

# Why Does the Health of Immigrants Deteriorate?

Osea Giuntella\*

Department of Economics, Boston University

Job Market Paper

January 12, 2013

## Abstract

Despite their lower socioeconomic status, Hispanic immigrants in the United States initially have better health outcomes than natives. However, while their socioeconomic status improves over time and across generations, their health deteriorates. This phenomenon is commonly known as the “Hispanic health paradox.” There is an open debate about whether the observed convergence is explained by selection on health or by the adoption of less healthy lifestyles. This paper uses a unique dataset linking the birth records of two generations of Hispanics born in California and Florida (1975–2009), to analyze the mechanisms behind the generational decline in birth outcomes. I calibrate a simple model to interpret the health trajectories of immigrant descendants, using country-level differences in health outcomes to pin down the degree of selection of the first-generation immigrants and existing estimates to account for the intergenerational transmission of health status. Accounting for socioeconomic differences between second-generation Hispanics and natives, the model not only explains, but actually reverses the paradox: the puzzle is not that immigrant relative health deteriorates so rapidly, but that it does not deteriorate rapidly enough. In order to quantify the relative importance of behaviors, I estimate the effect of health behaviors and assimilation on third-generation birth outcomes. Hispanics preserve a large advantage in health behaviors and present a lower incidence of risk factors. Higher incidence of risk factors and higher assimilation are associated with poorer third-generation birth outcomes. These effects hold true even in a subset of siblings, and holding constant grandmother-fixed effects. I conclude that the lower incidence of risk factors among Hispanics can explain 76% of the “reverse paradox.”

---

\*270 Bay State Road, 02215, Boston (MA); phone: 617-353-5616; fax: 617-353-4449; email: osea@bu.edu. I am grateful to Daniele Paserman, Randall Ellis, Kevin Lang, and Claudia Olivetti for their comments and advice. I would like to thank all the participants to the Northeast Universities Development Consortium Conference (Dartmouth, 2012), the American Society of Health Economists Conference (Minneapolis, 2012), the XII Doctoral Workshop in Economic Theory and Econometrics (Rome, EIEF, 2012), the EconCon Conference (Princeton U, 2012), the European Meeting of the Econometric Society (Malaga, 2012), the XV IZA Summer School (Buch-Ammersee, 2012), the European Association of Labor Economists/IZA Conference (Bonn, 2012), as well as all seminar attendees at the BC-BU Green-Line Labor Meeting (2012), 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.

# 1 Introduction

A substantial body of research has documented that immigrants are healthier than natives when they first arrive in the United States but that this initial advantage deteriorates with time spent in the United States and over generations. These findings are particularly striking when focusing on Hispanics. Because they are characterized by lower socioeconomic status than natives, they should be expected to be at higher risk for negative health outcomes. Furthermore, despite positive socioeconomic assimilation and a positive socioeconomic gradient in health, there is evidence of a downward convergence in health over time and across generations. Previous scholars have thus referred to these stylized facts as the “Hispanic health paradox.” This apparent paradox has been observed in 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). The goal of this paper is to analyze the mechanisms underlying these facts.

There is a general consensus that selection can explain the first-generation advantage (Palloni and Morenoff, 2001; Jasso et al., 2004; Antecol and Bedard, 2006); however, researchers are still puzzled about the possible explanations for the subsequent health convergence observed in the second generation. The observed health patterns may be explained by the fact that immigrants are positively selected on health status (Palloni and Morenoff, 2001) and 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 the initial advantage through a process of natural regression towards the mean (Jasso et al., 2004). However, other scholars emphasize the role of behaviors, providing evidence of fewer risk factors among immigrants at the time of emigration giving way to riskier behavior as more time is spent in the United States and across generations (Acevedo-Garcia et al., 2005; Antecol and Bedard, 2006; Fenelon, 2012). Overall, the lack of extensive longitudinal data and small sample sizes severely limited the ability to clarify

the possible channels behind the Hispanic paradox as observed in birth outcomes.

I contribute to these previous studies by taking advantage of a large longitudinal inter-generationally linked data set. In particular, this paper analyzes the birth outcomes of the second- and third-generation Hispanics born in California and Florida, two of the top immigrant destination states in the United States. Linking the birth records of two generations overcomes certain of the limits faced by previous studies and assists in the investigation of the factors affecting the generational decline of birth outcomes among Hispanic immigrant descendants.

To test whether the paradox can be explained entirely by selection and regression towards the mean, I develop a simple model of health transmission. I use country-level differences in health outcomes to pin down the degree of selection in the first generation and existing estimates from the literature to impute the intergenerational transmission of health status. While second-generation Hispanics improve their socioeconomic status with respect to the first generation, they still have lower socioeconomic status than non-Hispanic natives. Therefore, on average, they do not have the identical quality of care as non-Hispanic natives. Calibrating the differences in the quality of health care to match the differences in socioeconomic status, the model not only explains all of the paradox, but, everything else constant, it actually overpredicts convergence and results in a “reverse paradox.” Contrary to the nonsignificant difference observed between third-generation immigrants and natives, the calibration exercise predicts a fairly large health advantage for natives. Third-generation Hispanics show better birth outcomes than what we would expect, given the relatively low rate of intergenerational transmission observed in the data and the relatively low socioeconomic conditions they are in. Thus, the new puzzle is to ascertain how third-generation birth outcomes do not deteriorate as rapidly as predicted by the model.

In the paper, I show that first-generation immigrants have substantially lower incidence of both risky behaviors (such as smoking and alcohol consumption) and health risk factors (hypertension) that are known to seriously affect birth outcomes (Almond et al., 2005; Shireen

and Lelia, 2006; Gonzalez, 2011; Kaiser and Allen, 2002; Forman et al., 2009). Although risk-factor behavior worsens between first- and second-generation Hispanics, immigrants maintain a sizeable advantage in terms of lower incidence of health risk factors compared to white natives. The persistence of healthier risk-factor behavior explains 76% of the “reverse paradox.”

The importance of socioeconomic factors and risky behaviors is confirmed by the analysis of differences in the health convergence among second-generation Hispanics. I show that third-generation birth outcomes correlate significantly with quality of care, socioeconomic status, and risk-factor behavior. To address the potential endogeneity of these covariates, I follow the Currie and Moretti (2007) strategy of linking siblings, and I test whether the correlations are robust to the inclusion of grandmother-fixed effects. Analyzing within family variations in the patterns of socioeconomic and cultural assimilation of second-generation Hispanics, I can disentangle the contribution of these factors from the background characteristics that are common within a family at birth (including the migrant’s selectivity). Overall, the within-family analysis confirms that risky behaviors do matter and do significantly affect differences in the convergence rate among Hispanics. The convergence is more marked among those who are less likely to maintain the health-protective behaviors and conditions (such as low rates of smoking, alcohol consumption, and hypertension) that characterize the first-generation immigrants and, more generally, among those who are more likely to have culturally assimilated. In particular, among second-generation Hispanics, intramarried couples show higher resilience in healthy behaviors, health conditions, and birth outcomes. Using ethnic intermarriage as a metric of cultural assimilation, I show that third-generation children of intermarried Hispanic couples are 14% more likely to be of low birth weight than children of intramarried couples. This result is particularly striking because intermarriage is usually associated with positive socioeconomic outcomes (Wang, 2012).

The paper is organized in the following manner. Section 2 discusses previous literature. Section 3 describes the data and verifies the Hispanic paradox in birth outcomes. Section 4

discusses the possible mechanisms behind these health patterns. In Section 5, I examine the heterogeneity in health convergence within the Hispanic group, exploiting grandmother-fixed effects. Concluding remarks are in Section 6.

## 2 Hispanic paradox: Selection or worsening of behaviors?

A vast literature investigates the health differences between U.S. natives and immigrants. Most papers show that immigrants are healthier upon their arrival and that their advantage erodes as more time is spent in the United States.<sup>1</sup> As discussed above, there is consensus that the initial health advantage is the result of positive selection into the United States biasing immigrant–native health differences upward (Palloni and Morenoff, 2001; Jasso et al., 2004; Chiswick et al., 2008; Antecol and Bedard, 2006), but researchers continue to be puzzled by the mechanism underlying the ensuing convergence to native health status.

A first group of scholars (Palloni and Morenoff, 2001; Jasso et al., 2004; Chiswick et al., 2008) argue that the apparent deterioration may be largely attributed to a regression towards the mean following the initial selection, which is a statistical artifact. In particular, Palloni and Morenoff (2001) provide a simple model to show how even a moderate degree of selection at migration may explain the second-generation advantage in birth outcomes. Following this argument, Jasso et al. (2004) suggest that immigrants might select on transitory health traits and that their inability to fully forecast the evolution of their health might naturally revert towards the average health of the original population. These articles provide empirical support for the selection hypothesis as a plausible explanation of the initial health advantage observed among first-generation immigrants and their children compared to na-

---

<sup>1</sup>Gutmann et al. (1998) describe the origin of the epidemiologic paradox. Using data from the 1910 U.S. Census and the 1990 linked birth and death certificate file, the authors find that Hispanics did not suffer higher child mortality than non-Hispanic whites, but there was already evidence of a health advantage compared to the African–American population in the early 20th century. Historical data suggest that Hispanics did not show better birth outcomes than white non-Hispanic natives until the early 1960s.

tives. However, these studies do not test the implications of selection and regression towards the mean on the adult health of second-generation immigrants or on the birth outcomes of their children.<sup>2</sup>

A second strand of the literature emphasizes the importance of negative acculturation. According to these scholars, the unhealthy convergence is explained by the worsening of dietary styles, the adoption of risky behaviors, and the erosion of social and cultural protective factors such as familism and religiosity (Guendelman and Abrams, 1995; Acevedo-Garcia et al., 2004; Antecol and Bedard, 2006; Fenelon, 2012). These studies offer evidence of a protective effect on birth outcomes and infant health risk factors of foreign-born status, ethnic density, age at migration, and years since immigration (Acevedo-Garcia et al., 2005; Bates and Teitler, 2008; Guendelman and Abrams, 1995; Hummer et al., 2007; Finch et al., 2007; Osypuk et al., 2010; Shaw et al., 2010). However, these authors did not attempt to disentangle the causal effect of behaviors in accounting for selection and other potential confounding factors.

Indeed, most of these studies were limited in their scope, either by sample size, the cross-sectional nature of the data, or the lack of objective and reliable measures on nativity, ethnicity, and health. Previous research on obesity and other health outcomes relied on the use of synthetic cohorts of immigrants (Antecol and Bedard, 2006; Kaushal, 2008) to analyze the effects of time spent in the United States and of age at arrival. Because of the lack of information on parental nativity, researchers were often forced to use foreign-born status and self-reported ethnicity to analyze generational changes in health and health-related behaviors. While Jasso et al. (2004) note the need to analyze health trajectories of immigrants across generations, there is no study analyzing the Hispanic health paradox using individual linked data on two generations of immigrants, to the best of my knowledge.

---

<sup>2</sup>Previous scholars have also postulated that the low infant mortality observed among Hispanics might be explained by selective re-migration (Palloni and Arias, 2004). While this may be relevant, Hummer et al. (2007) show that women of Mexican origin are extremely unlikely to migrate to Mexico with newborn babies. Furthermore, as remarked by Palloni (2010), it is unlikely that the generational deterioration in birth outcomes could be explained by selective return migration, given the low rate of return migration among second-generation immigrants.

