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

Osea Giuntella*

University of Oxford, Blavatnik School of Government

December 31, 2013

Abstract

Despite their lower socioeconomic status, Mexican immigrants in the United States have similar or better health outcomes than natives. However, while second-generation Mexicans assimilate socio-economically, their health deteriorates. This phenomenon is commonly known as the "Hispanic health paradox". There is an open debate about whether this unhealthy assimilation 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 children born in California and Florida (1970–2009), to analyze the mechanisms behind the generational decline observed in birth outcomes. I show that a modest positive selection on health at the time of migration can account for the the initial advantage in birth outcomes of second-generation Mexicans. At the same time, a simple process of regression towards the mean reverses the apparent paradox, predicting a worse deterioration than the one observed in the data. Using a subset of siblings, and holding constant grandmother-fixed effects, I show that the persistence of healthier behaviors (e.g. smoking during pregnancy) among second-generation Mexican mothers can explain more than half of the difference between the model prediction and the observed birth outcomes of third-generation Mexicans.

Keywords: Hispanic health paradox, birth outcomes, risky behaviors
JEL Classification Numbers: I10, J15

*email: osea.giuntella@bsg.ox.ac.uk. I am grateful to Daniele Paserman, Randall Ellis, Kevin Lang, and Claudia Olivetti for their comments and advice. I am also thankful to Yasmine Serrano (Florida Department of Health) and Julie Turner (California Department of Public Health) who were extremely helpful in the data collection process. I would like to thank all the participants to the Society of Labor Economics Meetings (2013), the Association for the Study of Economics, Culture and Religion (2013), the Northeast Universities Development Consortium Conference (2012), the American Society of Health Economists Conference (2012), the XII Doctoral Workshop in Economic Theory and Econometrics (2012), the EconCon Conference (2012), the European Meeting of the Econometric Society (2012), the XV IZA Summer School (2012), the European Association of Labor Economists/IZA Conference (2012), as well as all seminar attendees at the University of Connecticut, University of Lugano, Northeastern University, University of Oxford, Universitat Pompeu Fabra, Syracuse University, University of Queensland, Singapore Management University, 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

At 31.8 million in 2010, Mexican-Americans comprise 63% of the U.S. Hispanic population and 10% of the total U.S. population, with births overtaking immigration as the main source of growth of the American population. Mexicans are the most disadvantaged immigrants in terms of socio-economic status and their earnings assimilation is considerably slower than for other immigrant groups. Because they are characterized by lower socioeconomic status than natives, and given the poorer average health conditions observed in Mexico (e.g. life expectancy, incidence of low birth weight), they would be expected to be at higher risk for negative health outcomes, as there is evidence of a positive socioeconomic gradient in health. However, despite being the most disadvantaged immigrant group in the country, a substantial body of research has documented that Mexicans are healthier than natives along several dimensions. Furthermore, despite positive socioeconomic assimilation, previous studies have shown that the initial advantage deteriorates with time spent in the United States and erodes in the next generation. For these reasons, previous scholars have referred to these stylized facts as the "Hispanic health paradox." This apparent paradox has been observed in general health status, life expectancy, mortality from cardiovascular diseases, cancer, age of puberty, and infant outcomes (Markides and Coreil, 1986; Antecol and Bedard, 2006; Bates and Teitler, 2008; Elder et al., 2012; Powers, 2013) capturing the attention of media and policy makers (Tavernise, 2013). The goal of this paper is to shed light on the mechanisms underlying these facts.

There is a general consensus that selection can explain the first-generation health advantage (Jasso et al., 2004; Antecol and Bedard, 2006; Riosmena et al., 2013). International migrants are not a random sample of their population of origin. Migrants move in search of better labor market opportunities and health status affects the perceived costs and benefits of migration. Crossing the border might be more costly for unhealthy individuals. At the same time, healthy migrants might have higher returns to migration, as health enhances their earning capacity. Despite most of the studies refer to health selectivity as one of the main explanation of the Hispanic health paradox, there has been little formal theoretical investigation of the relationship between health and the migration decision. A handful of studies have empirically tested the healthy migrant hypothesis and found evidence of positive, but mild, selection on health (Crimmins et al., 2005; Barquera et al., 2008; Rubalcava et al., 2008; Ullmann et al., 2011; Riosmena et al., 2013). 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 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). 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). Previous studies did 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. 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.

This paper contributes to the extant literature by taking advantage of a large longitudinal intergenerationally linked data set. In particular, I analyze the birth outcomes of the second- and third-generation Mexicans born in California and Florida, two of the top immigrant destination states in the United States, between 1970 and 2009. 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 Mexican immigrant descendants.[1]

[1] The paper focuses on birth-weight as this is the only health outcome that allows an intergenerational analysis using Vital Statistics linked data covering different generations.

Exploiting these data, I first verify the facts described as the apparent paradox in birth outcomes among immigrant descendants. Once these stylized facts have been established, I present a simple model of selection and intergenerational health transmission to interpret the health trajectories of Mexican immigrants in the United States. Using country-level differences in health outcomes, I show that a modest selection on health can explain the fact that second-generation Mexican have better birth outcomes than natives. This is consistent with the evidence of mild positive selection on health presented in studies based on information collected in the sending country prior to the time of migration (Crimmins et al., 2005; Barquera et al., 2008; Rubalcava et al., 2008; Ullmann et al., 2011; Riosmena et al., 2013) and also with the mild positive selection of women on education (Chiquiar and Hanson, 2005; Moraga, 2011).

