# Does Legal Status Affect Educational Attainment in Immigrant Families?

Zachary Liscow* William Gui Woolston

October 7, 2017

Preliminary - Please do not circulate or cite without permission.

**Abstract**

Of the estimated 11.1 million unauthorized immigrants in the U.S., 1.1 million are children. Due to differential treatment in the labor market, teenage undocumented immigrants face low returns to schooling. To measure the effect of legal status on the educational choices of Hispanic teenagers, we compare siblings who differ in their legal status due to their birth country. We find teenagers who were born in Mexico are 2.6 percentage points more likely to be out of school than their U.S. born siblings. Alternative explanations, such as differences in prenatal or childhood environment, appear largely unable to explain this result, suggesting that legal status has a significant impact on schooling decisions. Accounting for these alternative explanations to the extent possible and using proxies for legal status in U.S. census, we show that being undocumented reduces the number of years of high school by between 0.13 and 0.17 years. Back-of-the-envelope estimates suggest that providing legal status increases an undocumented worker's lifetime wages by $8,455 and his lifetime net government fiscal contribution by $4,257, with estimated aggregate impact across all undocumented Hispanics over 75 years of $20 billion in increased earnings and $9 billion in increased fiscal contribution.

**Keywords:** Demand for schooling, human capital, immigration
**JEL CLASSIFICATION:** I21, I28, J61, F22.

*Zachary Liscow, corresponding author: Associate Professor, Yale Law School, zachary.liscow@yale.edu. William Woolston, Nuna Health: william.woolston@gmail.com. This paper has benefited enormously from helpful comments from Ran Abramitzky, Leah Platt Boustan, David Card, Mark Duggan, Michael Greenstone, Caroline Hoxby, Ilyana Kuziemko, Max Schanzenbach, Luke Stein, and two anonymous referees. Daniel Giraldo provided excellent research assistance.

# 1 Introduction

> *Well, when I realized I was undocumented around like 15, like, that's when it really hit me. . . . And I was going like, what is the point of me even trying in school? What is the point of me doing anything if I'm not going to be able to have a career or be able to, I guess, be normal?*
>
> - Undocumented immigrant[1]

There are an estimated 1.1 million undocumented immigrants in the United States who are children (Passel and Cohn 2010). This paper seeks to shed light on the educational choices of these individuals and to explore the hypothesis that lack of legal status may discourage them from finishing high school. Due to differential treatment in the labor market and barriers to attending college, teenagers who are undocumented immigrants likely face lower net returns to attending high school. While being undocumented may not discourage work in industries where low-skilled immigrants are heavily represented, such as landscaping or agriculture (Cortes 2008), legal status may prevent work in highly skilled professions. Furthermore, the risk of deportation may decrease the returns to country-specific human capital investment (Chiswick 1984; Chiswick 1988). A third reason why undocumented immigrants may be more likely to leave high school is that undocumented immigrants may have higher costs, broadly construed, of attending college. If attending college is too costly, individuals who might otherwise finish high school to retain the option of going to college may instead choose to drop out (Chin and Juhn 2007).

Determining the effect of legal status on the educational choices of teenagers is important for at least three reasons. First, measuring this effect can inform optimal policy toward immigrants who currently live in the U.S. There is heated debate surrounding policies, such as the DREAM Act, that would give some undocumented teenagers legal status. If current laws decrease the net returns to education for undocumented immigrants, teenage undocumented immigrants may acquire less schooling and compete more directly in the labor market with low-skilled and poor natives. In addition, if current policies decrease the incentives for undocumented immigrants to invest in school, the government may recover less in taxes from undocumented immigrants

[1] Martin, M. (2017) "DACA, A Student's Story: 'They are The Types of Immigrants You Want in Your Country,"' *NPR*. Interview Transcript. http://www.npr.org/templates/transcript/transcript.php?storyId=551544757

and spend more on the children of these undocumented immigrants (Bouvier and Gardner 1986). Second, the lack of legal status may be one possible explanation for the low educational attainment of Hispanic adults, a fifth of whom are estimated to be undocumented.[2]

Third, the expected educational outcomes of this population are likely to influence the number and type of individuals who choose to move to the U.S. without authorization. Just as individuals are likely to compare the returns to their education and skills in their home and potential host countries (Borjas 1987; Roy 1951), parents are likely to take into account the prospects for their children when deciding to immigrate without authorization. If parents understand that U.S. policy discourages their non-U.S.-born children from finishing high school, parents who believe that the returns to education are likely to be high for their children may be less likely to immigrate. Moreover, the legal treatment of immigrants' foreign-born children can influence the speed of their economic assimilation.

Empirically assessing how legal status influences educational choices is complicated by the fact that legal status is not randomly assigned, with a bias that could go in either direction. On the one hand, being undocumented may be correlated with factors that increase the chance that an individual will drop out of high school, such as attending lower-performing schools or having parents who do not speak English. On the other hand, parents who immigrate may be especially motivated to ensure that their children succeed in school. Moreover, throughout much of the 1990s, some undocumented immigrants were able take steps (such as finding a sponsoring employer) that would allow both the individual and her family to become legal. Therefore, legal status may be a choice on the part of an individual or her family and hence may be endogenous. Although others have used policy changes like the 1986 Immigration Reform and Control Act, which conferred amnesty on many undocumented immigrants, to study college enrollment rates (Cortes 2013) and Deferred Action for Childhood Arrivals ("DACA"), President Obama's administrative grant of documentation on some individuals (Pope 2016; Amuedo-Dorantes and Antman 2017), the state-of-the-art for estimating the effect of citizenship status versus undocumented status on high school outcomes is measuring this correlation between legal status and school attainment (Diaz-Strong and Ybarra 2016; Bean, Leach, Brown, Bachmeier,

[2]In the U.S., 41 percent of Hispanic adults aged 20 and older do not have a regular high school diploma, versus 23 percent for Blacks and 14 percent of Whites (Fry 2010).

and Hipp 2011; Greenman and Hall 2013), which suffers from these endogeneity problems.

To help address the potential endogeneity of legal status across families, this paper exploits differences in legal status across siblings that arise from their different countries of their births. Since the end of the 19th century, the United States has extended citizenship (and hence legal status) to all individuals who are born in the U.S., even if their families moved without authorization. This paper compares the educational attainment of siblings who are close in age but differ in their legal status because one sibling was born in and the other outside of the U.S. This within-family analysis helps to ensure that family-level differences, such as parental or local residential characteristics, are not confounding the results.

To understand the identification strategy, consider two families, each with two children aged 15 and 16. In family A, the family moved from Mexico 15.5 years before the data was collected. The older sibling was born in Mexico and is an undocumented immigrant. Her younger sister was born in the U.S. and is eligible for U.S. citizenship. One strategy to measure the effect of legal status on educational enrollment would be to determine if the older sibling in family A is less likely to be enrolled in school than her younger sibling. However, typically (and in this example), if a family has siblings who differ in their country of birth, the older sibling is born outside the U.S. while the younger sibling is born in the U.S. The older sibling is therefore more likely to be an undocumented immigrant. Because older siblings are also less likely to be in school, and being older is correlated with being undocumented, failing to control for age would bias the estimated effect of being undocumented away from zero.

To control for age effects, this paper also studies individuals in otherwise similar families where the siblings do not differ in their countries of birth. Family B, for example, also has children ages 15 and 16. Unlike family A, family B moved from Mexico 14 years before the data was collected: both children are born in Mexico and are undocumented immigrants. To account for age effects, this paper compares the educational outcomes of the siblings in family A with those from family B.

Using this within-family comparison, this paper establishes two main facts. First, Mexican-born teenagers, who are far more likely to be noncitizens or undocumented than their U.S.-born siblings, are also approximately 2.7 percentage points less likely to be enrolled in school than their U.S.-born siblings.

Second, while differences in legal status are only one of many possible differences between the U.S.- and Mexican-born siblings, we consider four main alternative explanations for the pattern in school attendance. They appear largely unable to explain the pattern. First, we address the concern that the results arise from older siblings being more likely to be born abroad in two ways: first, by controlling for age fixed effects and, second, by focusing on atypical birth-order families in which the younger sibling was born abroad, which does not reduce the results. Second, we see what the data can say about the impact of potentially different childhood environments. We use several tests for indications that childhood environment alone could drive the results, including seeing whether Mexican-born children are more likely to be held back and whether controlling for health status changes the results. Neither test suggests that childhood impacts drive the results. More importantly, we measure whether children closer in age–and who therefore have less differential exposure to different childhood environments–exhibit different rates of school attendance. We show that this factor has at most a modest impact on the results. Third, controlling for English ability does not change the results. A fourth factor, pre-natal environment, is more difficult to test for. Although there is some suggestive evidence that they may not matter much (e.g., on the absence of differential rates of being held back in school, or impacts of health or English skills), we cannot rule out this possibility. We thus use estimates from the literature to approximate how much prenatal environments contribute, and subtract that from our main effect. Since this prenatal estimate is an order of magnitude smaller than our estimated effect, what remains is still a sizable impact from the difference in legal status between the U.S.- and Mexican-born siblings on being in school, at about 0.13 lost years of schooling, or about 86% of the unadjusted effect of being born in Mexico.

To scale the relationship between being born in Mexico, legal status, and school attendance, we embed our estimates in an instrumental variables framework with family fixed effects. Using proxies for legal status from the U.S. Census and American Community Surveys from 2000 – 2012, these results suggest that being undocumented makes a teenager of Mexican heritage 2.30 to 2.68 percentage points less likely to be in school, therefore accounting for more than three quarters of the cross-sectional differences in school attendance rates between undocumented and documented teenagers of Mexican heritage. These estimates show that about one-half of the baseline level of not attending school for this population of undocumented Mexican

immigrants can be attributed to lack of documentation. We produce conservative back-of-the-envelope calculations to estimate the size of these impacts on lifetime wages and net government fiscal impact. We estimate that documentation increases and individual's lifetime earnings by $8,456and increases his net fiscal contribution by $4,257in present discounted value. The back-of-the-envelope estimates suggest that, over the next 75 years, documentation increase earnings by $20 billion in aggregate and have a positive net fiscal impact of $9 billion.

The paper unfolds as follows. Section 2 discusses background on U.S. immigrants and immigration policy. Section 3 discusses the data and presents summary statistics. Section 4 describes the empirical strategy. Sections 5 and 6 present results and consider alternative explanations, respectively. Section 7 discusses the results, and Section 8 concludes with a discussion of the implications of the results for proposed legislation like the DREAM Act.

## 2 Background on U.S. Immigrants

The children of undocumented immigrants constitute a substantial fraction of America's youth. Passel and Taylor (2010) estimate that children of undocumented immigrants represent 7 percent of the under-18 population and 8 percent of the newborn population. While roughly four-fifths of the children of undocumented immigrants are U.S.-born and hence eligible for U.S. citizenship, in 2009 there were an estimated 1.1 million children who were themselves likely undocumented immigrants (Passel and Cohn 2010).

Understanding the educational choices of this population is important because, as a whole, undocumented immigrants are much less well educated than the U.S.-born population. While only 2 percent of the U.S.-born adult population aged 25 to 64 has less than a 9th grade education, an estimated 29 percent of undocumented immigrants do. Under 10 percent of U.S.-born adults are not high-school graduates, while 48 percent of undocumented immigrants are. Perhaps partly as a consequence, undocumented immigrants are disproportionately represented in low-wage sectors and are substantially more likely to be classified as poor (Passel and Cohn 2009). Moreover, these differences in educational attainment are present even for undocumented immigrants who moved to the U.S. early in their life. For example, Passel and Cohn (2009)

estimate that 28 percent of undocumented immigrants aged 18 – 24 who arrived in the U.S. before age 14 have not completed high school, versus 18 percent for the U.S.-born population.

One potential explanation for these differences in educational outcomes is that the inability to work legally may depress undocumented immigrants' returns to schooling. Simple regressions support the view that the returns to schooling are lower for individuals who are less likely to be able to work legally in the U.S.[3] Individuals are able to work legally in the U.S. if they meet any of the following three conditions. First, all citizens are able to work legally. Second, an individual can become a legal permanent resident and obtain a work visa, although these are generally difficult for undocumented immigrants to get. A work visa allows an individual and her family to enter the country for a specified period of time. The visas are generally given to inter-company transferees, foreign investors, and workers themselves. The majority of the visas are given to individuals who engage in highly skilled work. Others, who obtain family-based visas, typically by being a family member of a citizen, may also work. Finally, an undocumented immigrant can be granted amnesty. Of course, there are many other statuses as well–for example, those on student visas who cannot work for extended periods of time. In our analysis, we compare teenagers who are noncitizens with those who are not and teenagers who are likely undocumented with those who are not. Those who are noncitizens may later gain citizenship and those who are in the subcategory of noncitizens who are undocumented may gain citizenship or lawful permanent resident status and thus be able to work with authorization. [4]

This paper exploits differences in legal status that are driven by citizenship status originating from an individual's country of birth. Since 1898, the U.S. has followed the practice of *jus soli*

[3]Following Barrow and Rouse (2005), we run regressions of the logarithm of annual earnings on years of schooling, a quadratic in potential experience, and an indicator of whether the individual was female. We also add state fixed effects and run OLS regressions on a sample of Hispanic individuals aged 25 – 65. Data is drawn from the 2000 Decennial Census and the 2001 – 2009 American Community Surveys. We further divide the sample into two groups — citizens (who are able to work legally in the USA) and noncitizens (some of whom are undocumented immigrants) and estimate the return to schooling separately for each group. To ensure that differences in quality of formal schooling between the U.S. and other countries is not driving the results, we restrict attention to individuals who were either U.S.-born or who arrived in the U.S. by age 6. For citizens, an additional year of schooling is associated with 0.119 log point increase in annual earnings. For noncitizens, the return is substantially (0.026 log points) smaller ($p < 0.01$).

[4]In recent decades, there was one major episode of amnesty: the Immigration Reform and Control Act of 1986 (IRCA). Researchers have used the passage of the Immigration Reform and Control Act of 1986 to study the effect of amnesty on a wide range of outcomes. Chiswick (1988) provides a review of the major provisions of the act. Donato, Durand, and Massey (1992) study the effect of the law on the migration patterns of Mexicans. Cobb-Clark, Shiells, and Lowell (1995) and Kossoudji and Cobb Clark (2002) estimate the effect of legalization and employer sanction on the wages.

("right of soil"), granting citizenship to all U.S.-born children, including those of immigrants.[5]

Although the children of undocumented immigrants are eligible to become citizens, it is possible that many of them may not. Their parents could be afraid to interact with government agencies or could be unaware of their children's eligibility. There are three reasons that these concerns are unlikely to invalidate the results presented here. First, to ensure that endogenous take up of citizenship is not driving the results, this paper exploits the variation in citizenship originating from the country of a child's birth rather than relying on a parent's choice to obtain citizenship for her child. Second, becoming a U.S. citizen after being born in the U.S. is very easy, even for the children of undocumented immigrants.[6] Third, given the perceived attractiveness of U.S. citizenship, parents of children born in the U.S. are likely to be highly aware of the potential for their children to become citizens.[7]

# 3 Data and Summary Statistics

This paper's primary data source draws on the 5 percent sample of the 2000 U.S. decennial Census, as well as American Community Survey data from 2001 – 2012 (Ruggles et al. 2011).[8] These data have three attractive features. First, our empirical strategy relies on families with some teenagers born in the U.S. and some born in Mexico, and such families are relatively rare;

[5]Many countries follow a practice of *jus sanguinis* ("right of blood") where citizenship is determined by ethnic ties rather than by birth. China, Germany, India, Japan, and Turkey, among many other countries, reduce the requirements to obtain citizenship for people of favored ethnic groups. Others countries, especially in the New World, practice *jus soli* ("right of soil") where citizenship is largely determined by place of birth. Examples of New World countries that practice *jus soli* include Argentina, Brazil, Canada, Mexico, and Venezuela. In many of those countries, the original goal of this immigration policy was generally to encourage immigration. While the U.S. follows the *jus soli* doctrine (in addition to *jus sanguinis*) and awards citizenship to all individuals who are born in the U.S., it does so for a different reason from that of most other countries. In 1857 U.S. Supreme Court decision *Dred Scott v. Sandford,* the Court held that people of African descent imported into the United States and held as slaves and their descendants — regardless of whether they were slaves — could never be citizens of the United States. After the North won the Civil War, the U.S. adopted a constitutional amendment that reversed the *Dred Scott* decision. The 14th amendment to the U.S. Constitution declares that "all persons born or naturalized in the United States...are citizens of the United States." In 1898, the Court's decision in *United States v. Wong Kim Ark* clarified that the 14th amendment applied to U.S.-born children of foreigners.

