2013

Three essays on the economics of immigration and health

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

THREE ESSAYS ON THE ECONOMICS OF IMMIGRATION AND HEALTH

by

GIOVANNI OSEA GIUNTELLA

Laurea Triennale in Economia Europea, Universita' di Roma "Tor Vergata", 2004
Laurea Specialistica, Universita' di Roma "Tor Vergata", 2006
Master in Economics (MEI), Universita' di Roma "Tor Vergata", 2007

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2013

Approved by

First Reader

Daniele Paserman, Ph.D.
Professor of Economics

Second Reader

Claudia Olivetti, Ph.D.
Associate Professor of Economics

Third Reader

Kevin Lang, Ph.D.
Professor of Economics

to my father

Acknowledgments

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.

THREE ESSAYS ON THE ECONOMICS OF IMMIGRATION AND HEALTH

(Order No.)

GIOVANNI OSEA GIUNTELLA

Boston University, Graduate School of Arts and Sciences, 2013

Major Professor: Daniele Paserman, Professor of Economics

ABSTRACT

This thesis analyzes different sources of disparities in health and access to care among immigrants and ethnic minorities in the United States. The first chapter studies the generational worsening observed in the birth outcomes of Hispanics in the United States. Despite their lower socio-economic status, Hispanic immigrants in the United States initially have better health outcomes than natives. However, while their socio-economic 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. The second chapter investigates the role of ethnic networks and the effect of providing online information in foreign-language in facilitating Medicaid take-up among immigrants in the US.

Many low-income immigrants are uninsured yet eligible for public health insurance. In this paper, we examine whether language barriers and network effects can explain disparities in insurance Medicaid participation. Using the 2008 and 2009 American Community Survey, we show that linguistic networks facilitate Medicaid enrolment among non-English speaking adults. The third chapter analyzes the effect of food-environment on maternal weight gain and pregnancy outcomes. This paper studies how changes in the quality of food environment affect maternal and child health. Similarly to Currie et al. (2009), I exploit variation in the availability across births of the same mother in the proximity to different types of restaurants. Results show that proximity to Mexican restaurants is associated with a lower likelihood of excessive weight gain among US born mothers.

Contents

1 Why Does the Health of Immigrants Deteriorate? 1

1.1 Introduction 1

1.2 Hispanic paradox: Selection or worsening of behaviors? 5

1.3 Data 7

1.3.1 Matching and selection: Descriptive statistics 10

1.3.2 Verifying the Hispanic paradox in birth outcomes 12

1.4 A simple model of selection and health transmission 17

1.4.1 Theoretical framework 17

1.4.2 Empirical moments and calibration 21

1.4.3 Accounting for maternal risky behaviors 25

1.5 Heterogeneity among second-generation Hispanics: A within-family analysis 27

1.5.1 Socioeconomic status and quality of care 28

1.5.2 Maternal risk factors 30

1.5.3 The role of acculturation: Analyzing ethnic intermarriage 32

1.6 Conclusion 36

2 Medicaid and Ethnic Networks 50

2.1 Introduction 50

2.2 Data 55

2.3 Empirical Specification 61

2.4 Main Results 64

2.5 Robustness Checks 68

2.5.1 Restricting the Sample by Age, Sex, and Language 69

2.5.2 Citizenship, Language Fluency, and Length of Residence 73

2.5.3 Availability of Medicaid Information in Other Languages 74

2.5.4 Contact Availability Based on Country of Origin 77

2.5.5 Identification by *Ex Ante* Characteristics 79

2.6 Conclusion 79

3 Food Environment, Maternal Weight Gain and Pregnancy Outcomes 86

3.1 Introduction 86

3.2 Food-environment, maternal and child's health outcomes 89

3.3 Data and Empirical Specification 90

3.3.1 Data 90

3.3.2 Summary Statistics 91

3.3.3 Empirical Specification 92

3.4 Main results 93

3.4.1 Other health outcomes 94

3.4.2 Robustness checks 95

3.5 Conclusion 96

References 107

Curriculum Vitae **120**

List of Tables

1.1 Matching quality. Women born in California and Florida, 1975–1981 42
1.2 Hispanic Health Paradox in birth weight (BW) and low birth weight incidence (LBW) 43
1.3 Differences between 1st, 2nd generation Mexicans and U.S. white natives . 44
1.4 Differences between 1st, 2nd generation Mexicans and U.S. white natives - U.S. 45
1.5 Adequacy of care, SES and birth outcomes among 3rd generation Hispanics, within-family analysis 46
1.6 The effect of health risk factors on low birth weight, Florida 1989–2009, within-family analysis 47
1.7 Ethnic intermarriage and low birth weight risk factors among 2nd generation Hispanic mothers, within-family analysis 48
1.8 Ethnic intermarriage and birth outcomes among 3rd generation Hispanics, within-family analysis 49

2.1 Summary Statistics 56
2.2 Health Insurance Coverage by Language 60
2.3 Main Results 65

2.4 Changing the Control Variables 68

2.5 Robustness Checks: Sub-Samples by Sociodemographic Characteristics . . . 70

2.6 Robustness Checks: Years in U.S., Citizenship, and English Fluency 73

2.7 Foreign-Language Medicaid Information 75

2.9 Availability of Foreign-Language Medicaid Information Online 82

2.10 Network Effects by Country of Origin 83

2.11 Summary Results for First-Stage Regression 84

2.12 Alternative Measures of Network Density and Probit Estimation 85

3.1 Summary Statistics, Florida Birth Records (1989-2009), Miami CBSA . . . 98

3.2 Food Environment and Excessive Weight Gain (larger than 20 kg), Florida Birth Records(1989-2009), Miami CBSA 99

3.3 Food environment and excessive weight gain (larger than 20 kg), demographic subgroups, US-Born 100

3.4 Other maternal outcomes and food environment, US-Born 101

3.5 Birth outcomes and food environment, US-born 102

3.6 Other maternal outcomes and food environment, Florida-Born, linked-sample 103

3.7 Birth outcomes and food environment, Florida-born, linked sample 104

3.8 Placebo tests: food-environment and other determinants of weight gain . . 105

3.9 Excessive Weight Gain and Food-Environment, Sensitivity to the Addition of Controls 106

List of Figures

1·1 Birth weight (grams) distribution, California and Florida 39

1·2 Selection on health at migration and differences in the incidence of low birth weight betweeen 2^{nd} generation Mexicans and white natives ($\gamma = 0.75, \rho = 0.35$) 40

1·3 Regression towards the mean and differences in the incidence of low birth weight betweeen 3rd generation Mexicans and white natives ($\gamma = 0.75, \rho = 0.35$) 41

List of Abbreviations

ACS	American Community Survey
BMI	Body Mass Index
CPS	Current Population Survey
IV	Instrumental Variables
LBW	Low birth weight
NETS	National Establishment Time Series
OLS	Ordinary Least Squares
U.S.	United States of America

Chapter 1

Why Does the Health of Immigrants Deteriorate?

1.1 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 intergenerationally 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.

1.2 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.[1] 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 tran-

[1]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.

sitory 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 natives. 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.[2]

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,

[2]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.

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.

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.

1.3 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.[3] 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.[4]

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.[5] 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

[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.

[4]The Online Appendix is available on my personal web page: http://people.bu.edu/osea .

[5]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.

(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.

1.3.1 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 1.1 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.[6] [7]

1.3.2 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

---

[6]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.

[7]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.

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 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.[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 1.2 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.[9] I restricted the sample to children born between 1975 and

[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.

[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 ethnic group in the United States. However, when looking at the identical analysis for children of immigrants coming from

1981 to white mothers and Hispanic first-generation immigrant mothers coming from Mexico, Puerto Rico and Cuba.[10] The final sample includes 2,234,571 births for which information on birth weight and zip code 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. 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

---

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 born between 1975 and 1981.

children of non-Hispanic women sharing a similar socioeconomic background.[12] 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).[13] Figure 1·1 depicts the cumulative distribution of birth weight in California and Florida for the immigrant descendants of Hispanic origin and for white natives over the 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

[12]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.

[13]The predominant distribution is substantially equivalent to the distribution of birth weight observed for term births.

on average than children of white native mothers and are 50 grams heavier on average in the 10% quantile (see the online Appendix).[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 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 controls used in the analysis of second-generation birth outcomes.

Columns 5–8 in Table 1.2 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.[15] The estimates in

[14]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 1.2, 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 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.

columns 6 and 8 include the identical 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 (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.

1.4 A simple model of selection and health transmission

1.4.1 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 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

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

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).[17] 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:

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

$$h_1 = u_1$$

where h_1 is the health of the first generation at the time of migration, which is distributed

[17] 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).

as a random normal (μ_j, 1) reflecting the health distribution in the country of origin, μ_j is the average health in country j, and t_1 is the migration threshold. Thus, μ_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, σ_v^2) normal variable reflecting the effect of other unobservable factors on the birth weight of the second generation, γ 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 (μ_{j2}, σ_{u_2}) normal variable 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

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

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, σ_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.[18] 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

$$\rho = \frac{\text{Cov}(BW_3, BW_2)}{\gamma^2} \tag{1.1}$$

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 inci-

[18] 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 γ and ρ would instead depend on the extent of correlation between unobserved random shocks to health and birth weight.

dence among immigrant descendants.

1.4.2 Empirical moments and calibration

Panel A in Table 1.3 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 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.[19] 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 1.4).[20]

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

[19]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.

[20]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.

of their country of origin (Mexico or the United States). I use the native health as a benchmark and set μ_{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 μ_{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 ρ and γ, such that equation (1.1) would hold 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.[21] 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 γ must be $\in [0.58, 1]$.[22] Here, I use the case in which ρ is equal to 0.35 as a baseline, but the interpretation of the simulation exercise does not change significantly for different values of ρ in the defined range [0.2, 0.5]. Using

[21]Without loss of generality one can choose units of birth weight such that $\sigma^2_{BW} = \gamma^2_{h_2} + \sigma^2_v = 1$ and therefore $\text{Cov}(BW_3, BW_2) = \text{Corr}(BW_3, BW_2)$.

[22]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 γ 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.

this parametrization and under the assumption of health having identical effects on birth weight in the two populations, the model can be solved analytically for different level of selection on health at migration t_1. In particular, for the second-generation immigrants, the cumulative distribution function of birth weight will be given by the sum of a truncated normal at t_1 and a random normal variable v_2, while it will be the sum of two normal distribution for the native populations (see Online Appendix B for the formal solution and Turban (2010) and Azzalini (2005) on the convolution of a normal and a truncated normal variable).

Figure 1·2 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 1.3). 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.[23]

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

[23]This prediction is consistent with the findings of Rubalcava et al. (2008) who use the Mexican Family Life Survey and provide evidence that the health levels of migrants from Mexico to the United States differed only slightly from those who did not migrate.

of second-generation Mexican immigrants and children of white natives (y-axis) (Figure 1·3). The vertical solid line corresponds to the degree of selection explaining the second-generation advantage (see Figure 1·2). 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 1.3).

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 1.3, Panel C, column 2).[24] 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

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

than what observed in the data.[25]

1.4.3 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 1.3 illustrates the mean differences in the incidence of these risk factors between first-generation Hispanics and natives (column 3), and between second-generation

[25]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 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)).

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 1.4). 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 1.4.

Accounting for the observed risk factors (the upper dashed line in Figure 1·3), the model can explain approximately 76% of the reverse paradox found after accounting for socioeconomic differences.[26] 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

[26]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.

the importance of socioeconomic and behavioral assimilation, in the next section I analyze the heterogeneity of health convergence among Hispanics.

1.5 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 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 $H_{it,2}$ 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.[27] 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.

1.5.1 Socioeconomic status and quality of care

Table 1.5 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

[27]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.

birth). Column 2 includes grandmother-fixed effects exploiting within-family variation.[28]

The adequacy of prenatal care is here defined as starting prenatal care in the first trimester of pregnancy. The first row of Table 1.5 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 non-significant 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 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 1.5 suggests that the poverty of the environment at birth

[28] A more detailed analysis of the sensitivity of the coefficients to the addition of different batteries of controls is available in the online Appendix.

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.[29]

1.5.2 Maternal risk factors

Table 1.6 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.[30] Column 2 control for the same set of sociodemographic controls using in 1.5. 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 1.6 shows that tobacco 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

[29] 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).

[30] I do not have information on pre-pregnancy BMI before 2004.

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. [31]

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

[31] While not reported in Table 1.6, 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.[32]

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.

1.5.3 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

[32]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.

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 Rico or Mexico.[33] 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.[34] 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 1.7 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

[33]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.

[34]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.

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.[35]

The coefficients of intermarriage on birth weight and incidence of low birth weight reported in Table 1.8 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

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

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.[36]

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% 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 intermar-

[36] 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 that 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.

riage 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 1.8).

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.[37]

1.6 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

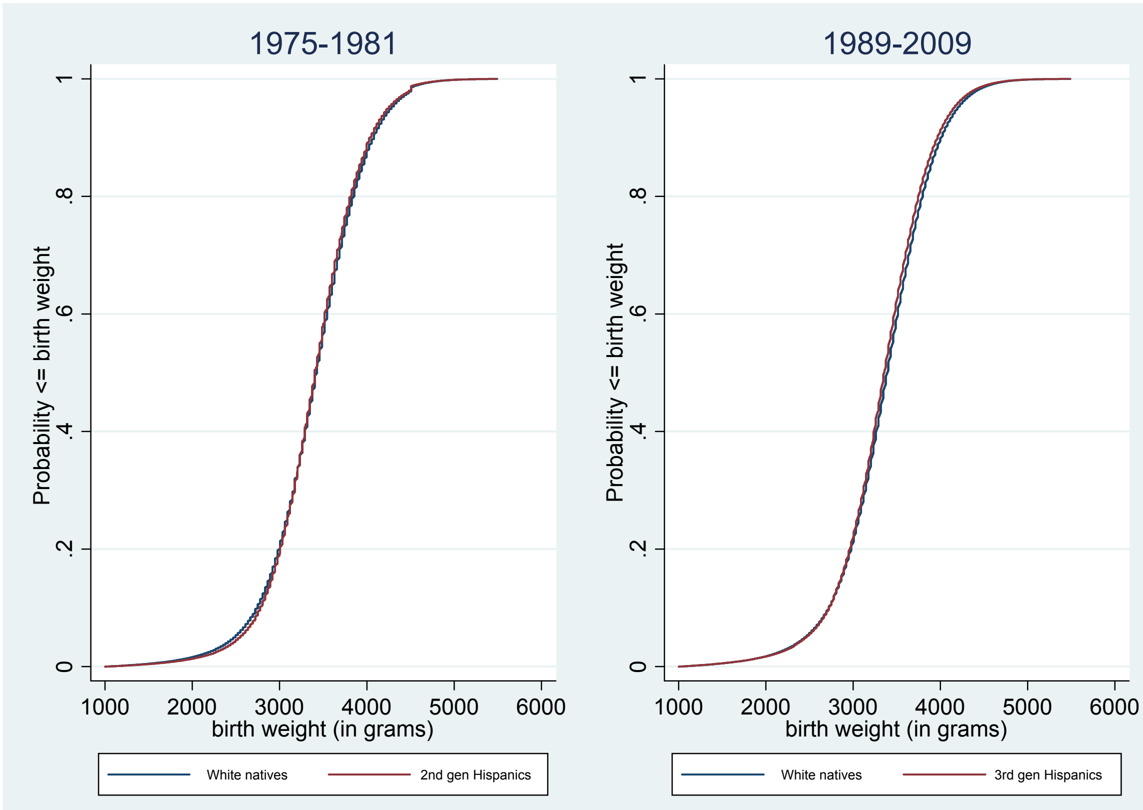
[37]The placebo test is robust not only to the addition of father's and mother's education, which are included in columns 2–4, but also to the separate analysis of women who married to equally, more, or less educated Hispanic men (results available upon request). Note also that when controlling for grandmother-fixed 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.

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.