The large size of the data set and the ability to link the records of two generations allow this study to address questions that other researchers have not. By exploiting the intergenerational nature of the data, I can verify whether the apparent deterioration in birth outcomes may be explained entirely by positive selection at migration and a subsequent regression towards the mean. Furthermore, the large size of the sample allows me to focus on a subsample of second-generation siblings and include grandmother-fixed effects. Using within family variation, I can partially isolate the original migrant’s selectivity to analyze heterogeneity in the path of convergence in the birth outcomes of the third generation. Finally, one of the significant advantages of this paper is that most of the health outcomes considered (pregnancy outcomes and maternal health characteristics) are recorded by medical officials and are therefore not subject to self-reporting bias.

### 3 Data

The main data used in this paper are drawn from the Birth Statistical Master File provided by the Office of Vital Records of the California Department of Health and from the Birth Master Dataset provided by the Bureau of Vital Statistics of the Florida Department of Health. These data contain information extracted from the birth certificates for all children born in California and in Florida for the years 1975–1981 and 1989–2009. For expositional ease, for both the immigrants and the natives in the sample, I refer to all the women delivering between 1975 and 1981 as first-generation (grandmothers, G1), to all the children born between 1975 and 1981 and who delivered between 1989 and 2009 as second-generation (mothers, G2), and to all the children born between 1989 and 2009 as third-generation (children, G3).

Information on mother’s country and state of birth, mother’s first and maiden name, child’s full name, date of birth, gender, parity, race, birth weight, hospital of birth, county of birth are available in both states for all the period considered. However, not all the variables

are available in each year and for each of the two states. For instance, mother's age is reported for the entire period in California, but only since 1989 in Florida, while mother's education is reported for the entire period in Florida, but only since 1989 in California. Data do not contain information on legal marital status, which is self-reported in Florida and is inferred by birth clerks in California. Information on birth weight is available for the entire period in both states, while unfortunately other important measures of health at birth (e.g. Apgar score, gestational length) are available only in the more recent years. While Almond et al. (2005) and Wilcox (2001) cast doubt on the causal effect birth weight might have on mortality and more generally on infant health, there is a general consensus that low birth weight is an important marker of health at birth and that is strongly associated with higher risk of mortality and morbidity (Currie and Moretti, 2007; Conley and Bennett, 2000). Since this study does not analyze the effects of birth weight and given that birth weight is the only measure of birth outcome available for the entire period, I will mostly focus on birth weight and incidence of low birth weight as indicators of health at birth.<sup>3</sup> A full description of the variables used in the paper and their availability in each of the two states for the period considered is provided in the online Appendix.<sup>4</sup>

As with the previous literature (Fryer and Levitt, 2004; Currie and Moretti, 2007; Royer, 2009) that used administrative birth records, I am able to link information available at a woman's birth to that of her children, if the woman is born in California (Florida) and also gave birth in California (Florida). To ensure the comparability of the analysis in the two states, I focus on women and in particular on the cohort of women born between 1975 and 1981.<sup>5</sup> One of the typical drawbacks of administrative vital statistics is the lack of information on individual income and occupation. However, the data contain certain information

---

<sup>3</sup>However, results go in the same direction when using alternative measures (e.g. Apgar scores, infant mortality) of infant health for the years in which other metrics are available.

<sup>4</sup>The Online Appendix is available on my personal web page: <http://people.bu.edu/osea> .

<sup>5</sup>Florida data contain information on the father's full name and date of birth, allowing me to conduct a parallel analysis using the father's information. However, because of the lesser quality of information about fathers and because they are less likely to become parents at an early age, the matching rate is considerably lower than that of women and the selectivity of the sample increases. The results are similar in that direction, but only marginally significant and are available upon request.



on parental education and on the mother’s residential zip code; this information is available from 1989 onwards in California and for the entire period in Florida. Therefore, with the data from Florida, I can use grandmother’s education, and the median income and poverty rate in her residential zip code. In California, I do not have information on the grandmother’s education and on the grandmother’s residential zip code, but I can use the socioeconomic characteristics of the zip code of the birth hospital as a proxy for the socioeconomic status of the grandmother, as in Currie and Moretti (2007). Data on zip code sociodemographic and economic characteristics are drawn from the U.S. Census (source: Social Explorer). In particular, I use the median family income and the poverty rate as of the 1980 Census for the zip code of the mother’s birth and grandmother’s residence and as of the 1990 Census for the zip code of the child’s birth and mother’s residence. To construct the intergenerational sample, I linked the records of all the infants born between 1989 and 2009 whose mother was born in California or Florida between 1975 and 1981 to the birth records of their mothers. I matched the child’s birth record to the mother’s record using the mother’s first and maiden name, exact date of birth, and state of birth. Whenever I was able to uniquely identify the mother’s birth record, I included them in the linked sample.

### **3.1 Matching and selection: Descriptive statistics**

The quality of matching for children born in California and Florida between 1989 and 2009 whose parents were born in the same states between 1975 and 1981 is high in Florida (96.6%) and only slightly lower in California (87.5%). I do not manage to match observations for names that were misspelled or changed across birth certificates, or for dates of birth that were misreported or could not be uniquely identified with the information available. Despite the high rate of matching, the linked sample is not representative of women (men) born between 1975 and 1981. The final sample includes 726,837 (44%) of the 1,643,865 female children born between 1975 and 1981 in California and Florida. This reflects the reality that not all the women born in Florida and California between 1975 and 1981 were still living in

those states between 1989 and 2009 and that not all of these women became mothers before 2009. In particular, the Natality Detail Data, which contains information on the mother's state of birth and state of birth of the child, shows that approximately 11.5% of women born in California and 13.5% of women born in Florida between 1975 and 1981 had a child in another U.S. state before 2004 (the last year for which both the information on the state of birth of the mother and the state of birth of the child are available in this database). By using the American Community Survey (2010), we know that approximately 42% of women born in California and 39% of women born in Florida between 1975 and 1981 had not had a child by 2009. Data problems such as misspelling or missing information account for the rest of the attrition. Table 1 shows the matching rates for the main race and ethnic groups in the sample. The matching is particularly high for African-Americans (70%) because of both their lower mobility and higher likelihood of having a child at a early age. Children of Hispanic immigrants have a higher rate of matching than natives. The matching rate among children of Mexican origin is 55% and 50% among second-generation immigrants of Cuban origin. Children of Puerto Rican immigrant women are less likely to be linked (37%). Differences in the density of ethnic networks and in the different types of migration are reflected in the different matching rates for Mexicans and Cubans in the two states. The rate of matching also depends on socioeconomic background, which is clearly associated with infant health, mobility, and age of the mother at first birth. Children of first-generation mothers who were residing in poor zip codes (in the lowest income quartile) are more likely to be linked to the records of their offspring than the children of first-generation mothers who were living in wealthier zip codes (in the highest income quartile). While these descriptive statistics show evidence of selection on sociodemographic characteristics (see column 3), the differences in initial health endowments between linked and nonlinked observations are not striking (see columns 4–9). If anything, they suggest that the linked sample has a slightly lower incidence of low birth weight. The differences in birth weight appear to be negligible and nonsystematic. A 100-gram increase in birth weight increases the probability of a later

observation only by 0.5%. However, if the mother was born with a weight below the 2,500 grams threshold, she is 13% less likely to be linked. The lower incidence of low birth weight (LBW) in the linked sample can be explained by higher rates of infant mortality, higher probabilities of returning to the family’s country of origin (“salmon bias”), or by a lower probability of having a child among those children born with poor health outcomes. Because the differences between the linked and nonlinked sample appear to be small, I present all my results without making any correction for potential selection bias. However, using a Heckman selection model with child’s year of birth as the excluded variable yields essentially identical results.<sup>6 7</sup>

### 3.2 Verifying the Hispanic paradox in birth outcomes

The focus of this paper is on the mechanisms behind the apparent deterioration in infant health of later generations of Hispanic immigrants. However, it is important to first verify the Hispanic paradox within the sample of birth records under analysis. To this end, I use a simple linear probability model that relies on a comprehensive set of individual and contextual controls to study the conditional differences in birth outcomes between immigrants and natives. Formally, I consider the following model:

$$H_{izt,2} = \alpha + \beta IMM_{izt,1} + \gamma X_{izt,1} + \tau_{t,2} + \xi_{z,2} + \epsilon_{izt,2}$$

where the subscripts 1 and 2 represent first and second generation.  $H_{izt}$  is the birth outcome (such as birth weight, incidence of low birth weight, etc.) of the second-generation child  $i$ , whose mother resided (or delivered) in zip code  $z$  at time  $t$ .  $IMM_{izt,1}$  is a dummy

---

<sup>6</sup>The year of birth of second-generation women is a significant predictor of later observations, while differences in birth outcomes by year of birth are negligible when considering children born fewer than 6 years apart.

<sup>7</sup>Palloni and Arias (2004) suggested that a large part of the lower mortality rates observed in the Mexican population can be explained by selective out-migration (the “salmon bias” effect). However, Hummer et al. (2007) argue that selective out-migration is unlikely to explain the advantages observed in the health outcomes of second-generation children, especially when looking at first-hour, first-day, and first-week mortality.

equal to one when the first-generation woman delivering between 1975 and 1981 was born outside the United States. The set of individual sociodemographic characteristics of the first-generation mothers is delineated in  $X_{izt}$ , including education (high school dropout, high school graduate, some college, and college or more), marital status, race, age dummies (in Florida, the mother’s age is not available for the period 1975–1981), an index of adequacy of prenatal care based on the month in which prenatal care started, father’s age (quadratic), father’s education (high school dropout, high school graduate, some college, and college or more) and father’s race.<sup>8</sup> I include indicators for missing information on parental education and age, marital status, and parity. Finally, I control for both time  $\tau_{t,2}$  and zip code  $\xi_{z,2}$  fixed effects.

Table 2 illustrates the Hispanic paradox in birth outcomes reporting the differences between children of first- and second-generation immigrants coming from the three largest Hispanic groups in the United States (Mexicans, Puerto Ricans and Cubans) and children of white U.S.-born mothers.<sup>9</sup> I restricted the sample to children born between 1975 and 1981 to white mothers and Hispanic first-generation immigrant mothers coming from Mexico, Puerto Rico and Cuba.<sup>10</sup> The final sample includes 2,234,571 births for which information on birth weight and zip code is not missing.<sup>11</sup>

---

<sup>8</sup>In Florida, the month in which prenatal care started is imputed using the number of visits and the usual relationship between the number of visits and the month in which prenatal care started. However, the results are similar when using the number of visits only.

<sup>9</sup>In this paper, I focus on immigrants of Hispanic origin, for which the paradox is particularly striking, given their socioeconomic background characteristics, and who are by far the largest ethnic group in the United States. However, when looking at the identical analysis for children of immigrants coming from other countries, I find that the incidence of low birth weight is 12% lower among children of Canadians than among U.S. natives, while it is 20% higher among children of Japanese and is nonsignificantly different among children of Chinese and Vietnamese mothers, although the coefficient is negative for the latter.

<sup>10</sup>The mother’s ethnicity is not consistently reported before 1989. Restricting the sample to the second-generation mothers that I am able to link to their offspring, I can use the ethnicity reported at the time of delivery to further restrict the sample of natives to non-Hispanics. The coefficients differ only slightly in the magnitude and are consistent with the patterns of convergence observed among immigrants of Hispanic origin. The results are similar when considering the samples of male and female children separately. These tables are available upon request.

<sup>11</sup>Notice that this number includes male and female births and therefore is approximately twice as large as the number of observations presented in Table 1, which includes only the birth records of women who could be potentially linked to the birth records of their offspring. Furthermore, in Table 1, the entire sample also includes black children. The results are similar when the data are restricted to women born between 1975 and 1981.