[6]If the birth takes place in the hospital (and by law undocumented immigrants cannot be barred from emergency rooms) or at home with a midwife, children are automatically given birth certificates. If a birth takes place unaided by professionals, obtaining a birth certificate generally requires the parent to deliver a form to the state/county vital records office. If a birth certificate is filed late ( $> 1$ year after birth), obtaining a birth certificate generally requires more work (documents showing that the mother was in state when the birth occurred, etc.) but is not especially burdensome, even for undocumented immigrants. The Social Security Administration, State Department, and other government agencies accept birth certificates as proof of citizenship.

[7]Moreover, Mexico and many other countries in Latin America also follow the doctrine of *jus soli*, so immigrants are more likely to be familiar with the legal doctrine.

[8]Because of the similarity in the structure and administration between the Census and the American Community Survey, we will refer to both as "census data."

in the baseline sample of teenagers of Mexican heritage, only 3.0 percent of individuals are in such families. Our empirical strategy therefore requires a large sample size, which the Census, unlike alternatives such as the Survey of Income and Program Participation, provides. Second, unlike highly specialized datasets such as the Children of Immigrants Longitudinal Study that only examine immigrants who live in a particular city or who immigrated at a specific time, the Census is designed to be nationally representative. Finally, the Census contains a relatively rich set of individual and family level covariates which are used to explore the robustness of the results.

There are, however, two primary disadvantages of the Census data. First, like all nationally representative datasets, the Census does not directly ask for a respondent's legal status (Chin and Juhn 2007). We will therefore rely on proxies of legal status, which introduce measurement error, a topic discussed in Section 4. Second, we know if an individual is enrolled in school and (for some years) if the student is the expected grade for her age, but the Census does not contain more detailed information on individuals' academic performance such as grades or test scores.

Using data from 2000 – 2012 strikes a balance between increasing the sample size and using only the most recent data on the other hand. We limit data to school years beginning before President Obama announced his Deferred Action on Childhood Arrivals policy in 2012, which could have separately impacted the results; given DACA's uncertain future, we wish to estimate a comparison of providing citizenship to undocumented immigrants and other noncitizens, ignoring the small movement in the direction of citizenship provided by DACA. Using data from before 2000 in the primary specification might be less informative of the effect of legal status on schooling decisions today, both because of changes in the labor markets and returns to schooling (Goldin and Katz 2007; Autor, Katz, and Kearney 2008) and because of the evolving nature of the immigrant population and immigration policy (Passel, Capps, and Fix 2004). Appendix Table 1 reports the number of observations by survey year, birthplace, and citizenship status.

We primarily focus on undocumented immigrants from Mexico because Mexico is both the largest source of immigrants, documented and undocumented, and also the large source with the highest share of undocumented immigrants. Since 1990, 34 percent of total immigrant arrivals to the U.S. have been from Mexico. The number of Mexican-born people in the U.S. equals

roughly a tenth of the population of Mexico (Hanson 2006). At the same time, the majority of undocumented immigrants are from Mexico (52%), with 25 percent coming from elsewhere in Latin America (Passel and Cohn 2016), and those of Mexican heritage have an especially high probability of being undocumented (an estimated 72% of Mexican noncitizens are undocumented vs. 59% of non-Mexican Hispanic noncitizens and 28% of non-Hispanic noncitizens).[9] Focusing on one country of origin helps us address concerns about differential prenatal and childhood environments, and focusing on Mexico, which has such a large undocumented share, helps us focus on undocumented immigrants.

The paper makes four selection criteria for the baseline sample. First, for these reasons, we consider only those of Mexican heritage. [10] In robustness checks, we consider all Hispanics, who with Mexican-born individuals constitute 77% of all undocumented immigrants (over 2/3 of which are of Mexican heritage). Second, the paper focuses on individuals who are aged 13 – 17, for whom dropping out of high school is a relevant choice.[11] Third, to ensure that observations are siblings rather than unrelated teenagers living in the same household, we consider only sons and daughters of the household head, although this sample restriction has little impact on the results. Finally, to focus on children who moved to the U.S. as children rather than as teenagers, we drop any families where any of the children arrived in the U.S. after age 12.

Table 1 presents summary statistics for the individuals in the sample by their country of birth. (Results are weighted by Census sample weights, like all results in the paper.) The baseline sample of only individuals who identify as being of Mexican heritage contains 262,721

[9]We estimated the fraction of noncitizen Mexican heritage individuals in the U.S. that were undocumented in 2009. Passel and Cohn (2016) estimated that the number of undocumented immigrants of Mexican heritage in the U.S. in 2009 were 6.35 million. From the weights in the ACS sample from 2009, we estimate that the number of Mexican heritage noncitizens in the U.S. were 8,817,936. Dividing one number by the other, we estimate that 72.0% of noncitizen Mexican heritage immigrants in the United States were undocumented in 2009. We calculated the figure for non-Mexican Hispanic noncitizens in a similar manner, with a slight wrinkle. Passel and Cohn (2016) reported estimates of undocumented immigrants from Mexico, Central America, South America and the Caribbean. The figure from the Caribbean likely includes people that do not usually identify as Hispanic, such as Jamaicans or Haitians. Without the figures from the Caribbean our estimate is that 67.7% of Hispanic noncitizens are undocumented. Finally, we calculated the figure for all non-Hispanic noncitizens by dividing the total estimated non-Hispanic undocumented immigrants in 2009 from Passel and Cohn (2016) by the total number of non-Hispanic noncitizens in 2009, which we estimated using tabulations of the 2009 ACS.

[10]In particular, the 2000 census long form asked respondents if they were "Spanish/Hispanic/Latino." This questions is asked in some way in every ACS. The baseline specification restricts attention to individuals who indicated that they were "Mexican, Mexican Am., [or] Chicano."

[11]The paper does not include 18 year olds because the Census only allows linking of siblings if they live in the same household, and individuals who are 18 or older are much more likely to leave their parent's home for reasons that may be correlated with school attendance. See Appendix Figure 1 for information on the probability that an individual is in the same household as his parent(s).

individuals. Of these, about 80 percent (217,073) were born in the U.S. About one fifth (44,019) were born in Mexico.[12] Across the three main groups (full sample, born in the U.S., and born in Mexico), the average age is close to 15, approximately half of each sample is female, and the number of siblings is broadly similar.

In addition, Table 1 explores how a sufficient condition for legal status — having U.S. citizenship — varies with country of birth. The Census, like all nationally representative data sets, does not contain direct information on an individual's legal status. As a first cut of the data, Table 1 uses a dummy variable for not having U.S. citizenship. While all citizens are able to work legally in the U.S., not all noncitizens are undocumented. In our sample, 87.9% of the Mexican-born individuals are not U.S. citizens. Mexican-born individuals are also less likely to be enrolled in school and are substantially less likely to speak English very well.

[Table 1: Summary Stats by Country of Birth]

Appendix Table 2 replicates Table 1, showing summary statistics based on proxies for legal status. Individuals who are classified as noncitizens are less likely to be enrolled in school and are substantially less likely to speak English very well. In addition, we attempt to classify individuals as undocumented (see Section 4). This method reclassifies a noncitizen as authorized if someone in her family participated in government programs such as public welfare that makes it less likely that members of the family are undocumented (though, of course, the measure is not exact). While this proxy for legal status is subject to measurement error, it likely correctly identifies some noncitizens as authorized immigrants. This definition reclassifies roughly 11% of the individuals who are labeled as undocumented. Along most observables, the two groups look similar.

Because the paper's empirical strategy relies on families with some teenagers born in the U.S. and some born in Mexico, Appendix Table 3 classifies individuals based on their birthplace and those of their siblings. The vast majority (81%) of the individuals in the baseline sample live in families in which every member of the household in the sample was born in the United States. A smaller number (15%) are from families with all members of the household included in the

[12] A small number (1,179) of individuals were born in neither the U.S. nor Mexico.

sample born in Mexico. Finally, roughly 3 percent of individuals are from families where at least one sibling was born in Mexico and at least one sibling was born in the United States ("mixed-country-of-birth families").[13] As before, individuals in these three groups have similar mean age and are approximately equally likely to be female. Along some dimensions, such as having U.S. citizenship, speaking English very well, and probability of being in school, individuals in mixed-country-of-birth families fall in between their counterparts in U.S.- or Mexican-born families.[14]

Finally, Table 2 presents summary statistics by country of birth for the 7,851 individuals living in families where some siblings in the sample are born in Mexico and some in the U.S. The fourth column of this table summarizes the results from eight separate regressions where the variable of interest is the outcome indicated by the row, and the regressors are family fixed effects and a dummy variable for being born in Mexico. Relative to their U.S.-born siblings, Mexican-born siblings are, on average, 84.2 percentage points more likely to be noncitizens and4.1 percentage points less likely to be enrolled in school. In addition, these Mexican-born siblings are approximately 1.3 years older and are 5.0 percentage points less likely to report speaking English very well. The differences in legal status and school enrollment preview the paper's results, The differences along other dimensions highlight the fact that, within a family, being born in Mexico is not randomly assigned. These summary statistics raise threats to identification: Mexican-born siblings are likely to be older and speak English less well. Section 6 explores the robustness of the paper's results in the face of these threats.

[Table 2: Summary Stats and Regression Coefficients for Mixed-Country-of-Birth Families]

[13] Note that these three groups ("all siblings born in the U.S.", "some siblings born in the U.S., some born in Mexico", and "all siblings born in Mexico") are not collectively exhaustive of the full sample. For example, an individual who is born in the U.S. and has a Canadian born sibling in the sample would not be included in any of these three groups. Similarly, individuals with all siblings born in Canada would not be included in any of these groups.

[14] One notable exception to this general pattern is with the number of siblings in the sample. All else equal, individuals are more likely to be classified in the mixed country of birth group if they have more siblings in the sample. Individuals from mixed country of birth families have, on average, approximately one more sibling in the sample than individuals in either the U.S. born or Mexican born families. That this pattern is not surprising, as individuals from families with mixed country of birth siblings will, by construction, not include any one-child families.

# 4 Empirical Framework

Interpreting a bivariate correlation between being undocumented and school enrollment as causal is misleading due to omitted variable bias: the families of undocumented and authorized immigrants are likely to differ along a number of dimensions that could influence their educational outcomes.[15] Although one could attempt to control for observable differences between undocumented immigrants and other Hispanics, such strategies are unlikely to yield consistent estimates for two reasons. First, the families, schools, and neighborhoods of the two populations are likely to differ in unobserved ways. Second, for some individuals, legal status may be endogenously determined. Under some circumstances, undocumented immigrants can, with significant difficulty and expense, obtain legal status. For example, during much of the 1990s, under Section 245(i) of the Immigration and Nationality Act, individuals who entered the U.S. without authorization but who would be eligible for immigrant visas based on their family relationships or job skills could become legal permanent residents without leaving the U.S., provided they pay a fee of several hundred dollars (Bruno 2002). This provision meant that individuals who were able to find a sponsoring employer or spouse and were willing to pay the fee are more likely to become citizens. Individuals who attempt this difficult process are likely to do so because believe that they (or their children) have high returns from becoming a legal worker.[16] Having a high return from work may be correlated with other determinants of an individual's optimal schooling choice, raising the prospect that legal status is endogenous to an individual's decision to remain in school.

The paper's empirical strategy seeks to address both empirical challenges by comparing the educational outcomes of siblings who are close in age and who share the same family and neighborhoods but differ in their legal status due to the country of their birth. To focus on within-family comparisons, the baseline results control for family fixed effects. To exploit the

[15] In particular, the parents of undocumented immigrants choose to come to the U.S., and this endogenous selection raises the possibility that parents of undocumented immigrants may differ along a range of difficult to observe dimensions: motivation, education, intelligence, and social value of education. In addition, English language and education may be complements; differences in educational attainment between undocumented immigrants and other Hispanics could be driven by differences in language ability. Finally, undocumented immigrants may live in different neighborhoods, attend different schools, go to different churches, and face differing amounts of discrimination than do authorized immigrants.

[16] Quantitatively, estimates suggest that the number of undocumented immigrants who obtained permanent resident status (a green card holder) is large; from 1998 – 2001**,** Vaughan (2003) reports that "at least 600,000 undocumented aliens have received green cards."

difference in legal status originating from a sibling's country of birth rather than the possibly endogenous legal status, the paper studies the relationship between not being enrolled in school and being born in Mexico.[17] The estimating equation for individual $i$ in family $j$ is

$$\mathbb{I}_{\text{Not in School},\, i\ \ j} = \alpha_1 + \alpha_2 \cdot \mathbb{I}_{\text{Born Mexico},\, i\ \ j} + \underbrace{\sum_{l=13}^{17} \alpha_{3l} \mathbb{I}_{\text{age}=\, l}}_{\text{Age FE}} + \underbrace{\sum_{j} \alpha_{4j} \mathbb{I}_{\text{family}\ j}}_{\text{Family FE}} + \varepsilon_{1i\ \ j}. \quad (1)$$

The impact of being born in Mexico is the focus of the paper. Nevertheless, to scale these reduced form results, we also present instrumental variables estimates where being born in Mexico is used as an instrument for legal status. This strategy is valid only if being born in Mexico influences school attainment only through an individual's legal status, a hypothesis that we explore extensively in Section 6. As such, we defer presenting those results until after Section 6. Implementing these estimates using two-stage least squares, the estimating equations are

$$\mathbb{I}_{\text{Undocumented}\ i\ \ j} = \beta_1 + \beta_2 \cdot \mathbb{I}_{\text{Born Mexico},\, i\ \ j} + \sum_{l=13}^{17} \beta_{3l} \mathbb{I}_{\text{age}=\, l} + \sum_{j} \beta_{4j} \mathbb{I}_{\text{family}\ j} + \varepsilon_{2i\ \ j} \quad (2)$$

$$\mathbb{I}_{\text{Not in School},\, i\ \ j} = \psi_1 + \psi_2 \cdot \mathbb{I}_{\widehat{\text{Undocumented}}\ i\ \ j} + \sum_{l=13}^{17} \psi_{3l} \mathbb{I}_{\text{age}=\, l} + \sum_{j} \psi_{4j} \mathbb{I}_{\text{family}\ j} + \varepsilon_{3i\ \ j} \quad (3)$$

where $\mathbb{I}_{\widehat{\text{Undocumented}}\ i\ \ j}$ are the predicted values from the first stage. The coefficients of interest, ($\alpha_2, \beta_2$, and $\psi_2$) are identified from siblings in mixed-country-of-birth families. All families in the sample identify control variables (age fixed effects and other control variables that are included in some specifications as robustness checks). To ease interpretation, all outcome variable are multiplied by 100 so that the coefficients can be interpreted as a representing percentage point changes in the outcome variables. Finally, despite the fact that the primary variable of interest (a dummy for not being in school) is binary, we estimate linear models because of the well-known incidental parameters associated with logit models with a small number of observations per fixed effect. Still, it is reassuring that when we estimate reduced form conditional logit regressions with family fixed effects, being born in Mexico has a large and highly statistically significant

[17] We focus on Mexican-born undocumented immigrants because, as discussed in Section 3, Mexican-born undocumented immigrants constitute more than half of all undocumented immigrants (Passel and Cohn, 2016), and most Mexican immigrants are undocumented (Passel, Capps, and Fix, 2004). Note, however, that results very similar if a dummy variable for being born outside the U.S. is used or the regressions are run on the subset of individuals who are born only in the U.S. or Mexico.

effect on the probability that an individual is not in school.[18]

While this identification strategy alleviates many of the concerns related to omitted variables and selection, there are three potential limitations. First, this approach does not estimate general equilibrium effects that may be associated with a change in the legal status of a large number of undocumented immigrants from, for example, passage of the DREAM Act. Second, if siblings born in different countries differ in ways other than legal status, then the estimates will reflect the direct effect of legal status on schooling and the effect of those other differences, a hypothesis that we explore in Section 6. Third, the effect estimated will be local to families with mixed-country-of-birth children; in particular, the estimates are local to families with more than one child.[19]

A second potential concern is measurement error in the Census data. There are two potential sources of measurement error. First, because the Census is conducted by the U.S. government, undocumented immigrants may be reluctant to answer truthfully questions from the Census Bureau. While non-truthful response is a concern, four facts about the Census partially alleviate concerns about non-truthful responses. First, privacy and confidentiality are major concerns of the Census Bureau. By law, the Census Bureau cannot share personal information with other government or law enforcement agencies, and the Census Bureau strives to make this fact clear to respondents.[20] Second, to increase response rate and accuracy, the Census engages in outreach to Hispanic organizations .[21] Third, because the Census does not directly ask about legal status, undocumented immigrants do not need to directly report that they are breaking the law. Fourth, the Census is largely conducted by mail, so respondents have the opportunity to talk with their neighbors and discover that responding does not endanger them. Moreover, the Census gives respondents an opportunity to obtain questionnaires in Spanish (and

[18]Specifically, in a conditional logit model version of Equation [1], the coefficient on the dummy variable for being born in Mexico is 1.42 ($p < 0.01$). Controlling for English language ability and health proxies yields very similar results.