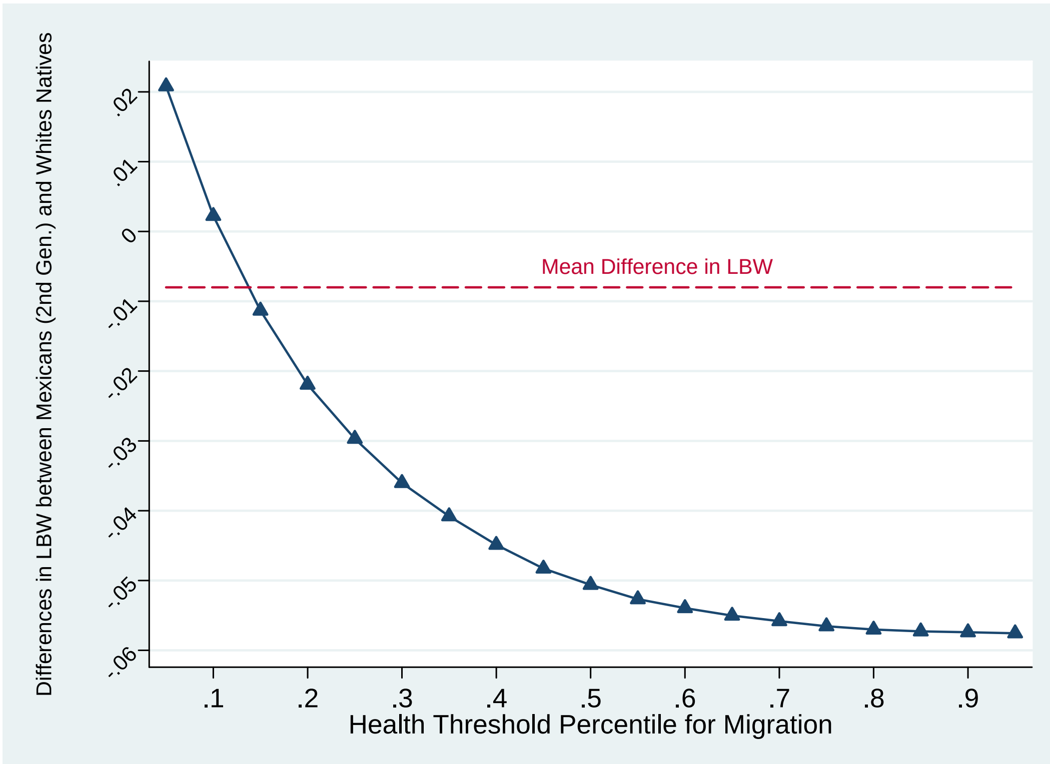
As a whole, these findings show that the health trajectories observed among Hispanic descendants cannot be entirely explained by a pure mechanical statistical process. While there is evidence of a natural regression towards the mean, socioeconomic and behavioral factors mediate the transmission of health 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.

Figure 1.1: Birth weight (grams) distribution, California and Florida



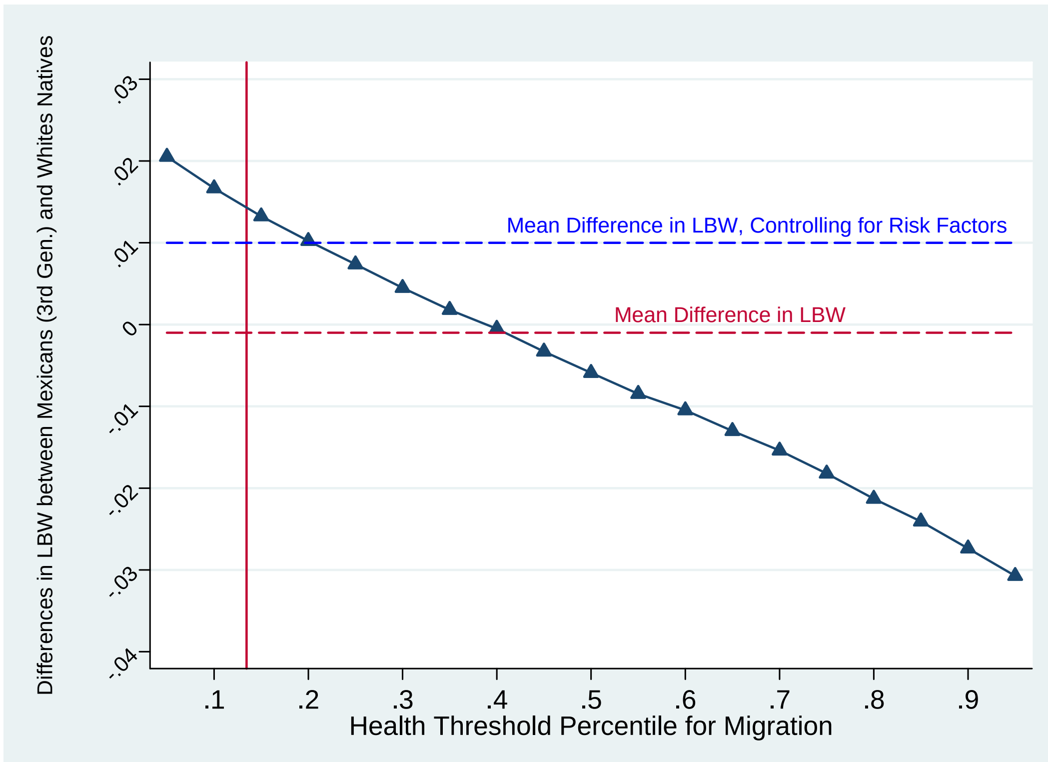
Notes – Source: California and Florida Vital Statistics, 1975–1981, 1989–2009.

Figure 1·2: 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). 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 1.3, col. 1).

Figure 1·3: 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, $\mu_{MX_2} = -0.1$). 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 1.3, 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 1.3, 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 1·2.

Table 1.1: Matching quality. Women born in California and Florida, 1975–1981

	Observations			**Birth Weight (grams)**			**Low Birth Weight (below 2500 grams)**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Sample:	Overall	Linked	Matching rate	Overall	Linked	Nonlinked	Overall	Linked	Nonlinked
Overall	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 1.2: Hispanic Health Paradox in birth weight (BW) and low birth weight incidence (LBW)

	2nd generation 1975–1981				**3rd generation 1989–2009**			
Dependent variable:	BW		LBW		BW		LBW	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Country of origin:								
Cuba	35.407*	−0.285	−0.004	−0.009***	−50.778***	−44.807***	−0.001	0.001
	(20.246)	(4.266)	(0.012)	(0.002)	(5.501)	(7.242)	(0.002)	(0.003)
Mexico	−2.129*	22.647***	−0.008***	−0.016***	−27.983***	−8.078***	−0.001**	−0.004***
	(1.123)	(1.326)	(0.000)	(0.001)	(1.735)	(1.927)	(0.001)	(0.001)
Puerto Rico	−74.826***	−91.191***	0.014	0.005	−172.970***	−158.849***	0.028***	0.022***
	(18.714)	(7.406)	(0.011)	(0.003)	(6.884)	(6.974)	(0.003)	(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 1.3: Differences between 1st, 2nd generation Mexicans and U.S. white natives

	(1)	(2)	(3)	(4)	(5)
	CA-FL		**Florida**		
	$MX_1 - N$	$MX_2 - N$	$MX_1 - N$	$MX_2 - N$	$MX_2 - N$
Panel A: CA-FL Vital Statistics					
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
Panel B: FL Vital Statistics					
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)	
Panel C: CA-FL CPS					
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 1.4: 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	−0.009***	0.001***	0.008***
	(0.000)	(0.000)	(0.000)
Control for risk factors	NO	NO	YES
Panel B: Natality Detail Data			
Tobacco consumption	−0.219***	−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)	
Panel C: CPS			
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 1.5: Adequacy of care, SES and birth outcomes among 3rd generation Hispanics, within-family analysis

	(1)	(2)
Panel A: Birth Weight (in grams)		
Prenatal care	182.003***	153.765***
started 1st trimester of pregnancy	(4.417)	(4.927)
High-poverty (zip code of residence)	−4.261	−8.954*
	(3.518)	(5.009)
High poverty (hospital zip code)	−10.671**	−11.073*
	(4.199)	(5.763)
Mean of Dep. Var	3326.377	
Std. dev.	560.681	
Panel B: Incidence of Low Birth Weight		
Prenatal care	−0.049***	−0.041***
started 1st trimester of pregnancy	(0.002)	(0.002)
High-poverty (zip code of residence)	0.003*	0.001
	(0.001)	(0.002)
High poverty (hospital zip code)	0.005***	0.006**
	(0.002)	(0.003)
Mean of Dep. Var	0.056	
Std. dev.	0.229	
Sociodemographic controls	YES	YES
Mother's birth weight	YES	YES
Grandmother F.E.	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 1.6: 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***	0.028***	0.019***	0.036***	0.029***	0.019***
	(0.001)	(0.002)	(0.003)	(0.002)	(0.003)	(0.007)
Alcohol consumption	0.041***	0.005	0.006	0.063***	0.021	0.005
	(0.006)	(0.008)	(0.010)	(0.013)	(0.014)	(0.028)
Gestational hypertension				0.143***	0.137***	0.088***
				(0.004)	(0.004)	(0.007)
Chronic hypertension				0.108***	0.086***	0.027**
				(0.006)	(0.007)	(0.013)
Diabetes				0.032***	0.018**	0.005
				(0.008)	(0.008)	(0.015)
Normal weight gain	−0.032***	−0.021***	−0.019***			
	(0.001)	(0.001)	(0.001)			
Normal weight gain (adjusted for U.S.)				0.001	0.002	−0.001
				(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 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 1.7: 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***	0.025***	0.020***	0.006**	0.006**
	(0.001)	(0.002)	(0.002)	(0.003)	(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***	0.001***	0.001***	0.001*	0.001
	(0.000)	(0.000)	(0.000)	(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***	0.006**	0.005	0.010	0.010
	(0.000)	(0.003)	(0.003)	(0.007)	(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 1.8: Ethnic intermarriage and birth outcomes among 3rd generation Hispanics, within-family analysis

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

Chapter 2

Medicaid and Ethnic Networks

2.1 Introduction

Uninsured immigrants present a challenge to U.S. efforts to expand health insurance coverage. Compared with native-born citizens, immigrants are more likely to be uninsured and less likely to have regular sources of health care (Kaiser Family Foundation, 2008; Long et al., 2010). Although the 2010 Patient Protection and Affordable Care Act establishes a mandate for insurance, universal coverage may remain elusive. Even in Massachusetts, the first state to set an individual mandate, an estimated 38% of potentially Medicaid-eligible immigrants remained uninsured two years after the state's 2006 reform.[1] It is important to understand differences in take-up within the adult population and, in particular, among those disproportionately uninsured—including racial minorities, non-citizens, and individuals of limited English proficiency. Previous studies have shown that communication problems hinder immigrant enrollment in welfare programs (Ponce et al., 2006; Aizer, 2007; Lavarreda et al., 2006). In this paper, we adopt methods developed by Bertrand et al.

[1]This estimate is our own using the 2008 American Community Survey. We consider "potentially eligible" those immigrants below the 300% of the federal poverty level and not covered by private insurance, the same definition as Long et al. (2010).

(2000) to investigate the role of social networks in immigrant participation in Medicaid.

In the context of social science, two common channels through which networks influence individual behavior are information and norms. We believe both play a role in the case of Medicaid. If immigrants rely on networks for information about public programs, then language barriers and distrust in authorities should be less problematic for those who have access to larger and more informed networks. By ruling out shared norms within ethnic groups, we can identify residual peer effects that can be attributed to differences at the local level, including exchange of information.

Beyond legal and informational barriers to Medicaid enrollment, immigrants may also hesitate to sign up because they distrust government authorities or because they fear deportation. Public program participation has been found to be correlated with the policy climate (Kaushal and Kaestner, 2005; Mazzolari, 2004; Kandula et al., 2004). In particular, Watson (2010) provides evidence that "chilling effects" from immigration policy may widen disparities in Medicaid enrollment across states. She shows that around the time of the 1996 welfare reform, non-citizen participation in Medicaid declined further in states where immigration laws were more strictly enforced.

A rich literature describes the influence of ethnic groups and cultural networks on individual preferences and economic behavior (Borjas, 1995; Edin et al., 2003; Munshi, 2004; Andersson et al., 2010; Jackson and Scheider, 2011). Ethnic networks can help immigrants overcome language barriers and their distrust of authorities, acting as conduits for information and encouraging participation in public programs (Figlio et al., 2011) and the use of the health care system (Devillanova, 2008; Deri, 2005). Bertrand et al. (2000) studied network effects on participation in the welfare system. The authors used U.S.

Census microdata to calculate the proportion of people receiving income from public assistance in various immigrant communities. They found that the receipt of welfare income was more common among immigrants who lived in enclaves and who were of an ethnicity characterized by a relatively high welfare-participation rate. Using similar methodology, Deri (2005) confirmed network effects on health-care utilization among immigrants in Canada. Aslund and Fredricksson (2005) studied a natural experiment for welfare participation among refugees in Sweden and claimed that it is mostly the quality of the contacts (measured as average welfare use in the group) that explains network effects.

The degree to which the peer effects in these studies arise from information-sharing, rather than from spurious correlation, remains debated. Aizer and Currie (2004) disputed the causal effects of ethnic networks on receipt of public assistance. Using data from a California pre-natal care program, the authors showed that the supposed network effects were larger among women who previously participated in the program than among first-time users. Some recent studies using microdata have confirmed the information-sharing hypothesis. Devillanova (2008) found network effects for health-care utilization among undocumented immigrants in Milan, using data from a survey that asked respondents whether they had obtained information from friends or relatives. Figlio et al. (2011) found strong evidence that social networks mitigated the "information shock" faced by immigrant women in Florida following the 1996 welfare reform. These findings suggest that network effects are weaker when information is widespread and have greater influence when information is scarce.

We contribute to this literature with evidence from the American Community Survey (ACS). With the ACS, which began collecting information on health insurance coverage in 2008, we have a new tool for understanding disparities in insurance rates among ethnic groups across the United States. While much past analysis of uninsured immigrants has been based on relatively small sample sizes or on the limited reach of state-based public programs, the ACS allows us to look at the Medicaid program nationwide while taking advantage of local variation and a large number of ethnic networks.

Consistent with previous literature on public assistance, we provide evidence that social networks are important in explaining variation in Medicaid enrollment among immigrants. Our study is of particular importance given changes in migration patterns. As new waves of immigrants are locating to areas without established immigrant communities, many may face challenges in becoming settled (Massey, 2008). Furthermore, we believe that it is valuable to focus on Medicaid for at least two reasons.

First, while we believe the ethnic networks have similar effects for both welfare and Medicaid, the rules governing public assistance have changed radically since 1990, the year studied by Bertrand et al. (2000). The 1996 reform, with which President Clinton promised to end "welfare as we know it," was designed to encourage recipients of government assistance to seek employment. Only 3% of our sample of potentially Medicaid-eligible immigrants reports having received welfare income in the past year.[2] In addition, the law limited immigrants' access to federally funded public assistance, including Medicaid (Borjas, 2002). We find evidence that among potentially eligible immigrants, the probability of Medicaid coverage is greater for individuals living in a linguistic enclave and whose language group is characterized by a high Medicaid enrollment rate. While previous studies focused on programs aimed at women or children (Aizer and Currie, 2004; Aizer, 2007),

[2]Of the Medicaid enrollees in our sample, 10% report receiving welfare. About 70% of welfare recipients report having Medicaid.

we confirm the presence of network effects among a broader population.

Second, Medicaid expansion is a key vehicle for expanding coverage to the uninsured in the Patient Protection and Affordable Care Act. At a minimum, state Medicaid programs are required to cover adults up to 133% of the federal poverty level. Implementation of this policy requires outreach to the uninsured and an understanding of how information is shared and of who is most likely to remain uninformed. In addition to expanding eligibility, PPACA is also expected to spur enrollment among those who were eligible even before the reform, pulling them "out of the woodwork" through information campaigns and the individual mandate (Sommers and Epstein, 2011).

In this paper, we look for a differential effect in participation across networks, modifying the methods developed by Bertrand et al. (2000). Our approach exploits variation in the concentration of immigrant communities and in the Medicaid enrollment rates of different ethnic groups. We also test instrumental variables to address any selection that would occur if immigrants sorted themselves differentially into enclaves. Our instrument is based upon Bertrand et al. (2000), but we slightly modify their strategy to better address the reflection problem and the possibility of random group shocks in the error term. We also discuss an alternative approach to these problems using a measure of *exante* peer characteristics—specifically, past data on Medicaid enrollment—in our robustness checks.

While the data we have do not allow us to pin down the exact contribution of norms and information, our analysis does support that information is a key channel through which networks influence Medicaid enrollment. Network effects are strong even when we control for states' publication of Medicaid information in other languages. Our results also suggest that network quality has the strongest influence on individuals who are less likely to obtain Medicaid information through formal channels and who are least assimilated. Depending on the specification, we find little or no effect among immigrants who have been in the U.S. longer, are proficient in English, or are U.S.-born.

