2016

Age of Immigration and Adult Labor Market Outcomes: Childhood Environment in the Country of Origin Matters

Aaron Watson McCartney
Oberlin College

Age of Immigration and Adult Labor Market Outcomes: Childhood Environment in the Country of Origin Matters

Aaron W. McCartney
Oberlin College Honors Seminar
2015-2016

This paper builds on previous studies that have examined the effect of age of immigration on adult labor market outcomes by considering the potential impact of the childhood environment in the country of origin. 2000 United States Census data and historical child mortality data is used to quantify the impact of the childhood environment in the country of origin on the effect of age of immigration on adult labor market outcomes. Results from children who immigrated to the United States between ages zero and ten indicate that the impact of age of immigration on adult labor market outcomes is more negative for immigrants arriving from countries with poor childhood environments.

I thank Martin Saavedra and Barbara Craig for their helpful comments on my initial draft.

I. Introduction

Immigration is one of the most hotly debated topics in the United States, and it undoubtedly has economic impacts on the people migrating to the United States, as well as on people already residing in the United States. The nature of this economic impact depends on a multitude of factors, including the type of people immigrating, where they are immigrating from, and at what age they immigrate. Unsurprisingly, economists have studied immigration extensively, with numerous studies having examined the effect of immigration on both immigrants and the native population.

This study examines the impact of age of immigration on adult wages among people who immigrated to the United States as children. This is one of the lesser-studied immigration related topics, and this paper builds on the existing literature by considering how the childhood environment in the country of origin contributes to the effect of age of immigration. The motivation for considering the potential impact of the childhood environment in the country of origin developed out of the existing literature which suggests that childhood health can significantly impact adult labor market outcomes. In the context of immigration, the effect of the childhood environment in the country of origin is interesting to consider as immigrants arrive to the United States from all over the world and from vastly different childhood environments. The existing literature suggests that exposure to different childhood environments could have a significant impact on the adult wages of immigrants who arrived to the United States as children, despite having spent the majority of their life in the United States. The purpose of this study is to answer the following:

How does the childhood environment in the country of origin impact the effect of age of immigration on adult labor market outcomes?

This study uses data on children who immigrated to the United States between the ages of zero and ten to answer this question. The next section contains a review of other studies that have examined age of immigration effects. The third section introduces the datasets used in this study. The fourth section presents the econometric models. The fifth section presents results, and the final section is a conclusion.

II. Literature Review

The majority of papers that examine the economic effects of immigration on immigrants focus on economic assimilation (see Chiswick 1978). The potential impact of age of immigration is briefly considered in some of these papers, but in general, the effect of age of immigration on labor market outcomes is not a topic that has been examined extensively.

Among those who considered age at immigration, Joseph Schaafsma and Arthur Sweetman focused exclusively on the impact of age of immigration in their paper "Immigrant earnings: age at immigration matters". This study used 1986, 1991, and 1996 Canadian Census data to examine what impact age of immigration had on the difference between observed immigrant earnings and the earnings predicted by the Canadian-born age-earnings profile. They found that people who immigrated after age 35 earned less than those who immigrated between the ages of

0 and 5, and that the magnitude of this effect was greater for people who arrived between ages 45 and 64[1]. They also observed that people who immigrated in their late teens had lower earnings than people who immigrated at slightly younger or older ages. The explanation they presented for this is that people who immigrated as late teens were less educated on average compared to those who immigrated at slightly younger or older ages. The authors argue that this is most likely due to the process of immigrating as a late teen prematurely ending formal schooling[2].

Olaf Åslund, Anders Böhlmark, and Oskar Nordström Skans conducted a study in which they investigated the effect of age of immigration among immigrants who arrived to Sweden as young children. Their study focused on examining the effect of age of immigration on social integration in the residential, labor, and marriage markets. Their dataset consisted of individuals born between 1960 and 1971 who had immigrated to Sweden before the age of 15 or whose parents had immigrated to Sweden at most 10 years before their birth, and they measured the average adult outcomes for these individuals between ages 31 and 34. They measured labor market outcomes by looking at employment and wages, and they measured social integration by looking at the proportion of foreign-born people within their environment (workplace, neighborhood, etc.). The strongest result they found was a positive relationship between age of migration and workplace, neighborhood, and marriage segregation. They also found a negative impact on employment that was significant for those arriving after age 6. The estimated magnitude of this effect was that

[1] Schaafsma and Sweetman, page 1077.
[2] Schaafsma and Sweetman, page 1083

increasing age of arrival by 10 years results in a 6% lower probability of being employed, and this effect tends to be larger among males[3].