To verify whether the erosion of the advantage in third-generation birth outcomes can be explained by a simple process of regression towards the mean, I predict what would be the expected incidence of low birth weight in the third-generation, based on the observed intergenerational correlation in birth weight and on existing estimates in the literature on the intergenerational transmission of health status. 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. 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 Mexicans 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. In other words, the calibrated model implies that the paradox is reversed: health should actually deteriorate much faster than what the data shows given the socioeconomic outcomes of second generation vis-a-vis natives. Thus, the new puzzle is to ascertain how third-generation birth outcomes do not deteriorate as rapidly as predicted by the model. Exploiting the matched birth records for California and Florida and cross-sectional Vital Statistics for the United States, I find that more than half the reverse paradox is explained by lower incidence of risky behaviors.

I show that first-generation Mexican mothers 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, Mexican mothers maintain a sizeable advantage in terms of lower incidence of health risk factors compared to white natives. The birth outcomes of third-generation Mexicans 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 Mexicans, 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). Conditioning on risk-factors and accounting for the persistence in healthy behaviors explains more than half of the difference between the model prediction and what we observe in the data.

The paper is organized in the following manner. Section 2 describes the data and verifies the Hispanic paradox in birth outcomes. Section 3 discusses the possible mechanisms behind these health patterns. Concluding remarks are in Section 4.

2 Data and Stylized Facts

2.1 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 the years 1970–1985 (1970–1981 in California and 1971-1985 in Florida) and 1989–2009.[2] For expositional ease, for both the immigrants and the natives in the sample, I refer to all the women delivering between 1970 and 1985 as first-generation (grandmothers, G1), to all the children born between 1970 and 1985 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. 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.[3]

As with the previous literature (Fryer and Levitt, 2004; Currie and Moretti, 2007; Royer,

[2]I obtained data from the California Department of Public Health for the years 1970-1981 and 1989-2009. Data for 1970 in Florida do not include information on mother's country of origin.

[3]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.

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).[4]

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 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.

2.2 Matching and selection: Descriptive statistics

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 1970 and 1985 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.

[4]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.

The quality of matching for children born in California and Florida between 1989 and 2009 whose parents were born in the same states between 1970 and 1985 is relatively high: 96.6% in Florida, 87.5% in California. 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 1970 and 1985. The final sample includes 1,355,896 (46%) of the 2,952,909 female children born between 1970 and 1985 in California and Florida. This reflects the reality that not all the women born in California and Florida between 1970 and 1985 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 13.2% of women born in California and in Florida between 1970 and 1985 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 37% of women born in California and Florida between 1970 and 1985 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 rate among children of Mexican origin is 58%. The rate of matching also depends on the 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 0.6%. 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.[5] [6]

To further address the concern of selection bias arising from a matching process that selects on a sample of women who were both born and have given birth in either California and Florida, I verify the external validity of the results using data from the Natality Detail Data, which collects detailed data on all births in the United States. Using these data allows me to conduct cross-sectional analysis for the entire United States for the years 1970–2004 and address the concern that the results obtained with the California and Florida data may suffer from selection bias because of the attrition in the matching process.[7]

[5]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.

[6]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.

[7]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.

2.3 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 Mexican immigrants. However, it is important to first verify the 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:

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

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. $MX_{izt,1}$ is a dummy equal to one when the first-generation woman delivering between 1970 and 1985 was born in Mexico. 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, parity, race, age dummies (in Florida, the mother's age is not available for the period 1970–1985), 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), child's gender and type of birth (singleton vs multiple birth).[8] 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 Mexicans and children of white U.S.-born

[8] 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.

mothers.[9] I restricted the sample to children born between 1970 and 1985 to white mothers and Hispanic first-generation immigrant mothers coming from Mexico.[10] The final sample includes 4,704,571 births for which information on birth weight is not missing.[11]

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. Children of Mexican mothers are only slightly heavier (approximately 23 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). In the online Appendix (Tables A1 and A2), 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.[12] 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.[13]

[9] 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 immigrant group in the United States. In an earlier version of the paper I included children of Cuban and Puerto Rican origin. 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). By contrast, Puerto Rican mothers are more likely to give birth to lighter babies (columns 2 and 4). 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.

[10] 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.

[11] 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.

[12] The Online Appendix is available on my personal web page: `http://www.bsg.ox.ac.uk/osea-giuntella`.

[13] 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.

There remains a "healthy immigrant effect" when considering the incidence of low birth weight. However, there is only a difference of 23 grams in the average birth weight. [14] 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.

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:

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

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. $MX_{izt,1}$ is a dummy equal to one if the first generation was born in Mexico. Note that the analysis sample here includes only second-generation mothers between 1970 and 1985 in CA and FL, who were babies in the second-generation sample. To ensure the comparability of the analysis, the model includes the identical set of 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 and third-generation white natives.[15] The estimates in columns 6 and 8 include the iden-

[14]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). The predominant distribution is substantially equivalent to the distribution of birth weight observed for term births. 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. 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 60 grams heavier on average in the 10% quantile (see the online Appendix, Tables A3 and A4). 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.