The structure of the paper is as follows. In Section 2, we describe the data. We explain our model in Section 3 and present the main results in Section 4. In Section 5, we conduct several robustness checks and some sensitivity analyses. Section 6 concludes our study.

2.2 Data

We use the 2008 and 2009 American Community Survey (ACS), which is available from the Integrated Public Use Microdata Series (Ruggles et al., 2010). The ACS, a random sample of 3 million households, is conducted annually by the U.S. Census Bureau. Compared to other data sets with health insurance information, the ACS provides the advantages of being a nationwide survey and having a large sample size. These facts allow us to examine the role of network effects across all states and dozens of ethnic groups. The ACS is appropriate for our purposes because it contains individual-level information on health-insurance coverage, location, and language.

The ACS asks respondents which language they speak at home, allowing us to identify linguistic enclaves and to test for network effects within them. Strictly speaking, the "language groups" we use are not ethnic categories. The Spanish group contains Spaniards as well as Mexicans and Salvadorans, while Indians are represented by a multitude of languages. In most cases, however, language- and culture-based groupings do coincide.

Using answers from the questions "Does this person speak a language other than English at home?" and "What is this language?", we build two alternate measures of

Table 2.1: Summary Statistics

Variable	Mean	Std. Dev.	Min	Max
enrolled in Medicaid	0.19	0.39	0	1
male	0.50	0.50	0	1
age	38.48	11.59	19	64
high school drop-out	0.47	0.50	0	1
high school graduate	0.26	0.44	0	1
some college	0.13	0.33	0	1
college or more	0.14	0.35	0	1
married, present	0.46	0.50	0	1
married, absent	0.07	0.26	0	1
widowed	0.02	0.14	0	1
divorced	0.08	0.27	0	1
separated	0.05	0.21	0	1
never married	0.33	0.47	0	1
number of children	1.28	1.43	0	9
child present	0.40	0.49	0	1
single mother	0.04	0.20	0	1
white	0.49	0.50	0	1
black	0.05	0.22	0	1
years in U.S.	15.06	10.72	0	65
more than 5 years in U.S.	0.80	0.40	0	1
fluent	0.48	0.50	0	1
MSA CA	3.14	4.30	0.03	85.30
PUMA CA	8.30	26.58	0.05	476.13
log(MSA CA)	0.69	0.92	-3.42	4.45
log(PUMA CA)	1.13	1.22	-3.02	6.17

Notes - Data source: 2008 and 2009 American Community Survey. For the top portion of the table, the sample is composed of non-English-speaking immigrants ages 19 to 64 below the 200% of FPL who belong to a language group with more than 50 observations remaining after imposing all other criteria and who live in a state that provides Medicaid-type coverage to immigrants (see Section 2). Individuals living in non-identified MSAs and the institutional population are excluded from the sample. The final sample consists of 118,803 individuals. In the lower portion of the table, the contact availability measures are calculated as described in Section 2. "Child present" is a dummy that equals 1 if an individual has any children. Our dummy variable for fluency equals 1 for individuals who speak English "well" or "very well." The means reported in the lower panel of the table refer to the CA for individuals in the final sample.

network availability. The first is based on metropolitan statistical areas (MSAs) and the second on public use microdata areas (PUMAs), the smallest geographic unit available to us in the ACS. A typical PUMA has a population of about 100,000 and can be thought of as a large neighborhood within an MSA. For example, the MSA that contains New York City also reaches into New Jersey and Pennsylvania. The New York MSA is broken down into dozens of PUMAs. In lower Manhattan alone, Greenwich Village, Chinatown, and the Financial District represent three separate PUMAs.

We capture network strength as "contact availability" (CA), adopting our definition from Bertrand et al. (2000). CA is the logarithm of the local proportion of people speaking a given language, relative to the national share of that language's speakers. Formally, network density is computed as follows:[3]

$$ln(\frac{C_{jkt}/A_{jt}}{L_{kt}/T_t}) \tag{2.1}$$

where C_{jkt} is the number of people in area j at time t speaking language k and A_{jt} is the total population of area j at time t. As Bertrand et al. (2000) do, we normalize the local density by the the national level density of language group. The denominator L_{kt}/T_t is the share of population living in the United States and speaking language k at time t.

Our calculation for CA is based on the entire ACS population because the relevant network for information dissemination is all speakers of a language, not just those who are Medicaid-eligible.[4] For all other parts of our analysis, we pare down the ACS to potentially Medicaid-eligible immigrants, broadly defined. We focus on first-generation immigrants

[3] Our results are not sensitive to the exact specification of CA. In Table 2.12 we provide robustness checks using different measures of concentration, including a non-log measure and a non-normalized measure.

[4] Summary statistics on contact availability measures are presented in the lower panel of Table 2.1. These measures are consistent with other findings in the literature (Bertrand et al., 2000).

by excluding anyone born in the United States.[5] Because we are interested in linguistic networks, we drop individuals whose primary language spoken at home is English. We then keep only individuals who are non-institutionalized, non-elderly adults (ages 19 to 64 years old). It is also necessary to drop those living outside identifiable MSAs for our identification procedure.

Describing the remaining criteria for our sample necessitates a brief discussion of Medicaid eligibility rules. Medicaid was created to assist the low-income, low-asset population. While each state designs and administers its own program, all must comply with the federal minimum levels of provision, based on the federal poverty level. The poverty level is not a single number but a sliding scale based on income and household size, and the threshold for eligibility also varies by demographic group. For example, in 2009 the federal minimum was 235% for children, 185% for pregnant women, 64% for working parents, 38% for non-working adults, and 75% for the elderly and disabled. While the bulk of Medicaid enrollees are women and children, a legacy of the historical link between Medicaid eligibility rules and those of Aid to Families with Dependent Children, the elderly and disabled account for 70% of Medicaid expenditures (Kaiser Family Foundation, 2010).

The welfare reform act of 1996 bars most immigrants from receiving federally supported Medicaid until they have resided in the United States for 5 years. Any state that wishes to cover immigrants during that initial period must do so with its own dollars. States are allowed to use federal money to provide Medicaid to qualified immigrants who have

[5]We find only small or no effects among second-generation immigrants, depending on the definition of the network. This finding, discussed later in the paper, is consistent with our intuition that linguistic networks should have less influence on the behavior individuals who can easily communicate in English or who have greater familiarity with American culture. We also tested our model on a more restrictive sample that included only foreign-born U.S. citizens, and our findings were similar.

been in the U.S. for at least 5 years. In 2008 and 2009, 7 states restricted Medicaid access to immigrants even after the 5-year bar: Alabama, Mississippi, North Dakota, Ohio, Texas, Virginia, and Wyoming (Broder and Blazer, 2009). We exclude these states from our analysis.[6] Unfortunately, the ACS does not have sufficient information to determine Medicaid eligibility with precision. In order to focus on those who *potentially* have the option of Medicaid as their primary form of insurance, we further restrict the sample to households below the 200% of the FPL and individuals who have been in the country for at least 5 years.[7] Finally, we use only language groups with more than 50 observations after imposing all other restrictions.[8]

Our final sample comprises 37 language groups, 164 MSAs, 1,218 PUMAs, and 118,803 individuals. Within the sample, 21% are covered by Medicaid. We report descriptive statistics for our main variables of interest and our controls in Table 2.1 in the Appendix. There is substantial variation in insurance status across language groups (see Table 2.2). The average Medicaid enrollment ranges from around 10% for speakers of Korean, German, and Gujarati to over 40% among the Armenian, Bengali, Cushite, Cantonese, and Miao language groups. Spanish speakers account for nearly three-quarters of final sample (83,740 people), and are far more numerous than the next-largest groups: Chinese (3,862),[9] Vietnamese (3,546), Korean (2,552), Filipino (2,282), and Arabic (2,146).

Table 2.2 shows descriptive statistics for insurance coverage by language group. Though the sample used for our model excludes individuals with private insurance, in Table 2.2,

[6]However, including the residents of these states in the sample has little impact on our results because any state-level policy discrepancies are absorbed by the PUMA and MSA fixed effects in our model.

[7]As discussed later in the paper, we conduct tests using other cutoffs of the FPL.

[8]Our results are not sensitive to raising or lowering this threshold for language group size.

[9]We follow the ACS convention and leave "Chinese" as a category distinct from Cantonese, Formosan, and Mandarin. Combining all dialects into a single Chinese-language group does not affect our results.

Table 2.2: Health Insurance Coverage by Language

Language	% Private insurance	% Medicaid	% Uninsured	Individuals
Spanish	0.22	0.17	0.62	83,740
Chinese	0.44	0.23	0.34	3,862
Vietnamese	0.33	0.31	0.36	3,546
Korean	0.41	0.11	0.49	2,552
Filipino, Tagalog	0.55	0.14	0.31	2,282
Arabic	0.29	0.33	0.38	2,146
French	0.32	0.16	0.52	2,071
Russian	0.33	0.32	0.36	1,776
Cantonese	0.32	0.43	0.28	1,367
Portuguese	0.29	0.20	0.52	1,268
French	0.46	0.16	0.38	1,197
Mandarin	0.37	0.25	0.41	1,028
Hindi	0.55	0.12	0.33	790
Polish	0.31	0.17	0.50	776
Urdu	0.27	0.30	0.44	755
Persian	0.32	0.29	0.39	726
Bengali	0.31	0.47	0.28	681
Armenian	0.21	0.43	0.35	649
Kru	0.38	0.17	0.45	605
Japanese	0.67	0.05	0.28	600
Mon-Khmer	0.32	0.40	0.26	597
German	0.58	0.12	0.28	537
Panjabi	0.32	0.23	0.45	535
Miao, Hmong	0.31	0.47	0.23	505
Amharic	0.34	0.19	0.46	498
Italian	0.55	0.16	0.27	463
Gujarati	0.42	0.12	0.47	445
Thai	0.49	0.08	0.44	340
Ukrainian	0.27	0.25	0.48	294
Turkish	0.42	0.23	0.37	294
Albanian	0.32	0.24	0.44	289
Cushite	0.19	0.60	0.23	283
Laotian	0.38	0.29	0.28	282
Rumanian	0.35	0.17	0.47	278
Hebrew	0.58	0.20	0.22	270
Serbo-Croatian	0.54	0.17	0.28	242
Greek	0.57	0.21	0.22	234
Total	0.26	0.19	0.55	118,803

Notes - Data source: 2008 and 2009 American Community Survey. The sample is defined as in Table 1.

we show the insurance status for all immigrants in the ACS (i.e., our sample for the calculation of CA) who live in identifiable MSAs and speak one of the 37 languages included in the restricted sample. Spanish-speaking immigrants have an especially high proportion of uninsured individuals (62%).

2.3 Empirical Specification

The network effect in our model is represented as interaction between the local density of people who share a language and their "culture" with respect to Medicaid. Bertrand et al. (2000) define a network measure that captures both the "quantity" and the "quality" of contacts in a given language group. Following the model from Bertrand et al. (2000), we express this as:

$$network_{jkt} = CA_{jkt} \times \overline{Medicaid_{kt(-i)}} \tag{2.2}$$

CA_{jkt}, the measure of network availability can be thought of as a network *quantity* measure. Network *quality* is represented by $\overline{Medicaid_{kt(-i)}}$, the Medicaid enrollment rate of the language group, expressed as a deviation from the the average of the entire sample over all groups. Although we are concerned with the peer effect that occurs at the local level, we use the nationwide Medicaid enrollment rate for each language as a proxy for cultural attitudes toward public assistance. Using the local Medicaid enrollment rate potentially creates an omitted variable bias because the average enrollment of language group k in area j could be correlated with unobservable characteristics individual i shares with others in the immediate vicinity (Bertrand et al., 2000).

Our main specification is a linear probability model:

$$Medicaid_{ijkt} = (CA_{jkt} \times \overline{Medicaid_{kt(-i)}})\alpha + X_{ijkt}\beta + CA_{jkt}\theta$$
$$+ \gamma_j + \delta_k + \tau_t + \epsilon_{ijkt} \tag{2.3}$$

where i is the subscript for individuals, j for geographic areas, k for languages, t for time. Dependent variable $Medicaid_{ijkt}$ is binary and equal to 1 for a person covered by Medicaid.[10]

Adopting the notation of Bertrand et al. (2000), we use fixed effects γ_j for geographic units and δ_k for languages, which allows us to control for potential unobserved differences between localities and between language groups. We also include a dummy variable for year, τ. The PUMA fixed effects control for institutional sources of variation in Medicaid enrollment, for any state-level differences in Medicaid rules, and for the effect of "welfare magnets" (Borjas, 1999). Using language group fixed effects allow us also to take into consideration possible omitted characteristics of the language group, such as cultural attitudes toward assimilation or welfare dependence. Through the contact availability measure CA_{jk} on the right-hand side, the model also accounts for any omitted individual-level characteristics that may be correlated with the likelihood of living in an ethnic enclave.

As in other peer-effects studies, this model would still suffer omitted variable bias if $CA_{jkt} \times \overline{Medicaid_{kt(-i)}}$ were correlated with unobserved individual characteristics (Bertrand et al., 2000). Living in an ethnic enclave could be associated with different traits in the

[10]The ACS asks, "Is this person currently covered by any of the following types of health insurance or health coverage plans?" We consider the respondent to be on Medicaid if they selected the multiple-choice answer "Medicaid, Medical Assistance, or any kind of government-assistance plan for those with low incomes or a disability."

culture of high- or low-enrollment groups. We illustrate differential selection with the following example. Suppose that for members of high-Medicaid language groups, living in an ethnic enclave is associated with being more dependent on the local community and also on public assistance. For low-Medicaid language groups, however, those in enclaves have a culture of self-reliance, and for them, enclaves have with relatively low use of public assistance. If such selection existed as described, it would bias upward our estimate of network effects because we cannot observe and control for cultural attitudes toward autonomy. We would be capturing spurious correlation because members of high-Medicaid ethnicities who place less value on self-reliance are more likely to live in enclaves.

Another concern with peer-effects modeling is the reflection problem (Manski, 1993), which occurs when a specification incorrectly attributes network effects to what are actually correlations due to unobserved group-level shocks. While our model does not allow us to distinguish endogenous effects (influenced by group behavior) from exogenous effects (influenced by group characteristics), we do address the reflection problem by excluding individual i from the computation of $\overline{Medicaid_{kt}}$.[11] Correlated errors within groups can be troublesome for statistical inference (see Angrist and Pischke (2008)). In our case, a shock that affects all members of a language group in a local area would bias our estimates through spurious correlation.

For these reasons, we test an instrumental variables specification in addition to the ordinary least squares model. Following Bertrand et al. (2000) and Evans et al. (1992), our IV involves contact availability at the two local geographic divisions available to us: the smaller PUMA and larger MSA. We use MSA-level CA as an instrument for both

[11] In an abuse of notation, from here onward we will refer to the mean Medicaid rate simply $\overline{Medicaid_{kt}}$ rather than $\overline{Medicaid_{kt(-i)}}$.

the PUMA-level CA and the PUMA-level $CA_{jkt} \times \overline{Medicaid_{kt}}$ interaction term. Unlike Bertrand et al. (2000), however, we exclude PUMA j itself when calculating the MSA-level concentration for the instrument for PUMA j. By doing so, we avoid contaminating in the instrumental variable with any common shock that affected PUMA j while still retaining MSA-level idiosyncrasies.

Our IV identification rests on the assumption that selection at the PUMA level is greater than selection at the MSA level (Bertrand et al., 2000). The concentration of an ethnic group in a PUMA is correlated to concentration of the group in the corresponding MSA, but if differential selection were driving the results, then the OLS would overestimate the network effects. The IV specification allow us to detect whether even in the presence of differential selection there are network effects among language groups. An alternative identification method to address the reflection problem, in which we use lagged variables as a proxy for current Medicaid use, is discussed in Section 5.

2.4 Main Results

The main parameter of interest in our linear probability model is α, the coefficient on the interaction of CA_{jkt} with $\overline{Medicaid_{kt}}$. The term captures the multiplier effect that is greatest when an enclave is dense and its members come from a culture of widespread Medicaid use.