The coefficients reported in columns 1 and 3 report the unconditional mean differences in birth weight and incidence of low birth weight, respectively. Column 2 and 4 include a broad set of sociodemographic controls. Among children of Cuban mothers there are no significant differences in birth weight (column 2), but there is evidence of a lower incidence of low birth weight (column 4). Children of Mexican mothers are only slightly heavier (approximately 22 grams, column 2), but show a significantly lower incidence of low birth weight compared to the children of white native mothers who share a similar socioeconomic background (column 4). By contrast, Puerto Rican mothers are more likely to give birth to lighter babies (columns 2 and 4). In the online Appendix I show the sensitivity of the magnitude of the coefficients to the addition of different sets of controls. It is important to note that the addition of geographic controls (county-, hospital- or zip code–fixed effects) is associated with a stronger advantage in terms of lower risk of low birth weight for children of Mexican origin. This is consistent with the original definition of the epidemiological paradox as the fact that children of Hispanic immigrants fare considerably better than children of non-Hispanic women sharing a similar socioeconomic background.<sup>12</sup> Taken as a whole, columns 2 and 4 show that children of Puerto Ricans fare considerably worse than their native counterparts, while there remains a “healthy immigrant effect” when considering the incidence of low birth weight for Mexicans and Cubans. This is consistent with the idea that Puerto Ricans in the sample might be less favorably selected because Puerto Rico is a U.S. territory. Even among children of Mexicans, for whom the advantage in low birth weight is highest, there is only a difference of 22 grams in the average birth weight. The differences between the continuous and the discrete outcome variables reflect the independence of the predominant and residual distribution of birth weight and, more generally, differences in the distribution of term and pre-term births (Wilcox, 2001).<sup>13</sup> Figure 1 depicts the cumulative distribution of birth weight in California and Florida for the immigrant descendants of Hispanic origin and for white natives over the

---

<sup>12</sup>When breaking down the analysis by state, the coefficient for children of Mexican mothers tends to be higher in Florida than in California, most likely reflecting higher selection.

<sup>13</sup>The predominant distribution is substantially equivalent to the distribution of birth weight observed for term births.

period analyzed in this study. The distributions are very similar. Previous studies have shown that the size and nature of the effects of covariates on the conditional mean might not capture the importance of the effects on the lower tail of the birth weight distribution (Koenker and Hallock, 2001). Indeed, quantile regression indicates that the advantage in birth weight (in grams) is more substantial in the left tail of the birth weight distribution. Children of Hispanic origin immigrants are on average 50 grams heavier than children of white natives in the 5% quantile of the distribution, while the differences are much smaller, and even become negative, in the upper tail of the distribution. In particular, in the 5% quantile of the distribution, the children of Mexican mothers weigh 70 grams more on average than children of white native mothers and are 50 grams heavier on average in the 10% quantile (see the online Appendix).<sup>14</sup> In summary, columns 2 and 4 document that the healthy immigrant effect in infant outcomes is mostly concentrated in the lower tail of the birth weight distribution and that it is heterogeneous across ethnic groups.

I then turn to the analysis of the linked sample and analyze whether these differences persist over time and are transferred to the children of third-generation immigrants. Formally, I estimate the following model:

$$H_{izt,3} = \alpha + \beta IMM_{izt,1} + \gamma X_{izt,2} + \tau_{t,3} + \xi_{z,3} + \epsilon_{izt,3}$$

where the subscripts 1, 2 and 3 represent first, second and third generation, respectively.  $H_{izt,3}$  is a birth outcome of the third-generation child, whose mother resided (or delivered) in zip code  $z$  at time  $t$ .  $IMM_{izt,1}$  is a dummy equal to one if the first generation was born outside the United States. Note that the analysis sample here includes only 2nd generation mothers between 1975 and 1981 in CA and FL, who were babies in the 2nd generation sample. To ensure the comparability of the analysis, the model includes the identical set of

---

<sup>14</sup>The 0.05 quantile roughly corresponds to the traditional threshold of low birth weight. In the quantile regression, I include gender, marital status, adequacy of prenatal care, parity, type of birth, year fixed effect, state fixed effect, maternal education (Florida), and a quadratic for age. This is substantially equivalent to the specification used in Table 2, without including zip code-fixed effects.

controls used in the analysis of second-generation birth outcomes.

Columns 5–8 in Table 2 illustrate the differences in birth weight and incidence of low birth weight between third-generation children whose grandmothers were born in Mexico, Puerto Rico or Cuba and third-generation white natives.<sup>15</sup> The estimates in columns 6 and 8 include the identical set of controls used in columns 2 and 4.<sup>16</sup> The deterioration in birth outcomes is mostly evident in the incidence of low birth weight; even when analyzing differences in birth weight, the coefficients are always negative and larger in magnitude compared to those of second-generation immigrants. The average incidence of low birth weight is relatively stable among second- and third-generation white natives (see the online Appendix), but the coefficient ( $-0.004$ ) for the third-generation children of Mexican origin (column 8) shrinks significantly (by approximately 65%) compared to the one observed among second-generation children in column 4 ( $-0.016$ ). However, the third-generation children of Mexican origin do conserve some of the initial health advantage. The deterioration with respect to native birth outcomes is even stronger when children of Cuban and Puerto Rican origin are analyzed.

## 4 A simple model of selection and health transmission

### 4.1 Theoretical framework

In the previous section, I confirmed the existence of an apparent paradox in the birth outcomes of Hispanic descendants. This section develops a theoretical model to analyze the mechanisms behind these health trajectories. As mentioned earlier in the paper, previous scholars have questioned the paradoxical nature of these stylized facts by arguing that they could be entirely explained by selection and a subsequent process of regression towards the

---

<sup>15</sup>Unfortunately, the data do not contain information on the country of origin of the father for the entire period. To be able to compare the results shown in columns 1–4, I included all grandchildren of U.S.-born white women. However, one could restrict the sample to grandchildren of U.S.-born white women whose mothers did not report Hispanic origin. The results (available upon request) are substantially similar.

<sup>16</sup>In the online Appendix I report the conditional mean differences obtained using different sets of control variables.

mean. To verify this hypothesis, I build on Palloni and Morenoff (2001) and introduce a simple model of selection on health at migration and intergenerational health transmission. Because of the limited information available on birth weight distribution in the country of origin, I am not able to provide a direct estimate of the original selection. However, I can calibrate the model using the observed differences in health outcomes between the United States and the countries of origin to pin down the degree of selection of first-generation immigrants. Similarly, I use existing estimates from the literature to capture the degree of intergenerational transmission of health. To keep the model simple and intuitive, I focus on the primary country of origin—Mexico—and compare the health distribution of Mexicans and natives in the United States.

The decision to migrate can be represented by a dichotomous variable that equals 1 when an individual migrates and 0 otherwise. The underlying idea is that, holding everything else constant, the cost of migration will be higher for those who are less healthy. This is consistent with intermediate or mildly positive selection being driven by higher costs of migration as argued by Chiquiar and Hanson (2005).<sup>17</sup> Immigrants who have health above a certain threshold,  $t_1$ , at the time of migration will be able to migrate, while the rest will stay in the country of origin. This may be represented formally as:

$$Imm_1 = \begin{cases} 1 & \text{if } h_1 \geq t_1 \\ 0 & \text{if } h_1 < t_1 \end{cases}$$

$$h_1 = u_1$$

---

<sup>17</sup>There is an open debate on whether Mexican migrants to the US tend to be negative selected from the Mexican distribution of education and earnings. In a seminal article Borjas (1987) concluded that Mexican migrants tend to be negatively selected on education and earnings. Chiquiar and Hanson (2005) provide evidence against the negative-selection hypothesis and suggesting that migrants are selected from the middle of Mexican earnings distribution. In particular, Chiquiar and Hanson (2005) find evidence of positive selection for Mexican-born women. Similarly, the findings of Orrenius and Zavodny (2010); McKenzie and Rapoport (2010); Kaestner and Malamud (2010) confirm positive or intermediate selection. However, other studies provide evidence in favor of the negative selection hypothesis (Ibarraran and Lubotsky, 2007; William and Peri, 2012; Moraga, 2011; Reinhold and Thom, 2012).



where  $h_1$  is the health of the first generation at the time of migration, which is distributed as a random normal  $(\mu_j, 1)$  reflecting the health distribution in the country of origin,  $\mu_j$  is the average health in country  $j$ , and  $t_1$  is the migration threshold. Thus,  $\mu_j$  is the composite effect of genes, quality of health care, socioeconomic environment, and risk-factor behavior on health. Individuals with  $h_1 \geq t_1$  will be able to migrate. The higher the threshold, the more selected is the sample of migrants. The incidence of low birth weight is determined as follows:

$$BW_2 = \gamma h_1 + v_2$$

$$LBW_2 = \begin{cases} 1 & \text{if } BW_2 \leq t_2 \\ 0 & \text{if } BW_2 > t_2 \end{cases}$$

where  $BW_2$  is the birth weight of the second generation,  $h_1$  captures maternal health at migration,  $v_2$  is distributed as a random  $(0, \sigma_v^2)$  normal variable reflecting the effect of other unobservable factors on the birth weight of the second generation,  $\gamma$  captures the effect of maternal health on the child's health, and  $t_2$  represents the low birth weight threshold. Similarly, third-generation birth outcomes can be described as a function of second-generation health characteristics and other factors. This may be described formally as

$$h_2 = \rho h_1 + u_2$$

where  $h_2$  is the health of second-generation mothers,  $u_2$  is distributed as a random  $(\mu_{j2}, \sigma_{u_2})$  normal variable reflecting the effect of other unobservable factors on the health of the second-generation mother, and  $\rho$  measures the degree of intergenerational correlation in health between the first and second generations. Then if the distribution of health is stable

$$\sigma_{h_1}^2 = \sigma_{h_2}^2 = \sigma_{u_2}^2 + \rho^2 = 1$$

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

health, with the following formal designation:

$$BW_3 = \gamma h_2 + v_3$$

$$LBW_3 = \begin{cases} 1 & \text{if } BW_3 \leq t_2 \\ 0 & \text{if } BW_3 > t_2 \end{cases}$$

where  $BW_3$  is the birth weight distribution in the third generation,  $v_3$  is distributed as a random normal  $(0, \sigma_v^2)$  variable reflecting the effect of other unobservable factors on the birth weight of the third generation, and  $t_2$  determines the amount of low birth weight in the third generation. Without loss of generality, I assume that the unobserved random shocks to health and birth weight are not correlated.<sup>18</sup> The covariance between the birth weight of the two generations may therefore be rewritten as the following:

$$\text{Cov}(BW_3, BW_2) = \text{Cov}(\gamma h_2 + v_3, \gamma h_1 + v_2) = \text{Cov}(\gamma \rho h_1 + \gamma u_2 + v_3, \gamma h_1 + v_2) = \gamma^2 \sigma_{h_1}^2 \rho = \gamma^2 \rho$$

which implies

$$\rho = \frac{\text{Cov}(BW_3, BW_2)}{\gamma^2} \tag{1}$$

Within this framework, I can estimate the extent to which selection and the estimated intergenerational correlation in health may explain the evolution of low birth weight incidence among immigrant descendants.