[15]Unfortunately, the data do not contain information on the country of origin of the father for the entire

tical set of controls used in columns 2 and 4.[16] 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, but the coefficient (-0.003) for the third-generation children of Mexican origin (column 8) shrinks significantly (by approximately 80%) compared to the one observed among second-generation children in column 4 (-0.015). However, the third-generation children of Mexican origin do conserve some of the initial health advantage.[17]

3 Theoretical framework

3.1 Immigrant self-selection

In the previous section, I confirmed the existence of an apparent paradox in the birth outcomes of Mexican descendants. 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 average health in the population of origin. Although health selectivity is one of the main explanations proposed in the literature to explain the Hispanic health paradox, there has been little theoretical and empirical investigation of this relationship. There is, instead, a wide and open debate on whether Mexican immigrants in the United States are more or less educated than nonmigrants in Mexico. In a seminal article, Borjas (1987) argued that immigrants from countries with relatively high returns to education and income inequali-

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.

[16] In the online Appendix I report the conditional mean differences obtained using different sets of control variables.

[17] The deterioration with respect to native birth outcomes is even stronger when children of Cuban and Puerto Rican origin are analyzed.

ties tend to be selected from the lower half of the skill distribution in the sending country. However, this result holds only if migration costs are constant across individuals. Chiquiar and Hanson (2005) show that if migration costs decrease with education, migrants might be negatively or positively selected, depending on the size of the costs and on the shape of the skill distribution. Using counterfactual wage densities Chiquiar and Hanson (2005) provide evidence against the negative-selection hypothesis. Their results suggest that migrants are selected from the middle of Mexican earnings distribution with evidence of positive selection for Mexican-born women. Similarly, the findings of Orrenius and Zavodny (2010); McKenzie and Rapoport (2010); Kaestner and Malamud (2010) confirm positive or intermediate selection. However, other scholars have provided new evidence in favor of the negative selection hypothesis (Ibarraran and Lubotsky, 2007; William and Peri, 2012; Moraga, 2011; Reinhold and Thom, 2012). The discrepancy in these set of results is explained by the fact that the negative selection found by the latter group of studies is driven by unobserved wage-earning characteristics. However, as this paper focuses on pregnant mothers, it is important to note that all of these studies find evidence of mild positive selection on education among immigrant women.

A few papers attempted to provide a framework to analyse the importance of health selection in explaining the Hispanic paradox. Palloni and Morenoff (2001) propose a simple model of selection on health at migration and show 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. More recently, a handful of recent studies has attempted to empirically estimate selection on health at migration finding evidence of positive, but mild selection on health (Crimmins et al., 2005; Barquera et al., 2008; Rubalcava et al., 2008; Ullmann et al., 2011; Riosmena et al., 2013). These findings are consistent with the idea that health will affect the migration decision by either enhancing earning capacity or reducing the actual cost of migration. 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 natives. However, these studies do not test the role of selection and regression towards the mean in explaining the unhealthy assimilation in adult second-generation and third-generation children.

Building on Palloni and Morenoff (2001), Jasso et al. (2004), and Chiquiar and Hanson (2005) this section develops a simple theoretical framework to analyze the mechanisms behind these health trajectories. 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 using birth records. However, I can calibrate the model using the observed differences in health outcomes between the United States and Mexico to pin down the degree of selection of first-generation immigrants. Once assessed the patterns of selection, I can estimate the expected intergenerational trajectories in birth weight using the observed intergenerational correlation in birth outcomes and existing estimates of the intergenerational correlation in health status.

3.2 The decision to migrate

Following Chiquiar and Hanson (2005), I consider migration as a one-time decision based on gains and costs of migration. Mexican residents face the following wage equation:

$$ln(w_0) = \nu_0 + \delta_0 s$$

where w_0 is the wage, ν_0 is the base wage, s is the level of education, and δ_0 are returns to education in Mexico. Similarly Mexican immigrants in the United States will face the wage equation

$$ln(w_1) = \nu_1 + \delta_1 s$$

where w_1 is the wage, ν_1 is the base wage, s is the level of education, and δ_0 are returns to education in the United States. Analogously to Chiquiar and Hanson (2005), I assume that $\delta_0 > \delta_1$ because of the scarcity of skills in Mexico. Costs are defined in time-equivalent units as $\pi = C/w_0$. Mexicans will migrate to the US if

$$ln(w_1) - ln(w_0 + C) \approx ln(w_1) - ln(w_0) - \pi$$

Costs of migration are assumed to be negatively correlated with schooling, such that

$$ln(\pi) = \nu_\pi - \delta_\pi s$$

Health can affect both returns and costs of migration, and it interacts with the other determinants of migration (e.g., education). Health will increase returns to migration by both being an important component of human capital (Grossman, 1972) and by increasing labor supply, or skill utilization (Jasso et al., 2004). In other words, health will enter the migration decision by increasing both the level of human capital (s) and the returns to human capital (δ_0 and δ_1), and by decreasing the cost of migration (ν_π). For these reasons, the extent of selection on health will be higher the greater are the costs of migration. Depending on the costs and returns to migration, and holding everything else constant, Mexican residents will migrate to the United States only if their health is above a minimum threshold.

3.3 A simple model of migration and intergenerational transmission of health

3.3.1 Health Selectivity and Migration

To avoid cluttering, following Palloni and Morenoff (2001), I assume that, everything else constant, health is the only factor affecting migration. Based on this assumption, I sketch a simple model of health selectivity at migration.