We show our estimates from our baseline model and some variations in Table 2.3. Standard errors are clustered at the PUMA or MSA level as appropriate for the specification.[12] Our first step was to look for network effects at the PUMA level in an ordinary

[12] All results were generated with Stata Statistical Software, Release 11, using the ACS person weights

Table 2.3: Main Results

Dependent Variable: Medicaid Enrollment	(1)	(2)	(3)
CA measure:	log PUMA	log MSA	log MSA
Methodology:	OLS	IV	OLS
CA$\times\overline{Medicaid}$	0.186***	0.343***	0.240***
	(0.027)	(0.047)	(0.032)
CA	0.008***	0.006	0.008***
	(0.003)	(0.006)	(0.003)
male	-0.064***	-0.064***	-0.064***
	(0.003)	(0.003)	(0.003)
fluent	-0.022***	-0.022***	-0.025***
	(0.003)	(0.003)	(0.003)
child present	-0.014**	-0.014**	-0.015***
	(0.006)	(0.006)	(0.006)
number of children in household	0.028***	0.028***	0.028***
	(0.002)	(0.002)	(0.002)
high school drop-out	0.057***	0.057***	0.060***
	(0.005)	(0.005)	(0.005)
high school graduate	0.038***	0.038***	0.040***
	(0.004)	(0.004)	(0.005)
some college	0.029***	0.030***	0.031***
	(0.005)	(0.005)	(0.005)
years in U.S.	0.003***	0.003***	0.003***
	(0.001)	(0.001)	(0.001)
years in U.S.^{2}	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
more than 5 years in U.S.	-0.008	-0.008*	-0.009*
	(0.005)	(0.005)	(0.005)
single mother	0.084***	0.090***	0.090***
	(0.008)	(0.008)	(0.008)
age	-0.009***	-0.009***	-0.009***
	(0.001)	(0.001)	(0.001)
$age^2/100$	0.014***	0.014***	0.014***
	(0.001)	(0.001)	(0.001)
white	0.004	0.004	0.001
	(0.004)	(0.004)	(0.004)
black	0.043***	0.044***	0.057***
	(0.012)	(0.012)	(0.012)
PUMA F.E.	YES	YES	YES
Language Group F.E.	YES	YES	YES
Year F.E.	YES	YES	YES
Citizenship status	YES	YES	YES
Marital status	YES	YES	YES
Adjusted R^2	0.180	0.180	0.166
Observations	118,803	118,803	118,803

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes - Data source: 2008 and 2009 American Community Survey. The sample is defined as in Table 2.1. MSA-level CA and MSA-level CA$\times\overline{Medicaid}$ are used as instruments for CA and CA$\times\overline{Medicaid}$ at the PUMA level, respectively. First-stage estimates are reported in Table 2.11 in the Appendix. Standard errors are clustered at the PUMA level.

least squares regression, the results of which are shown in Column 1. The network effect has a coefficient of 0.19 and is significant at the 1% level.

The model includes fixed effects for year, language groups, and geographic areas, as well as standard socioeconomic and demographic controls.[13] Specifically, our controls include age and its quadratic, and dummies for education, marital status, single-motherhood, the presence of a child in the household, the number of children ever born, race, and U.S. citizenship.[14] We also include the number of years since arrival in United States, and to account for non-linear effects of residency in the United States, dummy for having resided in the U.S. for more than 5 years and the quadratic of years since arrival. As expected, variables associated with lower socioeconomic status increase the likelihood of Medicaid coverage. High-school drop-outs are more likely to be on Medicaid, as are single mothers. Being white is associated with lower probability of being on Medicaid, while having more children and being divorced or never married are associated with a higher probability of enrollment.

If individuals could costlessly self-select geographically, and sorted themselves differentially across language groups, then OLS estimate would be biased from selection at both the MSA and the PUMA levels. Any bias that remains in the coefficient of the IV model, presented in Column 2 of Table 2.3, would come from MSA-level selection. Instrumenting with MSA-level contact availability produces a network coefficient of 0.34 and does not

---

with the *aweight* option. We also tried dual-level clustering by PUMA and language, by language only, and by PUMA and household. Our results were similar to the baseline.

[13]We also tested each year separately. The results for individual years do not substantially differ from the results using the pooled data.

[14]Our baseline specification does not include income as a control. We did, however, test what happens when we do include the logarithm of the sum of weekly income of all family members in the household. Only half our sample has non-missing wage data, and among those we find similar network effects.

change the sign of the coefficients on the control variables.[15] If anything, the increase in the size of the network effect suggests that OLS underestimates the network effect on Medicaid enrollment and that differential selection cannot completely explain our results. The IV results support the hypothesis that living in an area with a concentration of high-enrollment groups raises the probability of Medicaid coverage for immigrants. Consistent with the literature, we adopt as our baseline specification the OLS model; this estimate is the most conservative estimate of network effect and can be considered the lower bound of the true effect (Bertrand et al., 2000).

Although we can easily compare various specifications of our model by the relative magnitude of network effect coefficient α, the interpretation of α itself is not straightforward. We repeat the thought experiment presented in Bertrand et al. (2000). Imagine a policy shock that directly increases Medicaid participation by some amount λ. Mathematically, this effect is expressed in the model as:

$$Medicaid_{ijk} = \lambda + CA_{jk} \times \overline{Medicaid_k}\alpha + X_{ijk}\beta + CA_{jk}\theta + \gamma_j + \delta_k + \epsilon_{ijk}. \tag{2.4}$$

In this formulation, a 1 percentage point increase in λ would increase Medicaid enrollment by 1 percentage point if networks had no effect on participation. However, an increase in λ generates an additional indirect effect through the increase in $\overline{Medicaid_k}$. After taking the average with respect to k on both sides we can solve for the overall increase in Medicaid participation and $\overline{Medicaid_k}$ and identify the network effect by differentiating with respect to λ.

[15]Using the IV from Bertrand et al. (2000) generates a slightly larger estimate of 0.35 and a standard error of 0.04. We also try the lagged-variables approach suggested by Card (2001). Instrumenting CA for each PUMA with past CA levels based on the 1990 and 2000 Censuses produces similar results.

Table 2.4: Changing the Control Variables

Dependent Variable: Medicaid Enrollment	(1)	(2)	(3)	(4)	(5)
CA$\times\overline{Medicaid}$	0.176***	0.179***	0.179***	0.183***	0.186***
	(0.034)	(0.029)	(0.028)	(0.028)	(0.027)
PUMA F.E.	NO	YES	YES	YES	YES
Language group F.E.	YES	YES	YES	YES	YES
Year F.E.	YES	YES	YES	YES	YES
Exogenous controls	NO	NO	YES	YES	YES
Exogenous controls and education	NO	NO	NO	YES	YES
Adjusted R^2	0.037	0.131	0.153	0.155	0.180
Observations	118,803	118,803	118,803	118,803	118,803

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes - Data source: 2008 and 2009 American Community Survey. The sample is the same as in the baseline specification (Column 1 of Table 2.3). Standard errors are clustered at the PUMA level.

Formally:

$$\frac{d\overline{Medicaid_k}}{d\lambda} = 1 + \alpha CA_{jk} \times \frac{d\overline{Medicaid_k}}{d\lambda} \tag{2.5}$$

A policy that boosts Medicaid enrollment by 1 percentage point directly would increase enrollment in language group k by $\frac{1}{1-\alpha\overline{CA}_k}$ percentage points when the multiplier effect is taken into consideration. We plug in the coefficient α from our baseline OLS specification and take the mean of that expression weighted by language group size, we find that PUMA-level linguistic networks generate a marginal increase of 26 percent. In Table 2.4 we show the estimate is not sensitive to our choice of covariates.

2.5 Robustness Checks

In this section, we discuss the sensitivity of our results to other specifications. We test other batteries of control variables, strengthen our criteria for Medicaid eligibility, and run the model on subsets of the data based on sex and language. Our results are robust to these tweaks of our model. We also find network effects when using a specification that uses *exante* group characteristics to proxy for "welfare culture" and when controlling for the availability of Medicaid resources in foreign-languages, which has a positive impact on participation.

In Table 2.4 we test the sensitivity of our estimates to different sets of control variables. Adding and removing these exogenous control variables (e.g., age, gender, and race), as well as others that are more likely endogenous (marital status or family size), does not substantially affect the coefficient. Our baseline specification uses a linear probability model, and we obtain similar results using probit estimation (see Table 2.12, Column 3). Our results are also robust to different definitions of contact availability. For example, instead of using CA as our measure for people speaking language k in area j, we replaced it with the simple logarithm $ln(C_{jkt})$ and with the logarithm of the local share speaking that language, $ln(\frac{C_{jkt}}{A_jt})$, without normalizing by the national level (see Appendix Table 2.12, Columns 1 and 2).

2.5.1 Restricting the Sample by Age, Sex, and Language

We next investigate how the network effect varies for subsets of our sample. Because more than three-quarters are Spanish speakers, we verify our results are not driven solely by this group. Only 22% of Spanish speakers have private insurance, a proportion much lower than that of the other language groups. By comparison, the rate of private insurance is 46% among English speakers and around 60% for speakers of Japanese, German, Greek and Hebrew (Table 2.2). Running the same set of regressions from the previous sections but excluding Spanish language groups, we obtain a coefficient of 0.16, which is significant

Table 2.5: Robustness Checks: Sub-Samples by Sociodemographic Characteristics

	CA×$\overline{Medicaid}$	Sample size
Panel A: By Demographic Characteristics		
Overall sample	0.186***	118,803
	(0.027)	
Non-Spanish speakers	0.164***	35,063
	(0.029)	
Men	0.192***	55,344
	(0.033)	
Women	0.186***	63,459
	(0.032)	
Women<55	0.207***	54,662
	(0.036)	
Women≤ 45	0.195***	42,254
	(0.042)	
Women≤ 35	0.106**	23,839
	(0.054)	
Women w/children in household	0.241***	50,959
	(0.047)	
Panel B: By Percentage of Federal Poverty Line		
Below 150% FPL	0.185***	83,175
	(0.033)	
Below 100% FPL	0.152***	47,073
	(0.043)	
Below 50% FPL	0.137***	18,401
	(0.055)	
Panel C: By Fluency and Second-Generation Immigrants		
Non-Fluent	0.188***	59,765
	(0.045)	
Fluent	0.139***	59,038
	(0.029)	
U.S.-born speaking a language other than English at home	0.069*	31,189
	(0.004)	

Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Notes - Data source: 2008 and 2009 American Community Survey. Each line represents a separate regression with Medicaid enrollment as the dependent variable. The control variables are the same as in the baseline specification (Column 1 of Table 2.3). The sample is the same as baseline, plus additional restriction(s) as indicated. Our dummy variable for fluency equals 1 for individuals who report speaking English "well" or "very well." Standard errors are clustered at the PUMA level.

and similar to our estimate for the overall sample.

One might also be concerned that variation in our measure of CA might reflect not the density of immigrants but the size of their households. We address this issue with specifications using households, rather than individuals, as the unit of analysis. To ensure that each household is represented only once in the data, we eliminate all men and retain only one woman per household. Because Medicaid rules are more generous toward women in their childbearing years, we also try isolating various age groups. When we bring down the upper bound on age to 55 or 45 years old, instead of 64 years old, the effect is still strong and significant. However, when we limit the sample to the 19-35 age group, the sample size is greatly reduced and the network effect becomes weaker and loses statistical significance. We also find the network effect is slightly smaller among women than among men, although though not statistically different (Table 2.5, Panel A). The difference could be explained by the fact that women have a wider variety of opportunities to learn about Medicaid from out-of-network sources, while men are more dependent upon acquaintances for knowledge of the program. States target women in outreach campaigns, and Medicaid and CHIP advertisements tend to feature women and children rather than male parents (**?**).

One criticism of our analysis might be that the sample is overly broad. We consider "potentially Medicaid-eligible" all immigrants with family incomes below the 200% of the FPL. Ideally, we would want to use a sample of only those actually eligible for the program, but we cannot observe all the conditions necessary to impute Medicaid eligibility in the ACS data. While we do not have the information needed to determine eligibility for Medicaid exactly, we can further restrict our sample to low-income individuals (see

Table 2.5, Panel B).[16] Network effects remain present when we focus on individuals below 150%, 100%, and 50% of the FPL.

When we do attempt to impute eligibility, the results are inconsistent with the ACS data. While the main criterion for Medicaid eligibility is income, eligibility also depends on sex, age, disability status, pregnancy, and assets. Further complicating matters, most states provide Medicaid or SCHIP coverage beyond the federal minimum requirements. We take the 2009 state-level Medicaid criteria as summarized by the Kaiser Family Foundation (Ross and Horn, 2008) and apply those to the ACS respondents' information on state of residence, age, family size, income, and sex. The imputed eligibility differs greatly from self-reported Medicaid participation; almost half the individuals who reported Medicaid coverage are deemed not eligible by our imputation. Despite this inconsistency, characteristics of self-reported Medicaid recipients in our sample do generally match demographic characteristics of the actual Medicaid population. For example, we can use the ACS to confirm that enrollment rates in Medicaid are lower among immigrants. In Massachusetts, which has relatively liberal eligibility rules, we find that 38% of potentially eligible immigrants in our sample report having no insurance in 2008, compared with 28% of native-born residents.

[16] In the entire 2008-2009 ACS, half of individuals covered by Medicaid are above the 115% of the FPL, while 25% are above the 206% of the FPL. We also try limiting our sample to those not covered by private insurance (i.e., who report having Medicaid or being uninsured), but this does not affect our results substantially. We still find positive and significant results for samples with 150% and 100% FPL cutoffs. The sample shrinks considerably and we lose power, however, when looking at individuals below 50% of the FPL who lack private insurance.

Table 2.6: Robustness Checks: Years in U.S., Citizenship, and English Fluency

	(1)	(2)	(3)	(4)	(5)
CA$\times\overline{Medicaid}$	0.186***	0.274***	0.241***	0.315***	0.374***
	(0.027)	(0.033)	(0.029)	(0.030)	(0.094)
$CA \times \overline{Medicaid}\times$years in U.S.		-0.006***			
		(0.001)			
$CA \times \overline{Medicaid}\times$U.S. citizen			-0.118***		
			(0.027)		
$CA \times \overline{Medicaid}\times$English fluency				-0.235***	
				(0.021)	
$CA \times \overline{Medicaid}\times$ local avg. English fluency of language group					-0.274**
					(0.112)
Adjusted R^2	0.189	0.190	0.189	0.191	0.189
Observations	118,803	118,803	118,803	118,803	118,803

Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Notes - Data Source: 2008 and 2009 American Community Survey. The sample and control variables are the same as in the baseline specification (Column 1 of Table 2.3). Our dummy variable for fluency equals 1 for individuals who report speaking English "well" or "very well." Standard errors are clustered at the PUMA level.

2.5.2 Citizenship, Language Fluency, and Length of Residence

Based upon our main hypothesis that network effects reduce informational costs, we expect networks to be more influential for people of limited English proficiency or who arrived in the United States more recently. In general, we examine the effect on our results from altering the model in two ways: by changing the sample of individuals i and by changing our definition of CA. When we restrict our sample to the English-proficient, the network effect is weaker than for the whole sample. Effects are smaller among the U.S.-born who speak a language other than English at home (see Table 2.5, Panel C).[17] If networks affect

[17]If we restrict the CA definition to consider only the U.S.-born as members of the network, the network coefficient is not significantly different from zero. Due to the limitations of the ACS questionnaire, we are not able to identify as immigrant descendants those who report speaking English at home. Additionally, we tested our model using a sample of only U.S. citizens. This more restrictive sample produced a network

insurance choices by reducing language barriers and facilitating access to information, then it makes sense that those with the weakest English skills would benefit most from exposure to speakers of their native language.

We also tested a specification that takes into account individuals' degree of assimilation. We test additional network interaction terms that include the length of residency in United States, U.S. citizenship, and a dummy for English fluency (equal to 1 for those who report speaking English "well" or "very well" and 0 otherwise). We also test the influence of average English proficiency of the local network (Table 2.6). The network effects appear weaker for for fluent speakers, older cohorts of immigrants, and for U.S. citizens. The average English fluency of the network also reduces the network's effect. These results are consistent with the idea that the information advantage implied by network availability becomes less relevant for those who have been in the U.S. longer or who have greater skill in understanding English.

2.5.3 Availability of Medicaid Information in Other Languages

An alternative explanation for the phenomenon that we attribute to language networks would be that government authorities respond to immigrant populations with outreach campaigns. A simple and relatively low-cost policy solution to narrowing the participation gap between immigrants and natives is to create multilingual information campaigns. States can improve awareness of Medicaid by providing radio advertisements, billboards, posters, or pamphlets in the languages of their local immigrant communities. Aizer (2007) used data from California, which launched an advertising campaign and hired community-

---

effect that is of of the same sign and slightly smaller. The coefficient is 0.16 (standard error 0.04) for citizens, compared with 0.21 (standard error 0.03) for non-citizens.