[19]Appendix Table 4 examines the extent to which Mexican-born children in mixed country of birth families differ from the general Mexican-born population in the U.S.

[20]For example, the first substantive sentences on the 2001 American Community Survey form assures respondents that the Census Bureau cannot violate their privacy. The sentences read: "People are our most important resource. This Census Bureau survey collects information about education, employment, income, and housing — information your community uses to plan and fund programs. Your response is important, and we keep your answers confidential."

[21]The Census Bureau works with NALEO, MALDEF, LULAC, and the National Council of La Raza, faith-based leaders, and locally elected officials. Because representation in state and federal governments and appropriations are based on census, community organizations have an incentive to ensure that household heads complete the census.

other languages). In addition, this paper's fixed effects strategy will help prevent non-truthful response from biasing the estimates.[22]

A second source of measurement error is that, like all nationally representative individual level datasets, the Census does not contain direct information on an individual's legal status (Chin and Juhn 2007). Measurement error in legal status will not influence the reduced form relationship between birthplace and school enrollment, but it will influence how the reduced form results are scaled in the IV estimates. This paper uses two measures of legal status, both of which introduce measurement error if the question is being able to work in the U.S. legally on a permanent basis. In the basic specification, the variable *noncitizen* is a dummy variable for not being a U.S. citizen. Citizenship is a sufficient but not necessary condition for being able to work legally in the U.S.; while all citizens can work legally, individuals on green cards or work visas are noncitizens but also are able to work legally. The following table summarizes the extent of measurement error in the data.

| | U.S. Citizen/Documented | Not U.S. Citizen/Undocumented |
|---|---|---|
| U.S.-born | No measurement error | Not in the data |
| Foreign born | No measurement error | Measurement error - some should be documented |

In a just-identified system, the two stage least squares estimate is the ratio of the coefficients on the excluded instrument in the reduced form and the first stage. The measurement error in the endogenous variable does not change the instrument (a dummy variable for being born in Mexico), the outcome variable (a dummy variable for not being in school), or any of the controls used in the baseline specification (age fixed effects and a female dummy). Hence the measurement error does not alter the reduced form. The measurement error does, however, increases the strength of the first stage relationship between being born in Mexico and legal status. By inflating the coefficient on the first stage, the measurement error biases the IV results toward zero.[23]

[22]Suppose, for example, that some fraction of respondents with undocumented children falsely report that all of their children are citizens and born in the U.S. Because of the family fixed effects, all families where the reported country of birth and legal status are identical across siblings do not identify the parameter of interest. Therefore, while differential non-truthful response will render the estimates local to families that truthfully answer the census, it may not bias the coefficient of interest.

[23]Similarly, the OLS regressions will also be biased toward 0. If $noncitizen^*$ is an individual's true legal status, we have $noncitizen \geq noncitizen^*$. That is, some individuals whom we believe are treated ($noncitizen=1$)

Other specifications use *likely undocumented* as a measure of legal status. An individual is labeled as *likely undocumented* if two things are true: (1) she is not a U.S. citizen and (2) no one in her household has participated in government programs (since participation in the programs make it unlikely that the family is undocumented).[24] The intuition behind this refinement is simple; although there is evidence that undocumented immigrants have been able to collect supplemental Social Security income, especially before the 1996 Welfare Reform Act (Hayes 2003), families in which the parents are undocumented are plausibly less likely to register with the government for social insurance. Moreover, undocumented immigrants generally cannot serve in the U.S. military, and immigrants who serve are generally given a path to citizenship upon discharge. Hence individuals labeled as *likely undocumented* are more likely to be actually undocumented than individuals who are labeled merely as *noncitizen.* Consistent with this hypothesis, in the instrumental variable specifications, the sign and statistically significance of *noncitizen* and *likely undocumented* are similar, but the point estimate of *likely undocumented* is generally larger. Appendix B performs simulations that introduce *more* measurement error into the data and shows that the quantitative effect of this measurement error is modest.

# 5 Results

## 5.1 Cross-Sectional OLS Results

Here, we present the basic cross-sectional relationship between being a noncitizen and being undocumented and school attendance without family fixed effects. These results will serve as a baseline for later specifications. Standard errors are clustered at the family level, like all results in the paper. The estimating equation is given by

$$\mathbb{I}_{\text{Not in School},\, i\ \ j} \cdot 100 = \alpha + \beta \cdot \mathbb{I}_{\text{Undocumented}\ i\ \ j} + \sum_{l=13}^{17} \psi_l \mathbb{I}_{\text{age}=\, l} + \varepsilon_{i\ \ j} \tag{4}$$

are in fact not treated (*noncitizen*$^*$=0). In the cross section, the measurement error implies that the mean of the school attendance of individuals with *noncitizen* equal to 1 reflects the outcomes both of actually undocumented individuals and those who are not citizens but are authorized. Unfortunately, without more detailed information on the extent of the measurement error and on how the size of the measurement error differs in across versus within families, it is not possible to determine the size of the bias.

[24] These government programs consist of receiving money from the Social Security Administration (either regular Social Security or supplemental) or public welfare or being a veteran of the U.S. military.

The results from this regression are presented in Table 3.

[Table 3: OLS results]

The coefficients on *noncitizen* and *likely undocumented* are positive and precisely estimated in all models. Model (1) includes no controls except for age fixed effects, and (2) includes both age fixed effects and a control for being a female. These estimates indicate that individuals who are either noncitizens or *likely undocumented* are approximately 2.8 percentage points less likely to be in school than their peers. Due to measurement error, this estimate is a plausible lower bound for the true cross-sectional estimate. Omitted variables, in contrast, may cause this OLS estimate to be biased upward.

Model (3) interacts being female with either *noncitizen* or *likely undocumented*. The negative coefficient on the interaction terms suggest that the effect of legal status is stronger for men than for women. Model (4) adds family fixed effects to Model (2), slightly increasing the estimated coefficient to about 3.0.

Model (5) restricts attention to observations in families where all individuals in the sample arrived in the U.S. by age 6 or were born in the U.S.. The estimated coefficient is smaller than the baseline estimate.

## 5.2 Effect of Being Born in Mexico on School Attendance

This subsection presents the key results of the paper. It establishes that Mexican-born teenagers are more likely not to be enrolled in school than their U.S.-born siblings. Table 4 presents estimates of Equation [1], the reduced form within-family relationship between being born in Mexico and not attending school. Model (1) estimates the equation without fixed effects. The baseline estimates with family fixed effects are presented in Model (2). Being born in Mexico is associated with a 2.74 percentage point increase in the probability that a Hispanic teenager is not enrolled in school, off of a baseline probability of 2.87 percent. The effect is precisely estimated ($p < 0.01$) and is slightly higher than the coefficient of the specification without family fixed effects. Models (3) – (5) present the same robustness checks as in Table 3 and shows that the point estimate and statistical significance changes little across specifications.

When being female is interacted with being born in Mexico in Model (4), the coefficient of the interaction term is statistically significant ($p < 0.05$) and about -1.10. Being born in Mexico has a stronger effect on dropping out for men than women; women born in Mexico are estimated to be about 40% less likely to dropout than men born in Mexico.

[Table 4: Effect of Being Born in Mexico on Being Out of School]

### 5.3 Effect of Being Born in Mexico on Legal Status

One potential explanation for the difference in the school enrollment patterns between the U.S. and Mexican-born siblings is the difference in legal status. Table 5 presents results from the first stage regressions (Equation [2]) with family fixed effects where an individual's legal status is predicted by a dummy variable for being born in Mexico. Panel A shows results with *noncitizen*, while estimates with *likely undocumented* as a variable of interest are presented in Panel B. In both panels, the point estimate and standard error change little across the baseline set of robustness checks. In Panel A, the estimates suggest that a Mexican-born sibling is approximately 84 percentage points more likely to be classified *noncitizen* than her non-Mexican-born sibling. The estimates in Panel B show that a Mexican-born sibling is approximately 74 percentage points more likely to be classified as *likely undocumented* than her non-Mexican-born sibling. In all specifications, the effect is precisely estimated.

[Table 5: Effect of Being Born in Mexico on Legal Status]

The difference in magnitude between the results in the two panels is expected. All individuals who are born in the United States are recorded as citizens and hence are recorded as not being *noncitizen* or *likely undocumented*. In contrast, many of the individuals born in Mexico are noncitizens, but some of these individuals live in households where there is reason to think that the members of the household are not undocumented. These individuals are recorded as being *noncitizen* but not as *likely undocumented*. Therefore the difference in legal status between Mexican and non-Mexican-born individuals varies more when the outcome variable is *noncitizen* than when it is *likely undocumented*. These results could be seen as the first stage in an instrumental variables regression. But we defer any IV regressions until after discussing other

potential drivers of the relationship between being born in Mexico and educational attainment.

## 5.4 Sample Selection

The baseline sample includes Mexican heritage individuals aged 13 – 17. This section explores the sensitivity of the results by considering the effect by age or nationality. To begin, in column 1 of Table 6, we re-estimate the baseline model, restricting attention only to individuals 14 and older. If 13 year olds are relatively unresponsive, we would expect the estimated coefficient on legal status to increase in magnitude. In fact, the estimate is essentially the same as the baseline estimate.

[Table 6: Sample Selection Checks]

Second, we augment the baseline model by interacting the variable of interest with a dummy variable for an individual's being aged 16 or 17, allowing the effect to differ for individuals aged 13 to 15 from those aged 16 and 17. Results are reported in column 2 of Table 6. The uninteracted coefficient is positive but not statistically significant ($p > 0.05$), suggesting that we cannot reject the null that, for individuals aged 13 - 15, being undocumented has no effect on school attendance. The coefficient on the interaction between being *born in Mexico* and being aged 16 and 17 is large and highly statistically significant ($p < 0.01$); this estimate suggests that being born in Mexico increases the likelihood of not being in school by 3.90 percentage points for children 16 and 17 years old. The probability that 16 and 17 year old individuals who are classified as noncitizens are not in school is 8.25 percent. Taken at face value, these estimates suggest that being undocumented can account for slightly less than half of this baseline level for 16 and 17 year olds.

We also looked at the effect of being born abroad on the sample of all Hispanic immigrants aged 13-17 who arrived by age 12. Column 3 of Table 6 shows the results of a specification with age, female, and family fixed effects. The estimated coefficient of *born abroad* is highly statistically significant ($p < 0.01$) and suggests that being born outside of the U.S. increases the likelihood of not being in school by 2.76. This result is similar to that for our baseline sample.

# 6 Alternative Explanations for the Effect of Being Born in Mexico

Within a family, siblings born in Mexico are much less likely to be citizens or to be documented. However, other differences between siblings born in Mexico and those born in the U.S. may explain the pattern of Mexican-born individuals being less likely to be in school than their U.S.-born siblings. We explore four such explanations here and conclude that, to the best the evidence suggests, they largely do not explain the effect of being born in Mexico on school attainment. First, because many families give birth to their older children in Mexico and their younger children in the U.S., siblings born in Mexico are more likely to be the oldest sibling in the household, which some may worry could partially drive the results. Second, those born in the U.S. and Mexico may face different childhood environments with lasting effects on educational attainment. Third, and relatedly, the two groups may have different English abilities. Finally, they may have different prenatal environments. This section examines each potential threat to the identification and shows that, so far as the data indicates, they largely do not appear to be the drivers of the differences in educational outcomes between U.S.- and Mexican-born siblings. But they may modestly contribute and we adjust estimates in Section 7 accordingly.

## 6.1 Older Sibling and Age Effects

The typical mixed-country-of-birth family has the older children born in Mexico and the younger children born in the U.S. Because older children are less likely to be enrolled in school, failing to control for age introduces a spurious direct correlation between being born in Mexico and being in school. To address this concern, all regressions controlled for age fixed effects. However, if the schooling-age profile differs between the individuals in mixed-country-of-birth family and other individuals, then the impact of being born in Mexico will be partly driven by this factor even with age fixed effects. For example, if the age profile of school attendance is steeper for Mexican-born children than for U.S.-born children for reasons unrelated to legal status, then the fixed effects would fail to pick up this different profile and would therefore overstate the impacts.

Table 6 presents a variety of alternative robustness checks. Model (1) restricts attention only to individuals who live in families with total income less than or equal to the median family income in the sample, calculated separately for each sample year. If individuals from higher-income families have a more flat age-schooling profile and if immigrant families are less likely to be high-income, then excluding these individuals should decrease the estimated effect. The estimates in column (1) are slightly smaller (14%) than the baseline result, but are still positive and are precisely estimated.

A more direct test of the robustness of the results to improperly measured age effects comes from Models (2) and (3) which identify the coefficient of interest by restricting attention to families with atypical birth orders. For example, if parents encourage their oldest siblings to leave school early to provide financial support for their younger siblings, then the observed relationship between country of birth and educational attainment could partially be driven by this parental encouragement; looking at atypical birth order families helps alleviate this concern. In 4.09 percent of the cases where at least one member of the family is classified as a noncitizen, the oldest citizen child in the sample is older than the youngest noncitizen sibling. If the estimated age coefficients are inappropriate for the population of interest, looking at these families with atypical birth/noncitizen order should produce estimates that differ from the baseline results. Model (2) includes only individuals where all siblings in the sample are citizens or where the siblings in the sample have an atypical noncitizen status. This sample restriction sharply reduces the sample size of individuals with mixed-country-of-birth siblings. Nonetheless, the estimated coefficients in (2) are, larger in magnitude to the baseline specification and significant at the 1 percent level. Model (3) repeats this exercise for birthplace. Here, we restrict attention to individuals for whom all siblings in the sample are U.S.-born or the oldest U.S.-born child is older than the youngest Mexican-born child. The results in (3) are similar to those in (2).[25]

[25] As an additional check, we re-estimated the baseline model using panel data from the 2004 Survey of Income and Program Participation (SIPP), a nationally representative survey of the non-institutional resident population, including undocumented immigrants. Because we observe teenagers multiple times when they are at different ages, panel data allows me to include *family by age fixed effects*. Here, the parameter is identified by comparing the schooling choices of siblings at various points in time. For example, consider the hypothetical family A, which moved from Mexico 15.5 years before the data was collected. The older sibling (aged 16 in 2004) was born in Mexico and is an undocumented immigrant. Her younger sister (aged 15 in 2004) was born in the U.S. and is eligible for U.S. citizenship. By including family by aged fixed effects, this specification compares the educational choices of the older sibling in 2004 with the younger sibling in 2005, when both were age 16. Other families allow me to include year fixed effects. Therefore, this empirical strategy allows me to estimate a separate

A final threat to identification comes from the fact that, in families with mixed-country-of-birth children, the oldest sibling tends to be born in Mexico. If parents or teachers systemically treat oldest siblings differently, differences in schooling outcomes between Mexican and U.S.-born siblings may be contaminated by these birth order effects. Model (4) includes a control variable for being the oldest sibling *observed in the sample*; because we cannot follow individuals over time, we do not know if the oldest child in the sample is the oldest child in the family. Still, the fact that the inclusion of this control hardly changes the point estimate is reassuring.

[ Table 7: Age Robustness Results ]

## 6.2 Childhood Environment

A second concern is whether potentially different childhood environments generate part of our results. (We discuss prenatal environment separately.) Evidence suggests that childhood environment matters for subsequent performance (e.g., Hoynes, Schanzenbach, and Almond 2016; Chetty et al 2011; Almond, Currie, and Herrmann 2012; Currie 2009). We do not know though how different childhood environments are in Mexico and the U.S. for immigrant children. Of course, Mexico is substantially poorer, and there are a variety of ways in which Mexico has worse childhood environments than the U.S. For example, in Mexico, there may be less social assistance for the poor, less access to quality childhood education, more dangerous disease environments, and lower quality healthcare. However, evidence suggests that Mexican immigrants are drawn from the middle of the Mexican income distribution (Chiquiar and Hanson 2005; Hanson 2006), whereas Mexican immigrants are more likely to be at the bottom of the earnings distribution in the U.S.[26] In addition, undocumented immigrants face challenges in the U.S. that they do not face in their home countries that could hinder childhood development, such as reduced access to healthcare (Hacker, Anies, Folb, and Zallman 2015). If individuals at the bottom of the income distribution tend to have worse childhood environments, then country

age-schooling profile for each family. Unfortunately, the relatively small sample size of the SIPP is compounded by the fact that the panel is relatively short (2.5 years) and by the fact that the SIPP records information on school attendance starting only at age 15. These estimates from the SIPP are larger in magnitude to the census-based ones, but due to the small sample sizes, are not statistically significant from zero.

[26] See Borjas (1987) for other evidence on the selection of immigrants from Mexico. Mexican immigrants are more likely to earn the minimum wage and be high school dropouts than the median native born American (Passel, Capps, and Fix, 2004).

level averages may overstate the true difference in the quality of childhood environments between Mexico and the United States that is relevant for our sample. Thus, since we do not know how different childhood environments for U.S.-born and Mexican-born siblings—or even which country has better environments—we have four tests in which better childhood environments would predict a certain result; we show that the predictions that better childhood environment in Mexico would suggest largely do not attain in the data. In the next section, we try to purge the results of this factor to the extent it is an issue.