Let $h_{jt} \sim N(\mu_{jt}, 1)$ be the distribution of health in country j at time t, where h_{jt} is the health of the first generation at the time of migration, which is distributed as a random normal (μ_{jt}, 1) reflecting the health distribution in the country of origin, μ_{jt} is the average health in country j at time t, and t_1 is the migration threshold. μ_{jt} can be viewed as the composite effect of genes, quality of health care, socio-economic environment, and risk-factor behavior on health.[18]

The migration process is then defined as following. An individual from the source country ($j =$ e.g., Mexico) will be able to migrate to the destination country (US) only if their health is above a certain threshold t_1. This may be represented formally as:

$$Imm_j = \begin{cases} 1 & \text{if } h_{jt} \geq t_1 \\ 0 & \text{if } h_{jt} < t_1 \end{cases}$$

Individuals with $h_{jt} \geq t_1$ will be able to migrate. The higher the threshold, the more selected is the sample of migrants. These relations imply that the health of first generation immigrants in the US is, therefore, distributed as a truncated normal, $TN(\mu_{jt}, 1, t_1)$.

Hence, the birth weight of second-generation $(t+1)$ is determined as follows:

$$BW_{j,t+1} = \gamma h_{jt} + v_{t+1}$$

where $BW_{j,t+1}$ is the birth weight of the second generation, h_{jt} captures maternal health at migration, v_{t+1} is distributed as a random (0, σ_v^2) normal variable reflecting the effect of other unobservable factors on the birth weight of the second generation, and γ captures the effect of maternal health on the child's health. For the second-generation immigrants, the cumulative distribution function of $BW_{j,t+1}$ will be given by the sum of H_{jt}, a truncated normal at t_1, and E_{t+1}, a random normal variable. Respectively,

[18]As the variance σ_{jt} in the birth weight of the two populations differs by less than 0.6%, without loss of generality, we assume that the distribution of health has the same variance: in the two populations considered: $\sigma_{MX,t}$=$\sigma_{US,t}$=1.

$$H_{jt} \sim TN(\lambda, \delta^2, (\delta(t_1 - \mu_{jt}) + \lambda))$$

$$E_{t+1} \sim N(0, \varepsilon^2)$$

where $\lambda = \gamma\mu_{jt}$, $\delta^2 = \gamma^2$, and $\varepsilon^2 = \sigma_v^2$.

Following Turban (2010) and Azzalini (2005), $BW_{j,t+1} = H_{jt} + E_{t+1}$ is distributed according to the density:

$$f(bw) = \eta e^{-\frac{(bw-\lambda)^2}{2(\varepsilon^2+\delta^2)}}[\Phi(\frac{bw - t_1 - \alpha}{\beta})] \tag{1}$$

where $\alpha = \frac{\varepsilon^2(bw-\lambda)}{\varepsilon^2+\delta^2}$, $\beta^2 = \frac{\varepsilon^2\delta^2}{\varepsilon^2+\delta^2}$, $\eta = \frac{\sqrt{2\pi}\beta}{2\pi\varepsilon\delta(\Phi(d))}$, $d = \frac{\lambda - t_1}{\delta}$, and Φ is the c.d.f. of a standard normal distribution. Given the distribution of BW_{t+1}, the incidence of low birth weight is determined as follows:

$$\int_{-\infty}^{t_2} f(bw)dbw \tag{2}$$

where t_2 represents the low birth weight threshold:

$$LBW_{t+1} = \begin{cases} 1 & \text{if } BW_{t+1} \leq t_2 \\ 0 & \text{if } BW_{t+1} > t_2 \end{cases}$$

For the native population, the cumulative distribution function of birth weight will be given by the sum of two normal distribution. Once we set the parameters of the model (t_2, $\mu_{j,t}$, and γ) it is then straightforward to compute the expected gap in low birth weight between the second generation Mexicans and natives for any level of selection (t_1).

3.4 Unhealthy assimilation

I then turn to analyze whether a simple process of regression towards the mean of the population could explain the convergence observed in the data. For any level of selection, one can indeed predict the expected birth outcomes of the third-generation. Third-generation birth outcomes can be described as a function of second-generation health characteristics and other factors. Let the health of second-generation be defined as:

$$h_{j,t+1} = \rho h_{jt} + u_{t+1}$$

where $h_{j,t+1}$ is the health of second-generation mothers, u_{t+1} is distributed as a random normal variable ($\mu_{j,t+1}$, $\sigma_{u_{t+1}}$) reflecting the effect of other unobservable factors on the health of the second-generation mother, and ρ measures the degree of intergenerational correlation in health between the first and second generations. Then, if the distribution of health is stable

$$\sigma^2_{h_t} = \sigma^2_{h_{t+1}} = \sigma^2_{u_{t+1}} + \rho^2 = 1$$

Hence, the birth weight of the third generation can then be expressed as a function of maternal health, with the following formal designation:

$$BW_{t+2} = \gamma h_{t+1} + v_{t+2} = \gamma(\rho h_{jt} + u_{t+1}) + v_{t+2}$$

where BW_{t+2} is the birth weight distribution in the third generation, v_3 is distributed as a random normal (0, σ^2_v) variable reflecting the effect of other unobservable factors on the birth weight of the third generation. Without loss of generality, I assume that the unobserved random shocks to health and birth weight are not correlated.[19] The covariance between the birth weight of the two generations may therefore be rewritten as the following:

$$\mathrm{Cov}(BW_{t+2}, BW_{t+1}) = \mathrm{Cov}(\gamma h_{t+1}) + v_{t+2}), \gamma h_t + v_{t+1}) =$$