Table 2.7: Foreign-Language Medicaid Information

Language	Number of States
Spanish	30
Vietnamese	7
Russian	6
Hmong	5
Chinese	4
Korean	3
Cambodian/Mon-Khmer	2
Somali	2
Lao	2
Creole	1
Arabic	1
Farsi	1
Tagalog	1
Japanese	1
Armenian	1
French	1
German	1
Tigrigna	1
Burmese	1
Ukranian	1
Other languages in the sample	0

Notes - Websites were checked on November 30, 2010. States are counted as having online Medicaid information if the state Medicaid site has either a web page or a downloadable application form or pamphlet in that language.

based assistants to facilitate the Medicaid-application process, to show that outreach efforts explain 55% of the increase in Medicaid enrollment between 1997 and 2000. These increases were concentrated among Hispanics and Asians.

Imagine a state that decides it will provide Medicaid information in a foreign language, but only if some critical mass of immigrants either resides there or is potentially eligible for Medicaid. As a result, it would appear in the data that immigrants in high-concentration states would be more likely to receive Medicaid, even though the mechanism is not information-sharing but a “bureaucratic channel” (Bertrand et al., 2000). We do not, however, think this story about bureaucracy completely describes our findings. Our model uses PUMA fixed effects, which will pick up any state-level differences in policy. Moreover, not all immigrant waves have historically been drawn to the same cities or states.

We search for a possible bureaucratic response as distinct from the language network effect. While we have no way of measuring the volume of state-provided information, we do want to capture the availability of Medicaid information to non-English speakers. In some states, Medicaid agencies provide on their websites enrollment information and application forms in languages other than English. Although websites with English-language portals may not be the primary source of information for immigrants with limited language proficiency, our rationale is that the availability of Internet-based materials is correlated with the overall availability of foreign-language materials. Given that a state has already created a form or pamphlet in, for example, Vietnamese, the marginal cost of placing that resource online is negligible. Web links are also a convenient medium for passing on information to another person.

We visit each state's Medicaid website to check whether it includes either an application form or an informational page in a language other than English (see Table 2.7).[18] The vast majority of states do not provide any information in a language other than English or Spanish. A total of 30 states provide some sort of Spanish-language information, while 9 states have material in Vietnamese, the next most common language. Multilingual outliers are Washington state (12 languages besides English), California (10) and Kansas (9). In Table 2.9, we modify our baseline specification to control for the effect of having information in one's own language. We find a positive and significant coefficient for the effect of Internet-based information. We then interact the network measure with the own-language information dummy and find a positive effect.

In general, immigrants whose state provides Internet resources in their native tongue are more likely to enroll in Medicaid. While the results of our website survey may not be sufficient to disentangle the mechanism underlying the network effect, they do suggest that publishing information in the languages of the immigrant community might have positive effects on participation both directly and through the network's multiplier effect.

2.5.4 Contact Availability Based on Country of Origin

While a common language makes communication easier, immigrants may associate with people who speak their language because they also share cultural norms. For example, perhaps what matters for the Senegalese is not the local presence of other francophones but of other people from Senegal. In the case of Medicaid, relevant cultural norms include attitudes toward the welfare state and stigma related to government handouts. In times of need, immigrants from some cultures may be more inclined to turn to friends or family,

[18] All websites were visited on November 30, 2010.

rather than to the state.

We cannot definitively determine the importance of language-based networks relative to cultural networks, but we can test a model with a different network. We use data from the same 2008-09 ACS to create “ethnic” networks that are instead defined by country of origin. We exclude from the ACS all immigrants whose native countries are represented by fewer than 50 observations in our final sample and those from countries where English is spoken by a majority of the population.[19]

Our econometric specification remains the same as our main model, except we define contact availability using country of origin instead of language spoken at home, and we replace language fixed effects with country-of-origin fixed effects.[20] With these modifications, shown in Columns 1-3 of Table 2.10, the network effects are still positive and significant. Because our baseline results use language-based networks, we cannot directly compare the coefficient to the country-of-origin findings, but the effect is of a similar magnitude. A 1 percentage point direct increase in enrollment would raise participation within the average country-of-origin group by a total 1.22 percentage points.

If the network effect were driven by bureaucratic response alone we would not see any marginal effect when we consider country-of-origin *within* a given language group because most policies are implemented on the basis of linguistic differences and not ethnic differences. In Column 2, we show that country-based network effects exist when we limit the analysis to Spanish speakers. This result suggests that the network effect is some combination of information and culture within the language groups.

[19]We take our list of majority-English countries from the United Kingdom Border Agency's regulations for non-European Union migrants exempt from language examinations and add to it the U.K. and Ireland. The list is available at http://www.ukvisas.gov.uk/en/aboutus/features/englishlanguagetestforpartners.

[20]Our results here are not sensitive to our choice of fixed effects.

2.5.5 Identification by *Ex Ante* Characteristics

Another solution to the reflection problem, besides the instrumental variable used in Bertrand et al. (2000) and Evans et al. (1992), is to use *exante* peer characteristics (Angrist and Pischke, 2008). In our case, this means using past Medicaid enrollment levels as a proxy for "Medicaid culture" (Bertrand et al., 2000). While past Medicaid enrollment will capture information about cultural attitudes toward public assistance, it should be uncorrelated with shocks specific to group behavior in 2008-09.

While *exante* peer characteristics from past ACS data would be ideal, the ACS began recording health-insurance status only in 2008. The Current Population Survey asked about health insurance beginning in 1994. We use the CPS numbers on past average enrollment, $\overline{Medicaid_{k1994}}$, to capture the culture of each country-based network k. The 1994 CPS is a suitable proxy because it is recent enough to be correlated with 2008-09 enrollment numbers yet old enough to be pre-welfare reform and uncorrelated with unobservable contemporary shocks.[21]

We find that the network effect with the past-Medicaid proxy is positive and significant both among all immigrants and within the Spanish-speaking population, while there is no significant effect when we restrict the sample to languages other than Spanish (Table 2.10, Columns 4-6).

2.6 Conclusion

In this paper, we provide fresh evidence that network effects contribute to disparities in Medicaid enrollment among immigrants. The American Community Survey allows

[21] The correlation between ethnic groups' 2008-09 Medicaid enrollment and the 1994 level is 0.44.

us to investigate the role of ethnic networks across a broad population and to verify that network effects exist for insurance and not just for assistance programs that provide immediate, tangible benefits. With modifications to the instrumental variables strategy from Bertrand et al. (2000), we control for the effects of potential differential selection while also addressing the reflection problem.

We find that language networks are most influential for those who are less likely to obtain Medicaid information from sources outside the network. Alternate specifications in our sensitivity analysis support the hypothesis that the network effect within an ethnic group comes from information-sharing. This finding is supported by the fact that we find smaller language-network effects among second-generation immigrants and those fluent in English.

Our results suggest that future policies to expand Medicaid enrollment among immigrants will have a larger marginal effect for groups with stronger networks and higher contact availability. At the same time, variation in network density and cultural differences also represent a source of disparities in insurance coverage among immigrants. As migration patterns change (Massey, 2008) and as new waves of immigrants increasingly locate in states that do not provide Medicaid to newcomers and that have not historically been major immigrant destinations, these disparities could worsen.

Encouragingly, we find that states' efforts to provide multilingual Medicaid information are associated with greater participation. We believe further efforts to reduce language barriers, clarify eligibility criteria, and streamline the application process will be needed if policy-makers' goal is to minimize the number of people who remain uninsured. While prioritizing outreach toward large immigrant enclaves may be the most obvious policy initiative, the multiplier effect we detect among communities of shared language suggests that people from small, relatively dispersed ethnicities may be at the greatest risk for being uninsured yet Medicaid-eligible.

Table 2.9: Availability of Foreign-Language Medicaid Information Online

	(1)	(2)	(3)
$CA \times \overline{Medicaid}$	0.186***	0.184***	0.152***
	(0.027)	(0.027)	(0.030)
website in own language		0.020**	0.018*
		(0.009)	(0.009)
$CA \times \overline{Medicaid} \times$ website in own language			0.086**
			(0.042)
Adjusted R^2	0.189	0.189	0.189
Observations	118,803	118,803	118,803

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes - Data source: 2008 and 2009 American Community Survey. The sample and control variables are the same as in the baseline specification (Column 1 of Table 2.3). Standard errors are clustered at the PUMA level. We checked each state's Medicaid website in December 2010. A state is counted as having foreign-language information if its website contains a page or a downloadable pamphlet or application in a language other than English.

Table 2.10: Network Effects by Country of Origin

	1	2	3	4	5	6
Sample	Overall	Spanish Speakers	Other Languages	Overall	Spanish Speakers	Other Languages
CA$\times\overline{Medicaid_k}$	0.125***	0.112***	0.104***			
	(0.017)	(0.023)	(0.032)			
CA$\times\overline{Medicaid_{k1994}}$				0.067***	0.093***	0.025
				(0.015)	(0.020)	(0.020)
Adjusted R^2	0.199	0.196	0.252	0.198	0.196	0.215
Observations	115,696	82,290	33,406	112,067	82,276	29,791

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes - Data source: 2008 and 2009 American Community Survey and 1994 Current Population Survey. The variable $\overline{Medicaid_{k1994}}$ is calculated from CPS data on Medicaid enrollment and country of origin; all other variables come from the ACS. The sample is composed of immigrants from non-English-speaking countries, ages 19 to 64 who are below the 200% of the FPL and come from a country-of-origin with more than 100 observations. The overall sample in Column 1 contains 62 country-of-origin "ethnic" groups.

Table 2.11: Summary Results for First-Stage Regression

Instrumented Variables:	CA	CA$\times\overline{Medicaid}$
First-stage F (2, 1218)	485.00	152.05
Angrist-Pischke F (1, 1218)	969.99	298.44
Shea's R^2	0.24	0.32
Angrist-Pischke R^2	0.23	0.31

Notes - Data Source: 2008 and 2009 American Community Survey. These are the first-stage summary statistics for the IV estimate reported in Table 3.

Table 2.12: Alternative Measures of Network Density and Probit Estimation

Network Measure	(1) OLS	(2) OLS	(3) Probit
$log(C_{jkt}) \times \overline{Medicaid_{kt}}$	0.197***		
	(0.026)		
$log(\frac{C_{jkt}}{A_{jt}}) \times \overline{Medicaid_{kt}}$		0.151***	
		(0.024)	
$CA_{jkt} \times \overline{Medicaid_{kt}}$			0.280**
			(0.121)
PUMA F.E.	YES	YES	NO
Language group F.E.	YES	YES	YES
Year F.E.	YES	YES	YES
Observations	118,803	118,803	118,803

Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Notes - Data Source: 2008 and 2009 American Community Survey. The sample and control variables are the same as in the baseline specification (Column 1 of Table 2.3). Standard errors are clustered at the PUMA level.

Chapter 3

Food Environment, Maternal Weight Gain and Pregnancy Outcomes

3.1 Introduction

More than one third of the U.S. adult population are obese (35.7%).[1] Obesity is associated with higher risks of heart disease, stroke, type II diabetes and certain types of cancer. Despite the increasing attention to obesity in the academic and public arena, not much is known about its causes and remedies.

Fast-food restaurants and food-deserts have been blamed to be important factors of the obesity epidemic. While there is consensus that fast-food is less healthy and can importantly contribute to obesity (Rosenheck, 2008), it is less clear whether exposure to fast-food restaurants and food deserts affect health. A few studies have attempted to analyze the causal effects of proximity to fast-food restaurants, but different strategies led to different conclusions on the magnitude and the significance of the effects on weight

[1] Source: NCHS Data Brief, January 2012, `http://www.cdc.gov/obesity/data/adult.html`

gain and obesity rates (Anderson and Matsa, 2011; Currie et al., 2010; Lhila, 2011). The debate on the effects of fast-food exposure remains open. Furthermore, less is known about the role of different types of restaurants. Given the evidence on the healthier dietary habits of first-generation Mexicans (Vargas, 2012; Guendelman and Abrams, 1995) and on the unhealthy assimilation in weight gain and obesity among Hispanics (Antecol and Bedard, 2006), it is natural to ask whether proximity to ethnic restaurants (eg. Mexican restaurants) might provide a more traditional and healthier alternative to American-style fast-food restaurants.

Using survey data Duerksen et al. (2007) show that child and parent body mass index (BMI) are lowest in Mexican-American families who select Mexican restaurants. Hanni et al. (2009) found that taquerias in Salinas (CA) were more likely to offer non-fried carbohydrate offerings and fruit and vegetables. In this direction, a few community intervention programs in low-income neighbourhoods (e.g. "Salud Tiene Sabor" and "Steps to a Healthier Salinas") have focused on helping Mexican restaurants to promote their healthy menus and to further reduce their fat, while increasing fruit and vegetable availability (Hanni et al., 2009).[2] However, to the best of my knowledge there are no studies analyzing the effects of exposure to Mexican-restaurants on weight gain and obesity.

This paper contributes to previous studies by analyzing the relationship between proximity to different types of restaurants (fast-food, traditional Mexican restaurants etc.) and a broad set of maternal and child's health outcomes. In particular, I focus on excessive maternal weight gain which has been linked to post-partum obesity and adverse health outcomes (Derbyshire, 2009). Focusing on pregnant women allows me to exploit the large

[2]See also `http://www.salud-america.org/sites/www.salud-america.org`

sample of administrative records drawn from the Vital Statistics of Florida to use mother fixed effects. This allows me to account for time invariant individual characteristics that might be correlated with both residential location and health outcomes. Similarly to Currie et al. (2010), I use data on the exact geographic location of restaurants and analyze how the availability of fast-food and ethnically defined restaurants is related to maternal and children health outcomes. The restaurant data are merged with information drawn from the Vital Statistics data for the metropolitan areas of Miami, covering the universe of births occurring in this area between 1990 and 2009.

The results of this paper suggest that proximity to Mexican restaurants is associated with a lower likelihood of excessive weight gain among US born mothers during pregnancy. Effects appear to be larger among black, low-skilled and young mothers, though these differences are not precisely estimated. No protective effect of Mexican restaurants was found for the foreign-born. Fast-food proximity has no significant effects on maternal weight gain. Proximity to fast-food or Mexican restaurant are not significantly associated with other maternal and child's health outcomes. While the results on fast-food proximity are consistent with the non-significant effects found by Anderson and Matsa (2011), the differences with respect to the small but significant effects of fast-food proximity found by Currie et al. (2010) can be explained by the demographic differences of the samples studied. While my analysis is restricted to the metropolitan area of Miami, Currie et al. (2010) use Vital Statistics for the entire states of Michigan, New Jersey, and Texas.

3.2 Food-environment, maternal and child's health outcomes

Several papers have examined the causes of the obesity epidemic in the US. Changes in agricultural, food production and distribution technology have been shown to have played an important role in the observed rise in obesity rates (Lakdawalla and Philipson, 2002; Philipson and Posner, 2003; Cutler et al., 2003; Courtemanche and Carden, 2011). At the same time, there is evidence that the decline in the relative cost of eating out versus home has favored the increase in adult obesity (Chou et al., 2004).

Despite the evidence on negative health outcomes associated with fast food and eating out, there is debate on whether variations in the exposure to fast-food (or other types of restaurants) or food-deserts affect health. Powell and Bao (2009) and Beydoun et al. (2008) show that the supply of supermarkets, restaurants as well as the grocery prices are importantly related to healthier behaviors and better health outcomes. The causal interpretation of these correlations has been questioned by Anderson and Matsa (2011) who exploit interstate highways as an instrument for the availability of restaurants and found no evidence of significant effects on obesity. Currie et al. (2010) find different results using a large sample and more precise measures of proximity to fast-food restaurants than the one used by Anderson and Matsa (2011) and providing evidence of modest, but significant effects on maternal weight gain. In line with these results, Lhila (2011) using the Natality Detail Data provides evidence of a positive association of greater access to fast-food restaurants with excessive weight gain during pregnancy, but no significant effects on birth-outcomes. This paper contributes to this previous set of studies by examining a broader set of health outcomes and analyzing the role of different types of restaurants.