The study presented in this paper differs from the above studies by examining people who immigrated to the United States. George Borjas previously investigated the effect of age of immigration on adult wages among people who immigrated to the United States in his paper "Assimilation and Changes in Cohort Quality Revisited: What Happened to Immigrant Earnings in the 1980's?". Borjas focused primarily on immigrant assimilation in this study, but also considered age of immigration. The study was conducted using 1970, 1980, and 1990 Public Use Samples of the US Census, and found a significant negative effect of age of immigration on immigrant earnings. Borjas found that the earnings of an immigrant who arrives at age 30 are approximately 5% lower than someone who arrives at age 20.[4]

The study presented in this paper builds on previous studies that have examined the effects of age of immigration by considering the potential impact of the childhood environment in the country of origin among immigrants who arrived as young children. Childhood environment has a large influence on the health of children in a country, and the motivation for this study arose from the existing literature indicating that childhood health significantly impacts adult labor market outcomes. One such article is "The Impact of Childhood Health on Adult Labor Market Outcomes" by James Smith. This study used data from the Panel Survey of Income Dynamics, a survey that followed groups of siblings for upwards of thirty years and tracked

[3] Åslund, Bohlmark, and Skans, pages 19-22.
[4] Borjas, page 227.

income, education, wealth, and health. The results of this study indicate that poor childhood health can have significant negative effects on adult labor market outcomes, including 13% lower household income and 12% lower individual income. Additionally, they found that this disadvantage in adult labor market outcomes is attributable to both lower initial wages (measured at age 25), and slower income growth thereafter.[5]

III. Datasets

The dataset used for the main regressions in this study consists of data compiled from three sources. The majority of the data is 2000 United States Census data from the Integrated Public Use Microdata Series (IPUMS).[6] The dataset from IPUMS contains data on age, sex, race, educational attainment, income, number of weeks worked in the last year, employment status, birthplace, year of immigration, and number of years since immigrating to the United States.

This study focuses on individuals who immigrated to the United States within the first 10 years of their life, and who were between age 25 and 64 in the year 2000. As a result, non-immigrants, people who immigrated after age 10, and people outside of the designated age range were removed from the dataset. Additionally, consistent with both the Borjas, and Schaafsma and Sweetman studies, this study is examining the wages of people who are currently working, so people with zero income were also removed. After removing these observations, the final U.S. Census dataset contained 120,929 people who immigrated to the United States from 147 different countries.

---

[5] Smith, pages 483-486.
[6] Ruggles, Alexander, Genadek, Goeken, Schroeder, and Sobek.

The second main dataset used in this study contains child mortality data from Gapminder. This dataset contains historical records of the number of deaths before the age of five per 1000 live births from 261 countries. This data was used to create a variable corresponding to the child mortality rate for each year between 1936 and 1975 for each country with at least one immigrant in the dataset. This variable for child mortality in the year of birth is used as the measure of childhood environment. I choose to use child mortality as the measure of the childhood environment as the factors that contribute to the child mortality rate, such as access to healthcare or clean water, are the same factors that one would use to assess the childhood environment. The value of this variable ranged from 10.02 (Sweden in 1975) to 576.88 (Ukraine in 1943). Child mortality in Ukraine rose from 204.2 in 1940 to the high in 1943 during World War II, and then fell back to 158.5 in 1946.[7] Figure 1 is a plot of the minimum, 25th percentile, 50th percentile, 75th percentile, and maximum child mortality values across all the countries in the dataset for each year from 1936 to 1975. I merged this dataset with the U.S. Census dataset, and the number of observations in the resulting dataset was 87,508. The number of observations in this dataset is less than the number of observations in the U.S. Census dataset due to missing child mortality data for some years in some countries.