[19]Note that while this assumption might seem strong, in practice it does not affect the model predictions for the birth outcomes. Assuming perfect correlation brings qualitatively similar results, as 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 γ and ρ would instead depend on the extent of correlation between unobserved random shocks to health and birth weight.

$$\text{Cov}(\gamma\rho h_t + \gamma u_{t+1} + v_{t+2}, \gamma h_t + v_{t+1})\gamma^2\sigma^2_{h_t}\rho = \gamma^2\rho$$

which implies

$$\rho = \frac{\text{Cov}(BW_{t+2}, BW_{t+1})}{\gamma^2} \tag{3}$$

The cumulative distribution function of $BW_{j,t+2}$ is then given by the sum of a truncated normal at t_1, $H_{jt+1} \sim TN(\lambda, \delta^2, (\delta(t_1 - \mu_{jt}) + \lambda))$ and a random normal variable $E_{t+2} \sim N(0, \varepsilon^2)$ where $\lambda = \gamma\rho\mu_{jt} + \gamma\mu_{jt+1}$, $\delta^2 = \gamma^2\rho^2$, and $\varepsilon^2 = \gamma^2(1-\rho^2) + \sigma^2_v$

$$f(bw) = \eta e^{-\frac{(bw-\lambda)^2}{2(\varepsilon^2+\delta^2)}}[\Phi(\frac{bw - t_1 - \alpha}{\beta})] \tag{4}$$

Given the distribution of BW_{t+2}, the incidence of low birth weight is determined as follows:

$$\int_{-\infty}^{t_2} f(bw)dbw \tag{5}$$

where t_2 represents the low birth weight threshold. Similarly to the previous section, one can then easily derive the predicted gap in low birth weight for the third generation by subtracting the observed incidence of low birth weight among natives. Again the the cumulative distribution function of birth weight will be given by the sum of two normal distribution.[20]

Within this framework one can analytically derive the incidence of low birth weight for third-generation immigrants. For different level of health selectivity at migration (t_1) we compute the predicted gap in low birth weight between Mexican descendants and native third generation. Without loss of generality one can choose units of birth weight such that

$$\sigma^2_{BW} = \gamma^2_{h_2} + \sigma^2_v = 1 \tag{6}$$

and therefore

$$Cov(BW_{t+2}, BW_{t+1}) = Corr(BW_{t+2}, BW_{t+1}) \tag{7}$$

[20]Note that the incidence of low birth weight among natives is stable around 6%.

This implies that equation 3 can be rewritten as follows:

$$\rho = \frac{Corr(BW_{t+2}, BW_{t+1})}{\gamma^2} = \frac{\rho_{bw}}{\gamma^2} \tag{8}$$

While I do not observe the intergenerational correlation in health status (ρ) and the effect of maternal health on birth outcomes (γ), I can estimate directly the degree of intergenerational correlation in birth weight (ρ_{bw}). Furthermore, we can use existing estimates of intergenerational correlation in adult health outcomes to infer a plausible range of values for ρ.

In practice, I consider different values of ρ and γ such that equation (3) holds and impose the estimated intergenerational correlation in birth weight estimated by Currie and Moretti (2007) and confirmed in my data (ρ_{bw}=$Corr(BW_3, BW_2)$=0.2).In our model, these restrictions on ρ imply that γ must be $\in [0.58, 1]$ (see eq. 8). While I consider ρ equal to 0.35 as a baseline, Table 4 illustrates the predictions of the model for different values of ρ in the defined range [0.2, 0.5].

3.5 Calibration

3.6 Predicting the second-generation (G2) birth outcomes

By defining t_2, $\mu_{j,t}$ and γ we can solve the model for different levels of selection t_1. Native health is used as a benchmark, hence I set $\mu_{US,t}$ equal to 0. The low birth weight threshold, t_2, is set to be -1.555 to match the average incidence of low birth weight observed in the data (0.06) over the entire period studied (1970–2009) in the entire population of the United States (excluding African–Americans). The mean of unobservable factors affecting health $\mu_{MX,t}$ is set such to reflect the incidence of low birth weight in Mexico (10.6%, see Buekens et al. (2012)). γ is determined by eq. 8. Note that while I consider different values of γ (see Table 4), the discussion will focus on the predictions obtained using the preferred assumption of $\rho = .35$, which implies $\gamma = .76$.

Using this parametrization and under the assumption of health having identical effects on birth weight in the two populations, the model can be solved semi-analytically for different levels of selection on health at migration (t_1). Figure 1 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 13th percentile do not migrate to the US, positive selection can explain the lower incidence of low birth weight observed among second-generation Mexicans.

Unfortunately, there is no way to directly estimate the extent of selection on birth weight as, to the best of my knowledge, there is no dataset containing both information on conditions just prior to migration and allowing to compare birth outcomes of migrants to the US with those of non-migrants who gave birth in Mexico. However, the modest selection implied by the calibration exercise is consistent with the evidence on mild positive selection on education among women (Chiquiar and Hanson, 2005; Moraga, 2011) and the evidence of weak positive selection on health. Indeed, Rubalcava et al. (2008) used longitudinal data from the Mexican Family Life Survey and find some but weak support for the healthy immigrant hypothesis. Furthermore, evidence of mild positive selection on height and other health outcomes has also been found by Crimmins et al. (2005), Barquera et al. (2008), Ullmann et al. (2011), and Riosmena et al. (2013) using other anthropometric information for Mexican in the United States and in Mexico.