3.3 Data and Empirical Specification

3.3.1 Data

The main data used in this paper are drawn from Vital Statistics Natality Data from Florida and the National Establishment Time Series Database (NETS, Dun and Bradstreet). Specifically, the data on maternal and child's outcomes are drawn from the birth certificates of all children born in the metropolitan area of Miami between 1990 and 2009. I obtained confidential information on mother and child's name, exact date of birth, and full address of residence and use this information to link births to the same mother.

The NETS dataset provides time-series information on establishment mobility patterns, sales growth performance, job creation and destruction, changes in primary markets, and historical D&B ratings. I obtained a panel of virtually all the establishments in SIC 58 ("Eating and Drinking Places") from 1990 to 2009, with addresses, names and categorical classification allowing me to identify different types of restaurants and their exact geographical location. In addition, the NETS data contain information on the primary standard industrial classification of each establishment. Either through letter surveys, phone surveys or internet updates, the establishment chooses (or Dun and Bradstreet assigns) its primary (and secondary) SICs from a list of over 18,750 SIC8s developed by Dun and Bradstreet.

In the baseline specification, I considered only the top 10 fast-food chains as fast-food restaurants. The fast-food list includes McDonalds, Subway, Burger King, Taco Bell, Pizza Hut, Little Caesars, KFC, Wendy's, Domino's Pizza, and Jack in the Box. Using this classification, I define as Mexican all the restaurants that were in SIC 58120112: "Mexican restaurants".[3] Other restaurant include all the restaurants not classified as fast-food or Mexican.

Using *ArcInfo*, I merged these data with the information drawn from the universe of birth records of Florida. In particular, following Currie et al. (2010), I matched the data on weight gain during pregnancy and birth outcomes of the child with the proximity to fast-food, Mexican or other type of restaurant in the year that overlaps the most with the gestation period.

3.3.2 Summary Statistics

Table 3.1 shows the summary statistics for the main variables used in the analysis. Following Currie et al. (2010) and using data on restaurant and mother's address of residence I constructed indicators for whether there were fast food, Mexican or other restaurants located within .5 miles of the mother's residence. Column 2 reports the same statistics for the sample of mothers with at least 2 pregnancies observed in the sample, while in columns 3(4), I restrict the sample to mothers who reside within .5 mile from a fast-food (Mexican) restaurant. About 25% of pregnant mothers in the sample live within 0.5 miles of a fast-food restaurants and only 5% within a .5 miles from a Mexican restaurant.

Mothers who live in proximity of a fast-food tend to be less educated, are more likely to be Hispanic, and less likely to be married. Mothers who live in proximity of Mexican restaurants are less likely to be black, more likely to be Hispanic, are relatively older and less likely to smoke.

[3]As alternative classification for fast-food restaurants, I considered all the establishment in the fast-food 8digit SIC.

3.3.3 Empirical Specification

Following Currie et al. (2010), the baseline specification is:

$$Y_{izt} = \beta_1 F5_{izt} + \beta_2 MX5_{izt} + \beta_3 Other5_{izt} + \delta X_{izt} + Z_{zt} + d_i + \epsilon_{izt} \tag{3.1}$$

where Y_{it} is a health outcome of mother (or of the child) i at time t in zip-code z; $F5$ is an indicator equal to one if there is a fast-food within 0.5 miles of the mother's address of residence, $MX5$ is an indicator equal to one if there is a Mexican restaurant within 0.5 miles of the mother's address of residence, and $Other5$ is an indicator equal to one if other type of restaurants are available within 0.5 miles of the mother's address of residence. X_{izt} is a vector of time-varying maternal characteristics including age dummies, four dummies for education (high-school drop-out, high-school graduate, some college, college or more), tobacco consumption, child's gender, parity, marital status and year dummies, dummies for race and ethnicity; Z_{zt} is a set of time varying zip-code characteristics including share of high-school drop-outs, high-school, college graduate, and those with more than college degree, share of Hispanics and Blacks, share of Cuban, Puerto Rican and Mexican mothers, share of female population, income per capita, income per capita among Hispanics, poverty rate; d_i is a mother fixed effect. Standard errors are clustered by mother. In alternative specifications, I include zip-code fixed effects. As the variation in restaurants' supply across different pregnancies could be induced by changes in the local food environment or by mothers' relocation, I consider an alternative model focusing on mothers who kept living in the same place, therefore limiting the source of variation to openings and closings of different types of restaurants. Using within-mother analysis allows to control for individual unobservables that might affect both her own locational choices and the likelihood of negative health outcomes.

3.4 Main results

Table 3.2 analyzes the relationship between food-environment and excessive maternal weight gain. Following previous studies, I use as dependent variable a dummy taking value equal to 1 if weight gain is above 20 kg.[4] In Panel A, I estimate equation 3.1, in Panel B I add zip-code fixed effects, while in Panel C I focus on those mothers who did not change zip-code across different pregnancies. Panel A, columns 1-6, presents the results obtained using the baseline specification by nativity and ethnicity. When analyzing all the mothers with at least two pregnancies in the sample, I find no evidence of significant effects of proximity to fast-food or Mexican restaurants (column 1). [5] It is important to notice that the differences in sample size and in the populations analyzed can explain the differences with respect to the findings of Currie et al. (2010). [6]

Interestingly, when focusing on US-born mothers, the availability of a Mexican restaurant within .5 miles is associated with a 2.3 percentage points reduction in the likelihood of gaining excessive weight during the pregnancy (column 2). This corresponds to a 14.3% effect with respect to the average weight gain in the sample. The result is non-significantly different for US-born Hispanics and non-Hispanics. Yet, foreign born are not affected by the proximity to different type of restaurants. This finding could be explained by the fact

[4]The incidence of low APGAR scores is shown to increase significantly with weight gain above 20 kg (Currie et al., 2010)

[5]While OLS estimates confirm a positive correlation between fast-food exposure and maternal weight gain (see Table 3.9), when including socio-demographic controls the coefficient shrinks and becomes non-significant. Among US -mothers the positive correlation remains robust to the addition of individual socio-demographic controls and zip-code time varying characteristics, but becomes non-significant once including zip-code (mother) fixed effects (see Table 3.9).

[6]When using alternative specifications examining availability of fast-food restaurant within .25 miles, I find evidence of a positive and significant with excessive weight gain.

that first-generation Hispanics are less likely to eat out and have healthier diets (Vargas, 2012; Guendelman and Abrams, 1995).

Panel B and C show that when including zip-code fixed effects (Panel B) or focusing on stayers (Panel C) the coefficient on the availability of Mexican restaurant within .5 miles is always negative except that for the foreign-born mothers. While the coefficient remains statistically significant only for non-Hispanics natives, its magnitude is fairly stable across all the samples.

In Table 3.3, I focus on US-born individuals and show how the proximity to different types of restaurants has different effects on excessive maternal weight gain across different socio-demographic groups. The coefficient on the proximity to Mexican restaurants is larger for black, low-educated and younger mothers.[7] These differences are not precisely estimated, however, these results suggest that proximity to Mexican restaurants is associated with a lower likelihood of excessive weight gain among US born mothers, with stronger effects among groups who are at higher risk of unhealthy behaviors: low-educated, young and black women.

3.4.1 Other health outcomes

Table 3.4 reports the estimates of equation 3.1 for other maternal outcomes on the sample of US-born mothers. In particular, I consider different measures of maternal weight gain (weight gain larger than 15 kg and weight gain in kg), and other maternal health outcomes (pre-pregnancy obesity (BMI>30) and overweight status (BMI>25), hypertension, and diabetes). Column 3 in Table 3.4 suggests that the availability of a Mexican restaurant within 0.5 mile decreases the weight gain of mothers during pregnancy by about 150 grams,

[7] Results are similar when including zip-code fixed effects, or focusing on stayers.

equivalent to 0.55 grams per day. Using CDC estimates this corresponds to a 4.28 less calories per day, about 1,155 less calories over the pregnancy.

While most of the reported coefficients are not significant, there is some evidence that proximity to Mexican restaurant is associated with higher likelihood of gestational hypertension. However, the coefficients reported in columns 4-7 should be interpreted with caution, as information on pre-pregnancy BMI, hypertension and diabetes is only available 2004 onwards, reducing substantially the sample of pregnancies for mothers with at least two singleton births occurring between 2004 and 2009.

Similarly, Table 3.5 shows no evidence of significant effects of proximity to fast-food, or Mexican restaurants on different metrics of fitness at birth : birth weight in grams, incidence of low birth weight (birth weight below 2,500 grams), excessive birth weight (above 4 or 4.5 kg), low APGAR (APGAR < 8), and APGAR score.

3.4.2 Robustness checks

In Table 3.6-3.7, I focus on the sample of mothers born in Florida between 1971 and 1985 for whom I was able to link the information available at their time of birth to that one available on the birth certificates of their children born in Florida between 1989 and 2009. Similarly to Giuntella (2013), I also linked the records of all the infants born between 1989 and 2009, whose mother was born Florida between 1971 and 1985, to the birth records of their own 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. This allows me to analyze the role of changes in the availability of different types of restaurants on maternal and child's health outcomes, while controlling for family fixed effects and for the health at birth of the mothers. In practice, I include grand-mother (Panel A) or mother

(Panel B) fixed effects when analyzing the role proximity to different type of restaurants on maternal and child's health. Table 3.6 confirms the association between proximity to fast-food restaurants and excessive weight gain and the non significant effects on the other maternal outcomes. In Table 3.7, results show a negative association between proximity to Mexican restaurant and the incidence of low birth weight. However, there is no evidence of significant effects on other metrics of fitness at birth. Results are similar when using grand-mother or mother fixed effects.

In Table 3.8, I conducted a placebo test analyzing the relationship between the availability of different types of restaurants and time-varying individual characteristics, while controlling for mother fixed effects. I examine maternal smoking, alcohol consumption, marital status, an indicator for being younger than 20 years old during pregnancy, and quality of care. I found no evidence of significant correlation of these individual time varying characteristics with the availability of different types of restaurants, conditional on mother fixed effects. These results suggest that the coefficients reported in Tables 3.2-3.7 are not confounded by individual characteristics that might be correlated with both maternal health and proximity to different types of restaurants.

3.5 Conclusion

Exposure to fast-food restaurants and food deserts are often blamed to have importantly contributed to the rise in obesity rates that we have observed over the last decades. However, despite the voluminous evidence on the excessive calorie intake associated with fast-food, previous studies found only negligible or non-significant causal effects of proximity to fast-food restaurants on obesity and weight gain, while not much is known about the role of different types of restaurants.

This paper extends previous studies by analyzing the role of Mexican restaurants and examining a broad set of maternal and pregnancy outcomes. I exploit changes in restaurants' availability within half a mile from the mother's address of residence across births of the same mother. I do not find evidence of significant effects of fast-food exposure on excessive maternal weight gain, once controlling for mother fixed effects. However, mothers living within half a mile from a Mexican restaurant are less likely to gain excessive weight during pregnancy. Results suggest that this relationship might be stronger among those at higher risk of unhealthy behaviors and characterized by low socio-economic status: low-educated, young and black women. The protective effect of Mexican restaurant exposure is on average equivalent to a 4.28 less calories per day, about 1,155 less calories over the pregnancy. Interestingly, foreign-born are not affected by proximity to fast-food or Mexican restaurants. This might be explained by their healthier diets and by the fact that they are less likely to eat out. There is no evidence of significant effects on other maternal and child's health outcomes.

Mexican restaurants may provide healthy options for low-income neighbourhoods and that community based interventions to help ethnic restaurants promote their healthy menus might have non-negligible effects. However, more research is needed to pin-down the exact mechanisms underlying these results and their external validity.

Table 3.1: Summary Statistics, Florida Birth Records (1989-2009), Miami CBSA

	All births	Siblings	Siblings, Fast Food rest. w/i .5 mi	Siblings, Mexican rest. w/i .5 mi
Mother-year observations	560,928	302,631	74,926	14,448
Demographic characteristics				
Age of mother	27.897	27.383	27.228	27.841
Mother high school graduate	0.372	0.364	0.387	0.367
Mother attended some college	0.191	0.188	0.175	0.150
Mother at least college graduate	0.244	0.250	0.228	0.260
Black Mother	0.282	0.304	0.255	0.111
Hispanic mother	0.562	0.547	0.625	0.758
Smoking mother	0.023	0.019	0.017	0.013
Male child	0.488	0.487	0.487	0.487
Parity	0.987	1.155	1.108	1.097
Married Mother	0.580	0.593	0.572	0.585
Outcome variables				
Weight gain >20 kg	0.148	0.145	0.148	0.149
Weight gain >15kg	0.397	0.392	0.396	0.396
Weight gain (kg)	14.199	14.108	14.162	14.155
Fast Food rest. w/i .5 mi	0.252	0.248	1.00	0.670
Mexican rest. w/i .5 mi	0.051	0.047	0.128	1.00
Other rest. w/i .5 mi	0.518	0 .520	0.98	0.999

Notes - There are 135,068 mothers with greater than or equal two children in the sample. There are 94 zip-codes. There are 39,873 mothers who experience a change in fast-food availability within 0.5 miles, and 10,274 mothers who experienced a change in Mexican-restaurant availability within 0.5 miles.

Table 3.2: Food Environment and Excessive Weight Gain (larger than 20 kg), Florida Birth Records(1989-2009), Miami CBSA

	Sample: All mothers, mother fixed effects				
	(1)	(2)	(3)	(4)	(5)
Demographic Sub-samples:	Overall	US Born	US Born Hispanics	US Born Non-Hispanics	Foreign Born
Fast food rest. w/i .5 mi	-0.0013	-0.0006	-0.0055	0.0014	-0.0014
	(0.003)	(0.004)	(0.008)	(0.005)	(0.003)
Mexican rest. w/i .5 mi	-0.0064	-0.0234***	-0.0212*	-0.0210**	0.0038
	(0.004)	(0.007)	(0.012)	(0.011)	(0.005)
Other rest. w/i .5 mi	0.0043*	0.0076**	0.0090	0.0061	0.0010
	(0.002)	(0.003)	(0.007)	(0.004)	(0.003)
Mother Fixed Effects	YES	YES	YES	YES	YES
Average of dependent variable	0.136	0.149	0.161	0.142	0.125
Observations	282,462	132,434	43,176	89,258	150,028

	Sample: All mothers, mother and zip-code fixed effects				
	(1)	(2)	(3)	(4)	(5)
Demographic Sub-samples:	Overall	US Born	US Born Hispanics	US Born Non-Hispanics	Foreign Born
Fast food rest. w/i .5 mi	-0.0033	-0.0028	-0.0048	-0.0018	-0.0035
	(0.003)	(0.004)	(0.008)	(0.005)	(0.003)
Mexican rest. w/i .5 mi	-0.0074*	-0.0229***	-0.0180	-0.0221**	0.0020
	(0.004)	(0.007)	(0.012)	(0.011)	(0.005)
Other rest. w/i .5 mi	0.0017	0.0051	0.0040	0.0040	-0.0012
	(0.002)	(0.003)	(0.007)	(0.004)	(0.003)
Mother Fixed Effects	YES	YES	YES	YES	YES
zip-code Fixed Effects	YES	YES	YES	YES	YES
Average of dependent variable	0.136	0.149	0.161	0.142	0.125
Observations	282,462	132,434	43,176	89,258	150,028

	Sample: Stayers, mother fixed effects				
	(1)	(2)	(3)	(4)	(5)
Demographic Sub-samples:	Overall	US Born	US Born Hispanics	US Born Non-Hispanics	Foreign Born
Fast food rest. w/i .5 mi	-0.0012	-0.0023	0.0083	-0.0056	-0.0001
	(0.004)	(0.006)	(0.012)	(0.007)	(0.005)
Mexican rest. w/i .5 mi	-0.0081	-0.0262**	-0.0099	-0.0245*	0.0024
	(0.006)	(0.010)	(0.018)	(0.014)	(0.008)
Other rest. w/i .5 mi	0.0022	0.0037	-0.0070	0.0051	0.0017
	(0.003)	(0.005)	(0.009)	(0.006)	(0.004)
Mother Fixed Effects	YES	YES	YES	YES	YES
Average of dependent variable	0.124	0.139	0.148	0.135	0.118
Observations	164,835	80,408	25,054	55,354	84,427

Notes - The unit of observation is a pregnancy for mothers with at least two singleton births. The dependent variable is excessive maternal weight gain, defined as a dummy variable which equals 1 when the weight-gain during pregnancy is above 20 kg. Entries in rows 1, 2, and 3 are the coefficient on a dummy for the existence of respectively a fast-food, a Mexican, or a different type of restaurant within 0.5 miles from mother's residence. All estimates include controls for time varying mother's characteristics including child's gender, mother's education (4 groups), mother's age dummies, parity, marital status, child's year of birth dummies. Regression also included a set of controls at the zip-code level: share of adults (over 25) with high school degree or less, some college, college or more, share of female, black, white and Hispanic population, share of Cuban, Mexican and Puerto Rican mothers, poverty rate, per capita income and per capita in come in the Hispanic population. The zip-code level data are drawn from the 1990 US Census for all pregnancies that occurred before 2000 and from the 2000 US Census for the pregnancies that occurred since year 2000. Other controls are indicators for missing variables. Standard errors clustered by mother.