The last component of the final dataset is data on the most commonly spoken language in every country of origin. The CIA World Factbook lists the most commonly used language in each country, and I used this data to construct a dummy variable for

---

[7] Johansson, Lindgren, Johansson, and Rosling.

each country indicating whether English is the most commonly spoken language. Classifying countries as English or non-English speaking is not a straightforward task, as the most common language can vary within a country by region or class. For example, Hindi is the most commonly used language in India, with 41% of the population speaking Hindi, but English is an important language for political and commercial communication, and is widely spoken among the upper class.[8] In order

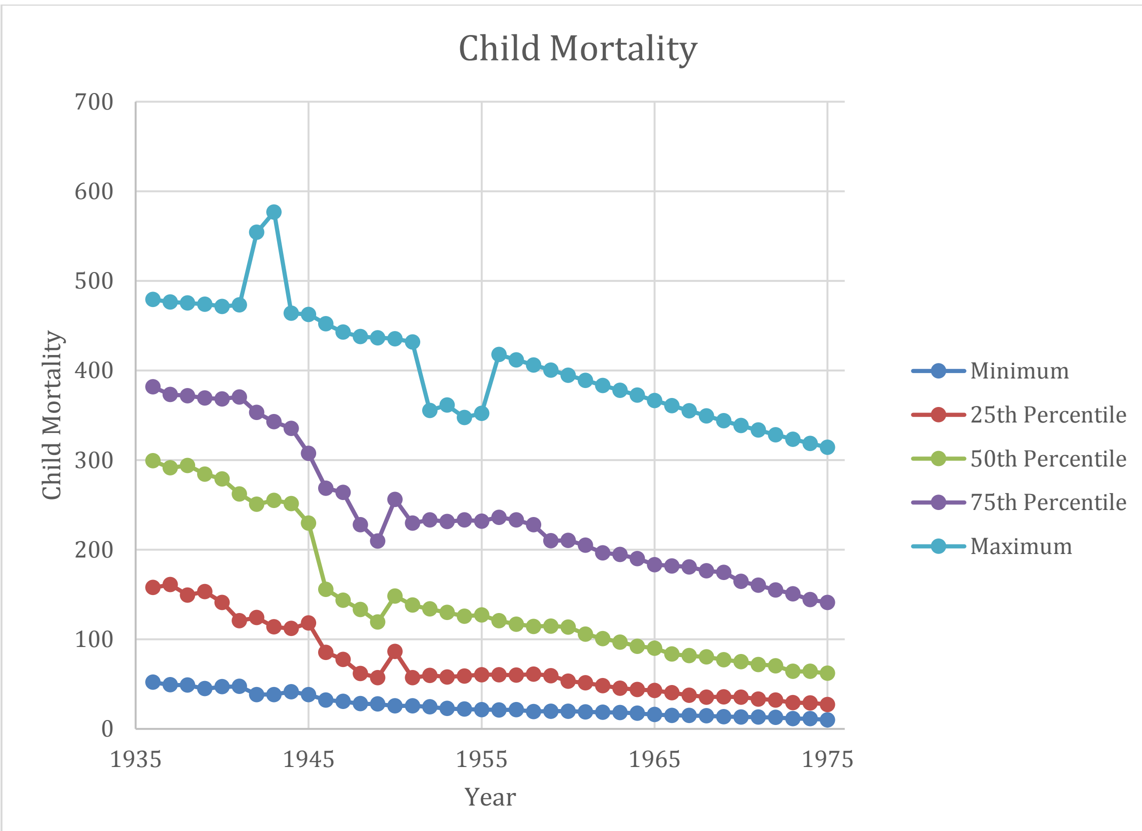

Figure 1. Historical Child Mortality Trends (measured as the number of deaths before the age of five per 1000 live births).[9]

[8] *The World Factbook* 2014-15.