3.6.1 Predicting the third-generation (G3) birth outcomes

Having assessed that a modest degree of selection on health would be sufficient to explain the initial advantage, I turn to analyze whether a simple process of regression towards the mean could explain the convergence observed in the data. Having parametrized ρ,$\rho_b w$ and γ, to solve semi-analitically the model I need to define the mean of unobservable factors affecting second generation health $\mu_{j,t=1}$, the mean of unobservable factors affecting second generation health u_{t+1}. As above, $\mu_{US,t+1}$ is used as a benchmark and set equal to 0. $\mu_{MX,t+1}$ is set such to reflect the lower socio-economic status of second-generation Mexican with respect to the natives.

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).[21] Using the Cramer (1995) estimate, with everything else constant, the birth weight of Mexicans should be on average 48 grams lower than that of natives. Based on these calculations, $\mu_{MX,t+1}$ is set equal to -0.1 to reflect the expected differences due to the lower socio-economic status of second generation Mexicans with respect to the native health benchmark. 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). In other words, this is equivalent to test the implication of a regression towards the average health of Americans with a similar socio-economic background. [22]

[21] The earliest year in which information on the birthplaces of the father and mother is available is 1994 in the CPS surveys.

[22] Note that accounting only for the relative weak intergenerational correlation in health and assuming no

Using this parametrization, one can solve the model semi-analyitically and obtain the incidence of low birth weight for the third generation. I can then predict the expected differences in low birth weight between children of second-generation Mexican immigrants and children of white natives (y-axis) (Figure 2). The vertical solid line corresponds to the degree of selection explaining the second-generation advantage (see Figure 1). 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). 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 1.4 percentage points higher than natives. Table 4 shows that for plausible values of $\rho \in [0.2, 0.5]$ the model always overpredicts convergence and the distance between the model prediction and the data ranges between 1 ($\rho = .5$) and 1.9 percentage points ($\rho = .2$). [23]

3.6.2 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

socioeconomic assimilation, the model would predict a much faster deterioration. On the contrary, assuming full assimilation in socioeconomics and accounting 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)).

[23] The results tend in the same direction if considering the entire Hispanic group or if using socioeconomic information at the zip code level.

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 A5). 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 A5.

Controlling for risky behaviors the coefficient increases to 0.01 (see column 5 of Table 3 and column 3 of Table A5), which is relatively close to the difference in the incidence of low birth weight predicted by the model. In other words, accounting for the observed risk factors (the upper dashed line in Figure 2), the model can explain approximately 80% of the reverse paradox found after accounting for socioeconomic differences.[24]

[24]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.

3.6.3 Within-family analysis

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 1970 and 1985 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.[25] 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.

Given that the conditional difference in the low birth weight incidence is time-invariant across individuals sharing the same grandmother, to account for grandmother fixed effects I follow Mundlak's approach (Mundlak, 1978). Column 6 in Table 3 shows that when partially controlling for family unobserved heterogeneity and other socio-demographic characteristics, accounting for the observed risk factors (the upper dashed line in Figure 2) can explain approximately 60% of the reverse paradox. Table 4, columns 8 and 9 show that when considering the entire range of plausible value of $\rho \in [0.2, 0.5]$ risky behaviors can explain at least 44% of the difference between the model prediction and the data.

[25] 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.

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. A modest selection on health at migration can account for the second-generation birth outcomes advantage. As a natural process of regression towards the mean and less than full socio-economic assimilation third-generation birth outcomes would be expected to deteriorate even more than what observed in the data. The persistence of healthy behaviors during pregnancy can explain the difference between the model predictions and the data.

4 Conclusion

This paper confirms that while second-generation Mexicans have lower incidence of low birth weight than children of native white mothers, this advantage shrinks substantially in the third generation. I show that a modest selection on health might explain the better birth outcomes of second-generation children compared to white natives and that, given the relatively weak intergenerational correlation in health status and birth weight, 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. I show that more than half of the difference between what predicted by the model and the actual data is explained by the differences in risk factors (such as tobacco and alcohol consumption during pregnancy and gestational hypertension). 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. This holds true even after accounting for potential confounding factors, 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, Jocelyn Pan, Hee-Jin Jun, Theresa L. Osypuk, and Karen M. Emmons (2005) 'The effect of immigrant generation on smoking.' *Social Science and Medicine* 61, 1223–1242

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

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

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

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

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

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

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

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

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

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

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

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

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

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`. Accessed Nov. 2012

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

Fenelon, Andrew (2012) 'Revisiting the Hispanic paradox in the United States: The role of smoking.' `http://paa2012.princeton.edu/papers/120602`. Accessed Nov. 2012

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

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

Grossman, Michael (1972) 'On the concept of health capital and demand for health.' *Journal of Political Economy* 80(2), 223–255

