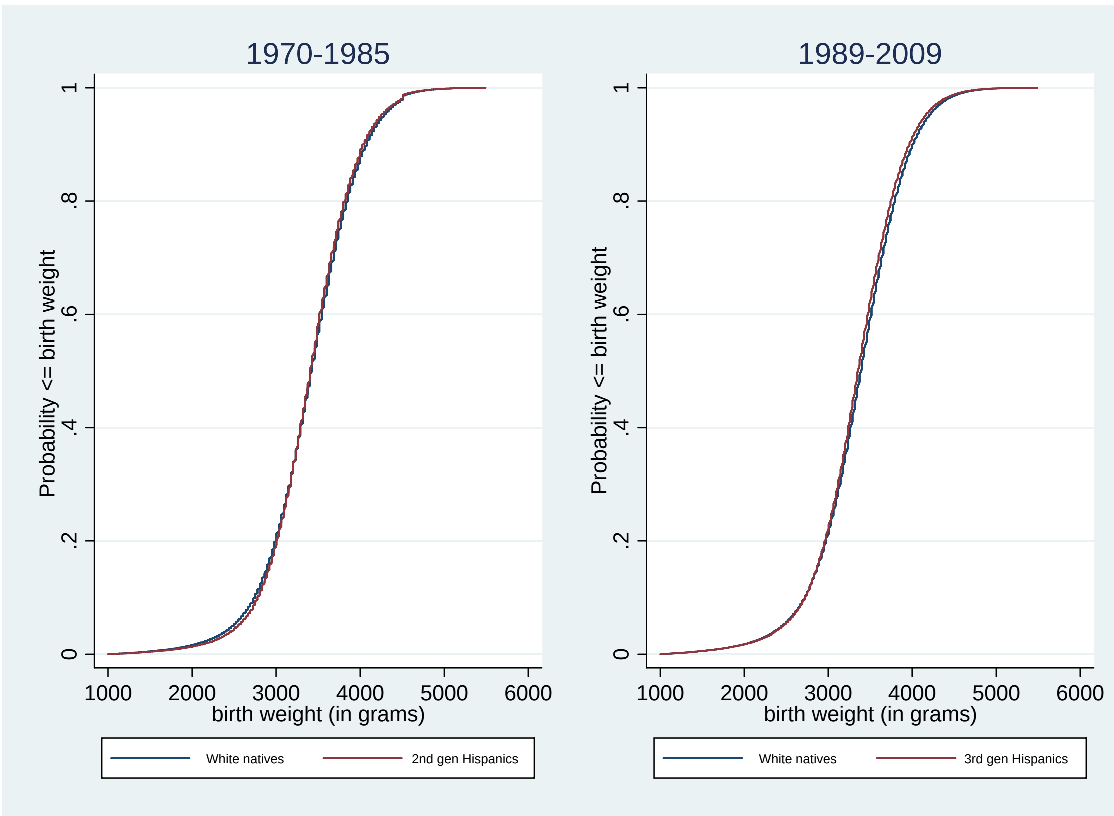
Table 4: Differences between 1st, 2nd generation Mexicans and white natives - U.S. Natality Detail Data

	(1) $MX_1 - N$	(2) $MX_2 - N$	$MX_2 - N$
Panel A: Natality Detail Data	G2	G3	G3
Low birth weight	−0.008***	0.001***	0.007***
	(0.000)	(0.000)	(0.000)
Observations	32,037,863	15,572,561	15,572,561
Control for tobacco use	NO	NO	YES
Panel B: Natality Detail Data	G2	G3	G3
Low birth weight	−0.008***	0.001***	0.008***
	(0.000)	(0.000)	(0.000)
Observations	32,037,863	15,521,502	15,521,502
Control for risk factors	NO	NO	YES
Panel C: Natality Detail Data	G1	G2	
Tobacco consumption	−0.143***	−0.120***	
	(0.000)	(0.000)	
	18,154,042	15,584,238	
Alcohol consumption	−0.007***	−0.003***	
	(0.000)	(0.000)	
	18,590,226	15,974,350	
Gestational hypertension	−0.014***	−0.008***	
	(0.000)	(0.000)	
Observations	22,159,647	18,081,805	
Panel D: CPS (1994-2009)	G1	G2	
Socioeconomic status log(family income/poverty)	−0.728***	−0.531***	
	(0.006)	(0.007)	
Observations	340,330	326,504	