[9] The minimum child mortality value was in either the Netherlands, Sweden, Switzerland, Norway, or Australia for every year in this study, and Sweden had the lowest child mortality in 35 out of 40 years. The maximum child mortality value was in either Sierra Leon, Ukraine, Afghanistan, Cameroon, or Egypt for every year in this study, and Sierra Leon had the highest child mortality 28 out of 40 years. There was more variability among the countries that had the child mortality rate at the 25th, 50th, and 75th percentile.

to be consistent, I classified countries as English speaking only if the CIA listed English as the most common language. One potential downfall to this approach is that the CIA does not have historical data on the most commonly used language by country, so I made all language classifications using 2015 data under the assumption that the most common language in a country is stable over time. An alternative approach would have been to classify countries as English speaking if the official language of the country was English. Historical data is available on official languages, but this classification method is flawed as some countries, such as Botswana, have English as the official language while only a small minority of the population speaks English. Tables 1, 2, 3, and 4 contain summary statistics for the final dataset used in this study.

Table 1. Summary Statistics on Child Mortality, Age of Immigration, Year of Birth, and Year of Immigration.

	Mean	Standard Deviation	Min	Max
Child Mortality	88.41	62.37	10.02	576.88
Age of Immigration	4.57	3.31	0	10
Year of immigration	1968.75	9.38	1936	1985
Year of birth	1964.18	8.80	1936	1975
Observations	87508			

Table 2. Sex Distribution.

Sex Distribution	Percent
Male	52.70
Female	47.30
Total	100
N	87508

Table 3. Racial Distribution Summary Statistics.

Race Distribution	Percent
White	57.26
Black/Negro	5.35
Chinese	2.14
Japanese	.83
Other Asian or Pacific Islander	12.17
Other race, nec	17.01
Two Major Races	4.97
Three or more major races	.27
Total	100
N	87508

Table 4. English vs. Non-English Speaking Distribution.

Distribution of Country Languages	Percent
Non-English	80.89
English	19.11
Total	100
N	87508

IV. Empirical Design

The objective of this study is to determine how the childhood environment in the country of origin impacts the effect of age of immigration on adult labor market outcomes. The empirical framework used in this study is based on the general difference-in-differences model. In the typical difference-in-differences model, outcomes are measured for two groups in two time periods, and one of the groups undergoes some treatment in the second period while the other group does not. The effect of that treatment can then be measured as the difference of the outcomes in the second period minus the difference of the outcomes in the first period. The model used in this study differs slightly from the general difference-in-differences framework as labor market outcomes are only measured in the year 2000, so the difference-in-differences is measured across age of arrival instead of time. To get a sense of the model used in this study, consider the hypothetical case of comparing the effect of age of arrival between immigrants arriving from two different counties. This scenario is depicted in Figure 2.

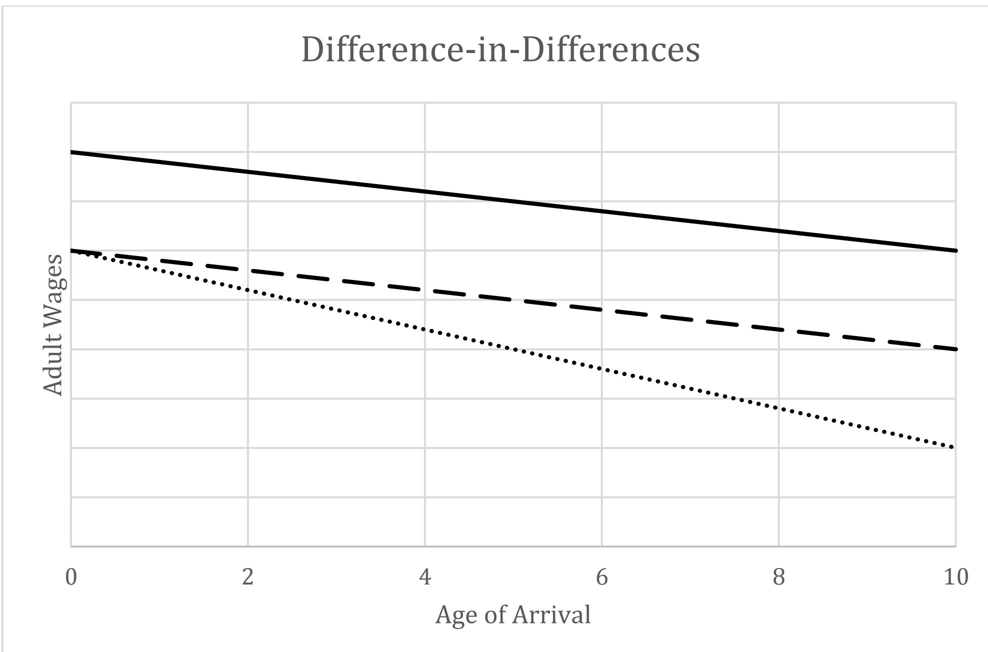

Figure 2. Difference-in-Differences Framework.