## 4.2 Empirical moments and calibration

Panel A in Table 3 presents the set of empirical moments targeted by the model. Column 1 reports the unconditional mean difference in the incidence of low birth weight between

---

<sup>18</sup>Note that while this assumption might seem strong, in practice it does not affect the model predictions for the birth outcomes, because the intergenerational correlation in birth weight is pinned down in the model using existing estimates (Currie and Moretti, 2007). While the focus of this study is on birth outcomes, it is important to note that the values of  $\gamma$  and  $\rho$  would instead depend on the extent of correlation between unobserved random shocks to health and birth weight.

second-generation Mexicans and white natives born between 1975 and 1981 in California and Florida. The incidence of low birth weight among children of first-generation Mexicans is 0.008 percentage points lower than among children of native white women, approximately a 15% difference with respect to the average incidence of low birth weight in the sample.<sup>19</sup> Column 2 reports the difference in the incidence of low birth weight between third-generation Mexicans and white natives born between 1989 and 2009 in California and Florida. These differences are in line with those observed across the United States, using the Natality Detail Data, which collects detailed data on all births in the United States (see Table 4).<sup>20</sup>

To verify whether a simple selection model can fit these moments, I use the white native population in the United States as a reference group, and calibrate the main parameters defined in the theoretical framework. I start with a population of 10,000 potential Mexican migrants and 10,000 U.S. natives. Individuals receive a random value for their health at migration that is drawn from a normal distribution with a mean identical to the mean of their country of origin (Mexico or the United States). I use the native health as a benchmark and set  $\mu_{US}$  equal to 0. The low birth weight threshold,  $t_2$ , is set to be  $-1.57$  to match the average incidence of low birth weight observed in the data (0.058) over the entire period studied (1975–2009) in the entire population of the United States (excluding African–Americans). The mean of unobservable factors affecting health  $\mu_{MX}$  is set to be such that the difference in the low birth weight of the two populations would be equivalent to that implied by the earliest available measure of incidence of low birth weight in Mexico (10.6%, see Buekens et al. (2012)) relative to the average incidence of low birth weight in the U.S. nonblack population (5.8%). I consider different values of  $\rho$  and  $\gamma$ , such that equation (1) would hold

---

<sup>19</sup>As mentioned in Section 3, the addition of geographic controls (county-, hospital- or zip code–fixed effects) is associated with a stronger advantage in terms of lower risk of LBW for children of Mexican origin.

<sup>20</sup>Note that the Natality Detail Data, in its public version, does not allow for cross-generational record linking because it does not release information on the names of the child and mother. Geographic data include state, county, city, standard metropolitan statistical area (SMSA, 1980 onwards), and metropolitan and non-metropolitan counties. From 2005 onwards, the data do not include any geographic variables such as state, county, or SMSA . However, I can use the Natality Detail Data data to conduct cross-sectional analysis for the entire United States for the years 1975–2004. This allows me to partially verify the external validity of the information in the main sample drawn from California and Florida.

for values of  $\text{Cov}(BW_3, BW_2)$  such that the estimated intergenerational correlation in birth weight would be around the 0.2 estimated by Currie and Moretti (2007) and confirmed in my data.<sup>21</sup> Previous studies estimated the intergenerational correlation in longevity and mortality to range between 0.2 and 0.3 (see Ahlburg (1998)), and the intergenerational correlation in BMI (Body Mass Index) to be approximately 0.35 (see Classen (2010)). Based on these estimates, I focus the analysis on values of  $\rho \in [0.2, 0.5]$ , with the assumption that intergenerational correlation in health status should be neither lower nor much different from the intergenerational correlation in longevity. The above restrictions imply that  $\gamma$  must be  $\in [0.58, 1]$ .<sup>22</sup> Here, I use the case in which  $\rho$  is equal to 0.35 as a baseline, but the interpretation of the simulation exercise does not change significantly for different values of  $\rho$  in the defined range  $[0.2, 0.5]$ . Using this parametrization and under the assumption of health having identical effects on birth weight in the two populations, the model can be solved analytically (see Online Appendix B for the formal solution).

Figure 2 shows the predicted differences in the incidence of low birth weight between children of first-generation Mexican immigrants and children of white natives ( $y$ -axis) by extent of selectivity at migration. The  $x$ -axis describes the percentiles of first-generation Mexican health distribution corresponding to different values of the selection threshold ( $t_1$ ). The dashed line marks the raw difference ( $-0.008$ ) in low birth weight in the data between second-generation Mexicans and white natives (column 1, Table 3). The figure suggests that the initial advantage can be explained entirely by a relatively moderate selection. If Mexicans with health below the 15th percentile do not migrate because of their health conditions, positive selection can explain the lower incidence of low birth weight observed among second-generation Mexicans.

To verify whether even the second part of the paradox—the deterioration of immigrant

---

<sup>21</sup>Without loss of generality one can choose units of birth weight such that  $\sigma_{BW}^2 = \gamma_{h_2}^2 + \sigma_v^2 = 1$  and therefore  $\text{Cov}(BW_3, BW_2) = \text{Corr}(BW_3, BW_2)$ .

<sup>22</sup>Alternatively, I can set the intergenerational correlation in health and analyze the relationship between selection and differences in low birth weight for different values of  $\gamma$  and for a range of values of intergenerational correlation in birth weight around the 0.2 estimated in the data and in the literature. The implications of the model do not change substantially.

health—can be entirely explained by selection and regression towards the mean, I then examine what would be the expected differences in low birth weight between children of second-generation Mexican immigrants and children of white natives (*y*-axis) (Figure 3). The vertical solid line corresponds to the degree of selection explaining the second-generation advantage (see Figure 2). The dashed line marks the raw difference ( $-0.001$ ) in low birth weight in the data between third-generation Mexicans and white natives (column 2, Table 3).

To pin down the effects of socioeconomic assimilation and account for the socioeconomic gradient in health, I rely on previous estimates on the causal effect of income on birth weight. Cramer (1995) finds that a 1% change in the income-to-poverty ratio increases birth weight by approximately 1.05 grams. More recently, Almond et al. (2009) find similar marginal effects analyzing the effect of food stamps on birth outcomes. Using CPS data (1994-2009), I estimate that on average the family income-to-poverty ratio among Mexicans is 42% lower than among U.S. natives (see Table 3, Panel C, column 2).<sup>23</sup> Using the Cramer (1995) estimate, with everything else constant, the birth weight of Mexicans should be on average 45 grams lower than that of natives. I can then impute the difference between the health distribution of second-generation Mexicans and that of U.S. natives, assuming full assimilation to white natives on other unobservable characteristics affecting health (including behavioral risk factors). Accounting for socioeconomic gradient in health and the positive, but less than full, socioeconomic assimilation observed among second-generation Mexicans, the model not only explains the paradox, but it reverses it: 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 about 1.5 percentage points higher than natives and 1.6 percentage points higher than what observed in the data.<sup>24</sup>

---

<sup>23</sup>The earliest year in which information on the birthplaces of the father and mother is available is 1994 in the CPS surveys.

<sup>24</sup>The results tend in the same direction if considering the entire Hispanic group or if using socioeconomic information at the zip code level (see online Appendix). Note that accounting only for the relative weak intergenerational correlation in health and assuming no socioeconomic assimilation, the model would predict a much higher convergence. On the contrary, assuming full assimilation in socioeconomics and accounting

### 4.3 Accounting for maternal risky behaviors

So far, I did not consider the role of risky behaviors. However, there is abundant literature showing that risky behaviors affect health and birth outcomes. Administrative records provide only limited information on health behavior during pregnancy and only for the more recent years. Therefore, I am not able to verify directly how the 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 U.S.-born second-generation immigrants of Hispanic origin and first-generation immigrants. Information on adult behaviors and health conditions is very limited in California, while the Florida data report tobacco use, alcohol consumption, and weight gain during pregnancy from 1989 onwards, and on pre-pregnancy U.S. (weight and height), chronic hypertension, gestational hypertension, and diabetes from 2004 onwards. For this reason to analyze the role of behavioral assimilation, I focus on the Florida sample but I integrate the analysis using the information on behaviors and risk factors contained in the Natality Detail Data for the entire United States.

Panel B in Table 3 illustrates the mean differences in the incidence of these risk factors between first-generation Hispanics and natives (column 3), and between second-generation Hispanics and natives (column 4). First-generation immigrants have substantially lower incidence of risk factors compared to non-Hispanic white natives. Second-generation immigrants show some convergence towards the less healthy behaviors and higher incidence of risk factors of natives, but they retain a fairly sizeable advantage over natives. Overall, these differences are similar when analyzing the Natality Detail Data (see Table 4). Note, however that in the Natality Detail Data I cannot distinguish second from later generation immigrants and this is likely to explain the more marked worsening in behaviors observed in column 2 of Table

---

for the persistent differences observed in behaviors, the model confirms the paradox that third-generation Hispanic children would be expected to show better statistics than natives for low birth weight, but they do not. However, second-generation Mexicans are not likely to be exposed to the identical quality of care, environment and socioeconomic characteristics of the “average non-Hispanic white” (see Duncan and Trejo (2011)).

4.

Accounting for the observed risk factors (the upper dashed line in Figure 3), the model can explain approximately 76% of the reverse paradox found after accounting for socioeconomic differences.<sup>25</sup> Note that I am able to account for the contribution of only a limited set of behaviors for which information is available in the data. Dietary practices have been shown to be significant determinants of birth outcomes. In particular, fruit and vegetable intake has been shown to be important (Guendelman and Abrams, 1995). Therefore, the unexplained part of the “reverse paradox” is likely to be related to other types of behavior, such as dietary habits, for which I do not have data but that are known to significantly affect birth outcomes.

Taken together, the model suggests that a combination of selection, alongside positive but less than full socioeconomic assimilation and persistence in lower incidence of health risk factors, can explain fairly well the Hispanic paradox in low birth weight. To confirm the importance of socioeconomic and behavioral assimilation, in the next section I analyze the heterogeneity of health convergence among Hispanics.

## **5 Heterogeneity among second-generation Hispanics: A within-family analysis**

Examining the differences in birth outcomes of third-generation immigrants of Hispanic origin may further clarify the extent to which the deterioration in infant health is inevitable. I explore the intergenerational pathways of infant health following migration restricting the sample to the descendants of first-generation immigrants. In particular, I focus on the role of socioeconomic characteristics, quality of care and health risk factors, and cultural

---

<sup>25</sup>More specifically, depending on whether we consider the low birth weight differences in the United States or in the California and Florida samples, controlling for behavior and health conditions helps us to explain between 66% and 83% of the reverse paradox. Despite these differences, these results show that the model fits fairly well with the observed pattern in the data once we account for both the persistence in healthy types of behavior and less-than-full socioeconomic assimilation.

assimilation. Formally, I estimate the following linear probability model:

$$H_{it,3} = \alpha + \beta_1 SES_{it,2} + \beta_2 Acc_{it,2} + \gamma X_{it,2} + \lambda H_{i,2} + \epsilon_{it,3}$$

where  $H_{it,3}$  is a metric of infant health of the third generation,  $SES_{it,2}$  is an indicator of socioeconomic status of the second generation,  $Acc_{it,2}$  captures the effect of cultural assimilation or risky behaviors,  $X_{it,2}$  are a set of the sociodemographic controls, and  $H_{i,2}$  is a metric of infant health of the second generation. To account for potential omitted variable bias, I include grandmother-fixed effects and exploit differences among siblings (within a family) in the covariates under analysis. I identify siblings born between 1975 and 1981 using information on the maternal grandmother (the mother’s mother). To match grandmothers (the first-generation immigrants) across the different birth certificates of their children (second-generation immigrants), I use information on the grandmother’s name, child’s last name, mother’s race, and mother’s state of birth. This implies that children born to the same mother but from different fathers would not be considered in my sample of siblings. I drop individuals for whom the matching variables are missing.<sup>26</sup> Controlling for the birth weight of the second-generation mother and including grandmother-fixed effects allows partially capturing the initial selectivity associated with the migration process. In particular, comparing the birth outcomes of third-generation cousins eliminates the bias introduced by genetic and environmental factors that are constant within the family and, in particular, for the common characteristics of mothers (sisters) who grew up in the same family.