Table 3.3: Food environment and excessive weight gain (larger than 20 kg), demographic subgroups, US-Born

	(1) Overall	(2) Blacks	(3) Whites	(4) Low-Educ	(5) High-Educ	(6) < 25	(7) 25-35	(8) 35-45
Fast-food restaurant	-0.0009	-0.0032	0.0008	-0.0060	0.0049	-0.0025	0.0020	0.0126
	(0.004)	(0.006)	(0.005)	(0.005)	(0.006)	(0.006)	(0.007)	(0.018)
Mexican restaurant	-0.0234***	-0.0305**	-0.0183**	-0.0270***	-0.0182	-0.0308***	-0.0227*	0.0298
	(0.007)	(0.015)	(0.008)	(0.010)	(0.011)	(0.012)	(0.012)	(0.028)
Other restaurant	0.0078**	0.0058	0.0065	0.0089*	0.0052	0.0107**	0.0113**	0.0005
	(0.003)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.015)
Mother F.E.	YES	YES	YES	YES	YES	YES	YES	YES
Observations	132,434	54,999	77,435	73,260	59,174	64,702	65,510	12,740

Notes - The unit of observation is a pregnancy for mothers with at least two singleton births. The dependent variable is excessive maternal weight gain, defined as a dummy variable which equals 1 when the weight-gain during pregnancy is above 20 kg. Entries in rows 1, 2, and 3 are the coefficient on a dummy for the existence of respectively a fast-food, a Mexican, or a different type of restaurant within 0.5 miles from mother's residence. All estimates include controls for time varying mother's characteristics including child's gender, mother's education (4 groups), mother's age dummies, parity, marital status, child's year of birth dummies. Regression also included a set of controls at the zip-code level: share of adults (over 25) with high school degree or less, some college, college or more, share of female, black, white and Hispanic population, share of Cuban, Mexican and Puerto Rican mothers, poverty rate, per capita income and per capita in come in the Hispanic population. The zip-code level data are drawn from the 1990 US Census for all pregnancies that occurred before 2000 and from the 2000 US Census for the pregnancies that occurred since year 2000. Other controls are indicators for missing variables. Standard errors clustered by mother.

Table 3.4: Other maternal outcomes and food environment, US-Born

Dependent variable:	(1) Weight gain >20 kg	(2) Weight gain >15kg	(3) Weight gain (kg)	(4) Obese	(5) Overweight	(6) Gestational hypertension	(7) Diabetes
Fast-food rest. w/i .5 mi	-0.0009	-0.0009	-0.0286	0.0018	0.0086	0.0048	-0.0004
	(0.004)	(0.005)	(0.063)	(0.008)	(0.009)	(0.004)	(0.001)
Mexican rest. w/i .5 mi	-0.0234***	-0.0001	-0.1578	0.0103	-0.0190	0.0125*	0.0021
	(0.007)	(0.010)	(0.119)	(0.011)	(0.013)	(0.007)	(0.003)
Other rest. w/i .5 mi	0.0078**	0.0098**	0.0737	-0.0125*	-0.0112	0.0010	0.0009
	(0.003)	(0.004)	(0.054)	(0.007)	(0.008)	(0.004)	(0.001)
Mother F.E.	YES	YES	YES	YES	YES	YES	YES
Observations	132,434	132,434	132,434	39,148	39,148	44,722	44,722

Notes - The unit of observation is a pregnancy for mothers with at least two singleton births. Entries in rows 1, 2, and 3 are the coefficient on a dummy for the existence of respectively a fast-food, a Mexican, or a different type of restaurant within 0.5 miles from mother's residence. All estimates include controls for time varying mother's characteristics including child's gender, mother's education (4 groups), mother's age dummies, parity, marital status, child's year of birth dummies. Regression also included a set of controls at the zip-code level: share of adults (over 25) with high school degree or less, some college, college or more, share of female, black, white and Hispanic population, share of Cuban, Mexican and Puerto Rican mothers, poverty rate, per capita income and per capita in come in the Hispanic population. The zip-code level data are drawn from the 1990 US Census for all pregnancies that occurred before 2000 and from the 2000 US Census for the pregnancies that occurred since year 2000. Other controls are indicators for missing variables. Standard errors clustered by mother.

Table 3.5: Birth outcomes and food environment, US-born

	(1) Birth weight (gm)	(2) Birth weight < 2500 gm	(3) Birth weight >4000 gm	(4) Birth weight >4500 gm	(5) Excessive/low birth weight	(6) APGAR< 8	(7) APGAR
Fast-food rest. w/i .5 mi	-0.7150	0.0014	-0.0020	-0.0012	0.0002	-0.0026	-0.0229
	(5.117)	(0.003)	(0.002)	(0.001)	(0.003)	(0.002)	(0.046)
Mexican rest. w/i .5 mi	-8.7048	-0.0067	0.0006	0.0002	-0.0065	0.0031	0.0080
	(9.486)	(0.005)	(0.005)	(0.002)	(0.005)	(0.003)	(0.070)
Other rest. w/i .5 mi	0.3891	0.0004	0.0010	0.0001	0.0005	0.0016	0.0516
	(4.325)	(0.002)	(0.002)	(0.001)	(0.003)	(0.001)	(0.043)
Mother F.E.	YES	YES	YES	YES	YES	YES	YES
Observations	145,006	145,006	145,006	145,006	145,006	144,714	144,996

Notes - The unit of observation is a pregnancy for mothers with at least two singleton births. Entries in rows 1, 2, and 3 are the coefficient on a dummy for the existence of respectively a fast-food, a Mexican, or a different type of restaurant within 0.5 miles from mother's residence. All estimates include controls for time varying mother's characteristics including child's gender, mother's education (4 groups), mother's age dummies, parity, marital status, child's year of birth dummies. Regression also included a set of controls at the zip-code level: share of adults (over 25) with high school degree or less, some college, college or more, share of female, black, white and Hispanic population, share of Cuban, Mexican and Puerto Rican mothers, poverty rate, per capita income and per capita in come in the Hispanic population. The zip-code level data are drawn from the 1990 US Census for all pregnancies that occurred before 2000 and from the 2000 US Census for the pregnancies that occurred since year 2000. Other controls are indicators for missing variables. Standard errors clustered by mother.

Table 3.6: Other maternal outcomes and food environment, Florida-Born, linked-sample

Dependent variable:	(1) Weight gain >20 kg	(2) Weight gain >15kg	(3) Weight gain (kg)	(4) Obese	(5) Overweight	(6) Gestational hypertension	(7) Diabetes
Panel A:							
Fast-food rest. w/i .5 mi	0.00142	-0.00627	-0.109	-0.000935	0.0152	0.00869*	0.000130
	(0.00536)	(0.00714)	(0.0955)	(0.0115)	(0.0134)	(0.00514)	(0.00202)
Mexican rest. w/i .5 mi	-0.0275***	-0.0175	-0.136	0.0136	-0.0345	0.00647	0.00286
	(0.0104)	(0.0141)	(0.186)	(0.0176)	(0.0211)	(0.00921)	(0.00465)
Other rest. w/i .5 mi	0.0121**	0.0214***	0.190**	-0.00810	-0.0143	-0.00450	0.000578
	(0.00477)	(0.00638)	(0.0842)	(0.0103)	(0.0118)	(0.00472)	(0.00166)
Grandmother F.E.	YES	YES	YES	YES	YES	YES	YES
Observations	61,889	61,889	61,938	21,383	21,383	24,512	24,512
Panel B:							
Fast-food rest. w/i .5 mi	-0.000433	-0.0101	-0.124	-0.00365	0.00447	0.00364	-0.00125
	(0.00546)	(0.00726)	(0.0966)	(0.0111)	(0.0127)	(0.00555)	(0.00199)
Mexican rest. w/i .5 mi	-0.0355***	-0.0165	-0.187	0.0202	-0.0276	0.00777	0.00392
	(0.0109)	(0.0143)	(0.190)	(0.0164)	(0.0197)	(0.00964)	(0.00514)
Other rest. w/i .5 mi	0.0116**	0.0251***	0.217***	-0.00898	-0.0145	-0.00303	0.000238
	(0.00485)	(0.00645)	(0.0839)	(0.00974)	(0.0110)	(0.00504)	(0.00191)
Mother F.E.	YES	YES	YES	YES	YES	YES	YES
Observations	60,861	60,861	60,910	21,171	21,171	24,266	24,266

Notes - The unit of observation is a pregnancy for mothers with at least two singleton births. Entries in rows 1, 2, and 3 are the coefficient on a dummy for the existence of respectively a fast-food, a Mexican, or a different type of restaurant within 0.5 miles from mother's residence. All estimates include controls for time varying mother's characteristics including child's gender, mother's education (4 groups), mother's age dummies, parity, marital status, child's year of birth dummies. Regression also included a set of controls at the zip-code level: share of adults (over 25) with high school degree or less, some college, college or more, share of female, black, white and Hispanic population, share of Cuban, Mexican and Puerto Rican mothers, poverty rate, per capita income and per capita in come in the Hispanic population. The zip-code level data are drawn from the 1990 US Census for all pregnancies that occurred before 2000 and from the 2000 US Census for the pregnancies that occurred since year 2000. Other controls are indicators for missing variables. Standard errors clustered by mother.

Table 3.7: Birth outcomes and food environment, Florida-born, linked sample

	(1) Birth weight (gm)	(2) Birth weight < 2500 gm	(3) Birth weight >4000 gm	(4) Birth weight >4500 gm	(5) Excessive/low birth weight	(6) APGAR< 8	(7) APGAR
Panel A:							
Fast-food rest. w/i .5 mi	-2.114	0.00278	-0.00122	-0.000675	0.00217	-0.00214	0.0112
	(2.216)	(0.00403)	(0.00162)	(0.00124)	(0.00421)	(0.00232)	(0.01000)
Mexican rest. w/i .5 mi	-3.776	-0.0150**	0.00209	0.00314	-0.0122	0.00446	-0.0136
	(4.237)	(0.00713)	(0.00316)	(0.00239)	(0.00752)	(0.00402)	(0.0180)
Other rest. w/i .5 mi	1.907	-0.00152	0.00134	0.000220	-0.00132	1.80e-05	3.18e-05
	(1.775)	(0.00352)	(0.00144)	(0.00113)	(0.00369)	(0.00202)	(0.00851)
Grandmother F.E.	YES	YES	YES	YES	YES	YES	YES
Observations	67,725	67,712	67,712	67,712	67,725	67,597	67,597
Panel B:							
Fast-food rest. w/i .5 mi	-2.037	0.00522	-0.00161	-0.000533	0.00483	-0.00179	0.00797
	(1.594)	(0.00404)	(0.00145)	(0.00125)	(0.00423)	(0.00233)	(0.0100)
Mexican rest. w/i .5 mi	1.367	-0.0154**	0.00147	0.00263	-0.0133*	0.00520	-0.0152
	(2.953)	(0.00717)	(0.00250)	(0.00234)	(0.00757)	(0.00390)	(0.0171)
Other rest. w/i .5 mi	2.108	-0.00191	0.00179	0.000597	-0.00137	0.000564	-0.000131
	(1.403)	(0.00353)	(0.00128)	(0.00110)	(0.00369)	(0.00206)	(0.00874)
Mother F.E.	YES	YES	YES	YES	YES	YES	YES
Observations	67,720	67,707	67,707	67,707	67,720	67,592	67,592

Notes - The unit of observation is a pregnancy for mothers born in Florida between 1971 and 1985 with at least two singleton births. Entries in rows 1, 2, and 3 are the coefficient on a dummy for the existence of respectively a fast-food, a Mexican, or a different type of restaurant within 0.5 miles from mother's residence. All estimates include controls for time varying mother's characteristics including child's gender, mother's education (4 groups), mother's age dummies, parity, marital status, child's year of birth dummies. Regression also included a set of controls at the zip-code level: share of adults (over 25) with high school degree or less, some college, college or more, share of female, black, white and hispanic population, share of Cuban, Mexican and Puerto Rican mothers, poverty rate, per capita income and per capita in come in the Hispanic population. The zip-code level data are drawn from the 1990 US Census for all pregnancies that occurred before 2000 and from the 2000 US Census for the pregnancies that occurred since year 2000. Other controls are indicators for missing variables. Standard errors clustered by mother.

Table 3.8: Placebo tests: food-environment and other determinants of weight gain

	(1) Smoking during pregnancy	(2) Alcohol during pregnancy	(3) Married mother	(4) Teenage mother	(5) Adequate prenatal care
Panel A: Overall Sample					
Fast-food rest. w/i .5 mi	-0.000201	0.000327	-0.000654	-0.00138	-0.00320
	(0.000853)	(0.000354)	(0.00229)	(0.00183)	(0.00252)
Mexican rest. w/i .5 mi	0.00183	0.000168	-0.00422	0.00368	0.00150
	(0.00138)	(0.000540)	(0.00401)	(0.00297)	(0.00432)
Other rest. w/i .5 mi	0.000120	-0.000900***	0.00348*	-0.00151	-0.00446*
	(0.000776)	(0.000326)	(0.00201)	(0.00160)	(0.00228)
Observations	305,943	305,533	305,943	305,943	300,038
Panel B: US-born Sample					
Fast-food rest. w/i .5 mi	-3.72e-05	-3.22e-05	-0.000507	-0.00202	-0.00513
	(0.00146)	(0.000651)	(0.00326)	(0.00304)	(0.00377)
Mexican rest. w/i .5 mi	0.00224	0.000251	-0.00687	-0.00193	0.00437
	(0.00282)	(0.00111)	(0.00619)	(0.00559)	(0.00683)
Other rest. w/i .5 mi	-0.000414	-0.000915*	-5.53e-05	7.64e-05	-0.00842**
	(0.00130)	(0.000551)	(0.00276)	(0.00259)	(0.00334)
Observations	145,073	144,862	145,073	145,073	142,003

Notes - The unit of observation is a pregnancy for mothers with at least two singleton births. Entries in rows 1, 2, and 3 are the coefficient on a dummy for the existence of respectively a fast-food, a Mexican, or a different type of restaurant within 0.5 miles from mother's residence. All estimates include controls for time varying mother's characteristics including child's gender, mother's education (4 groups), mother's age dummies, parity, marital status, child's year of birth dummies. Regression also included a set of controls at the zip-code level: share of adults (over 25) with high school degree or less, some college, college or more, share of female, black, white and hispanic population, share of Cuban, Mexican and Puerto Rican mothers, poverty rate, per capita income and per capita in come in the Hispanic population. The zip-code level data are drawn from the 1990 US Census for all pregnancies that occurred before 2000 and from the 2000 US Census for the pregnancies that occurred since year 2000. Other controls are indicators for missing variables. Standard errors clustered by mother.

Table 3.9: Excessive Weight Gain and Food-Environment, Sensitivity to the Addition of Controls

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Overall Sample						
Fast-food rest. w/i .5 mi	0.006***	0.001	0.002	0.002	-0.001	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Mexican rest. w/i .5 mi	0.001	-0.003	-0.003	-0.005	-0.007	-0.007*
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Other rest. w/i .5 mi	-0.001	-0.000	-0.002	0.000	0.004*	0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Observations	282,462	282,462	282,462	282,462	282,462	282,462
Panel B: US-born Sample						
Fast-food rest. w/i .5 mi	0.008***	0.004	0.005*	0.003	-0.001	-0.003
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Mexican rest. w/i .5 mi	-0.002	-0.009*	-0.009	-0.012	-0.023***	-0.023***
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)
Other rest. w/i .5 mi	-0.001	0.002	0.000	0.002	0.008**	0.005
	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)
Observations	132,434	132,434	132,434	132,434	132,434	132,434
Standard socio-demographic controls	NO	YES	YES	YES	YES	YES
Time varying Zipcode characteristics			YES	YES	YES	YES
Zipcode F.E				YES	NO	YES
Mother F.E.					YES	YES

Notes - The unit of observation is a pregnancy for mothers with at least two singleton births. Entries in rows 1, 2, and 3 are the coefficient on a dummy for the existence of respectively a fast-food, a Mexican, or a different type of restaurant within 0.5 miles from mother's residence. Standard socio-demographic controls include controls for time varying mother's characteristics including child's gender, mother's education (4 groups), mother's age dummies, parity, marital status, child's year of birth dummies. Time-varying zip-code characteristics include share of adults (over 25) with high school degree or less, some college, college or more, share of female, black, white and hispanic population, share of Cuban, Mexican and Puerto Rican mothers, poverty rate, per capita income and per capita in come in the Hispanic population. The zip-code level data are drawn from the 1990 US Census for all pregnancies that occurred before 2000 and from the 2000 US Census for the pregnancies that occurred since year 2000. Other controls are indicators for missing variables. Standard errors clustered by mother.