For this example, let the solid line indicate the relationship between age of arrival and adult wages for immigrants arriving from a country with low child mortality, and let the dotted line indicate the relationship between age of arrival and adult wages for immigrants arriving from a country with high child mortality. The lines for the two countries would be parallel if the childhood environment in the country of origin had no impact on the effect of age of immigration on adult wages (represented by the dashed line), while a difference in slopes would indicate the impact of the childhood environment. Thus the impact of the childhood environment in the country of origin on the effect of age of immigration on adult wages can be measured as the difference between the lines for the two countries at a given age of arrival minus the difference between the lines at age 0. The main identifying assumption implicit in this model is that the trends would be parallel if the childhood environment in the country of origin has no impact on the effect of age of immigration on adult wages.

The specification of this model is in the following form:

$$\ln(incwage_i) = \alpha + \beta_1[sex_i] + \beta_2[race_i] + \beta_3[bpl_i] + \beta_4[ageimmig_i] + \beta_5[byr_i] + \beta_6[cmort_{it}] + \beta_7[(cmort_{it} \cdot ageimmig_i)] + \beta_8[(English \cdot ageimmig)_i] + u_i \quad (1)$$

Where:

- *incwage* is total pre-tax wage and salary income of person i.
- *sex* is a dummy variable indicating the sex of person i.
- *race* is the vector of race dummy variables of person i.
- *bpl* is the vector of birthplace dummy variables for person i.
- *ageimmig* is the age of immigration of person i.
- *byr* is the vector of birth year dummy variables for person i.
- *cmort* is the child mortality in the birthplace of person i in their year of birth.
- *cmort*ageimmig* is an interaction term between the child mortality in the birthplace of person i in their year of birth and the age of immigration of person i.
- *English*ageimmig* is an interaction term between a dummy variable indicating if person i is from a country where English is the most commonly spoken

language and the age of immigration of person i. The dummy variable *English* is assigned a value of one if the country of origin is an English speaking country.

- u_i is the error term for person i.

Year of immigration and age are collinear in this model as all the results are measured in the year 2000. Age is collinear with birth year, and year of immigration is collinear with birth year and age of immigration. Additionally, I clustered standard errors by birthplace to account for the fact that error terms may not be independent for immigrants from the same country (possibly due to country specific characteristics).

The coefficient of interest is β_7. The literature on the effects of childhood health on adult wages suggests that exposure to poor childhood conditions will hinder adult labor market outcomes, and thus this coefficient is expected to be negative and significant. Including the interaction term between a dummy variable indicating if English is the most commonly spoken language in the country of origin and age of arrival aims to control for the effects of language skills on adult wages. Bleakley and Chin demonstrated that English-language skills have substantial positive effects on education and labor market outcomes.[10] This interaction term controls for the possibility that countries with high child mortality may be less likely to be English speaking countries, so the effects of immigrating later may be due to lacking language skills rather than longer exposure to a poor childhood environment. That being said,

[10] Bleakley and Chin, page 487.

it would not be surprising if the coefficient on this interaction term is not significant. This is due to the fact that this study examines people who arrived to the United States as young children, and the critical period hypothesis suggests that children will likely be able to reach native ability in a language if they are exposed to it at a young age.[11]

One drawback of this model is that data on the parents was not available, so this model is unable to control for parental education or income. This is potentially problematic as a parent's education or income is almost certainly correlated with their child's adult labor market outcomes. If these variables are also correlated with the interaction between age of arrival and child mortality then the parallel trend assumption implicit in the difference-in-differences model may be violated.

V. Results

The results of the main regression are presented in Table 5 on page 15. The first column of table 5 is the main regression without either of the child mortality variables, and the second column is the main regression with all of the variables. The coefficients on birthplace and birth year are not presented in the table due to the large number of variables, and due to the fact that the coefficients on these variables did not present any particularly unexpected results. Mexico was the category that was excluded for the set of birthplace dummy variables because it was the country that had by far the most immigrants in this sample. The coefficients on the majority of the birthplace dummy variables were significant and positive, indicating that immigrants from most countries experience better adult labor market outcomes when compared to otherwise identical immigrants from Mexico. Bosnia, Laos, and Tonga were among

[11] Bleakley and Chin, page 482.