---

<sup>26</sup>Regarding the matching of mothers to grandmothers, in California I matched only one daughter in 84% of the cases, I matched two daughters in 12% of the cases, and I matched three or more daughters to each grandmother in 4% of the cases. In Florida, I matched only one daughter in 80% of the cases, I matched two daughters in 17% of the cases, and I matched 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, the average number of grandchildren linked to each grandmother is 2.50, which number is 4.20 if conditioned on linking at least two second-generation sisters to their offspring.



## 5.1 Socioeconomic status and quality of care

Table 5 summarizes the effects of quality of care and socioeconomic status on the birth outcomes of third-generation of Hispanic origin. All estimates include an extensive set of sociodemographic characteristics available at the time of birth of second generation (including second-generation birth weight, the poverty share in the zip code of birth of the second-generation mothers, first-generation grandmother’s age dummies, and dummies capturing the interaction of county of residence and year of birth at the time of second-generation’s birth) and a set of covariates available at the time of birth of third-generation children comprised (containing information on marital status, parity, father’s education (four educational groups), a quadratic in father’s age, and dummies capturing the interaction of county of residence and year of birth at the time of the third-generation birth). Column 2 includes grandmother-fixed effects exploiting within-family variation.<sup>27</sup>

The adequacy of prenatal care is here defined as starting prenatal care in the first trimester of pregnancy. The first row of Table 5 shows that children of mothers who started care later in the pregnancy or had no prenatal care show significantly lighter birth weights (Panel A) and are at higher risk of low birth weight (Panel B). The coefficient on adequacy of prenatal care remains strong and significant to the addition of sociodemographic controls (column 1) and grandmother-fixed effects (column 2).

In the second row, I use the poverty level of the zip code of residence of second-generation mothers at the time of birth of their children to identify mothers who were living in the poorest zip code (i.e., the lowest quartile of income distribution and highest quartile of poverty distribution). The marginal effects with respect to the mean correspond to a 3.5% increase among Hispanic descendants. However, coefficients become smaller and nonsignificant when controlling for grandmother-fixed effects (column 2). Notably, the coefficients are larger and more robust when using the poverty rate in the zip code of the hospital (the third row) rather

---

<sup>27</sup>A more detailed analysis of the sensitivity of the coefficients to the addition of different batteries of controls is available in the online Appendix.

than the poverty rate of the residential zip code of the mother. Third-generation immigrants of Hispanic origin born in a hospital located in a high-poverty zip code are 11% more likely to be low birth weight than their counterparts born in hospitals located in wealthier areas. The effect remains relevant and significant when including grandmother-fixed effects and exploiting differences in the adult socioeconomic background of second-generation siblings (see column 2). These results might reflect selection of hospital choice and differences in the quality of care received. Overall, the evidence presented in Table 5 suggests that the poverty of the environment at birth is associated with poorer birth outcomes. The results are similar when using the median income of the zip code rather than the zip code poverty rate to define socioeconomic status.<sup>28</sup>

## 5.2 Maternal risk factors

Table 6 illustrates the relationship between different risk factors and the incidence of low birth weight. The sample is composed of all the mothers born between 1971 and 1985 in Florida, including both natives and Hispanics, since here I am interested in providing further evidence of the causal effect of behaviors. However, including a set dummies capturing the interaction between risk factors and country of origin does not affect the results. Columns 1–3 analyze the effect on low birth weight of risk factors for which information is available since 1989: tobacco and alcohol consumption during pregnancy, normal weight gain which is defined gaining between 24 and 40 pounds.<sup>29</sup> Column 2 control for the same set of sociodemographic controls using in 5. Column 3 includes grandmother-fixed effects.

Smoking during pregnancy has been widely recognized as the most modifiable risk factor for LBW (Almond et al., 2005; Currie and Schmieder, 2009). Table 6 shows that tobacco

---

<sup>28</sup>Using as an alternative metric of socioeconomic status education I find that while the raw correlation between graduating from high school and the incidence of low birth weight is negative as one would expect, when controlling for socioeconomic characteristics and introducing grandmother-fixed effects the coefficient becomes negative. However, given the collinearity of maternal education with adequacy of care, poverty of the neighborhood of residence, and father’s education, this coefficient should interpreted with caution (see online Appendix for the entire set of results).

<sup>29</sup>I do not have information on pre-pregnancy BMI before 2004.

use during pregnancy increases the incidence of low birth weight by 2 percentage points, which is more than 20% of the average incidence of low birth weight in the Florida sample. The coefficient is relatively robust to the addition of sociodemographic controls (column 2), grandmother-fixed effects (column 3) and other risk factors for which information is available 2004 onwards (columns 4–6). Alcohol use during pregnancy is associated with a 4 percentage points increase in the incidence of low birth weight, but the coefficient becomes nonsignificant once controlling for sociodemographic characteristics. Having a normal weight gain (not adjusted for BMI) is associated with lower likelihood of low birth weight (columns 1–3). Mother’s chronic and gestational hypertension are shown to be important determinants for the incidence of low birth weight. The coefficients are important in magnitude and robust to the addition of sociodemographic controls (column 5) and grandmother-fixed effects (column 6). Diabetes increases the risk of low birth weight, but the coefficient is not significant after controlling for grandmother-fixed effects. Since 2004 onwards the Florida data contain information on pre-pregnancy BMI, I can compute a better measure of adequate gain using the Institute of Medicine (IOM) criteria which define as adequate weight gain between 28 and 40 pounds for women with a BMI lower than 18.5, between 25 and 35 pounds for women with a BMI between 18.5 and 24.9, between 15 and 25 pounds for women with a BMI between 25 and 29.9, between 10 and 20 for women with BMI equal to or higher than 30. When adjusting for pre-pregnancy BMI, the coefficient on adequate weight gain becomes nonsignificant.<sup>30</sup>

Given the cross-sectional differences in risk factors between first- and second-generation immigrants and the net effect of each risk factor on low birth weight, one can estimate, with a back-of-the-envelope calculation, that the incidence of low birth weight would have

---

<sup>30</sup>While not reported in Table 6, the coefficients of pre-pregnancy obesity is nonsignificant. If anything the sign of the coefficient shows that obesity might have some protective effect on the risk of low birth weight. However, the analysis of the relationship between maternal pre-pregnancy BMI and risk of low birth weight is problematic, given the nonlinearity of the relationship between pre-pregnancy BMI and birth weight. Indeed, when considering as an alternative measure of infant health the likelihood of having a low Apgar score, I find evidence of lower Apgar scores among children of obese mothers. Results are available upon request.

been 2% lower (with respect to the average low birth weight incidence), if the behaviors did not worsen. This confirms that behaviors may have worsened, but too little, compared to natives, to explain the Hispanic paradox.<sup>31</sup>

These results show that policies aimed at preventing the adoption of less healthy lifestyles that affect adult and child health might have nonmarginal benefits on the birth outcomes of future Americans. This is particularly significant with respect to smoking. While tobacco use has declined dramatically in developed countries, it has increased and become more acceptable in less developed countries and among immigrants living in the United States, particularly among women. These facts explain the growing attention that the tobacco industry has devoted to target U.S. immigrants (Acevedo-Garcia et al., 2004), who now represent an appealing market.

### **5.3 The role of acculturation: Analyzing ethnic intermarriage**

To overcome the limits imposed by the lack of extensive information on risky behaviors and health risk factors, I also use a broad metric of cultural assimilation: ethnic intermarriage. Acevedo-Garcia et al. (2005) provide evidence that ethnic intermarriage has an important role in influencing the adoption of native lifestyle behaviors, affecting, for instance, anti-smoking socialization among Latinos. More generally, intermarriage has been defined as “the last stage of assimilation” and it is known to significantly affect the process of adaptation in the new country (Furtado and Theodoropoulos, 2009; Furtado and Trejo, 2012). In practice, I define intermarriage as a dummy equal to one if the father is a non-Hispanic white and zero if the father is Hispanic. The sample is therefore restricted to second-generation immigrant women born between 1975 and 1981, whose mothers migrated from Cuba, Puerto

---

<sup>31</sup>To conduct the back-of-the-envelope calculation, I used generational differences, and, for each behavior, I estimated the net effect on LBW by using a unique regression, including smoking, alcohol consumption, normal weight gain, gestational hypertension, chronic hypertension, chronic diabetes and controlling gender, race, fixed effects for child’s year of birth and an interaction of the grandmother’s county of residence and the year of the mother’s birth. Most of the change is explained by changes in smoking and gestational hypertension.

Rico or Mexico.<sup>32</sup> Because I do not have information on the country of origin of the father for the entire period under study, I use information on the father's ethnic origin to define intermarriage.<sup>33</sup> I include all the observations for which I had information on paternal ethnicity. Hereafter, I will use intermarriage to indicate the different ethnicity of mother and father, regardless of the reported or inferred marital status.

Table 7 confirms that intermarriage is importantly related to behavioral assimilation. Column 1 reports raw correlations for the entire United States, while column 2 shows the same correlation using California and Florida data. Second-generation intramarried pregnant women are less likely to smoke and drink during pregnancy and they show lower rates of chronic and gestational hypertension. The rate of smoking among second-generation intramarried pregnant women is identical to the one observed among first-generation immigrants (see the online Appendix). Columns 2–4 test the robustness of these correlation to the inclusion of sociodemographic characteristics and grandmother-fixed effects. Panel A shows that the coefficient of intermarriage on smoking is always positive and significant. It shrinks from 2 percentage points to 0.6 percentage points after including grandmother-fixed effects, but the effect remains a large and meaningful effect. With respect to smoking, being intermarried is associated with a 30% higher likelihood of tobacco use during pregnancy. The results tend in the same direction for other risk factors, such as alcohol use and gestational hypertension (Panels B and C). However, in this case the magnitude and the significance of the coefficient are less robust to the addition of grandmother-fixed effects. Intermarriage is of course only a rough measure of acculturation, but this evidence suggests that it might be significantly correlated to many other important behaviors that are observed to change across generations and are known to significantly affect the health transmission process.<sup>34</sup>

---

<sup>32</sup>The data contain information on the country of origin only for major countries of origin. However, when considering individuals whose mothers were born abroad and who report Hispanic origin, the results are similar.

<sup>33</sup>Please note that Baumeister et al. (2000) found a surprisingly high coincidence (95%) between the ethnicity reported by birth clerks in the California birth certificates and the ethnicity reconstructed in personal interviews.

<sup>34</sup>For instance, using the American Time Use Survey, I find evidence that among immigrants of Hispanic origin, intramarriage is positively associated with time spent in food preparation. While this goes out of the

The coefficients of intermarriage on birth weight and incidence of low birth weight reported in Table 8 suggest the existence of a paradox within the paradox, in that the children of second-generation Hispanic mothers who intermarried have worse birth outcomes despite their higher socioeconomic status. In particular, children of Hispanic mothers who intermarried with a non-Hispanic white native weigh on average 88 grams less than children of intramarried couples (Panel A, column 8). The effects are larger in magnitude when focusing on the lower tail of the birth weight distribution. Intermarriage is associated with a significantly higher risk of low birth weight. The incidence of low birth weight for children of intermarried couples is 1.8 percentage point higher (Panel B, column 1). In light of the previous literature (Furtado and Theodoropoulos, 2009; Furtado and Trejo, 2012; Wang, 2012) that shows evidence of a positive relationship between intermarriage and socioeconomic outcomes, the fact that the children of intermarried couples have worse birth outcomes is particularly striking. Clearly intermarriage is not an exogenous decision, but the unobservables that are usually associated with the likelihood of marrying a native would be likely to downward bias the estimated coefficient. Wang (2012) reports that Hispanic–White couples on average earn approximately \$20,000 more than Hispanic–Hispanic couples.<sup>35</sup>

The intermarriage coefficient on birth weight (in grams) shrinks by 30% when including sociodemographic controls and previous generation birth weight, and becomes relatively small when including grandmother-fixed effects. The latter corresponds to an effect of approximately  $-0.5\%$  of the average birth weight in the sample. The pattern is similar when analyzing low birth weight. In particular, adding grandmother-fixed effects (column 3) and focusing on within-family variation, the protective effect of intermarriage shrinks by 20%

---

scope of this paper, it is important to notice that recent research has shown important correlations between health and consumption of more processed food.