Notes - Data used in Panel A are drawn from the Natality Detail Data (1970–1985; 1989–2004). Information on parental birth place is available in the CPS only since 1994. All estimates include fixed effects for child's state and year of birth. Note that Natality Detail Data does not allow to distinguish second or higher generation since it does not contain information on parental nativity of the mothers. Data on socioeconomic assimilation are drawn from the Current Population Survey (1994–2009). All estimates in Panel D include state and year fixed effects.

Appendix A

Figure 2: Birth weight (grams) distribution, California and Florida



Notes - Source: California and Florida Vital Statistics, 1970–1985, 1989–2009.

Table A.1: Differences between 1st, 2nd generation Mexicans and U.S. born whites (excluding Mixed Couples)

	(1)	(2)	(3)	(4)	(5)	(6)
	CA-FL		**Florida**			
	$MX_1 - N$	$MX_2 - N$	$MX_1 - N$	$MX_2 - N$	$MX_2 - N$	$MX_2 - N$
Panel A: CA-FL Vital Statistics	G2	G3	G2	G3	G3	G3
Low birth weight	-0.007***	-0.002***	-0.004	0.002	0.010**	0.011**
	(0.000)	(0.001)	(0.003)	(0.004)	(0.004)	(0.005)
Control for tobacco use	NO	NO	NO	NO	YES	YES
Grandmother QFE	NO	NO	NO	NO	NO	YES
Socio-demographic characteristics	NO	NO	NO	NO	NO	YES
Observations	4,050,035	1,569,384	1,249,493	349,059	349,059	349,059

Notes - All estimates include fixed effects for child's state and year of birth. Data are drawn from the California and Florida Birth Records (1970−1985, 1989−2009).

Table A.2: Differences between 1st, 2nd generation Mexicans and non-Hispanic white natives

	(1)	(2)	(3)	(4)	(5)	(6)
	CA-FL		**Florida**			
	$MX_1 - N$	$MX_2 - N$	$MX_1 - N$	$MX_2 - N$	$MX_2 - N$	$MX_2 - N$
Panel A: CA-FL Vital Statistics	G2	G3	G2	G3	G3	G3
Low birth weight	-0.008***	0.001	-0.005**	0.003	0.011***	0.009**
	(0.000)	(0.002)	(0.001)	(0.004)	(0.004)	(0.004)
Control for tobacco use	NO	NO	NO	NO	YES	YES
Grandmother QFE	NO	NO	NO	NO	NO	YES
Socio-demographic characteristics	NO	NO	NO	NO	NO	YES
Observations	3,560,302	1,498,611	1,109,223	324,412	324,412	324,412

Notes - All estimates include fixed effects for child's state and year of birth. Data are drawn from the California and Florida Birth Records (1970−1985, 1989−2009).

Table A.3: Differences between 1st, 2nd generation Mexicans and non-Hispanic black natives

	(1)	(2)	(3)	(4)	(5)	(6)
	CA-FL		**Florida**			
	$MX_1 - N$	$MX_2 - N$	$MX_1 - N$	$MX_2 - N$	$MX_2 - N$	$MX_2 - N$
Panel A: CA-FL Vital Statistics	G2	G3	G2	G3	G3	G3
Low birth weight	-0.074***	-0.061***	-0.071***	-0.055***	-0.053***	-0.108
	(0.001)	(0.003)	(0.001)	(0.004)	(0.006)	(0.007)
Control for tobacco use	NO	NO	NO	NO	YES	YES
Grandmother QFE	NO	NO	NO	NO	NO	YES
Socio-demographic characteristics	NO	NO	NO	NO	NO	YES
Observations	1,314,446	847,082	447,887	268,145	268,145	268,145

Notes - All estimates include fixed effects for child's state and year of birth. Data are drawn from the California and Florida Birth Records (1970−1985, 1989−2009).