First, we test for whether foreign-born children exhibit higher rates of being held back in school than U.S.-born children, as one indication of childhood environment and human capital investment even for children who are younger than 13 and at little risk of not being enrolled in school. If being born in Mexico means that a child was exposed to a worse childhood environment (by virtue of being in Mexico or by virtue of parental human capital investment), we would expect being undocumented to increase the probability that a child is not enrolled in the correct grade for her age. Alternatively, the childhood inputs and environment could be better for Mexican-born children either because the environment in Mexico is actually better than in the U.S. or because parents (or their children) invest *more* in the human capital of their undocumented children, perhaps due to concerns of equity or because of the tradition of compulsory pre-school education in Mexico. In this case, being undocumented should decrease the probability that a child is not enrolled in the correct grade for her age. If foreign-born children enter the U.S. with worse childhood experiences, they should be held back in school more than those U.S.-born children.

We construct an "age-in-grade" dummy variable that determines if an individual is in less than the expected grade for her age and re-estimate the baseline regressions with this outcome variable. Unfortunately, the Census data does not ask sufficiently detailed questions about age attendance for 2000 – 2007, so these regressions are estimated on the 2008 – 2012 samples only.[27] To construct this age-in-grade measure, we group observations into bins based on their age, birth quarter, and state of residence. For each bin, we calculate the modal grade attended. If an individual is attending a grade that is less than this grade, she is said to be not in the

[27]Specifically, for 2000 – 2007, the census only asks if an individual is in middle school or high school rather than asking the specific grade that a student is enrolled. This information is not sufficiently detailed to know if, for example most 17 year olds are in the correct grade for their age.

correct age-in-grade. To ensure that these results do not simply pick up the baseline result that being undocumented increases the probability of not being enrolled in school at, we include only individuals beneath the age of 16, the minimum age of mandatory school attendance in the U.S. We also include all individuals older than 5.[28] According to this definition, 21 percent of teenagers in the sample are below their expected grade.

Column 6 of Table 8 shows that, if anything, the U.S.-born children are held back in school more than the foreign-born children. When the outcome variable is a dummy variable (scaled by 100) for not being in the correct age-in-grade, the point estimate on *born in Mexico* is -1.553, which is significant at the 10% level, meaning that Mexican-born immigrants are actually less likely to be held back. This is evidence against a childhood environment that favors U.S.-born siblings, since–if the childhood environment did favor U.S.-born siblings–we might expect the estimate to be positive, reflecting an inferior childhood environment for Mexican-born siblings that leads them to be held back in school later in life.

[ Table 8: English and Health Checks ]

Second, we test whether health impacts resulting from different childhood experiences could drive the results. [29] Within-family differences in health that are correlated with an individual's country of birth might therefore explain the paper's reduced form results. The fourth column of Table 2 showed that Mexican-born siblings tend to be in slightly *worse* health than their U.S.-born siblings, at least as measured by the health outcomes present in the U.S. Census: having a personal care problem or a physical difficulty. However, those coefficients are not statistically significantly different from 0.[30] While these proxies are crude, these health variables are highly

[28] At age 15, 98.39% of children in the sample are in school, a similar proportion to earlier ages; attendance starts dropping off at age 16, to 97.58% and then to 94.65% at age 17. An alternative approach to addressing the endogeneity of school attendance is to define an individual who is not in school as being in the correct age-in-grade, but include the entire sample. When we estimate these regressions, the results are very similar to those produced here.

[29] For reviews, see Glewwe (2005) and Glewwe and Miguel (2008). Studies of calories and protein in four Guatemalan villages (Pollitt *et al.*, 1993), deworming in Kenya (Miguel and Kremer 2004), and iron supplements in India (see Glewwe 2005) suggest that health matters for education. Moreover, evidence from outside of developing countries points to similar results. For example, the 1918 influenza pandemic (Almond 2006), the eradication of hookworms in the American South (Bleakley 2007), and the fallout from radiation from Chernobyl (Almond, Edlund, and Palme 2009) all point to an effect of health on education.

[30] One might be concerned that physical disabilities would generate selection, since families with Mexican-born children with disabilities may be less likely to migrate as a family unit. In particular, the Mexican-born child has a physical disability may be less likely to immigrate to the U.S. with the rest of his family, thereby leading selection in the sample of at least two-child families that produce the variation that the regressions exploit.

predictive of cross-sectional school enrollment patterns.[31] Not surprisingly, adding these health variables as controls has essentially no impact on the baseline result (see Models (2) – (4) in Table 8).[32]

Third, we test whether restricting the results to sibling pairs of relatively small age differences—thus limiting the difference in childhood exposure between foreign-born children who arrived in the U.S. at young ages—and thus had less exposure to the Mexican childhood environment—affects the results we test whether siblings of a test compares siblings of a given age difference: 1, 2, 3, etc. years. To do so for families with more than two siblings, we create combinations of all pairs of siblings, thereby duplicating the data, and include sibling pair fixed effects; we, of course still cluster at the family level. If childhood experience is a driver of the results, we should see the results decline as we zoom in on children of similar ages, since they are exposed to substantially similar (and increasingly similar as the age difference declines) childhood environments. Table 9 shows the results. Again, it turns out that the results decline only modestly as we zoom in on smaller age differences, from 2.816 in the full sample to 2.498 for those with at most one year of age difference (there is not sufficient data for a precise estimate of those with 0 age difference).

[ Table 9: Age Difference Table ]

The fourth test is similar to the third, but instead of including only sibling pairs of small age differences, we test whether restricting the sample to those who arrived in the U.S. at young ages—and thus had less exposure to the Mexican childhood environment—affects the results. This test differs from the test comparing those of different ages by including in the regression

Such selection would mean that we under-estimate the impacts, since we would be including pairs of siblings when a disabled sibling (who is presumably more likely to be out of school) is born in the U.S., but not when a disabled sibling is born in Mexico. However, as Table 2 shows, there is little reason to be concerned about this under-estimation, since there is virtually no difference between the probability of having a physical difficulty between U.S.-born and Mexican-born siblings.

[31] In the baseline sample, the probability that an individual with neither a personal care problem nor a physical difficulty is not enrolled in school is 2.65; for individuals with both a personal care problem and a physical difficulty, the probability is 7.30, and this difference is highly statistically significant. In addition, previous work has used these variables as proxies for health (see Almond 2006).

[32] Model (2) in Table 8 re-estimates the baseline model, controlling for a dummy variable for having a personal care problem, model (3) adds a dummy variable for having a physical difficulty, and model (4) includes both health proxies as well as the language control. Across these specifications, the coefficients on *born in Mexico* are almost identical to the baseline specification and are statistically significant at the 1% level. As an additional robustness check, we also control for a dummy variable for having a cognitive difficulty; the results are essentially unchanged.

those of ages relatively far apart, but who arrived around the same time. It also has family fixed effects, not pair fixed effects. If childhood conditions truly are worse in Mexico than in the U.S., and that is what is driving the results, then the results should be larger for families in which the foreign-born sibling arrived in the U.S. at a later age and thus had more exposure to the foreign childhood environment. Table 10 shows these results. The columns from left to right limit the sample to individuals who arrived in the U.S. at progressively younger ages. Individuals who were born in the U.S. are included in all samples. If early childhood impacts drive the results, the results should shrink when going from left to right. Table 10 shows that, though the results decline somewhat, the effect is still statistically significant and around 2 when restricting to children who arrived in the U.S. by age 1 ($p < 0.01$) and by age 0 ($p < 0.05$). We also conducted three hypothesis tests to see if the coefficients of the different columns were equal. Let $H_i$ be the coefficient of *born in Mexico* for the sample that restricts to siblings who arrived in the United States by age $i$. We tested the null hypotheses $H_{12} = H_2, H_{12} = H_{1,}$and $H_{12} = H_0$. We failed to reject the null for each of the three hypotheses ($p > 0.1$) again suggesting that childhood environment is not an important driver of the results.

[ Table 10: Age of Arrival Estimates ]

Between these four tests, especially the last two, we think that—to the best that we can test in the data—we can largely rule out the possibility that childhood environment is a main driver of the results. Of course, we cannot definitively rule it out; the standard errors are not zero. But we can to the best we can test in the data, there is little evidence that childhood environment is the primary driver of the results. In the discussion section, we use estimates from these tables, zooming in on those siblings with similar ages and ages arrival, to try to purge the results of the impact of childhood environment.

## 6.3 English Ability

Table 2 demonstrated that Mexican-born siblings are 5 percentage points less likely to report speaking English very well than their non-Mexican-born siblings. Part of the difference in English ability may reflect the fact that Mexican-born siblings are more likely to have spent the

first years of life in an environment with limited exposure to English. If the inability to speak English decreases a student's return from schooling, then English ability would drive part of the effect of being born in Mexico on school attendance.[33]

To determine if differences in English ability can account for the schooling attendance results, we first note that the evidence from the previous subsection provides some reassurance that differences in English ability do not drive the results. Previous research suggests that children who received their first exposure to English at an earlier age attain a higher level of English-language proficiency than those who received it later (Bleakley and Chin 2004). But, when we restrict attention to individuals in families where all siblings in the sample moved to the U.S. early in life or were U.S.-born, Model (5) of Table 4 showed that the baseline result dropped by less than 7 percent when it was re-estimated on the sample of individuals where all siblings in the sample moved to the U.S. by age 6, and Table 10 explores the robustness of this relationship to alternative age cutoffs. Overall, the point estimate and statistical significance of the result is stable to "zooming in" on these families that arrived early in the U.S. For example, restricting attention just to families where all siblings in the sample arrived by age 3 reduces the point estimate by less than 15 percent.

More directly, we include a proxy for English ability (a dummy variable for an individual's speaking English "very well" or speaking "only English").[34] Results from this regression are reported in column (1) of Table 8. We see that, as expected, controlling for English ability decreases the coefficient of interest. However, the change in magnitude is small; the point estimate on *born in Mexico* drops by 7 percent relative to the baseline specification. The result remains precisely estimated.

[33]On the other hand, differences in English skills may partly reflect the endogenous choice not to learn English because an individual knows that he will not finish high school (Chiswick and Miller 1995). If education and English skills are complements on the labor market and if Mexican-born siblings understand that they will acquire less schooling than their U.S. born peers, then differences in English ability may reflect the indirect effect of citizenship on schooling rather than a threat to identification. In this case, accounting for differences in English ability will understate the effects of citizenship on school attendance.

[34]The Census contains no direct tests of English ability. Respondents report only an individual's language ability along a subjective scale. In general, such results from these subjective measurements are difficult to interpret because definitions of speaking English "very well" may differ across respondents. Because the specifications include family fixed effects, however, these estimates reflect differences in how the same respondent reports the English speaking ability of teenagers in his household. Controlling for speaking English "well", "very well", or speaking "only English" yields essentially identical results.

## 6.4 Prenatal Environment

There is another set of concerns for which we cannot test: those relating to the prenatal environment and the environment surrounding birth. This difference is inherent to the analysis: Mexican-born siblings have a Mexican prenatal environment and U.S.-born siblings have a U.S. one. As is the case for childhood environment, we do not know how different childhood environments are for immigrants from Mexico to the U.S. We do know about average differences, and conditions tend to be worse in Mexico. Again, though we do not know how conditions vary for those who immigrate, since evidence suggests that they are at the middle of the middle of the Mexican income distribution and arrive at the bottom of the U.S. income distribution, suggesting that average differences may overstate the differences that immigrants face between the U.S. and Mexico and may even get the sign wrong: that is, conditions may be better in Mexico than in the U.S. Indeed, the evidence on childhood environment suggests that early childhood is not a main channel driving the results, despite documented substantial impacts of early childhood. This evidence suggests that early childhood environments are in fact not much worse in Mexico and may even be better–though, of course, improved life outcomes, which could include not only economic but also health-related conditions, presumably drive the move to the U.S. These results on childhood may suggest that the prenatal environment also does not primarily drive the differences in school attainment between U.S.-born and Mexico-born individuals. And, in particular, the evidence on Mexican-born students not being held back in school more or having worse health outcomes suggests that the prenatal or birth environment is not playing a significant role in driving the results. However, we cannot know for sure, and tests are unavailable to exclude the possibility, since–while immigrants may have had different amounts of childhood experience in Mexico–they all were, by definition, born in Mexico, and thus experienced a Mexican prenatal environment. Since we do not know how much prenatal environment differs and cannot test its importance, we instead review the literature on the impact of prenatal environment, develop an estimate of how much those impacts might matter for high school attainment, and compare that to the results discussed in Section 5.

The standard measure of prenatal environment is birthweight, and we produce a calculation using the literature's estimates of prenatal environment (as measured by birthweight) on number

of years of education, given observed differences in birthweight. We explain methods in Appendix B and report results in Appendix Table 6. The results are modest. We find that, as a likely upper bound estimate based on studies on the impact of prenatal environments and average differences in environment between Mexico and the U.S., that such differences could lead to at most 0.015 of a year of reduced educational attainment. By contrast, our results imply that being an undocumented immigrant reduces the number of years of education by about 0.112, roughly an order of magnitude lower than the prenatal estimates. Again we emphasize that we do not know the true magnitude of the differences in prenatal environment, and we think it very likely that this is an upper bound of the impact, since undocumented immigrants are likely at the bottom of the U.S. distribution and higher in the Mexican distribution. The next section discusses the main results in light of these estimates of the potential impacts of prenatal environment differences.

# 7 Discussion

We begin by discussing the IV estimates that parallel our baseline specification, before turning to more conservative estimates that take into account the concerns raised in the previous section. Table 11 presents the baseline IV regressions from by Equations [2] and [3] . In a just-identified system with one endogenous variable, the IV results are simply the ratio of the reduced form to the first stage; hence, these results should be viewed as a convenient way to scale the reduced form results. Baseline results are summarized in Table 11. The estimated coefficients on *noncitizen* (Panel A) and *likely undocumented* (Panel B) are 3.254 and 3.713 respectively, and the estimates are highly statistically significant.

[ Table 11: IV Estimates ]

Ignoring threats to identification and measurement error, these results suggest that being an undocumented immigrant makes a Mexican-born teenager of Mexican heritage between 3.25 – 3.71 percentage points less likely to be in school. Models (3) – (5) submit these results to the standard robustness checks. Adding gender controls or restricting attention to families

where all children in the sample arrived by age 6 or were U.S.-born leaves the results essentially unchanged.

However, Section 6 gave some reason to be concerned about taking these IV results literally. Thus, we refine them here, trying to account for prenatal and childhood environments, before turning to back-of-the-envelope implications for loss in human capital and fiscal impacts. We begin by discussing the magnitude of the results in light of the potential impact of childhood and prenatal environment. We address these concerns by doing two things: (1) to address the potential impact of childhood environment, we focus on estimates where sibling are close in age or arrived in the U.S. at close ages and (2) to address the potential impact of the prenatal environment, we subtract off of the results an estimate of prenatal impacts using results from the existing literature. As a a result, we develop bands of estimated impacts. The baseline estimates from Section 5 suggest that being born in Mexico reduces the likelihood of being in school by 2.73 percentage points. A more conservative estimate of the effect averages the estimate from the sample that restricted to siblings with an age difference of one or less (2.50, Table 9, column (4)) and the coefficient from the sample that restricted to age of arrival of one or less (1.98, Table 10, column (11)) to suggest that being born in Mexico reduces the likelihood of being in school by 2.24 percentage points. Using this number helps address concerns about the impact of childhood environment. To get the average reduction in the number of years of education, we multiply these results by 5, since we measure the impact on being in school starting at age 13, and the results represent the impact *each year* through age 17. Thus, the total reduction in years of education is 5 times the estimates from Section 5: that being born in Mexico reduces schooling by an average of between 0.112 (2.24 percentage points times 5) and 0.137 (2.73 percentage points times 5) of a year. In our main estimate, we take the average of these two numbers.

Second, we adjust our results by an estimate of the impact of prenatal environment. The details on our use of the existing literature to develop estimates of prenatal environmental on schooling are explained in Appendix C. Our estimates of the impact of prenatal environment suggest that the U.S.-Mexico difference could contribute 0.015 to these measured effects, roughly an order of magnitude less than even our conservative estimate taking into account the impact of childhood environment by restricting the sample to siblings who were born or were arrived

at similar ages. So, subtracting this estimate from our results suggests that the impact of being being born in Mexico on years of schooling is between 0.097 and 0.121 years.