<sup>35</sup>In the data, intermarriage is positively correlated with median family income (+0.15) in the zip code, with mother’s education (0.10) and father’s education (.14), while negatively related to the zip code poverty share ( $-.17$ ). A relevant concern is that selection in the marriage market might be substantially different for men and women. However, while there are significant gender differences in intermarriage rates between blacks and whites, Wang (2012) shows that there are no significant gender differences in the intermarriage rates of Hispanics and whites; white men who married Hispanic women are not less educated than those who married white women. In particular, 32.3% of white men married to white women completed college education, compared to 33.1% of white men who married Hispanic women.

when adding sociodemographic controls and by a further 40% when including grandmother-fixed effects. However, the coefficient of intermarriage on low birth weight is still large and significant. The incidence of low birth weight among children of intermarried couples is 0.008 percentage points higher than among intermarried couples (+14% with respect to the mean of the dependent variable). It is noteworthy that this result is not sensitive to the addition of controls for the mother's and father's educational dummies, which are included in columns 2–4.

The results presented in column 3 might still reflect the selectivity of migrants on the father's side. By controlling for grandmother-fixed effects and mother's birth weight, I am able to eliminate the selection effects coming from the mother's side, but the intermarriage coefficient might nonetheless be the result of the genetic advantage carried by the immigrant-descendant father. To test for the role of the father's selectivity I conduct a placebo test analyzing the effect of intermarriage among second-generation non-Hispanic white women and focusing on the birth outcomes of their offspring. If the intermarriage coefficient is capturing father's selectivity, one should expect a protective role of having a Hispanic father even when analyzing the effects of intermarriage on birth weight among third-generation native children. When conducting this test, I find that marrying a man of Hispanic origin has no protective effect and if anything increases the risk of low birth weight (see column 4 in Table 8).

There may be other unobserved factors affecting both selection in the marriage market and birth outcomes. However, the extensive set of controls and the overall robustness of the coefficient to the inclusion of grandmother-fixed effects reduce the concern that these confounding factors might significantly alter the primary finding and the validity of the falsification test.<sup>36</sup>

---

<sup>36</sup>The placebo test is robust not only to the addition of father's and mother's education, which are included in columns 2–4, but also to the separate analysis of women who married to equally, more, or less educated Hispanic men (results available upon request). Note also that when controlling for grandmother-fixed effects and the previous generations birth weight, marrying a high school dropout has no significant effect on the risk of low birth weight of third-generation white natives.

## 6 Conclusion

This paper confirms that while second-generation Hispanics have lower incidence of low birth weight than children of native white mothers, this advantage shrinks substantially in the third generation. With the help of a simple model of selection and intergenerational health transmission, I show that selection might explain the better birth outcomes of second-generation children compared to white natives; however, if the only factor were selection, third-generation birth outcomes would be worse than the ones observed in the data.

Accounting for socioeconomic differences between second-generation Hispanics and natives, the model not only explains, but actually reverses the paradox: the puzzle is not that immigrant relative health deteriorates so rapidly, but that it does not deteriorate rapidly enough. However, accounting for the differences in risk factors (such as tobacco and alcohol consumption during pregnancy and gestational hypertension) the model explains approximately 76% of the reverse paradox. While there is evidence of a generational worsening in undertaking 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 little compared to natives. The importance of risky types of behavior is confirmed by the analysis of differences in the health convergence among second-generation Hispanics. Children of Hispanic women who show higher incidence of risk factors and are more assimilated are more likely to have poor birth outcomes. This holds true even after accounting for potential confounding factors, focusing on a subsample of second-generation siblings and controlling for grandmother-fixed effects.

As a whole, these findings show that the health trajectories observed among Hispanic descendants cannot be entirely explained by a pure mechanical statistical process. While there is evidence of a natural regression towards the mean, socioeconomic and behavioral factors mediate the transmission of health across generations. Policies aimed at reducing disparities in access to and quality of health care, and at maintaining healthy behaviors can significantly affect these health patterns. Because second-generation births are overtaking



migration as the main source of growth in the American population, such policies could have important effects.

## References

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

- Baumeister, Lisa, Kristen Marchi, Michelle Pearl, Ronald Williams, and Paula Braveman (2000) ‘The validity of information on “race” and “Hispanic ethnicity” in California birth certificate data.’ *Health Services Research* 35(4), 869–883
- Borjas, George J. (1987) ‘Self-selection and the earnings of immigrants.’ 77(4), 531–553
- 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
- Chiswick, Barry R., Yew Liang Lee, and Paul W. Miller (2008) ‘Immigrant selection system and immigrant health.’ *Contemporary Economic Policy* 26(4), 1074–3529
- 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
- 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

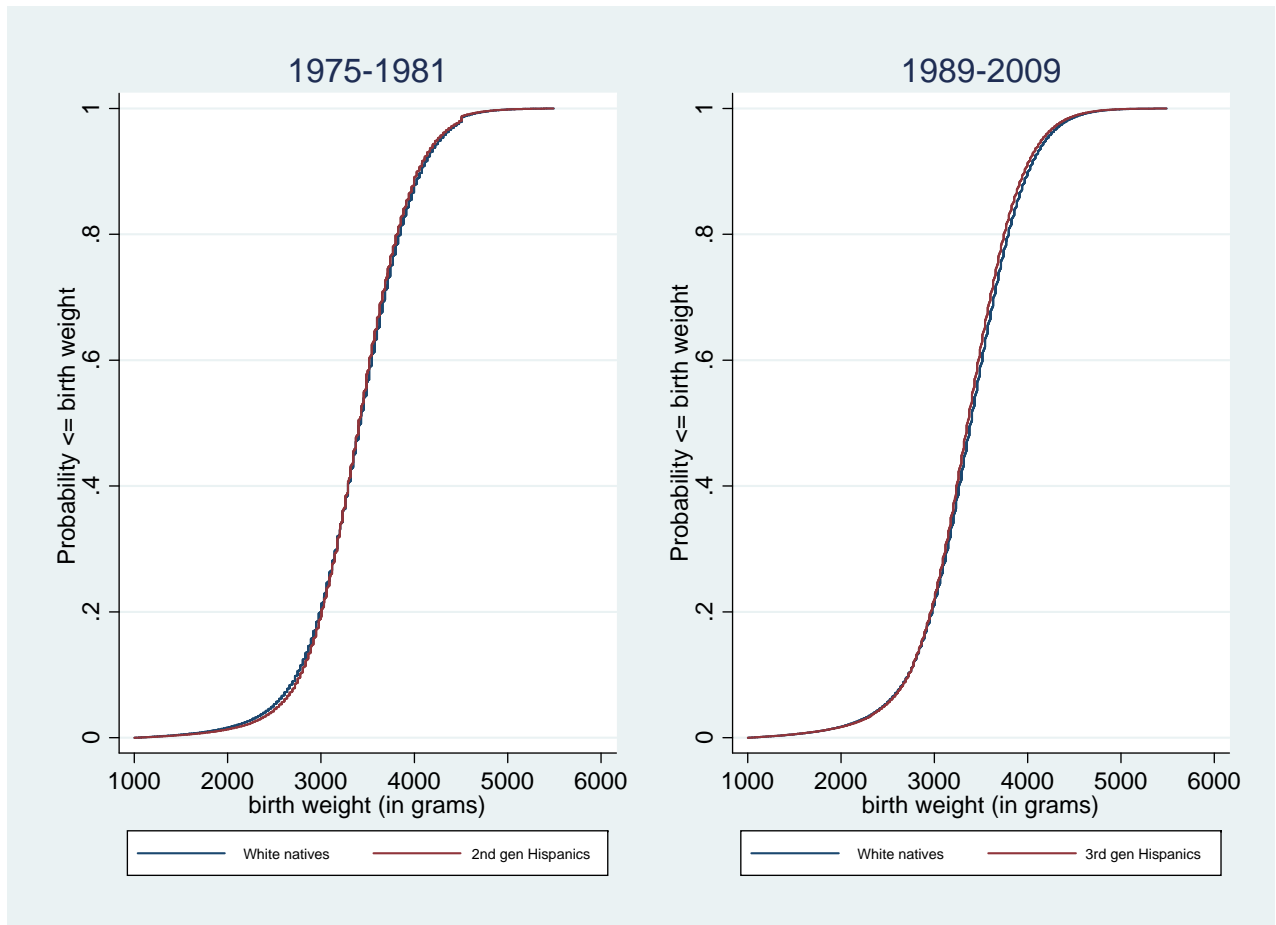
- Duncan, Brian, and Stephen Trejo (2011) ‘Low-skilled immigrants and the U.S. labor market.’ *IZA Discussion Paper No. 5964*. [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1929664](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1929664). Accessed Nov. 2012
- Elder, Todd E., John H. Goddeeris, and Steven J. Haider (2012) ‘Disparate disparities: Understanding differences in infant mortality across racial and ethnic groups.’ [https://www.msu.edu/~telder/DD\\_Current.pdf](https://www.msu.edu/~telder/DD_Current.pdf). Accessed Nov. 2012
- Fenelon, Andrew (2012) ‘Revisiting the Hispanic paradox in the United States: The role of smoking.’ <http://paa2012.princeton.edu/papers/120602>. Accessed Nov. 2012
- Finch, Brian Karl, Nelson Lim, William Perez, and D. Phuong Do (2007) ‘Toward a population health model of segmented assimilation: The case of low birth weight in Los Angeles.’ *Sociological Perspectives* 50(3), 445–468
- 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
- Furtado, Delia, and Nikolaos Theodoropoulos (2009) ‘Intermarriage and immigrant employment the role of networks.’ *The B.E. Journal of Economic Analysis and Policy*
- Furtado, Delia, and Stephen Trejo (2012) ‘Interethnic marriages and their economic effects.’ CReAM Discussion Paper Series 1205, Centre for Research and Analysis of Migration (CReAM), Department of Economics, University College London, Feb. <http://ideas.repec.org/p/crm/wpaper/1205.html>. Accessed Nov. 2012
- 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 non-Hispanic women.' *American Journal of Public Health* 85(1), 20–25
- Gutmann, Myron P., W. Parker Frisbie, Peter DeTurk, and K. Stephen Blanchard (1998) 'Dating the origins of the epidemiologic paradox among Mexican Americans.' *Texas Population Research Center Papers*. [http://www.utexas.edu/cola/centers/prc/\\_files/pdf/workingpapers/97-98-07.pdf](http://www.utexas.edu/cola/centers/prc/_files/pdf/workingpapers/97-98-07.pdf). Accessed Nov. 2012
- 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 (2010) 'Self-selection and international migration: New evidence from Mexico.' *NBER*. <http://www.nber.org/papers/w15765.pdf>. Accessed Nov. 2012
- 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
- Kaushal, Neeraj (2008) 'Adversities of acculturation? Prevalence of obesity among immigrants.' *Health Economics* 18, 291–303