Table A.4: Baseline Model Parameters

Parameters	Description	Parameter identification	Value
t_2	Low birth weight threshold	Based on observed incidence of low birth weight in overall sample	-1.555
μ_{US,G_1}	Unobservable factors affecting G1 native women	Normalized	0
μ_{MX,G_1}	Unobservable factors affecting the health of G1 Mexican mothers	Based on observed incidence of low birth weight in Mexico	-0.42
$\mu_{US,G2}$	Unobservable factors affecting G2 native women	Normalized	0
μ_{MX,G_2}	Unobservable factors affecting the health of G2 Mexican mothers	Based on estimated socio-economic gradient in birth weight (see section 5.3)	-0.1
ρ	Intergenerational correlation in health	Estimates from the literature (see also Table 3)	0.35
ρ_{bw}	Intergenerational correlation in birth weight	Directly estimated on full sample	0.2
γ	effect of maternal health on the child's health	$\rho = \frac{\rho_{bw}}{\gamma^2}$ (eq.11)	0.76

Table A.5: Risk Factors and Assimilation (%), Florida Vital Statistics (1989-2009)

Smoking during pregnancy	Alcohol use during pregnancy	Gestational Hypertension
	U.S. born whites	
16.96	0.42	5.32
(37.53)	(6.47)	(22.45)
	Non-Hispanic U.S. born whites	
17.57	0.43	5.38
(38.05)	(6.54)	(22.57)
	Non-Hispanic U.S. born blacks	
4.18	0.41	5.93
(20.01)	(6.45)	(23.63)
	First Generation Mexicans	
0.30	0.07	2.79
(5.55)	(2.76)	(16.47)
	Second Generation Mexicans	
1.30	0.048	4.69
(11.35)	(2.20)	(11.35)

Notes - Data on tobacco use during pregnancy, alcohol use during pregnancy, and gestational hypertension are drawn from Florida Birth Records (1989−2009). Information on gestational hypertension is available from 2004 onwards.

Table A.6: Risk Factors and Assimilation (%), Natality Detail Data (1989-2004)

Smoking during pregnancy	Alcohol use during pregnancy	Gestational Hypertension
	Non-Hispanic U.S. born whites	
21.05	1.01	4.32
(40.77)	(10.02)	(20.33)
	Non-Hispanic U.S. born blacks	
8.86	1.00	3.36
(28.42)	(10.12)	(18.80)
	First Generation Mexicans	
0.70	0.23	2.06
(8.39)	(4.77)	(14.20)
	*Second Generation Mexicans**	
5.35	0.87	2.95
(22.49)	(9.30)	(16.92)

Notes - Data are drawn from the Natality Detail Data (1989-2004). The sample is restricted to mothers born between 1970 and 1985 and includes 15,813,352 non-Hispanic US born mothers, 3,127,410 first-generation Mexican mothers, and 1,240,315 second-generation Mexican mothers. All numbers are expressed as percentages of total population. *In the Natality Detail Data, we cannot distinguish second-generation Mexicans from higher order generation immigrants.