To estimate the impact of being undocumented, we need to scale these results. We can do so either using the first stage (the effect of being born in Mexico on being a noncitizen or likely undocumented) or estimates from the literature on the share of Mexican-born immigrants that are undocumented. We produce both here. We first scale the effect of being born in Mexico on years of schooling by the the first stage estimate of the effect of being born in Mexico on being a noncitizen (Table 5, Panel A, Column 3). This suggest that being a noncitizen reduced years of schooling by about 0.114 to 0.144 years. We do the same for the first stage estimate of *likely undocumented* (Table 5, Panel B, Column 3), which suggests that the impact of being undocumented on years of schooling is between 0.130 and 0.164 years. Finally, we can check this last figure against estimates from other sources of the percentage of Mexican-born individuals in the United States who are undocumented. Using other sources, we calculate that the percentage of Mexican immigrants in the United States who are undocumented is 72%.[35] (See the discussion in Section 3, footnote 10.) Scaling our results for being born in Mexico by this number suggests that being undocumented reduces educational attainment by between 0.134 and 0.168 years.

To interpret the magnitude of these results, we produce back-of-the-envelope calculations of the implications of the results for the wage impacts and fiscal impacts. We base the estimates off of the average of (1) the estimates of the impact of being born in Mexico if one's sibling is born at most one year apart and (2) the parallel impact if one's sibling arrived in the U.S. at most one year apart, subtracting off the estimate of prenatal impacts, and scaling by the first-stage estimates of the impact of being born in Mexico on our "likely undocumented" measure (which, in any case, is similar to leading existing estimates). The calculations are detailed in Appendix C, but we briefly describe the calculations here. For both, we produce estimates for one undocumented teenager, one cohort of undocumented teenagers, and the 75-year impact of the policy (following National Academy of Sciences (2017)). All estimates are in present discounted value. We do not account for potentially changed immigration patterns as a result of a new policy. For the wage impacts, we combine our estimates with those on the impact on

[35]Note that these estimates are for all Mexican-born individuals, not just teenagers, which was unavailable beyond our own estimates.

earnings of attending more years of school. We find that being an undocumented immigrant reduces lifetime earnings by $8,455. For one cohort of undocumented immigrants, the total reduction in earnings is $728 million dollars. And over a 75-year time horizon, with many cohorts, the total reduction in earnings is $19.9 billion.

We then develop fiscal cost estimates using estimates from the National Academies of Sciences, Engineering, and Medicine (2017) estimates of the fiscal costs and benefits of immigrants with high school versus non-high school education. We find that one undocumented teenager will produce $4,257 less net revenue for the government and one cohort produces $366 million less. The 75-year cost is $9.3 billion.

[ Table 12: Net Present Value Wage and Fiscal Impacts ]

These results suggest that adopting a law like the DREAM Act, which would provide a path to citizenship for undocumented children, may have a substantial and positive impact on both individuals' wages and the public fisc. Ours is the first paper to study the impact of permanent legalization on high school education outcomes that goes beyond correlations. There are earlier studies with mixed results on the impact of DACA, a temporary measure, on college enrollment. [36] One of those did not find any effect of the work permits provided by DACA on college enrollment or obtaining a GED (Pope 2016) and one showed a negative impact on college enrollment (Amuedo-Dorantes and Antman 2017). Our results contrast with these estimates of the impact of DACA.[37] Though there could be many reasons for this disparity, it could be

[36] Announced by President Obama in June 2012, the Department of Homeland Security began accepting applications for the DACA program in August 2012. DACA allows undocumented immigrants who fit certain criteria to obtain a two-year reprieve from deportation and an Employment Authorization Document. To qualify for DACA, an undocumented immigrant must: (1) have been younger than 31 years old on June 15, 2012; (2) have been younger than 16 by the time they came to the U.S.; (3) have lived continuously in the U.S. since June 15, 2007; (4) have been in the U.S. on June 15, 2012 and at the time of filing the application; (5) have had no lawful status on June 15, 2012; (6) be currently in school, have graduated from high school, obtained a GED, or be an honorably discharged veteran of the Coast Guard or the U.S. Armed Forces; (7) not have been convicted of a felony, significant misdemeanor, or more than three misdemeanors, and not otherwise pose a threat to national security or public safety (Capps, Fix and Zong, 2017).

[37] Studies have also examined the impact of DACA on other outcomes. Amuedo-Dorantes and Antman (2016) found that DACA was associated with a reduction in the incidence of poverty for the eligible population. Amuedo-Dorantes and Antman (2017) used a differences-in-differences method to find that DACA decreased school enrollment and increased employment among the eligible population. Using a much larger sample, Pope (2016) found that DACA increased employment for people who qualified for the program and did not affect the likelihood of being in enrolled in college or obtaining a GED.

Also, though it does not look at the effects of a specific policy program, this paper is related to the large literature that has evaluated the effects of government policies towards undocumented immigrants. One such policy

explained by the fact that DACA was temporary and did not give a path to citizenship, thus reducing incentives to stay in school, since the grant was uncertain and was of a limited bundle of rights. A temporary reprieve from deportation and the granting of work authorization lower barriers to employment, but their effect on the returns to schooling are less clear, since the payoffs to education take time realize. By contrast, the temporary and easily revocable nature of DACA unequivocally raises the immediate returns to employment. Since the DREAM Act would not be temporary and would give citizenship, our results arguably are more relevant for guessing at the ultimate effects of the DREAM Act, or at least are a useful complementary piece of evidence, since we study years in high school rather than attainment of a college degree. The results from the other studies may be more appropriate for evaluating the effects of the proposed BRIDGE Act, which would essentially extend DACA protections by three years.[38]

# 8 Conclusion

This paper seeks to estimate the effect of being an undocumented immigrant on the educational attainment of the 1.1 million undocumented immigrants who are children. Because of labor market restrictions that make working in the formal and skilled labor market more difficult and because of their limited ability to finance higher education, undocumented immigrants may have worse job matches (Kossoudji & Cobb-Clark 2002) and have decreased returns to finishing high school.

To help address the endogeneity of legal status, this paper compares the educational outcomes of siblings who differ in their legal status due to the country of their birth. We show that Mexican-born siblings, who are more likely to lack legal status, are also more likely to not be enrolled in school. We look to the data to try to account for potential differences in prenatal and childhood environments; the estimates suggest that the results are largely not driven by

was the 1986 Immigration Reform and Control Act, which gave amnesty to many undocumented immigrants in the United States. Cortes (2013) found that being granted legal status increased the likelihood of college enrollment for a previously undocumented youth. Researchers have also evaluated state policies that allowed undocumented immigrants to pay in-state tuition at state colleges. Kaushal (2008) and Flores (2010) found that such policies increased postsecondary enrollment for undocumented Mexican and Latino youth, respectively.

[38] https://www.congress.gov/bill/115th-congress/house-bill/496

these factors, but rather by differences in legal status.

Using the data to the extent possible to account for different prenatal and childhood environments, we scale the reduced form relationship between country of birth, legal status, and school enrollment decisions by estimating IV regressions where legal status is treated as endogenous. Using proxies for legal status from U.S. Census data, we show that being undocumented reduces educational attainment by between 0.13 and 0.17 years.

The results of this paper suggest that legislation like the DREAM Act that would give teenage undocumented immigrants a path to citizenship is likely to alter their educational choices. This finding raises three additional questions for future research. First, how would the general equilibrium effects of the DREAM Act differ from the partial equilibrium estimates in this paper? In a static model without international trade, changing the legal status of a large number of individuals has the potential to alter the fraction of individuals who drop out of high school and decrease the returns to a high school diploma. In this case, the general equilibrium effects may be smaller than the partial equilibrium ones identified here. On the other hand, changing the legal status of a large number of individuals may encourage schools and families in immigrant neighborhoods to make investments that increases the net returns to finishing high school. Second, how would legalizing a large number of undocumented teenagers influence the incentives for natives and the wage structure of the United States? Finally, allowing teenage undocumented immigrants a path to citizenship may change the number and type of families who move without authorization. Legal status for one's children likely makes undocumented immigration more attractive for all potential immigrants, but especially those that believe that their children are likely to have high returns in the legal and formal labor market.

## References


Abramitzky, R. & Lavy, V. (2014). How Responsive Is Investment in Schooling to Changes in Redistribution Policies and in Returns? *Econometrica*, 82(4), 1241-1272.

Almond, D. (2006). Is the 1918 Influenza Pandemic Over? Long-term Effects of In Utero Influenza Exposure in the Post-1940 U.S. Population. *Journal of Political Economy*, 114 (4), 672–712.

Almond, D., Currie, J., & Herrmann, M. (2012). From Infant to Mother: Early Disease Environment and Future Maternal Health. *Labour Economics,* 19, 475-483.

Almond, D., Edlund, L., and Palme, M. (2009). Chernobyl's Subclinical Legacy: Prenatal Exposure to Radioactive Fallout and School Outcomes in Sweden. *The Quarterly Journal of Economics*, 124(4), 1729–1772.

Amuedo-Dorantes, C. & Antman, F. (2016). Can Authorization Reduce Poverty Among Undocumented Immigrants? Evidence from the Deferred Action for Childhood Arrivals Program. *Economics Letters,* 147, 1-4.

Amuedo-Dorantes, C. & Antman, F. (2017). Schooling and Labor Market Effects of Temporary Authorization: Evidence from DACA. *Journal of Population Economics,* 30, 339-373.

Batalova, J., Ruiz Soto, A., & Mittelstadt, M. (2017) Protecting the DREAM: The Potential Impact of Different Legislative Scenarios for Unauthorized Youth. *Migration Policy Institute Fact Sheet.*

Bean, F., Leach, M., Brown, S., Bachmeier, J. & Hipp, J. (2011) The Educational Legacy of Unauthorized Migration: Comparisons Across U.S.-Immigrant Groups in How Parents' Status Affects Their Offspring. *The International Migration Review,* 45(2), 348-385.

Bleakley, H. & Chin, A. (2004) Language Skills and Earnings: Evidence from Childhood Immigrants. *Review of Economics and Statistics*, 86 (2), 481–496.

Bruno, A. (2002). Immigration: Adjustment to Permanent Resident Status Under Section 245(i). *CRS Report for Congress.*

Autor, D., Katz, L., & Kearney, M. (2008). Trends in U.S. Wage Inequality: Re-Assessing the Revisionists. *Review of Economics and Statistics*, 90 (2), 300–323.

Barrow, L., & Rouse, C. (2005). Do Returns to Schooling Differ by Race and Ethnicity? *American Economic Review Papers and Proceedings*, 95 (2), 83–87.

Black, S., Devereux, P., & Salvanes, K. (2007). From Cradle to Labor Market? The Effect of Birth Weight on Adult Outcomes. *Quarterly Journal of Economics*, 122 (1), 409–439.

Borjas, G. (1987). Self-Selection and the Earnings of Immigrants. *American Economic Review,* 77 (4), 531–553.

Bleakley, H. (2007). Disease and Development: Evidence from Hookworm Eradication in the American South. *The Quarterly Journal of Economics*, 122 (1), 73–117.

Bleakley, H., & Chin, A. (2004). Language Skills and Earnings: Evidence from Childhood Immigrants. *Review of Economics and Statistics.* 86 (2), 481–496.

Bouvier, L., & Gardner, R. (1986). Immigration to the U.S.: The Unfinished Story. Population Reference Bureau Working Paper.

Capps, R., Fix, M., & Zong, J. (2017). The Education and Work Profiles of the DACA Population. *Migration Policy Institute Issue Brief.*

Chetty, R., Friedman, J., Hilger, N., Saez, E., Schanzenbach, D., & Yagan, D. (2011) How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR. *Quarterly Journal of Economics,* 126 (4), 1593-1660

Chin, A., & Juhn, C. (2007). Does Reducing College Costs Improve Educational Outcomes for Undocumented Immigrants? *James A Baker Institute for Public Policy Working Paper.*

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 (2), 239–281.

Chiswick, B. (1984). Illegal Aliens in the United States Labor Market: Analysis of Occupational Attainment and Earnings. *International Migration Review*, 18 (3), 714–732.

Chiswick, B. (1988). Illegal Immigration and Immigration Control. *Journal of Economic Perspectives,* 2 (3), 101–115.

Chiswick, B., & Miller, P. (1995). "The Endogeneity between Language and Earnings: International Analyses." *Journal of Labor Economics*, 13 (2), 246–288.

Cobb-Clark, D., Shiells, C., & Lowell, L. (1995). Immigration Reform: The Effects of Employer Sanctions and Legalization on Wages. *Journal of Labor Economics*, 13 (3), 472–498.

Cortes, P. (2008). The Effect of Low-Skilled Immigration on U.S. Prices: Evidence from CPI Data. *Journal of Political Economy*, 116 (3), 381–422.

Cortes, K. (2013). Achieving the DREAM: The Effect of IRCA on Immigrant Youth Postsecondary Educational Access. *American Economic Review: Papers & Proceedings,* 103(3), 428-432.

Currie, J. (2009) Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development. *Journal of Economic Literature,* 47 (1), 87-122.

Currie, J., & 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.

Diaz-Strong, D. & Ybarra, M. (2016). Disparities in High School Completion Among Latinos: The Role of the Age-at-Arrival and Immigration Status. *Children and Youth Review Services,* 71, 282-289.

Donato, K., Durand, J., & Massey, D. (1992). Stemming The Tide? Assessing the Deterrent Effects of the Immigration Reform and Control Act. *Demography*, 29 (2), 139–157.

Flores, S. (2010). State Dream Acts: The Effect of In-State Resident Tuition Policies and Undocumented Latino Students. *The Review of Higher Education*, 33(2), 239-283.

Fry, R. (2010). Hispanics, High School Dropouts, and the GED. *Pew Hispanic Center Report*.

Goldin, C., & Lawrence K. (2007). Long-Run Changes in the Wage Structure: Narrowing, Widening, Polarizing. *Brookings Papers on Economic Activity*, 2, 135–167.

Gould, E., Lavy, V., & Paserman, D. (2009). Sixty Years after the Magic Carpet Ride: The Long Run Effect of The Early Childhood Environment on Social and Economic Outcomes. *Review of Economic Studies,* 78 (3), 938-973.

Glewwe, P. (2005). The Impact of Child Health and Nutrition on Education in Developing Countries: Theory, Econometric Issues, and Recent Empirical Evidence. *Food and Nutrition Bulletin*, 26 (2) , S235–S250.

Glewwe, P. & Miguel, E. (2007). The impact of child health and nutrition on education in less developed countries. In T. P. Schultz and J. Strauss, *Handbook of Development Economics*, Volume 4, Amsterdam: North Holland Press.

Greenman, H. & Hall, M. (2013). Legal Status and Educational Transitions for Mexican and Central American Immigrant Youth. *Social Forces,* 91(4), 1475-1498.

Hacker, K., Anies, M., Folb, B., & Zallman, L. (2015). Barriers to Health Care for Undocumented Immigratns: A Literature Review. *Risk Management and Healthcare Policy,* 8, 175-183.

Hanson, G. (2006). Illegal Migration from Mexico to the United States. *Journal of Economic Literature*, 44 (4), 869–924.

Hayes, M. (2003). Illegal Immigrants Eligible for Social Security Benefits. Fox News. < http://www.foxnews.com/story/0,2933,79013,00.html >.

Hoynes, H., Schanzenbach, D., & Almond, D. (2016). Long-Run Impacts of Childhood Access to the Safety Net. *American Economic Review,* 106(4), 903-934.

Kaushal, N. (2008). In-State Tuition for the Undocumented: Education Effects on Mexican Young Adults. *Journal of Policy Analysis and Management,* 27(4), 771-792.

Kling, J., Liebman, J., & Katz, L. (2007). Experimental Analysis of Neighborhood Effects. *Econometrica*, 75 (1), 83–119.

Kossoudji, S., & Cobb-Clark, D. (2002). Coming out of the Shadows: Learning about Legal Status and Wages from the Legalized Population. *Journal of Labor Economics*, 20 (3), 598–628.

Miguel, E., & Kremer, M. (2004). Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities. *Econometrica*, 72 (1), 159–217.

National Academies of Sciences, Engineering, and Medicine. (2017). The Economic and Fiscal Consequences of Immigration. Washington, DC: The National Academies Press.

Passel, J. (2005). Estimates of the Size and Characteristics of the Undocumented Population. *Pew Hispanic Center Report.*

Passel, J., Capps, R., & Fix, M. (2004). Undocumented Immigrants: Facts and Figures. *Urban Institute Immigration Studies Program.*

Passel, J., & Cohn, J. (2010). U.S. Unauthorized Immigration Flows Are Down Sharply Since Mid-Decade. *Pew Hispanic Center Report.*

Passel, J., & Cohn, J. (2016). Overall Number of U.S. Unauthorized Immigrants Holds Steady Since 2009. *Pew Hispanic Center Report.*

Passel, J., & Taylor, P. (2010). Unauthorized Immigrants and Their U.S.-Born Children. *Pew Hispanic Center Report.*

Pollitt, E., Gorman, K., Engle, P., Martorell, R., J., Wachs, T., & Scrimshaw, N. (1993) "Early Supplementary Feeding and Cognition: Effects over Two Decades." *Monographs of the Society for Research in Child Development*, 1993, 58 (7235), pp. 1–118.