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

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

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

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

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

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

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

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

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

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

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

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

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

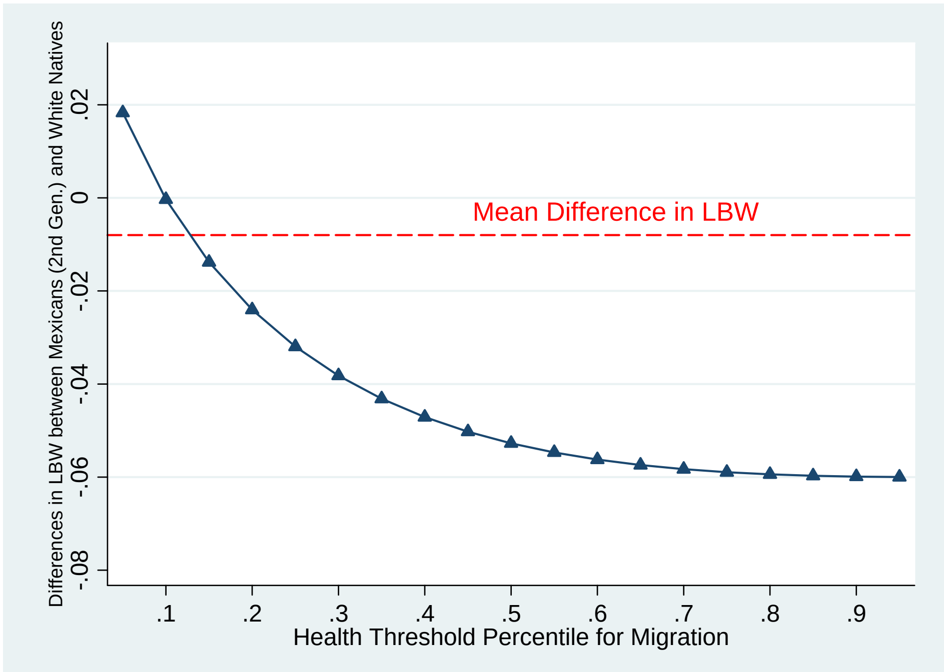
Turban, Sebastien (2010) 'Convolution of a truncated normal and a centered normal variable.' `http://www.columbia.edu/~st2511/notes/Convolution%20of%20truncated%20normal%20and%20normal.pdf`. Accessed Nov. 2012

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

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

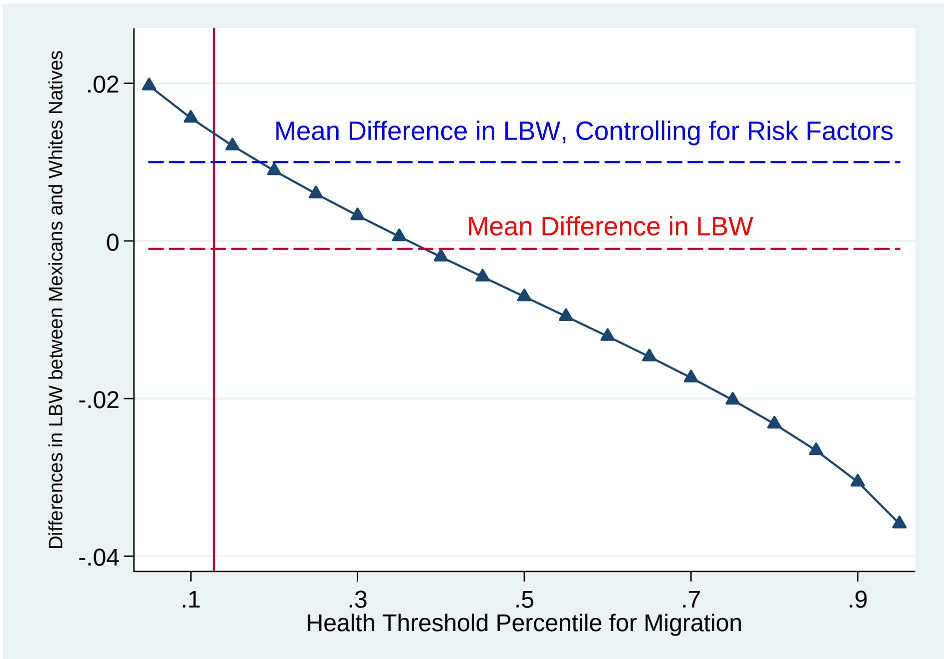
William, Ambrosini J., and Giovanni Peri (2012) 'The determinants and selection of mexico-us migrants.' *The World Economy (forthcoming)*

Figure 1: Selection on health at migration and differences in the incidence of low birth weight betweeen 2^{nd} 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 ρ is equal to 0.35 and the effect of maternal health on birth weight, γ, is equal to 0.75 (baseline). μ_{MX_t} is set equal to -0.42 to be such that the incidence of low birth weight in Mexico (10.6%, see Buekens et al. (2012)). The dashed line describes the observed raw difference in the incidence of low birth weight between 2nd generation Mexicans and white natives born between 1970 and 1985, in California and Florida (see Table 3, col. 1).

Figure 2: Regression towards the mean and differences in the incidence of low birth weight betweeen 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 ρ is equal to 0.35 and the effect of maternal health on birth weight, γ, 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, $\mu_{MX_{t+1}} = -0.1$). 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 1.