Table A.7: Differences between 1st, 2nd generation Mexicans and U.S., Linked Sample

	(1)	(2)	(3)	(4)	(5)	(6)
	CA-FL		**Florida**			
	$MX_1 - N$	$MX_2 - N$	$MX_1 - N$	$MX_2 - N$	$MX_2 - N$	$MX_2 - N$
Panel A: CA-FL Vital Statistics	G2	G3	G2	G3	G3	G3
Low birth weight	-0.012***	-0.001***	-0.015***	0.003	0.011***	0.009**
	(0.001)	(0.000)	(0.002)	(0.005)	(0.004)	(0.004)
Control for tobacco use	NO	NO	NO	NO	YES	YES
Grandmother QFE	NO	NO	NO	NO	NO	YES
Socio-demographic characteristics	NO	NO	NO	NO	NO	YES
Observations	2,023,959	2,023,959	384,373	384,373	384,373	384,373

Notes - All estimates include fixed effects for child's state and year of birth. Data are drawn from the California and Florida Birth Records (1970−1985, 1989−2009).

Table A.8: Hispanic Health Paradox in birth weight (BW) and low birth weight incidence (LBW)

	2nd generation (G2) 1970–1985				**3rd generation (G3) 1989–2009**			
Dependent variable:	BW		LBW		BW		LBW	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Country of origin:								
Mexico	0.891	22.742***	-0.008***	-0.015***	-32.675***	-11.814***	-0.001**	-0.003***
	(0.895)	(1.044)	(0.000)	(0.000)	(1.939)	(1.496)	(0.001)	(0.001)
Sociodemographic controls	NO	YES	NO	YES	NO	YES	NO	YES
Mean of Dep. Var.	3,377.666	3,377.666	0.057	0.057	3,358.716	3,358.716	0.061	0.061
Std.Dev.	568.435	568.435	0.232	0.232	577.542	577.542	0.240	0.240
Observations	4,613,690	4,009,198	4,613,690	4,009,198	2,023,959	1,931,056	2,023,959	1,931,056

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Notes - Data are drawn from the California and Florida Birth Records (1970−1985, 1989−2009). All estimates include state and year fixed effects. Sociodemographic controls include child's gender, parity, type of birth, year of birth-fixed effects, mother's age dummies, father's age (quadratic), mother's marital status, an indicator of adequacy of prenatal care, mother's education (4 groups dummies), father's education (4 group dummies), zip code–fixed effects, and indicators for missing variables: mother's age, father's age, mother's education, father's education, marital status, parity.

Table A.9: Socio-economic differences between 1st, 2nd generation Mexicans and non-Hispanic white natives

	(1)	(2)	(3)	(4)	(5)	(6)
	US		**CA-FL**		**Florida**	
	$MX_1 - N$	$MX_2 - N$	$MX_1 - N$	$MX_2 - N$	$MX_1 - N$	$MX_2 - N$
	G1	G2	G1	G2	G1	G2
Socioeconomic status	-0.858***	-0.575***	-0.883***	-0.532***	-0.730***	-0.541***
log(family income/poverty)	(0.007)	(0.010)	(0.014)	(0.019)	(0.053)	(0.092)
Observations	325,169	311,343	32,973	27,985	10,043	9,701

Notes - Data on socioeconomic assimilation are drawn from the Current Population Survey (CPS, 1994−2009). All estimates include state and and year fixed effects. Information on parental birth place is available in the CPS only since 1994.

Table A.10: Socio-economic differences between 1st, 2nd generation Mexicans and white natives, 15-65

	(1)	(2)	(3)	(4)	(5)	(6)
	US		**CA-FL**		**Florida**	
	$MX_1 - N$	$MX_2 - N$	$MX_1 - N$	$MX_2 - N$	$MX_1 - N$	$MX_2 - N$
	G1	G2	G1	G2	G1	G2
Socioeconomic status	-0.780***	-0.564***	-0.828***	-0.546***	-0.765***	-0.549***
log(family income/poverty)	(0.003)	(0.004)	(0.006)	(0.008)	(0.024)	(0.041)
Observations	1,267,802	1,213,118	136,030	114,552	41,265	39,737

Notes - Data on socioeconomic assimilation are drawn from the Current Population Survey (CPS, 1994−2009). All estimates include state and and year fixed effects. Information on parental birth place is available in the CPS only since 1994.