Pope, N. (2016). The Effects of DACAmentation: The Impact of Deferred Action for Childhood Arrivals on Unauthorized Immigrants. *Journal of Public Economics,* 143, 98-114.

Roy, A. (1951). Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers*, 3 (2), 135–43.

Ruggles, S., Alexander, J., Genadek, J., Goeken, R., Schroeder, M., and Sobek, M. (2011). Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]. Minneapolis: University of Minnesota.

Table 1: Summary Statistics by Country of Birth

| | Full Pop | Born in U.S | Born in Mexico |
|---|---|---|---|
| Age | 14.914 | 14.900 | 14.978 |
| | (1.407) | (1.409) | (1.395) |
| Female | 0.487 | 0.489 | 0.478 |
| | (0.500) | (0.500) | (0.500) |
| Noncitizen | 0.150 | 0.000 | 0.879 |
| | (0.357) | (0.000) | (0.326) |
| Likely Undocumented | 0.138 | 0.000 | 0.808 |
| | (0.345) | (0.000) | (0.394) |
| Speaks English Very Well | 0.882 | 0.918 | 0.705 |
| | (0.323) | (0.274) | (0.456) |
| Not in School | 0.029 | 0.024 | 0.051 |
| | (0.167) | (0.153) | (0.220) |
| Num Obs. Per Family in Sample | 1.511 | 1.512 | 1.508 |
| | (0.669) | (0.669) | (0.671) |
| Personal Care Problem | 0.007 | 0.007 | 0.006 |
| | (0.082) | (0.083) | (0.075) |
| Physical Difficulty | 0.010 | 0.010 | 0.009 |
| | (0.100) | (0.101) | (0.094) |
| Age At Arrival | 0.876 | 0.000 | 5.075 |
| | (2.444) | (0.000) | (3.650) |
| Observations | 262,271 | 217,073 | 44,019 |

*Note:* Mean and standard deviation reported. Data come from the 2000 decennial Census and the 2001-2012 American Community Survey. The unit of observation is an individual who identifies as being of Mexican heritage aged 13-17. An individual is classified as "Born in U.S." if she was born in the U.S. Those who were born in Mexico are classified as "Born in Mexico." In addition, note that the sample sizes for "Born in U.S." and "Born in Mexico" do not sum to the sample size for the "Full Pop" group because some individuals in the sample are not born in either the U.S. or Mexico. Age at arrival is calculated as the difference between a respondent's age and the number of years that she has been in the U.S. This variable is set to 0 for all U.S.-born individuals, including those born in Guam, Puerto Rico, and the Virgin Islands. All individuals who are not citizens of the U.S. are classified as "noncitizen." An individual is classified as "likely undocumented" if she is both not a citizen and if no one in her household is receiving money from the Social Security Administration, on public welfare or a veteran of the U.S. military.

Table 2: Summary Statistics for Individuals in Mixed Country of Birth Families

| | Summary Statistics | | | Regression Estimates with Family FE |
|---|---|---|---|---|
| | Full Pop in this Sample | Born U.S. | Born in Mexico | Born in Mexico |
| Age | 14.955 | 14.311 | 15.619 | 1.339*** |
| | (1.470) | (1.303) | (1.330) | (0.055) |
| Female | 0.509 | 0.509 | 0.510 | -0.005 |
| | (0.500) | (0.500) | (0.500) | (0.017) |
| Noncitizen | 0.415 | 0.000 | 0.842 | 0.842*** |
| | (0.493) | (0.000) | (0.364) | (0.009) |
| Likely Undocumented | 0.364 | 0.000 | 0.739 | 0.739*** |
| | (0.481) | (0.000) | (0.439) | (0.012) |
| Speaks English Very Well | 0.809 | 0.833 | 0.784 | -0.050*** |
| | (0.393) | (0.373) | (0.412) | (0.009) |
| Not in School | 0.033 | 0.013 | 0.053 | 0.041*** |
| | (0.178) | (0.114) | (0.223) | (0.006) |
| Num Obs. Per Family in Sample | 2.314 | 2.328 | 2.300 | - |
| | (0.541) | (0.551) | (0.531) | - |
| Personal Care Problem | 0.007 | 0.006 | 0.008 | 0.003 |
| | (0.081) | (0.074) | (0.087) | (0.003) |
| Physical Difficulty | 0.010 | 0.010 | 0.011 | 0.001 |
| | (0.100) | (0.097) | (0.103) | (0.002) |
| Age At Arrival | 1.272 | 0.000 | 2.582 | 2.499*** |
| | (2.634) | (0.000) | (3.272) | (0.087) |
| Observations | 7,851 | 3,964 | 3,887 | |

*Note:* Mean and standard deviation reported for the first three columns. Regression coefficient and standard error reported for fourth column. Regression outcome variable is given by the row heading. Regression includes a dummy variable for being born in Mexico and a family fixed effect. Note that the variable "number of siblings in the the sample" does not vary at the family level; the estimate of being born in Mexico in the number of siblings in the sample therefore is not identified in a regression with family fixed effects. Data come from the 2000 decennial Census and the 2001-2012 American Community Survey. The unit of observation is an individual who identifies as being of Mexican heritage aged 13-17 who lives in a family with at least one sibling in the sample who is born in the U.S. and at least one sibling in the sample who is born in Mexico. An individual is classified as "Born in U.S." if she was born in the U.S. Those who were born in Mexico are classified as "Born in Mexico." In addition, note that the sample sizes for "Born in U.S." and "Born in Mexico" do not sum to the sample size for the "Full Pop" group because some individuals in the sample included in this table are not born in either the U.S. or Mexico. For example, consider an individual who is born in Canada but has Mexican born and U.S. born siblings. This individual would be included in the sample for this table but would not fit into either of the two groups ("Born in U.S." or "Born in Mexico"). Age at arrival is calculated as the difference between a respondent's age and the number of years that she has been in the U.S. This variable is set to 0 for all U.S.-born individuals, including those born in Guam, Puerto Rico, and the Virgin Islands. All individuals who are not citizens of the U.S. are classified as "noncitizen." An individual is classified as "likely undocumented" if she is both not a citizen and if no one in her household is receiving money from the Social Security Administration, on public welfare or a veteran of the U.S. military.

Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the family level.

Table 3: OLS - Legal Status Effect on Being Out of School

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| *Panel A: Noncitizens* | | | | | |
| Noncitizen | 2.822*** | 2.818*** | 3.146*** | 3.044*** | 2.543*** |
| | (0.206) | (0.206) | (0.292) | (0.718) | (0.268) |
| Female | | -0.296*** | -0.193* | -0.444** | -0.223** |
| | | (0.103) | (0.103) | (0.173) | (0.104) |
| Noncitizen * Female | | | -0.686* | | |
| | | | (0.379) | | |
| *Panel B: Likely Undocumented* | | | | | |
| Likely Undocumented | 2.744*** | 2.740*** | 3.081*** | 3.204*** | 2.436*** |
| | (0.216) | (0.216) | (0.305) | (0.800) | (0.281) |
| Female | | -0.297*** | -0.198* | -0.443** | -0.222** |
| | | (0.103) | (0.104) | (0.173) | (0.104) |
| Likely Undocumented | | | -0.715* | | |
| * Female | | | (0.394) | | |
| Family FE | | | | X | |
| All Sibs Arrived in US By Age 6 | | | | | X |
| Mean Dep. Var. | 2.872 | 2.872 | 2.872 | 2.872 | 2.667 |
| Observations | 262,271 | 262,271 | 262,271 | 262,271 | 242,597 |

*Note:* Data come from the 2000 decennial Census and the 2001-2012 American Community Survey. The unit of observation is an individual who identifies as being of Mexican heritage aged 13-17. The outcome variable is a dummy for not being in school * 100. All individuals who are not citizens of the U.S. are classified as "noncitizen." An individual is classified as "likely undocumented" if she is both not a citizen and if no one in her household is receiving money from the Social Security Administration, on public welfare or a veteran of the U.S. military. The fifth column restricts to siblings in the sample that had moved to the U.S. by age 6. All regressions include age fixed effects.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the family level.

Table 4: Effect of Being Born in Mexico on Being Out of School

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Born in Mexico | 2.618*** | 2.735*** | 2.734*** | 3.290*** | 2.548*** |
| | (0.190) | (0.627) | (0.627) | (0.686) | (0.685) |
| Female | | | -0.442** | -0.253 | -0.376** |
| | | | (0.173) | (0.181) | (0.175) |
| Born Mexico * Female | | | | -1.096** | |
| | | | | (0.548) | |
| Constant | 3.033*** | 5.654*** | 5.868*** | 5.769*** | 5.499*** |
| | (0.135) | (0.255) | (0.272) | (0.273) | (0.257) |
| Family FE | | X | X | X | X |
| All Sibs Arrived in US By Age 6 | | | | | X |
| Mean Dep. Var. | 2.872 | 2.872 | 2.872 | 2.872 | 2.667 |
| Observations | 262,271 | 262,271 | 262,271 | 262,271 | 242,597 |

*Note:* Data come from the 2000 decennial Census and the 2001-2012 American Community Survey. The unit of observation is an individual who identifies as being of Mexican heritage aged 13-17. The outcome variable is a dummy for not being in school * 100. The fifth column restricts to siblings in the sample that had moved to the U.S. by age 6. All regressions include age fixed effects.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the family level.

Table 5: Effect of Being Born in Mexico on Legal Status

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| *Panel A: Effect on Being a Noncitizen* | | | | | |
| Born in Mexico | 87.711*** | 84.066*** | 84.066*** | 83.833*** | 83.624*** |
| | (0.275) | (0.880) | (0.880) | (0.899) | (0.984) |
| Female | | | 0.073 | -0.006 | 0.006 |
| | | | (0.082) | (0.073) | (0.081) |
| Born Mexico * Female | | | | 0.460 | |
| | | | | (0.421) | |
| Mean Dep. Var. | 15.022 | 15.022 | 15.022 | 15.022 | 9.329 |
| *Panel B: Effect on Being Likely Undocumented* | | | | | |
| Born in Mexico | 80.686*** | 73.674*** | 73.674*** | 73.461*** | 74.168*** |
| | (0.344) | (1.171) | (1.171) | (1.185) | (1.269) |
| Female | | | 0.035 | -0.037 | -0.110 |
| | | | (0.105) | (0.095) | (0.099) |
| Born Mexico * Female | | | | 0.419 | |
| | | | | (0.526) | |
| Mean Dep. Var. | 13.802 | 13.802 | 13.802 | 13.802 | 8.551 |
| Family FE | | X | X | X | X |
| All Sibs Arrived in US By Age 6 | | | | | X |
| Observations | 262,271 | 262,271 | 262,271 | 262,271 | 242,597 |

*Note:* Data come from the 2000 decennial Census and the 2001-2012 American Community Survey. The unit of observation is an individual who identifies as being of Mexican heritage aged 13-17. The outcome variable is a dummy for not being a citizen of the U.S. * 100 (Panel A) or a dummy variable for being classified "likely undocumented" * 100 (Panel B). All individuals who are not citizens of the U.S. are classified as "noncitizen." An individual is classified as "likely undocumented" if she is both not a citizen and if no one in her household is receiving money from the Social Security Administration, on public welfare or a veteran of the U.S. military. The fifth column restricts to siblings in the sample that had moved to the U.S. by age 6. All regressions include age fixed effects.

Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the family level.

Table 6: Effect on Being Out of School, Sample Selection Checks

| | (1) | (2) | (3) |
|---|---|---|---|
| Born in Mexico | 2.719*** | 1.351* | |
| | (0.777) | (0.701) | |
| Born in Mexico * Age ≥16 | | 2.548*** | |
| | | (0.589) | |
| Born Abroad | | | 2.759*** |
| | | | (0.526) |
| Just Individuals Age ≥ 14 | X | | |
| All Hispanics | | | X |
| Mean Dep. Var. | 3.265 | 2.872 | 2.848 |
| Observations | 206,572 | 262,271 | 400,497 |

*Note:* Data come from the 2000 decennial Census and the 2001-2012 American Community Survey. The unit of observation is an individual who identifies as being of Mexican heritage aged 13-17. The outcome variable is a dummy for not being in school * 100. The third column includes all Hipanic individuals aged 13-17. Born Abroad is a dummy variable for not being born in the U.S. All regressions include family, age and female fixed effects.
Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors are clustered at the family level.

Table 7: Effect of Being Born in Mexico on Being Out of School, Robustness Checks

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Born in Mexico | 2.345*** | 3.535*** | 4.215*** | 2.735*** |
| | (0.752) | (0.958) | (1.284) | (0.627) |
| Family Income < Median | X | | | |
| Atypical Legal Status | | X | | |
| Atypical Born Mexico Status | | | X | |
| Controls for Being Oldest Sib | | | | X |
| Mean Dep. Var. | 3.341 | 2.474 | 2.444 | 2.872 |
| Observations | 135,324 | 222,301 | 215,155 | 262,271 |

*Note:* Data come from the 2000 decennial Census and the 2001-2012 American Community Survey. The unit of observation is an individual who identifies as being of Mexican heritage aged 13-17. The outcome variable is a dummy for not being in school * 100. The first column restricts attention to individuals living in families where the total family income is less than the median family income for individuals in the sample. Median family income is calculated separately for each census year. The second column restricts attention to individuals where either all siblings are U.S.-born or where there is a citizen sibling who is older than a noncitizen sibling. The third column restricts attention to families where all siblings are born in the U.S. or where there is a Mexican-born sibling who is younger than a U.S.-born sibling. The final column includes a dummy variable for being the oldest sibling observed in the sample. All regressions include age, family, female fixed effects.
Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors are clustered at the family level.

Table 8: Effect of Being Born in Mexico on Being Out of School, English and Health Checks

| | Not in School | | | | | Held Back |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Born in Mexico | 2.551*** | 2.728*** | 2.732*** | 2.548*** | 2.903** | -1.533* |
| | (0.613) | (0.627) | (0.627) | (0.613) | (1.440) | (0.837) |
| Speaks English Very Well | -3.307*** | | | -3.292*** | | |
| | (0.698) | | | (0.699) | | |
| Personal Care Problem | | 1.694 | | 0.544 | | |
| | | (1.833) | | (2.473) | | |
| Physical Difficulty | | | 1.824 | 1.523 | | |
| | | | (1.328) | (1.852) | | |
| Born in Mexico * Year $>$ 2004 | | | | | -0.254 | |
| | | | | | (1.556) | |
| Mean Dep. Var. | 2.872 | 2.872 | 2.872 | 2.872 | 2.872 | 19.158 |
| Observations | 262,271 | 262,271 | 262,271 | 262,271 | 262,271 | 226,967 |

*Note:* For the fist five columns, data come from the 2000 decennial Census and the 2001-2012 American Community Survey. The unit of observation is an individual who identifies as being of Mexican heritage aged 13-17. The outcome variable is a dummy for not being in school * 100. Personal Care Control includes a dummy variable for having a personal care problem. Physical Difficulty Control controls for having a physical difficulty. English Ability Control includes a dummy variable for speaking English very well or as a sole language. For the sixth column, the sample is limited to children below age 16 and children who have not dropped out. Data come from the 2008-2012 American Community Survey. The unit of observation is an individual who identifies as being of Mexican heritage aged 6-15. The outcome variable of the regression is a dummy variable for not being in the correct grade for one's age*100. To construct this age-in-grade measure, we group observations into bins based on their age, birth quarter, and state of residence. For each bin, we calculate the modal grade attended. If an indivdual is attending a grade less than this grade, she is said to be not in the correct age-in-grade. All regressions include family, age, and female fixed effects. All regressions include age, family, female fixed effects.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the family level.

Table 9: Effect of Being Born in Mexico on Not Being in School by Sibling Age Difference

| | (1) ≤ 4 Years | (2) ≤ 3 Years | (3) ≤ 2 Years | (4) ≤ 1 Year | (5) 0 years |
|---|---|---|---|---|---|
| Born in Mexico | 2.816*** | 2.833*** | 2.290*** | 2.498** | 3.172 |
| | (0.675) | (0.668) | (0.690) | (1.175) | (4.942) |
| Mean Dep. Var. | 2.870 | 2.845 | 2.858 | 2.937 | 2.993 |
| Observations | 285,188 | 272,690 | 244,936 | 198,260 | 159,678 |

*Note:* Data come from the 2000 decennial Census and the 2001-2012 American Community Survey. The unit of observation is an individual who identifies as being of Mexican heritage aged 13-17. For these samples, we included all possible pairings of siblings. For example, in a family with four siblings, there are six possible pairings of siblings so we expanded the sample by creating two more copies of each sibling and then made a variable that identified sibling pairings. We then calculated the absolute value of the age difference between a pair of siblings. The first column restricts the sample to siblings pairings with at most 4 years of age difference. Each subsequent column reduces this number by one. For example, the second column includes all sibling pairings with 3 years or less of age difference. The outcome variable is a dummy for not being in school * 100. Regressions include sibling fixed effects, age and female fixed effects.
Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors are clustered at the family level.