Table 1: Matching quality. Women born in California and Florida, 1970–1985

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

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

Table 2: Hispanic Health Paradox in birth weight (BW) and low birth weight incidence (LBW)

	2nd generation 1970–1985				**3rd generation 1989–2009**			
Dependent variable:	BW		LBW		BW		LBW	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Country of origin:								
Mexico	0.980	23.132***	-0.008***	-0.015***	-32.442***	-11.993***	-0.001**	-0.003***
	(0.895)	(1.043)	(0.000)	(0.000)	(1.937)	(1.494)	(0.001)	(0.001)
Sociodemographic controls		YES		YES		YES		YES
Mean of Dep. Var.	3,377.666	3,377.666	0.057	0.057	3,358.716	3,358.716	0.061	0.061
Std.dev.	568.435	568.435	0.232	0.232	577.542	577.542	0.240	0.240
Observations	4,704,571	4,086,426	4,704,571	4,086,426	2,075,837	1,979,723	2,076,649	1,979,723

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

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

Table 3: Differences between 1st, 2nd generation Mexicans and U.S. white natives

	(1)	(2)	(3)	(4)	(5)	(6)
	CA-FL		**Florida**			
	$MX_1 - N$	$MX_2 - N$	$MX_1 - N$	$MX_2 - N$	$MX_2 - N$	$MX_2 - N$
Panel A: CA-FL Vital Statistics						
Low birth weight	-0.008***	-0.001***	-0.005**	0.003	0.011***	0.008**
	(0.000)	(0.000)	(0.002)	(0.004)	(0.004)	(0.004)
Control for risk factors	NO	NO	NO	NO	YES	YES
Grandmother F.E.	NO	NO	NO	NO	NO	YES
Panel B: FL Vital Statistics						
Tobacco consumption			-0.159***	-0.151***		
			(0.001)	(0.006)		
Alcohol consumption			-0.003***	-0.004***		
			(0.000)	(0.001)		
Gestational hypertension			-0.026***	-0.005**		
			(0.001)	-0.002		
Panel C: CA-FL CPS						
Socioeconomic status	-0.855***	-0.457***	-0.736***	-0.537***		
log(family income/poverty)	(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 (1970−1985, 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 and year fixed effects. Note that data drawn from the California Vital Statistics do not contain information on these risk factors for the period under analysis.

Table 4: Model parameters and predictions

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Model Parameters		Second Generation (G2)	Third Generation (G3)	Third Generation (G3)				
		Selection Matching Initial Advantage	Predicted Gap	Distance between Model and the Observed Data				
				Unconditional	(Conditional on risky behaviors)			
					OLS	%	QFE	%
ρ	γ_{bw}	$F(h_{MX,t} < t_1^*)$	$LBW_{MX} - LBW_{US}$	β=-0.001	$\beta = 0.011$	Explained	β=0.008	Explained
0.20	1.00	0.082	0.018	0.019	0.007	63%	0.010	44%
0.25	0.89	0.095	0.017	0.018	0.006	67%	0.009	47%
0.30	0.82	0.112	0.015	0.016	0.004	75%	0.007	53%
0.35	0.76	0.126	0.014	0.015	0.003	80%	0.006	57%
0.40	0.71	0.138	0.012	0.013	0.001	92%	0.004	67%
0.45	0.67	0.151	0.011	0.012	0.000	103%	0.003	75%
0.50	0.63	0.162	0.009	0.010	-0.002	120%	0.001	89%

Notes - For this calibration exercise the intergenerational correlation in birth weight $\rho_b w$ is set equal to 0.2 (see Section 3.4.1). μ_{US_t} and $\mu_{US_{t+1}}$ are set equal to 0 as a benchmark. μ_{MX_t} is set equal to -0.42 to be such that the incidence of low birth weight in Mexico (see Section 3.3.1). $\mu_{MX_{t+1}}$ is set equal to -0.1 to reflect the lower socio-economic conditions of second-generation Mexicans (see Section 3.4.1). Column 1 reports different values of the intergenerational correlation in health $\rho \in [0.2, 0.5]$. The effect of maternal health on birth weight ($\gamma_{bw} = \frac{\rho_{bw}}{\rho})^{1/2}$) is reported in column 2. Column 3 reports the health threshold percentile for which the model matches the observed data for G2 explaining their initial health advantage (t_1^*). Column 4 presents the difference in low birth weight between third-generation Mexicans (G3) and natives predicted by the model when using the parameters of columns 1-3. In column 5, I report the difference between the model prediction and the unconditional mean difference presented in column 2 of Table 3. Columns 6 and 8 illustrate the distance between the model prediction and the data once we condition on observable behaviors and include grandmother quasi-fixed effects. Column 7 and 9 show how much of the gap reported in column 5 is explained by conditioning on risky behaviors and including grandmother quasi-fixed effects.

Table 5: Differences between 1st, 2nd generation Mexicans and U.S. white natives - U.S.

	(1) $MX_1 - N$	(2) $MX_2 - N$	$MX_2 - N$
Panel A: Natality Detail Data			
Low birth weight	-0.008***	0.001***	0.008***
	(0.000)	(0.000)	(0.000)
Control for risk factors	NO	NO	YES
Panel B: Natality Detail Data			
Tobacco consumption	-0.144***	-0.125***	
	(0.000)	(0.000)	
Alcohol consumption	-0.007***	-0.003***	
	(0.000)	(0.000)	
Gestational hypertension	-0.014***	-0.008***	
	(0.000)	(0.000)	
Panel C: CPS			
Socioeconomic status			
log(family income/poverty)	-0.837***	-0.456***	
	(0.003)	(0.009)	

Notes - Data are drawn from the Natality Detail Data (1970–1985; 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 and year fixed effects. 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.