- Koenker, Roeger, and Kevin F. Hallock (2001) ‘Quantile regression.’ *Journal of Economic Perspectives* 15(4), 143–156
- 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
- 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
- Orrenius, Pia M., and Madeline Zavodny (2010) ‘Self-selection among undocumented immigrants from mexico.’ *Journal of Development Economics* 78, 215–250
- Osypuk, Theresa L., Lisa Bates, and Dolores Acevedo-Garcia (2010) ‘Another Mexican birth weight paradox? The role of residential enclaves and neighborhood poverty in the birth weight of Mexican-origin infants.’ *Social Science and Medicine* 70, 550–560
- 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
- Palloni, Miguel Ceballos Alberto (2010) ‘Maternal and infant health of Mexican immigrants in the United States: the effects of acculturation, duration, and selective migration.’ *Ethnicity and Health* 15(4), 377–396
- Reinhold, Steffen, and Kevin Thom (2012) ‘Migration experience and earnings in the Mexican.’ <https://files.nyu.edu/kt44/public/MigExperienceAndEarnings.pdf>. Accessed Nov. 2012

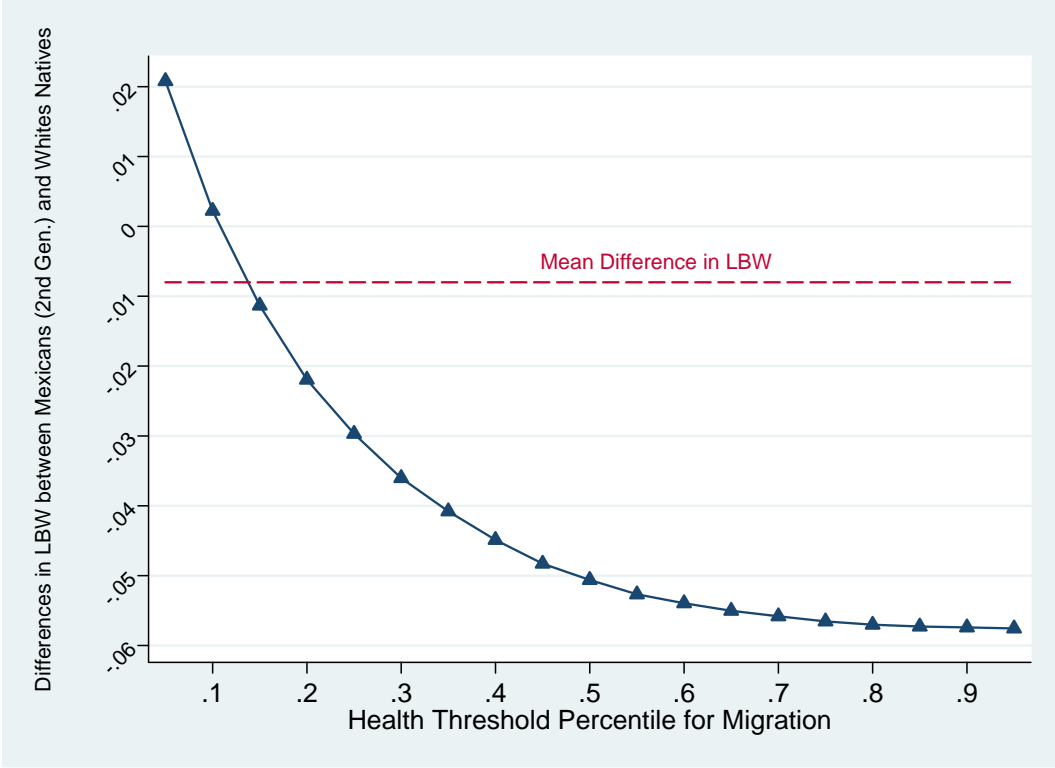
- Royer, Heather (2009) 'Separated at girth: US twin estimates of the effects of birth weight.' *American Economic Journal: Applied Economics* 2(2), 49–85
- Shaw, Richard J., Kate E. Pickett, and Richard G. Wilkinson (2010) 'Ethnic density effects on birth outcomes and maternal smoking during pregnancy in the US linked birth and infant death data set.' *American Journal of Public Health* 100(4), 707–713
- 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
- Wang, Wendy (2012) 'The rise of intermarriage. rates, characteristics vary by race and gender.' *Pew Social and Demographic Trends*. <http://www.pewsocialtrends.org/2012/02/16/the-rise-of-intermarriage/>. Accessed Nov. 2012
- 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 mexican migrants.' *The World Economy* (forthcoming)

Figure 1: Birth weight (grams) distribution, California and Florida



Notes - Source: California and Florida Vital Statistics, 1975–1981, 1989–2009.

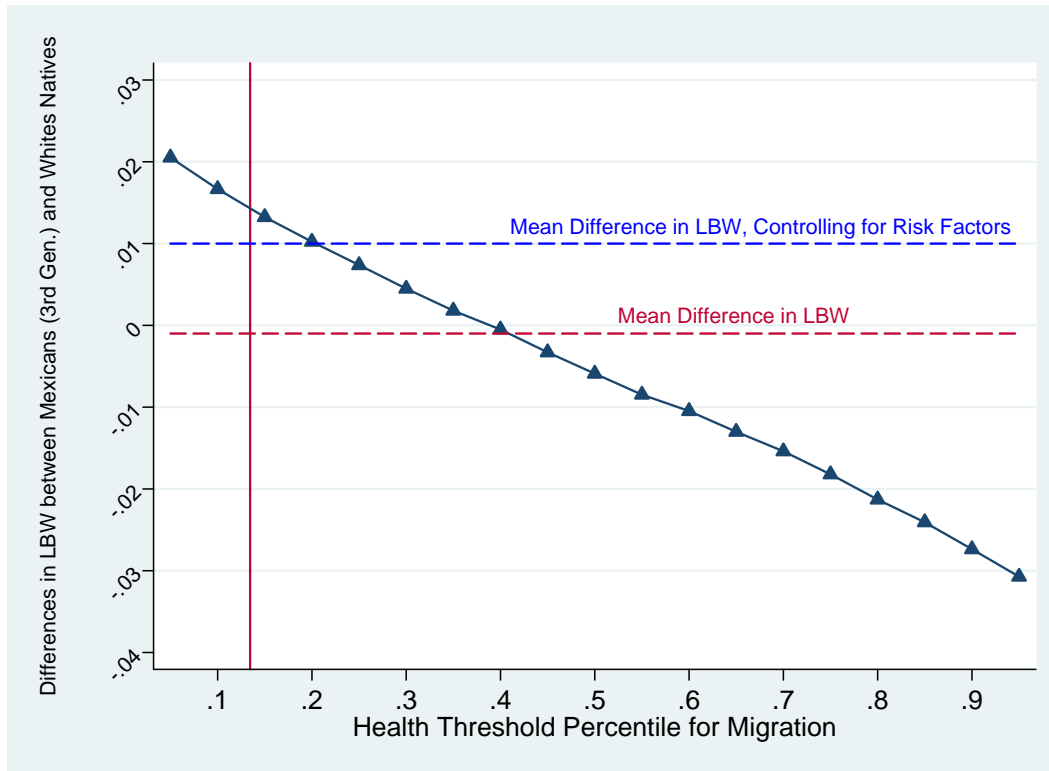
Figure 2: Selection on health at migration and differences in the incidence of low birth weight between 2<sup>nd</sup> generation Mexicans and white natives ( $\gamma = 0.75, \rho = 0.35$ )



Notes - The plotted curve reports the predicted low birth weight differences between 2nd generation Mexicans and white natives for each level of selection on health at migration, assuming that the intergenerational correlation in health  $\rho$  is equal to 0.35 and the effect of maternal health on birth weight,  $\gamma$ , is equal to 0.75 (baseline). The dashed line describes the observed raw difference in the incidence of low birth weight between 2nd generation Mexicans and white natives born between 1975 and 1981, in California and Florida (see Table 3, col. 1).



Figure 3: Regression towards the mean and differences in the incidence of low birth weight between 3rd generation Mexicans and white natives ( $\gamma = 0.75, \rho = 0.35$ )



Notes - The plotted curve reports the predicted low birth weight differences between 3rd generation Mexicans and white natives for each level of selection on health at migration, assuming that the intergenerational correlation in health  $\rho$  is equal to 0.35 and the effect of maternal health on birth weight,  $\gamma$ , is equal to 0.75 (baseline). The scenario considered assumes that Mexicans fully assimilate in behaviors but incorporates the estimated effect on birth weight of the observed socioeconomic differences between second-generation Mexicans and white natives (less than full socioeconomic assimilation,  $\mu_{MX_2} = -0.1$ ). The lower dashed line ( $y = -0.001$ ) describes the observed raw difference in the incidence of low birth weight between 3rd generation Mexicans and white natives born between 1989 and 2009 in California and Florida (see Table 3, col. 2). The upper long-dashed line ( $y = 0.011$ ) describes the observed raw difference in the incidence of low birth weight between 3rd generation Mexicans and white natives born between 1989 and 2009, after controlling for tobacco and alcohol consumption during pregnancy and gestational hypertension (see Table 3, col. 5). The vertical solid line represents the level of selection (0.135) that would explain the low birth weight difference observed in the data between 2nd generation Mexicans, see Figure 2.

Table 1: Matching quality. Women born in California and Florida, 1975–1981

Sample:	Observations			Birth Weight (grams)			Low Birth Weight (below 2500 grams)		
	(1) Overall	(2) Linked	(3) Matching rate	(4) Overall	(5) Linked	(6) Nonlinked	(7) Overall	(8) Linked	(9) Nonlinked
Overall	1,643,865	726,837	0.44	3,288	3,289	3,288	0.069	0.065	0.072
U.S. born	1,273,023	558,921	0.44	3,283	3,277	3,288	0.073	0.070	0.075
U.S. born black	175,493	123,472	0.70	3,072	3,081	3,050	0.130	0.124	0.144
U.S. born white	1,097,530	435,449	0.40	3,317	3,333	3,307	0.064	0.055	0.070
Foreign born	370,842	167,916	0.45	3,304	3,328	3,285	0.055	0.046	0.061
Hispanic	231,741	124,267	0.54	3,327	3,345	3,305	0.051	0.044	0.060
Cuban	17,290	8,695	0.50	3,301	3,309	3,293	0.056	0.047	0.065
Mexican	198,264	109,661	0.55	3,341	3,356	3,323	0.049	0.042	0.057
Puerto Rican	16,187	5,911	0.37	3,175	3,186	3,168	0.074	0.064	0.080
Zip code level income:									
1st income quartile	273,285	131,932	0.48	3,265	3,266	3,264	0.074	0.069	0.078
2nd income quartile	293,879	140,486	0.48	3,261	3,260	3,262	0.077	0.073	0.081
3rd income quartile	442,946	194,460	0.44	3,289	3,291	3,287	0.068	0.064	0.071
4th income quartile	367,580	144,510	0.39	3,321	3,324	3,320	0.060	0.056	0.062

Notes - Data are drawn from the California and Florida Vital Statistics, 1975–1981. The linked sample is composed of all the women born between 1975 and 1981 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: Hispanic Health Paradox in birth weight (BW) and low birth weight incidence (LBW)