Table 10: Effect of Being Born in Mexico on Not Being in School By Minimum Age of Arrival

| | (1) 12 years | (2) 11 years | (3) 10 years | (4) 9 years | (5) 8 years | (6) 7 years | (7) 6 years |
|---|---|---|---|---|---|---|---|
| Born in Mexico | 2.734*** | 2.798*** | 2.872*** | 2.764*** | 2.815*** | 2.687*** | 2.557*** |
| | (0.627) | (0.634) | (0.651) | (0.661) | (0.673) | (0.680) | (0.679) |
| Mean Dep. Var. | 2.691 | 2.652 | 2.619 | 2.580 | 2.552 | 2.524 | 2.498 |
| Observations | 262,271 | 260,298 | 258,005 | 255,407 | 252,353 | 249,123 | 245,591 |

| | (8) 5 years | (9) 4 years | (10) 3 years | (11) 2 years | (12) 1 year | (13) 0 years |
|---|---|---|---|---|---|---|
| Born in Mexico | 2.540*** | 2.442*** | 2.340*** | 1.657** | 1.980*** | 2.307** |
| | (0.693) | (0.712) | (0.737) | (0.662) | (0.751) | (1.010) |
| Mean Dep. Var. | 2.475 | 2.447 | 2.436 | 2.417 | 2.408 | 2.390 |
| Observations | 241,899 | 237,962 | 234,150 | 230,147 | 226,417 | 222,502 |

*Note:* Data come from the 2000 decennial Census and the 2001-2012 American Community Survey. The title of each column refers to the minimum age of arrival allowed for that column's sample. For example, column 3 restricts to individuals who were 10 years or younger when they arrived. The unit of observation is an individual who identifies as being of Mexican heritage aged 13-17. The outcome variable is a dummy variable for not being in school * 100. Age at arrival is calculated as the difference between a respondent's age and the number of years that she has been in the U.S. This variable is set to 0 for all U.S.-born individuals, including those born in Guam, Puerto Rico, and the Virgin Islands. All regressions include age, family, and female fixed effects.
Significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors are clustered at the family level.

Table 11: IV - Effect of Legal Status on Being Out of School

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| *Panel A: Noncitizen* | | | | | |
| Noncitizen | 2.985*** | 3.254*** | 3.252*** | 3.890*** | 3.047*** |
| | (0.216) | (0.731) | (0.730) | (0.799) | (0.803) |
| Female | | | -0.445** | -0.251 | -0.376** |
| | | | (0.175) | (0.183) | (0.176) |
| Noncitizen * Female | | | | -1.247** | |
| | | | | (0.629) | |
| *Panel B: Likely Undocumented* | | | | | |
| Likely Undocumented | 3.244*** | 3.713*** | 3.711*** | 4.411*** | 3.435*** |
| | (0.235) | (0.832) | (0.832) | (0.905) | (0.904) |
| Female | | | -0.443** | -0.249 | -0.372** |
| | | | (0.175) | (0.183) | (0.176) |
| Likely Undocumented | | | | -1.369** | |
| * Female | | | | (0.691) | |
| Family FE | | X | X | X | X |
| All Sibs Arrived in US By Age 6 | | | | | X |
| Mean Dep. Var. | 2.872 | 2.872 | 2.872 | 2.872 | 2.667 |
| Observations | 262,271 | 262,271 | 262,271 | 262,271 | 242,597 |

*Note:* Data come from the 2000 decennial Census and the 2001-2012 American Community Survey. The unit of observation is an individual who identifies as being of Mexican heritage aged 13-17. All individuals who are not citizens of the U.S. are classified as "noncitizen." An individual is classified as "likely undocumented" if she is both not a citizen and if no one in her household is receiving money from the Social Security Administration, on public welfare or a veteran of the U.S. military. The excluded instrument is a dummy variable for being born in Mexico. In the third column the sample is restricted to individuals who identify as being of Mexican heritage. The fourth column restricts to siblings in the sample that had moved to the U.S. by age 6. All regressions include age fixed effects.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the family level.

Table 12: Net Present Value Wage and Fiscal Impacts

| Category | Wage Losses | Fiscal Losses |
|---|---|---|
| Per-person | $8,455 | $4,257 |
| Single cohort | $728 million | $366 million |
| 75-year cost | $19.9 billion | $9.3 billion |

# Appendix A: Measurement Error Simulations (Not Intended For Publication)

The primary endogenous variable used in this paper, legal status as proxied by U.S. citizenship, is subject to measurement error. While this measurement error does not affect the reduced form results, it will influence the scaling of these results in the IV estimates.

For U.S.-born individuals, there is no measurement error in legal status. For foreign-born individuals,[39] there is measurement error only if they are recorded as noncitizens (*noncitizen=0*). To better understand the sensitivity of the results to measurement error, in this section, we introduce *more* measurement error into the data and explore by how much the results change.

To do this, Method 1 takes individuals who were born in Mexico and U.S. citizens and reassigns a random fraction of them as noncitizens. This method introduces extra measurement error that is, by construction, uncorrelated with other family or individual level characteristics, conditional on birthplace and reported legal status. In reality, however, the measurement error present in the Census data is likely to be correlated with some family-level characteristics. Measurement error is present when an individual is (or will likely be) able to work legally when he enters the labor force but is not a citizen. For example, children of permanent residents ("green card holders") who are not U.S. citizens can generally become permanent residents. Because some permanent residents gain their status through an employer and because doing so is easier for highly educated individuals, there may be more measurement error for children of high earning parents. Method 2 takes a first attempt at accounting for the correlation between family level characteristics and measurement error. In this procedure, we consider only individuals who were born in Mexico and classified as legal *and the household having above median family income for the sample*, calculated year-by-year, and randomly reassigns a random fraction of them as noncitizens. To ensure comparability between the two methods, for each simulation, we reassign an identical number of observations in both methods.

Both procedures do not change the instrument, the outcome variable, or any of the controls; these simulations do not alter the reduced form. Both methods do, however, change the first stage and tend to increase the magnitude of the first stage coefficient on the dummy variable

[39] To keep the discussion as simple as possible, in this section, we use the terms "Mexican born" and "foreign born" interchangeably. Results are essentially identical if we re-run the baseline regressions, restricting attention to only U.S. or Mexican-born individuals.

for being born in Mexico. Hence both methods result in two stage least squares estimates that are smaller than the baseline specification.

In both cases, however, the quantitative estimates are influenced relatively little by the introduction of this additional measurement error. For example, reassigning 40 percent of the observations that are born in Mexico and U.S. citizens as noncitizens decreases the baseline IV estimate by less than 10 percent (less than 5 percent) on average for Method 1 (for Method 2). While these simulations are crude, they do suggest that while the IV estimates are biased toward 0, the magnitude of the bias may be relatively small.

# Appendix B: Impact of Prenatal Environment on Schooling

Estimating the impact of birth environment on schooling required first measuring differences in the prenatal environment and then estimating from the existing literature how much of an impact those differences have on school attendance. Subsection 1 covers how we measure differences in the prenatal environment between Mexico and the U.S. Subsection 2 uses the existing literature to estimate how much these potential differences in the prenatal environment may affect educational attainment.

## B.1 Measuring Prenatal Environment for Mexico and the United States

Turning first to the difference in the prenatal environment between the U.S. and Mexico, we use birthweight, "a commonly used marker of infant health that can be seen as a summary proxy for prenatal conditions, [since] [l]ow birth weight is strongly associated with infant mortality and subsequent morbidity for infants who survive" (Currie and Rossin-Slater, 2015, pg. 212-213). We calculated mean birthweight and low birthweight rate for Mexico and the U.S. to estimate the percentage of our coefficient that might be explained by differences in prenatal conditions in the two countries. A child is classified as low birthweight if it weighs less than 2,500 grams.

***Births in Mexico***: Our Census sample has children born as early as 1983 and as late as 1999. The best available data for Mexican births in this year range came from a 1987 Demographic and Health Survey report conducted by the Health Ministry of Mexico and the

Institute for Resource Development/ Macro Systems, Inc. (1989, pg. 101). The survey contained data on 5,311 births from a representative sample of women of child-bearing age in Mexico. The births in the survey sample occurred between the years 1982 and 1987. While we could not find the original data set with the birthweights of these children, the report did contain a table listing the frequency of births that occurred at different ranges of birthweight. The bins that were not top or bottom-coded had widths of 500 grams; for these, we multiplied by midpoint of the bin by the fraction in the bin. For bottom-coded and top-coded bins, we assumed that their width was 500 g as well–and also imposed the constraint that there was no more of the population in that bin than the bin next to it. If there was more population in the top/bottom-coded bin than the one next to it, we created another bin of 500 g width with the excess density and used the midpoint of that.

Using this method, we calculated a mean birthweight for Mexico of 3,166 g. Schlaepfer and Infante (1995) used the original data set from the study above to estimate that the low birthweight rate in Mexico in that time was 11.1%.

***Births in the U.S.***: To calculate the mean birthweight of U.S. births, we used the birth certificate data from the National Vital Statistics System of the National Center for Health Statistics.[40] This data contains information from almost all births in the United States for each calendar year. For years after 1984, it contains birth certificate information for all births. The documentation file for each year's data contained a table similar to that of the Mexico data; it also reported bins of birthweight with a width of 500 g. We use the same method explained above to calculate mean birthweight. The mean birthweight calculated for the United States was 3,352 grams. Using these same tables, we calculated that the low birthweight rate in the United States at this time was 6.8%.

## B.2 Effect of Prenatal Environment on School Attendance

To estimate how much of our results might be explained by worse prenatal conditions in Mexico, we explored the literature on the association of prenatal conditions with later-life outcomes. To find studies, we first turned to literature reviews of studies that examined how prenatal

[40] Available at https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm.

conditions affected later life outcomes: Almond and Currie (2011), Currie and Vogl (2013), Currie and Rossin-Slater (2015), and Almond, Currie, and Duque (2017). Most economic studies have either used exogenous prenatal shocks (such as a pandemic or famine) or birth indicators of fetal health and conditions (almost always birthweight) to identify the effect of prenatal conditions on later life outcomes.

From these literature reviews, we include the studies that examined how prenatal conditions affected schooling outcomes and met two criteria. The first was that the study measured the relationship between birthweight and educational outcomes. We chose to focus on studies that used birthweight rather than prenatal shocks because it is difficult to extrapolate the findings from a particular famine or disease event to calculate how much of the our effect may be explained by persistent and more moderate differences in prenatal conditions between Mexico and the United States. Second, we included only studies that examined the effect between birthweight and years of school completed because this was the educational outcome our estimates were best equipped to address. This ruled out studies that measured birthweight's effect on a dummy for completing high school. Five studies met our criteria. We used these to produce estimates of how much of our estimated effect may be explained by the differences in the prenatal environments between Mexico and the United States.

[ [Appendix Table 5: Effect of Prenatal Environment on Years of Schooling ]

## Appendix C: Wage and Fiscal Impacts

We estimated the wage losses, fiscal impacts, and fiscal impacts taking descendants into account from our estimated effect of being undocumented on schooling. For each of these three categories of effects, we estimated: (1) present value effect for an individual undocumented 16-year old of Mexican heritage, (2) present value effect for the entire cohort of 2012 undocumented Mexican-born 16-year-olds, and (3) cost for all cohorts over the course of 75 years. We also estimated (2) and (3) using the 2012 cohort of undocumented Hispanic 16-year-olds. We calculated these figures starting from 2014, when the 16-year olds in the sample would have been 18. In the

following sections, we detail the steps taken and assumptions made to produce these estimates. All present value calculations used a discount rate of 3% following the National Academy of Sciences, Engineering and Medicine's *The Economic and Fiscal Consequence of Immigration* (2017). To convert all dollars to 2016 dollars, we used the Federal Reserve Bank of Minneapolis's yearly annual average CPI estimates in the United States. We estimated that number of likely undocumented 16-year olds of Mexican heritage and of Hispanic ethnicity in the United States in 2012 were 59,284 and 86,084, respectively, by using the sampling weights from the 2012 American Community survey and our measure of *likely undocumented*. (For the purpose of these calculations, we assume that the impact is similar for Mexican-heritage and non-Mexican-heritage Hispanics, which is borne out by our robustness check using all Hispanics, column (3) of Table 6.)

## C.1 Wage Losses

To estimate the wage losses from being undocumented, we multiply our estimate of the impact of being undocumented on years of schooling by an estimate of how much a year of schooling increases earnings. To estimate in turn how much a year of schooling increases earnings, we multiply an estimated baseline wage by a return to schooling. For the baseline wage, we use the the Bureau of Labor Statistics' average median weekly earning of full-time, Hispanic, wage and salary workers who were 25 years and older and who were less than high school graduates; this was $466 in 2014 dollars. We use this as our baseline wage for a less than high school graduate in all time periods. The literature has a consensus estimate of a return to schooling of a 10 percent increase in wages per year of schooling (Card 1999; Card 2001). This is consistent with estimates of the return to schooling for Hispanics (Barrow and Rouse 2005).

The expression $52 * 466 * 0.1$ represents the gains from an additional year of schooling for Hispanics with less than a high school degree. It evaluates to $2,430.20. Our estimated loss in years of schooling for Hispanics from noncitizenship is 0.130 years. As explained in the Section 7, to obtain this number we first averaged the *born in Mexico* coefficient from the age difference of 1 or less specification (Table 9, column 4) and the maximum age of arrival of 1 specification (Table 10, column 11). We then multiplied this number by five, the number of ages represented

by our sample of 13 to 17-year-olds, to obtain a years of schooling effect. From this number we subtracted the prenatal environment effect (0.015) then scaled it by the first stage estimate of *likely undocumented* on *born in Mexico* (Table 5, Panel B, column 3). As a result, we calculate $2,812 * 0.130 = 317.48$ as the mean yearly wage loss from being undocumented per undocumented Hispanic. This is the central figure used to calculate all three wage loss effects.

To calculate the present value of the lifetime wage loss from noncitizenship for an undocumented 16-year-old of Mexican heritage, we assumed the individual began working at age 18, lost $367.01 from lack of legal status every year in 2014 dollars, and stopped working at age 67. The first period of the calculation is 2014, the year in which the 16-year olds in 2012 turned 18. It is from 2014 that we start discounting. The present value of the wage loss per noncitizen Hispanic 16-year old in 2016 dollars is $8,455. Multiplying this figure by the number of 16-year-olds in 2012 yields a present value total loss from the lack of legal status for undocumented 16-year-olds of Mexican heritage of about $501 million in 2016 dollars. For all Hispanics, this figure was $727 million.

We calculated the present value 75-year cost by assuming that each subsequent cohort of people turning 18 would have the same number of people as the 2012 cohort of undocumented 16-year-olds of Mexican heritage and that each cohort would only work until age 67. The first period of the calculation is the year 2014, and the 75th and final period in the calculation is the year 2088. We assumed that each individual loses $317.48 each year from noncitizenship, discounting appropriately by the year. We multiplied this discounted wage loss by the number of cohorts of working age (1 in 2011, 2 in 2012, and so on until the year 2058, after which every year has 49 cohorts of working age) and by the number of undocumented 16-year olds of Mexican heritage in 2012. That gave the present value loss from lack of legal status for undocumented youth of Mexican heritage each year. Summing up the amount for every year yields a present value 75-year cost of $13.7 billion. The corresponding estimate for all Hispanics was $19.9 billion.

## C.2. Fiscal Impacts

We retrieved the figure for the difference in fiscal flows between an immigrant with less than a high school degree and an immigrant with a high school degree from *The Economic and Fiscal Consequences of Immigration* by the National Academies of Sciences, Engineering, and Medicine (2017, Table 8-12). The authors calculated the average 75-year net present value flows for consolidated federal, state, and local governments for immigrants by age of arrival and education level. We used the panel that calculated the averages using the assumptions in the CBO long-term budget outlook, did not include public goods, that used all available immigrants in their sample, and that were calculated for immigrants who arrived between the ages of 0 and 24.

In the averages that did not take the fiscal impact of descendants into account, the 75-year net present value fiscal benefit for an immigrant that arrived between the ages of 0 and 24 with less than a high school education was $32,000 in 2012 dollars. For an immigrant with only a high school education, this number was $157,000 in 2012 dollars, leading to a difference of $125,000 in 2012 dollars. When taking into account the fiscal impacts of descendants, an immigrant who arrived between the ages of 0-24 and had less than a high school degree had on average a 75-year net present value fiscal flow of $49,000 in 2012 dollars, while the corresponding immigrant with only a high school education had a 75-year net present value fiscal flow of $217,000 in 2012 dollars. The difference in the net flows, taking descendants into account, is $222,000 in 2012 dollars. On the one hand, we are de facto treating the observed difference in wages between non-high-school graduates and high school graduates as causal, which might be accurate and might overstate the effects. On the other hand, comparing only those without a high school degree with those with only a high school degree understates the impact of graduating from high school, since some of those graduates will go on to have some college or a college degree, and therefore earn substantially more.