References

Acevedo-Garcia, D., Barbeau, E., Bishop, J. A., Pan, J., and Emmons, K. (2004). Undoing an epidemiological paradox: The tobacco industry's targeting of US industry. *American Journal of Public Health*, 14(12):2188–2193.

Acevedo-Garcia, D., Pan, J., Jun, H.-J., Osypuk, T. L., and Emmons, K. M. (2005). The effect of immigrant generation on smoking. *Social Science and Medicine*, 61:1223–1242.

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

Aizer, A. (2007). Public health insurance, program take-up, and child health. *The Review of Economics and Statistics*, 89(3):400–415.

Aizer, A. and Currie, J. (2004). Networks or neighborhoods? Correlation in the use of publicly-funded maternity care in California. *Journal of Public Economics*, 88:2573–2585.

Almond, D., Chai, K. Y., and Lee, D. S. (2005). The cost of low birth weight. *Quarterly Journal of Economics*, 120(3):1031–1083.

Almond, D., Hoynes, H. W., and Schanzenbach, D. W. (2009). Inside the war on poverty:

The impact of food stamps on birth outcomes. *Review of Economics and Statistics*, XCIII(2):387–403.

Anderson, M. L. and Matsa, D. A. (2011). Are restaurants really supersizing America? *American Economic Journal: Applied Economics*, 3(1):152–188.

Andersson, F., Garcia-Perez, M., Haltiwanger, J. C., McCue, K., and Sanders, S. (2010). *Workplace Concentration of Immigrants*. National Bureau of Economic Research working paper.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press, Princeton, NJ.

Antecol, H. and Bedard, K. (2006). Unhealthy assimilation: Why do immigrants converge to American health status levels? *Demography*, 43(2):337–360.

Aslund, O. and Fredricksson, P. (2005). Ethnic enclaves and welfare culture: Quasi-experimental evidence. *CESifo Working Paper Series*, 115.

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

Bates, L. M. and Teitler, J. O. (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. `http://ideas.repec.org/p/pri/crcwel/1085.html`. Accessed Nov. 2012.

Baumeister, L., Marchi, K., Pearl, M., Williams, R., and Braveman, P. (2000). The validity of information on "race" and "Hispanic ethnicity" in California birth certificate data. *Health Services Research*, 35(4):869–883.

Bertrand, M., Luttmer, E. F. P., and Mullainathan, S. (2000). Network effects and welfare cultures. *The Quarterly Journal of Economics*, 115(3):1019–1055.

Beydoun, M. A., Powell, L. M., Wang, Y., et al. (2008). The association of fast food, fruit and vegetable prices with dietary intakes among US adults: is there modification by family income? *Social science & medicine (1982)*, 66(11):2218.

Borjas, G. (1995). Ethnicity, neighborhoods and human capital externalities. *American Economic Review*, 85(3):364–390.

Borjas, G. (1999). Immigration and welfare magnets. *Journal of Labor Economics*, 17(4):607–37.

Borjas, G. (2002). Welfare reform and immigrant participation in welfare programs. *International Migration Review*, 36(4):1093–1123.

Borjas, G. J. (1987). Self-selection and the earnings of immigrants. 77(4):531–553.

Broder, T. and Blazer, J. (2009). *Overview of Immigrant Eligibility for Federal Programs*. National Immigration Law Center.

Buekens, P., Canfield, C., Padilla, N., Lara Lona, E., and Lozano, R. (2012). Low Birthweight in Mexico: A systematic review. *Maternal and Child Health Journal*, pages 1–7.

Card, D. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics*, 19(1):22–64.

Chiquiar, D. and Hanson, G. (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, B. R., Lee, Y. L., and Miller, P. W. (2008). Immigrant selection system and immigrant health. *Contemporary Economic Policy*, 26(4):1074–3529.

Chou, S.-Y., Grossman, M., and Saffer, H. (2004). An economic analysis of adult obesity: results from the Behavioral Risk Factor Surveillance System. *Journal of health economics*, 23(3):565–587.

Classen, T. 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, D. and Bennett, N. G. (2000). Is biology destiny? Birth weight and life chances. *American Sociological Review*, 65(3):458–476.

Courtemanche, C. and Carden, A. (2011). Supersizing supercenters? the impact of walmart supercenters on body mass index and obesity. *Journal of Urban Economics*, 69(2):165–181.

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

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

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

Currie, J., Vigna, S. D., Moretti, E., and Pathania, V. (2010). The effect of fast food restaurant on obesity and weight gain. *American Economic Journal: Applied Economics*, 2(3):32–63.

Cutler, D., Glaeser, E., and Shapiro, J. (2003). Why have Americans become more obese? *Journal of Economic Perspectives*, 17(3).

Derbyshire, E. (2009). Implication of high maternal weight gain during pregnancy. *Nursing Standard*, 22(49).

Deri, C. (2005). Social networks, information and health care utilization. *Journal of Health Economics*, 24(6):1076–1107.

Devillanova, C. (2008). Social networks, information and health care utilization: Evidence from undocumented immigrants in Milan. *Journal of Health Economics*, 27(2):265–286.

Duerksen, S. C., Elder, J. P., Arredondo, E. M., Ayala, G. X., Slymen, D. J., Cambell, N. R., and Baquero, B. (2007). Family restaurant choices are associated with child and adult overweight status in Mexican-American families. *Journal of the American Dietetic Association*, 107:849–853.

Duncan, B. and Trejo, S. (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.

Edin, P., Fredrikson, P., and Aslund, O. (2003). Ethnic enclaves and the economic success of immigrants: Evidence from a natural experiment. *The Quarterly Journal of Economics*, 118(1):329–357.

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

Evans, W. N., Oates, W. E., and Schwab, R. M. (1992). Measuring peer group effects: A study of teenage behavior. *Journal of Political Economy*, 119(C):966–991.

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

Figlio, D. N., Hamersma, S., and Roth, J. (2011). Information shocks and social networks. (16930). National Bureau of Economic Research working paper.

Finch, B. K., Lim, N., Perez, W., and Do, D. P. (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., Stampfer, M., and Curhan, G. (2009). Diet and lifestyle risk factors associated with incident hypertension in women. *Journal of the American Medical Association*, 302(4):401–411.

Fryer, R. G. and Levitt, S. D. (2004). The causes and consequences of distinctively black names. *Quarterly Journal of Economics*, 119(3):767–805.

Furtado, D. and Theodoropoulos, N. (2009). Intermarriage and immigrant employment the role of networks. *The B.E. Journal of Economic Analysis and Policy*, 10(1).

Furtado, D. and Trejo, S. (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. `http://ideas.repec.org/p/crm/wpaper/1205.html`. Accessed Nov. 2012.

Giuntella, O. (2013). Why does the health of immigrants deteriorate. *Working Paper, Boston University.*

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

Guendelman, S. and Abrams, B. (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, M. P., Frisbie, W. P., DeTurk, P., and Blanchard, K. S. (1998). Dating the origins of the epidemiologic paradox among Mexican Americans. *Texas Population Research Center Papers*, (97). `http://www.utexas.edu/cola/centers/prc/_files/pdf/workingpapers/97-98-07.pdf`. Accessed Nov. 2012.

Hanni, K. D., Garcia, E., Ellemberg, C., and Winkleby, M. (2009). Targeting the taqueria: Implementing healthy food options at Mexican American Restaurants. *Health Promotion Practice*, 10(2).

Hummer, R. A., Powers, D. A., Pullum, S. G., Gossman, G. L., and Frisbie, W. P. (2007).

Paradox found (again): Infant mortality among the Mexican-origin population in the United States. *Demography*, 44(3):441–457.

Ibarraran, P. and Lubotsky, D. (2007). Mexican Immigration and Self Selection: New Evidence from the 2000 Mexican Census. In Borjas, G. J., editor, *Mexican Immigration to the United States*. University of Chicago Press, Chicago, IL.

Jackson, C. and Scheider, H. (2011). Do social connections reduce moral hazard? Evidence from the New York City taxi industry. *American Economic Journal: Applied Economics*, 3(3):244–267.

Jasso, G., Massey, D. S., Rosenzweig, M. R., and Smith, J. P. (2004). Immigrant health: Selectivity and acculturation. *The Institute for Fiscal Studies Working Paper*, 4(23):1–46.

Kaestner, R. and Malamud, O. (2010). Self-selection and international migration: New evidence from mexico. *NBER*, (15765). `http://www.nber.org/papers/w15765.pdf`. Accessed Nov. 2012.

Kaiser, L. L. and Allen, L. (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.

Kaiser Family Foundation (2008). *Summary: Five basic facts on immigrants and their health care*. Kaiser Commission on Medicaid and the Uninsured.

Kaiser Family Foundation (2010). *Medicaid: A Primer*. Kaiser Commission on Medicaid and the Uninsured.

Kandula, N. R., Grogan, C. M., Rathouz, P. J., and Lauderdale, D. S. (2004). Unintended impact of welfare reform on the Medicaid enrollment of eligible immigrants. *HSR: Health Sciences Research*, 39(5):1509–1526.

Kaushal, N. (2008). Adversities of acculturation? Prevalence of obesity among immigrants. *Health Economics*, 18:291–303.

Kaushal, N. and Kaestner, R. (2005). Immigrant and native responses to welfare reform. *Journal of Population Economics*, 18(1):69–92.

Koenker, R. and Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives*, 15(4):143–156.

Lakdawalla, D. and Philipson, T. (2002). The growth of obesity and technological change: a theoretical and empirical examination. Technical report, National Bureau of Economic Research.

Lavarreda, S. A., Brown, E. R., Yoon, J., and Glenn, S. (2006). *More than half of California's uninsured children eligible for public programs but not enrolled*. UCLA Center for Health Policy Research.

Lhila, A. (2011). Does access to fast food lead to super-sized pregnant women and whopper babies? *Economics and Human Biology*, 9(4):364–380.

Long, S. K., Phadera, L., and Lynch, V. (2010). *Massachusetts health reform in 2008: Who are the remaining uninsured adults?* State Health Access Reform Evaluation, Robert Wood Johnson Foundation.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, 60:531–542.

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

Massey, D. S. (2008). *New faces in new places: The changing geography of American immigration*. The Russel Sage Foundation, New York, NY.

Mazzolari, F. (2004). Effects of welfare reform on immigrants' welfare participation: How chilling are they? *Mimeo*.

McKenzie, D. and Rapoport, H. (2010). Self selection patterns in mexico-u.s. migration: The role of migration networks. *The Review of Economics and Statistics*, 92:811–821.

Moraga, J. F.-H. (2011). New evidence on emigrant selection. *The Review of Economics and Statistics*, 93(1):72–96.

Munshi, K. (2004). Networks in the modern economy: Mexican migrants in the U.S. labor market. *Quarterly Journal of Economics*, 118(2):549–599.

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

Osypuk, T. L., Bates, L., and Acevedo-Garcia, D. (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, A. and Arias, E. (2004). Paradox lost: Explaining the Hispanic adult mortality advantage. *Demography*, 41(3):385–415.

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

Palloni, M. C. A. (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.

Philipson, T. J. and Posner, R. A. (2003). The long-run growth in obesity as a function of technological change. *Perspectives in biology and medicine*, 46(3):S87–S107.

Ponce, N. A., Ku, L., Cunningham, W. E., and Brown, E. R. (2006). Language barriers to health care access among Medicare beneficiaries. *Inquiry*, 43(1):66–67.

Powell, L. M. and Bao, Y. (2009). Food prices, access to food outlets and child weight. *Economics and Human Biology*, 7(1):64–72.

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

Rosenheck, R. (2008). Fast food consumption and increased caloric intake: a systemating review of trajectory toward weight gain and obesity risk. *Obesity Reviews*.

Ross, D. C. and Horn, A. (2008). *Health Coverage for Children and Families in Medicaid and SCHIP: State efforts face new hurdles.* Kaiser Commission on Medicaid and the Uninsured.

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

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

Ruggles, S., Alexander, J. T., Genadek, K., Goeken, R., Schroeder, M. B., and Sobek, M. (2010). *Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]*. University of Minnesota, Minneapolis.

Shaw, R. J., Pickett, K. E., and Wilkinson, R. G. (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, M. and Lelia, D. (2006). Exercise or other physical activity for preventing pre-eclampsia and its complications. *Cochrane Database of Systematic Reviews*, (3):1233–1241.

Sommers, B. D. and Epstein, A. M. (2011). Why states are so miffed about Medicaid: Economics, politics, and the 'woodwork effect'. *New England Journal of Medicine*, 356(2):100–102.

Turban, S. (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.

Vargas, A. J. (2012). Eating habits among Mexicans immigrants to the Unites States: A time use perspective. *Mimeo, Texas Tech University.*

Wang, W. (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.

Watson, T. (2010). Inside the refrigerator: Immigration enforcement and chilling effects in Medicaid participation. (16278). National Bureau of Economic Research working paper.

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

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

Curriculum Vitae

GIOVANNI OSEA GIUNTELLA

Education

Ph.D., Economics, Boston University, 2013

Dissertation Title: *Three essays on the economics of immigration and health*

Committee: Daniele Paserman, Randall Ellis, Kevin Lang, Claudia Olivetti

Master in Economics (MEI), Quantitative Methods, University of Rome Tor Vergata, Italy, November 2007

M.Sc., Economics (with distinction and special mention for excellence), University of Rome Tor Vergata, Italy, October 2006

B.A., Economics (with distinction), University of Rome Tor Vergata, Italy, October 2004

Fields of Interest Health Economics, Labor Economics, Economics of Immigration

Teaching Experience

Instructor, Topics in Modern Labor Economics, Department of Economics, Boston University, Spring 2013, Fall 2012, Summer 2010 Head Teaching Fellow, Department of Economics, Boston University, Spring 2011

Teaching Fellow, Principles of Microeconomics, Department of Economics, Boston University, Fall 2010

Teaching Fellow, Principles of Macroeconomics, Department of Economics, Boston University, Spring 2010

Other Experience

Research Assistant, Boston University, Department of Economics, 2011-2012

Boston University Political Economy Reading Group, coordinator, 2011-2012

Mentor, RA-Mentor Program, Boston University, Department of Economics 2010-2012

Research Internship at the Bank of Italy, 2009

Publication/Submitted papers

"Do Immigrants Squeeze Natives out of Bad Schedules?" IZA Journal of Migration, 1:7, 2012

"Medicaid and Ethnic Networks." The B.E. Journal of Economic Analysis & Policy: Vol. 11: Iss. 1 (Contributions), Article 77, 2011. (with Emily R. Gee)

"The Effects of Age and Job Protection on the Welfare Costs of Inflation and Unemployment: a Source of ECB Anti-Inflation Bias?" European Journal of Political Economy, 26: 137-146, 2010. (with Leonardo Becchetti and Stefano Castriota)

Working Papers

Why Does the Health of Immigrants Deteriorate? (Job Market Paper)

Nuns and Vatican II: Reassessing the Evidence on the Effect of Catholic Schools (with Rania Gihleb)

Food environment, maternal and pregnancy outcomes

Fellowships and Awards

Summer Research Grant, Boston University, Summer 2011

Student Research Award, Institute for Economic Development, Boston University, 2011

Teaching-RA Fellowship, Boston University, Spring 2010-Spring 2012

Marco Fanno Scholarship, Medio Credito Centrale, Fall 2007-Spring 2009

Raeli Scholarship, University of Rome Tor Vergata, Spring 2007

Student Career Scholarship, University of Rome Tor Vergata, 2001-2006

Computer Skills: ARCGIS, ABLE2EXTRACT, LaTeX, MATLAB, Office, Rapid-Weaver, SAS, STATA

Citizenship: Italian