Dependent variable:	2nd generation 1975–1981			3rd generation 1989–2009				
	BW (1)	(2)	LBW (3)	(4)	BW (5)	(6)	LBW (7)	(8)
Country of origin:								
Cuba	35.407* (20.246)	-0.285 (4.266)	-0.004 (0.012)	-0.009*** (0.002)	-50.778*** (5.501)	-44.807*** (7.242)	-0.001 (0.002)	0.001 (0.003)
Mexico	-2.129* (1.123)	22.647*** (1.326)	-0.008*** (0.000)	-0.016*** (0.001)	-27.983*** (1.735)	-8.078*** (1.927)	-0.001** (0.001)	-0.004*** (0.001)
Puerto Rico	-74.826*** (18.714)	-91.191*** (7.406)	0.014 (0.011)	0.005 (0.003)	-172.970*** (6.884)	-158.849*** (6.974)	0.028*** (0.003)	0.022*** (0.003)
Sociodemographic controls		YES		YES		YES		YES
Mean of Dep. Var.	3394.289	3394.289	0.054	0.054	3347.184	3347.184	0.058	0.058
Std.dev.	572.415	572.415	0.225	0.225	575.201	575.201	0.233	0.233
Observations	2,560,258	2,234,571	2,560,258	2,234,571	1,102,271	1,043,636	1,102,271	1,043,636

Standard errors in parentheses

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Notes - Data are drawn from the California and Florida Birth Records (1975–1981, 1989–2009). All estimates include state FE (except when data are available for Florida only). 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 3: Differences between 1st, 2nd generation Mexicans and U.S. white natives

	(1)	(2)	(3)	(4)	(5)
	<b>CA-FL</b>			<b>Florida</b>	
	$MX_1 - N$	$MX_2 - N$	$MX_1 - N$	$MX_2 - N$	$MX_2 - N$
<i>Panel A: CA-FL Vital Statistics</i>					
Low birth weight	-0.008***	-0.001**	-0.009**	0.004	0.011***
	(0.000)	(0.000)	(0.003)	(0.004)	(0.004)
Control for risk factors	NO	NO	NO	NO	YES
<i>Panel B: FL Vital Statistics</i>					
Tobacco consumption			-0.159***	-0.150***	
			(0.001)	(0.002)	
Alcohol consumption			-0.003***	-0.003***	
			(0.000)	(0.000)	
Gestational hypertension			-0.020***	-0.016***	
			(0.001)	(0.002)	
<i>Panel C: CA-FL CPS</i>					
Socioeconomic status					
log(family income/poverty)	-0.798***	-0.401***	-0.721***	-0.529***	
	(0.006)	(0.011)	(0.023)	(0.054)	

*Notes* - Data used in the first row of Panel A are drawn from the California and Florida Birth Records (1975–1981, 1989–2009). Data on risk factors (rows 2–5) are drawn from Florida Birth Records (1989–2009). Information on gestational hypertension (row 5) is available from 2004 onwards. Data on socioeconomic assimilation are drawn from the Current Population Survey (CPS) (1994–2011). Information on parental birth place is available in the CPS only since 1994. All estimates include state FE (except when data are available for Florida only). Note that data drawn from the California Vital Statistics do not contain information on these risk factors for the period under analysis.

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

	(1)	(2)	
	$MX_1 - N$	$MX_2 - N$	$MX_2 - N$
<i>Panel A: Natality Detail Data</i>			
Low birth weight	-0.009***	0.001***	0.008***
	(0.000)	(0.000)	(0.000)
Control for risk factors	NO	NO	YES
<i>Panel B: Natality Detail Data</i>			
Tobacco consumption	-0.219***	-0.166***	
	(0.000)	(0.000)	
Alcohol consumption	-0.006***	-0.001***	
	(0.000)	(0.000)	
Gestational hypertension	-0.024***	-0.016***	
	(0.000)	(0.000)	
<i>Panel C: CPS</i>			
Socioeconomic status			
log(family income/poverty)	-0.726***	-0.459***	
	(0.003)	(0.004)	

*Notes* - Data are drawn from the Natality Detail Data (1975–1981; 1989–2004). Data on socioeconomic assimilation are drawn from the Current Population Survey (1994–2011). Information on parental birth place is available in the CPS only since 1994. All estimates include state FE (except when data are available for Florida only). 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.

Table 5: Adequacy of care, SES and birth outcomes among 3rd generation Hispanics, within-family analysis

	(1)	(2)
<b>Panel A: Birth Weight (in grams)</b>		
Prenatal care started 1st trimester of pregnancy	182.003*** (4.417)	153.765*** (4.927)
High-poverty (zip code of residence)	-4.261 (3.518)	-8.954* (5.009)
High poverty (hospital zip code)	-10.671** (4.199)	-11.073* (5.763)
Mean of Dep. Var	3326.377	
Std. dev.	560.681	
<b>Panel B: Incidence of Low Birth Weight</b>		
Prenatal care started 1st trimester of pregnancy	-0.049*** (0.002)	-0.041*** (0.002)
High-poverty (zip code of residence)	0.003* (0.001)	0.001 (0.002)
High poverty (hospital zip code)	0.005*** (0.002)	0.006** (0.003)
Mean of Dep. Var	0.056	
Std. dev.	0.229	
<b>Sociodemographic controls</b>	YES	YES
<b>Mother's birth weight</b>	YES	YES
<b>Grandmother F.E.</b>	NO	YES
Observations	201,754	201,754

*Notes* - Data are drawn from the California and Florida Vital Statistics (1975–1981, 1989–2009). The sample is restricted to third-generation Hispanics born between 1989 and 2009. Sociodemographic controls include 3rd generation child's gender, mother's birth weight (LBW), dummies for the interaction of county and year of birth of second-generation children (mothers), first generation's (grandmothers) age (at delivery) dummies, second generation's age (at delivery) dummies, second generation parity, poverty rate in zip code of birth of second generation, marital status, father's and mother's education (4 groups), parity, dummies for the interaction of county and year of birth of third-generation children, indicators for missing information on first generation's (grandmother's) age (at delivery) (FL), father's education, and age. Column 2 includes grandmother-fixed effects. Standard errors are clustered at the grandmother level.

Table 6: The effect of health risk factors on low birth weight, Florida 1989–2009, within-family analysis

Sample period	1989–2009			2004–2009		
	(1)	(2)	(3)	(4)	(5)	(6)
Tobacco consumption	0.018*** (0.001)	0.028*** (0.002)	0.019*** (0.003)	0.036*** (0.002)	0.029*** (0.003)	0.019*** (0.007)
Alcohol consumption	0.041*** (0.006)	0.005 (0.008)	0.006 (0.010)	0.063*** (0.013)	0.021 (0.014)	0.005 (0.028)
Gestational hypertension				0.143*** (0.004)	0.137*** (0.004)	0.088*** (0.007)
Chronic hypertension				0.108*** (0.006)	0.086*** (0.007)	0.027*** (0.013)
Diabetes				0.032*** (0.008)	0.018** (0.008)	0.005 (0.015)
Normal weight gain	-0.032*** (0.001)	-0.021*** (0.001)	-0.019*** (0.001)			
Normal weight gain (adjusted for U.S.)				0.001 (0.001)	0.002 (0.001)	-0.001 (0.002)
Sociodemographic controls	No	Yes	Yes	No	Yes	Yes
Grandmother F.E.	No	No	Yes	No	No	Yes
Observations	793,005	542,517	542,517	284,528	214,711	214,711
Avg. Incidence of LBW	0.07					
Std. dev.	0.25					

*Notes* - Data are drawn from Florida Birth Master Dataset (1989–2009). The sample is restricted to the children of mothers born between 1971 and 1985 in Florida. Sociodemographic controls include 3rd generation child's gender, race, country of origin dummies, mother's birth weight (LBW), dummies for the interaction of county and year of birth of second-generation children (mothers), second generation's age (at delivery) dummies, second-generation parity, poverty rate in zip code of birth of second generation, marital status, father's and mother's education (4 groups), parity, dummies for the interaction of county and year of birth of third-generation children, father's education and age. Standard errors are clustered at the grandmother level.

Table 7: Ethnic intermarriage and low birth weight risk factors among 2nd generation Hispanic mothers, within-family analysis

	(1) U.S.	(2) FL	(3) FL	(4) FL	(5) FL
Panel A, dependent variable: tobacco consumption					
Non-Hispanic white father	0.067*** (0.001)	0.025*** (0.002)	0.020*** (0.002)	0.006** (0.003)	0.006** (0.003)
Mean of dep. Var.	0.065	0.022			
Std. dev.	0.246	0.143			
Observations	626,017	39,662	38,952	38,952	37,243
Panel B, dependent variable: alcohol consumption					
Non-Hispanic white father	0.007*** (0.000)	0.001*** (0.000)	0.001*** (0.000)	0.001* (0.001)	0.001 (0.001)
Mean of dep. Var.	0.007	0.001			
Std. dev.	0.088	0.037			
Observations	630,675	39,608	38,900	38,900	37,368
Panel C, dependent variable: gestational hypertension					
Non-Hispanic white father	0.004*** (0.000)	0.006** (0.003)	0.005 (0.003)	0.010 (0.007)	0.010 (0.007)
Mean of dep. Var.	0.030	0.037			
Std. dev. .	0.169	0.190			
Observations	956,882	19,249	19,044	19,044	18,346
Child's gender, race			YES	YES	YES
Mother low birth weight			YES	YES	YES
Mother's and child's year of birth F.E.			YES	YES	YES
Mother's county of residence F.E.			YES	YES	YES
Grandmother's age at delivery			YES	YES	YES
Grandmother F.E.				YES	YES
County-year at birth of mother and child F.E.					YES
Mother's education, age, parity, marital status					YES
Adequate care					YES
Poverty of residence (zip code of mother's birth)					YES

Notes - Data in column 1 are drawn from the Natality Detail Data (1989–2004). Data in columns 2–5 are drawn from the Florida Birth Master File (1971–1985, 1989–2009). The sample is restricted to second-generation Hispanic mothers born in Florida between 1971 and 1985. Standard errors are clustered at the grandmother level.



Table 8: Ethnic intermarriage and birth outcomes among 3rd generation Hispanics, within-family analysis

Sample	(1) Hispanics	(2) Hispanics	(3) Hispanics	(4) Non-Hispanic Whites <i>Placebo</i>
<b>Panel A: Birth Weight (in grams)</b>				
Non-Hispanic white father	−88.901*** (5.321)	−61.047*** (5.706)	−20.743** (8.268)	50.689*** (14.215)
Mean of Dep. Var	3326.377			
Std. dev.	560.681			
<b>Panel B: Incidence of Low Birth Weight</b>				
Non-Hispanic white father	0.018*** (0.002)	0.014*** (0.003)	0.008** (0.004)	−0.011 (0.007)
Mean of Dep. Var	0.056			
Std. dev.	0.229			
<b>Sociodemographic controls</b>	NO	YES	YES	YES
<b>Mother's birth weight</b>	NO	YES	YES	YES
<b>Grandmother F.E.</b>	NO	NO	YES	YES
Observations	233,440	201,754	201,754	336,439

*Notes* - Data are drawn from the California and Florida Vital Statistics (1975–1981, 1989–2009). The sample is restricted to third-generation Hispanics born between 1989 and 2009 in columns 1–3 and to third-generation white natives born between 1989 and 2009 in column 4. sociodemographic controls include 3rd generation child's gender, mother's birth weight (LBW), dummies for the interaction of county and year of birth of second-generation children (mothers), first generation's (grandmothers) age (at delivery) dummies, second generation's age (at delivery) dummies, second-generation parity, poverty rate in zip code of birth of second generation, marital status, father's and mother's education (4 groups), parity, dummies for the interaction of county and year of birth of third-generation children, indicators for missing information on first generation's (grandmothers) age (at delivery) (FL), father's education and age. Columns 3 and 4 include grandmother-fixed effects. Standard errors are clustered at the grandmother level.