To get the present value effect per undocumented 16-year old of Mexican heritage, we multiplied the difference in net flows between terminal high school graduates and less than high school graduates by an estimated probability of not completing high school. Up to now, we have produced figures on being in high school, but the available figures are on *graduating* from

high school. To produce a lower bound estimate of the impact on graduating from high school, we look just at those greater than 16 years or older, as more representative of graduating from high school than the entire sample including the impact of being in school as a 13-year-old. Thus, to estimate a probability of not completing high school, we use a specification that looked at sibling age differences of 1 or less, similar to that in Table 9, but also include a term that interacted being born in Mexico with being 16 and older, similar to column (1) of Table 6. We included the same interaction term in a regression that restricted the sample to children who had arrived by age 1 or less. We averaged the sum of the main effect and the interaction effect of being born in Mexico across these two regressions to obtain a conservative estimate for the effect of being born in Mexico on the probability of finishing high school. This estimated probability was 0.0326, which implied a per-undocumented-youth-of-Mexican-heritage, 75-year, net present value fiscal loss from lack of legal status in 2016 dollars of $7,560 when taking descendants into account and $4,257 without taking them into account.

Multiplying these numbers by the number of undocumented 16-year olds of Mexican heritage in 2012 gave the estimated net present value fiscal loss from lack of documentation for the cohort of $488 million and $252 million 2016 dollars with descendants and without descendants, respectively. For all undocumented Hispanics, the corresponding estimates were $651 million and $366 million. Finally, we calculated the 75-year total cost by once again assuming that each subsequent cohort of undocumented youth of Mexican heritage turning 16 would have the same number of people as the 2012 cohort 16-year-olds, and that they would have the same 75-year net present value impact of the 2012 cohort. We also assumed that the fiscal impacts of the immigrants were constant each year; that is, we ignored life cycle differences in fiscal flows and assumed that the $222,000 figure for descendants, for example, represented the sum of a constant flow appropriately discounted each year of the 75-year period. Formally, we assumed that in the following 75-year summation of fiscal flow impact, where $x_t$ represents the net fiscal flow of the immigrant at year $t$,

$$\sum_{t=0}^{74} \frac{x_t}{(1.03)^t}$$

$x_t = x$ for all $t$. To find $x$ we solved for it using formula for the partial sum of a geometric

series. Using the $222,000 figure for descendants,

$$x = \frac{1 - \frac{1}{1.03}}{1 - \left(\frac{1}{1.03}\right)^{75}} * 222,000 = 7,257$$

The corresponding figure using the estimate that does not include descendants is $4,086. So we assumed that the effects from lack of legal status per year and per 16-year old including descendants and not including descendants were $7,257 and $4,086, respectively and in 2012 dollars. We calculated the total present value loss each year for the 75-year period by multiplying the per year loss, discounted appropriately, by the number of undocumented 16-year olds of Mexican heritage in 2012, by our estimated undocumented effect of 0.00354, and by the number of relevant cohorts. The number of relevant cohorts increases by one each year. In 2014, there is one; in 2015, there are two, and so on until there are 75 cohorts in 2088. By adding up the present value of the loss for each year, we reach the total net present value 75-year cost of being undocumented estimates of $11.4 billion with descendants and $6.4 billion without descendants in 2016 dollars. For all Hispanics, the estimated costs are of $16.6 billion and $9.3 billion, respectively

## Appendix References


Almond, D. & Currie, J. (2011). Killing Me Softly: The Fetal Origins Hypothesis. *Journal of Economic Perspectives*, 25(3), 153-172.

Almond, D., Currie, J., & Duque, V. (2017). Childhood Circumstances and Adult Outcomes: Act II. *NBER Working paper,* 23017

Behrman, J., & Rosenzweig, R. (2004). Returns to Birth Weight. *Review of Economics and Statistics,* 86 (2), 586–601.

Card, D. (1999). The Causal Effect of Education Earnings. In O. Ashenfelter and D. Card, *The Handbook of Labor Economics,* Volume 3, Amsterdam: North Holland Press.

Card, D. (2001). Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems. *Econometrica,* 69 (5), 1127-1160.

Currie, J. & Rossin-Slater, K. (2015). Early-Life Origins of Life-Cycle Well-Being: Research and Policy Implications. *Journal of Policy Analysis and Management,* 34 (1), 208-242.

Currie, J. & Vogl. (2013). T. Early-Life Health and Adult Circumstance in Developing Countries. *Annual Review of Economics,* 5, 1-36.

Figlio, D., Guryan, J., Karbownik, K., & Roth, J. (2014). The Effects of Poor Neonatal Health on Children's Cognitive Development. *The American Economic Review,* 104 (12), 3921-3955.

Health Ministry of Mexico and the Institute for Resource Development/Macro Systems, Inc. (1989). Encuesta Nacional sobre Fecundidad y Salud 1987. Mexico, D.F.

Rosenzweig, M. (2013). Economic Growth, Comparative Advantage, and Gender Differences in Schooling Outcomes: Evidence from the Birthweight Differences of Chinese Twins. *Journal of Development Economics,* 104, 245-260.

Royer, H. (2009). Separated at Girth: US Twin Estimates of the Effects of Birth Weight. *American Economic Journal: Applied Economics*, 1 (1), 49-85

Schlaepfer, L. & Infante, C. (1995) Bajo Peso al Nacer en Mexico: Evidencias a Partir de Una Encuesta Retrospectiva a Nivel Nacional. *Boletin Medico del Hospital Infantil de Mexico,* 52 (3), 168-179

Appendix Table 1: Number of Observations by Year Country of Birth and Citizenship Status

| | Born USA | | Born Mexico | | Born Other | | Total |
|---|---|---|---|---|---|---|---|
| Year | Citizen | Noncitizen | Citizen | Noncitizen | Citizen | Noncitizen | |
| 2000 | 56,954 | 0 | 2,224 | 13,801 | 276 | 96 | 73,351 |
| 2001 | 4,010 | 0 | 130 | 815 | 21 | 11 | 4,987 |
| 2002 | 3,773 | 0 | 133 | 744 | 21 | 15 | 4,686 |
| 2003 | 4,808 | 0 | 147 | 907 | 29 | 9 | 5,900 |
| 2004 | 4,767 | 0 | 128 | 840 | 28 | 7 | 5,770 |
| 2005 | 14,820 | 0 | 409 | 2,585 | 74 | 30 | 17,918 |
| 2006 | 15,781 | 0 | 344 | 2,687 | 60 | 24 | 18,896 |
| 2007 | 16,377 | 0 | 427 | 2,771 | 59 | 21 | 19,655 |
| 2008 | 18,297 | 0 | 377 | 2,612 | 55 | 33 | 21,374 |
| 2009 | 18,702 | 0 | 432 | 2,622 | 50 | 18 | 21,824 |
| 2010 | 19,331 | 0 | 371 | 2,582 | 50 | 38 | 22,372 |
| 2011 | 19,611 | 0 | 386 | 2,663 | 53 | 25 | 22,738 |
| 2012 | 19,842 | 0 | 400 | 2,485 | 46 | 30 | 22,803 |
| Total | 217,073 | 0 | 5,908 | 38,114 | 822 | 357 | 262,274 |

*Note:* Sample sizes by year, country of birth, and legal status are reported. Data comes from the 2000 decennial Census and the 2001-2012 American Community Survey. The unit of observation is an individual who identifies as being of Mexican heritage aged 13-17.

Appendix Table 2: Summary Statistics by Legal Status, Cross Section

| | Full Pop | Citizen of U.S. | Noncitizen | Likely Undocumented |
|---|---|---|---|---|
| Age | 14.914 | 14.907 | 14.954 | 14.950 |
| | (1.407) | (1.409) | (1.394) | (1.393) |
| Female | 0.487 | 0.489 | 0.476 | 0.475 |
| | (0.500) | (0.500) | (0.499) | (0.499) |
| Noncitizen | 0.150 | 0.000 | 1.000 | 1.000 |
| | (0.357) | (0.000) | (0.000) | (0.000) |
| Likely Undocumented | 0.138 | 0.000 | 0.919 | 1.000 |
| | (0.345) | (0.000) | (0.273) | (0.000) |
| Speaks English Very Well | 0.882 | 0.916 | 0.690 | 0.690 |
| | (0.323) | (0.277) | (0.463) | (0.462) |
| Not in School | 0.029 | 0.024 | 0.053 | 0.053 |
| | (0.167) | (0.154) | (0.224) | (0.223) |
| Num Obs. Per Family in Sample | 1.511 | 1.511 | 1.512 | 1.503 |
| | (0.669) | (0.667) | (0.678) | (0.667) |
| Personal Care Problem | 0.007 | 0.007 | 0.006 | 0.005 |
| | (0.082) | (0.083) | (0.074) | (0.072) |
| Physical Difficulty | 0.010 | 0.010 | 0.009 | 0.009 |
| | (0.100) | (0.101) | (0.094) | (0.093) |
| Age At Arrival | 0.876 | 0.106 | 5.232 | 5.254 |
| | (2.444) | (0.861) | (3.641) | (3.645) |
| Observations | 262,271 | 223,803 | 38,468 | 34,139 |

*Note:* Mean and standard deviation reported. Data come from the 2000 decennial Census and the 2001-2012 American Community Survey. The unit of observation is an individual who identifies as being of Mexican heritage aged 13-17. All individuals who are not citizens of the U.S. are classified as "noncitizen." An individual is classified as "likely undocumented" if she is both not a citizen and if no one in her household is receiving money from the Social Security Administration, on public welfare or a veteran of the U.S. military. Age at arrival is calculated as the difference between a respondent's age and the number of years that she has been in the U.S. This variable is set to 0 for all U.S.-born individuals, including those born in Guam, Puerto Rico, and the Virgin Islands.

Appendix Table 3: Summary Statistics by Place of Birth of Siblings, Cross Section

| | Full Pop | All Sibs born U.S. | Some Sibs Born U.S., Some Mexico | All Sibs Born Mexico |
|---|---|---|---|---|
| Age | 14.914 | 14.912 | 14.955 | 14.914 |
| | (1.407) | (1.408) | (1.470) | (1.386) |
| Female | 0.487 | 0.488 | 0.509 | 0.475 |
| | (0.500) | (0.500) | (0.500) | (0.499) |
| Noncitizen | 0.150 | 0.000 | 0.415 | 0.883 |
| | (0.357) | (0.000) | (0.493) | (0.322) |
| Likely Undocumented | 0.138 | 0.000 | 0.364 | 0.815 |
| | (0.345) | (0.000) | (0.481) | (0.388) |
| Speaks English Very Well | 0.882 | 0.920 | 0.809 | 0.697 |
| | (0.323) | (0.271) | (0.393) | (0.460) |
| Not in School | 0.029 | 0.024 | 0.033 | 0.051 |
| | (0.167) | (0.154) | (0.178) | (0.220) |
| Num Obs. Per Family in Sample | 1.511 | 1.495 | 2.314 | 1.428 |
| | (0.669) | (0.660) | (0.541) | (0.630) |
| Personal Care Problem | 0.007 | 0.007 | 0.007 | 0.006 |
| | (0.082) | (0.083) | (0.081) | (0.074) |
| Physical Difficulty | 0.010 | 0.010 | 0.010 | 0.009 |
| | (0.100) | (0.101) | (0.100) | (0.093) |
| Age At Arrival | 0.876 | 0.000 | 1.272 | 5.324 |
| | (2.444) | (0.000) | (2.634) | (3.593) |
| Observations | 262,271 | 212,891 | 7,851 | 40,102 |

*Note:* Mean and standard deviation reported. Data come from the 2000 decennial Census and the 2001-2012 American Community Survey. The unit of observation is an individual who identifies as being of Mexican heritage aged 13-17. An individual is classified as "All Sibs born U.S." if all individuals in the family that are included in the sample are born in the USA. "Some Sibs born U.S., some Mexico" means that at least one member of his family in the sample is born in the U.S. and at least one member of the family in the sample is born in Mexico. "All Sibs born Mexico" means that all members of the family in the sample are born in Mexico. Note that these three groups are not mutually exhaustive, and the sample sizes from the first column need not equal the sum of the sample sizes from the last three columns; there are some individuals who are not classified as "All sibs born U.S.", "Some sibs born U.S., some Mexico", and "All sibs born Mexico". In particular, an individual in a family where one sibling in the sample was born outside of Mexico and the United States and no sibling was born in the U.S. would be included in the full sample but not any of the three subgroups. All individuals who are not citizens of the U.S. are classified as "noncitizen." An individual is classified as "likely undocumented" if she is both not a citizen and if no one in her household is receiving money from the Social Security Administration, on public welfare or a veteran of the U.S. military. Age at arrival is calculated as the difference between a respondent's age and the number of years that she has been in the U.S. This variable is set to 0 for all U.S.-born individuals, including those born in Guam, Puerto Rico, and the Virgin Islands.

Appendix Table 4: Summary Statistics and Regression Coefficients for Mexican born Individuals by Birthplace of Siblings

| | Summary Statistics | | | Regression Estimates with Family FE |
|---|---|---|---|---|
| | Full Pop | Some Sibs U.S. Born | No Sibs U.S. Born | Some Sibs U.S. Born |
| Age | 14.978 | 15.619 | 14.914 | - |
| | (1.395) | (1.330) | (1.386) | - |
| Female | 0.478 | 0.510 | 0.475 | 0.035*** |
| | (0.500) | (0.500) | (0.499) | (0.013) |
| Noncitizen | 0.879 | 0.842 | 0.882 | -0.033*** |
| | (0.326) | (0.364) | (0.322) | (0.010) |
| Likely Undocumented | 0.808 | 0.739 | 0.815 | -0.069*** |
| | (0.394) | (0.439) | (0.388) | (0.013) |
| Speaks English Very Well | 0.705 | 0.784 | 0.697 | 0.073*** |
| | (0.456) | (0.412) | (0.460) | (0.011) |
| Not in School | 0.051 | 0.053 | 0.051 | -0.015*** |
| | (0.220) | (0.223) | (0.220) | (0.006) |
| Num Obs. Per Family in Sample | 1.508 | 2.300 | 1.429 | 0.862*** |
| | (0.671) | (0.531) | (0.631) | (0.018) |
| Personal Care Problem | 0.006 | 0.008 | 0.006 | 0.002 |
| | (0.075) | (0.087) | (0.074) | (0.002) |
| Physical Difficulty | 0.009 | 0.011 | 0.009 | 0.002 |
| | (0.094) | (0.103) | (0.093) | (0.003) |
| Age At Arrival | 5.075 | 2.582 | 5.324 | -2.926*** |
| | (3.650) | (3.272) | (3.593) | (0.098) |
| Observations | 44,019 | 3,887 | 40,132 | |

*Note:* Mean and standard deviation reported for the first three columns. Regression coefficient and standard error reported for fourth column. Regression outcome variable is given by the row heading. Regressions include a dummy variable for having some sibling in the sample born in the U.S. and age fixed effects. Because these regressions control for age effects, the mean difference in age is not identified. Data come from the 2000 decennial Census and the 2001-2012 American Community Survey. The unit of observation is an individual who identifies as being of Mexican heritage aged 13-17. All individuals who are not citizens of the U.S. are classified as "noncitizen." An individual is classified as "likely undocumented" if she is both not a citizen and if no one in her household is receiving money from the Social Security Administration, on public welfare or a veteran of the U.S. military.

Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are clustered at the family level.

Appendix Table 5: Estimates of Effect of Birthweight on Years of Education

| Study | Predicted Effect | Sample Size | Sample, Years of Birth |
|---|---|---|---|
| Behrman and Rosenzweig (2004) | 0.104 | 804 | Minnesota, 1936-1955 |
| Currie and Moretti (2007) | 0.0043 | 536,077 | California, 1970-1974 |
| Royer (2009) | 0.0288 | 6,792 | California, 1960-1982 |
| Rosenzweig and Zhang (2013) | 0.0497 | 1,872 | China, 1973-1984 |
| Figlio, Guryan, Karbownik, and Roth (2014) | 0.0281 | 42,212 | Florida, 1992-2002 |
| Mean (Weighted by square root of sample size) | 0.0154 | | |

Notes: The predicted effect is the average amount of more years of schooling that we would expect to see in the United States than Mexico given the study's findings and the average birthweight and low birthweight rates in the two countries. The difference in average birthweights between the two countries was about 180 grams for children born in 1982-1987. The difference in low birthweight rate between the two countries was about 4.3 percentage points in the period 1982-1987

Appendix Figure 1