Table 5. Main results.

	(1) Without Child Mortality	(2) With Child Mortality
Female	-0.427***	-0.427***
	(0.0261)	(0.0261)
Black/Negro	-0.00508	-0.00380
	(0.0272)	(0.0270)
Chinese	0.225***	0.223***
	(0.0511)	(0.0504)
Japanese	0.137***	0.131***
	(0.0168)	(0.0169)
Other Asian or Pacific Islander	0.161***	0.161***
	(0.0255)	(0.0255)
Other race, nec	-0.0282	-0.0287
	(0.0207)	(0.0202)
Two major races	-0.0526*	-0.0524**
	(0.0202)	(0.0200)
Three or more major races	-0.133	-0.132
	(0.0748)	(0.0747)
Age of Arrival	-0.00874*	-0.00329
	(0.00422)	(0.00459)
Child Mortality		0.0000918
		(0.000345)
Child Mortality*Age of Arrival		-0.0000528*
		(0.0000207)
English Speaking Country*Age of Arrival	0.0106*	0.00735
	(0.00477)	(0.00457)
Birthplace Controls	Yes	Yes
Birth Year Controls	Yes	Yes
Observations	87508	87508
Adjusted R^2	0.116	0.116

Clustered standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

the countries that had coefficients that were negative and significant. 1936 was the category that was excluded for the set of birth year dummy variables. Figure 3 plots the coefficients for birth year with age on the horizontal axis, which was calculated as 2000-birth year. This plot shows that the birth year coefficients demonstrate the expected age-earnings profile.

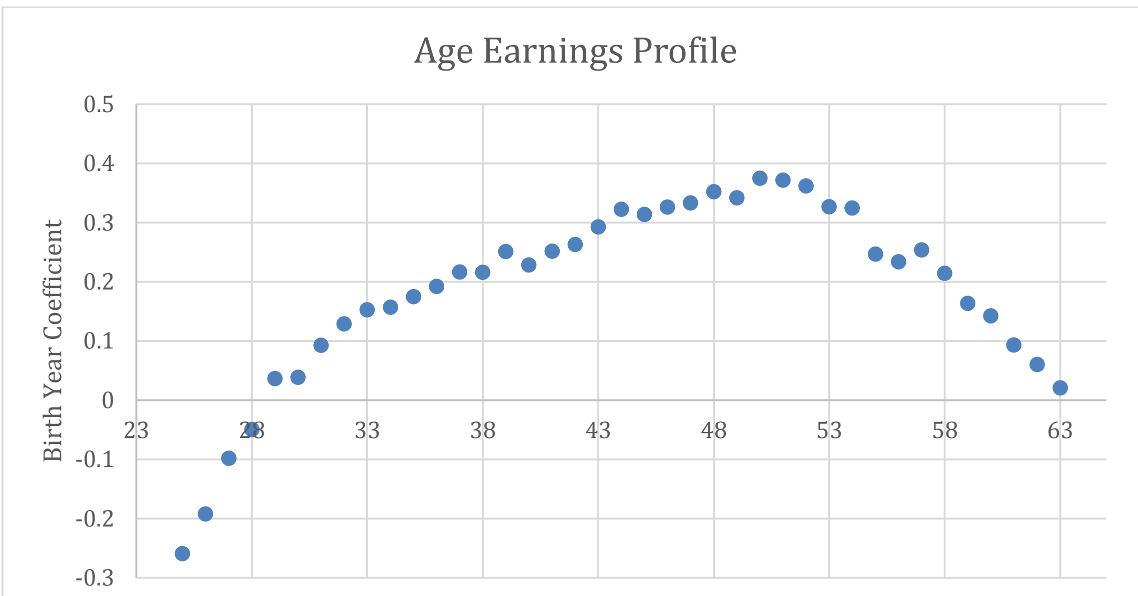

Figure 3. Age-earnings profile demonstrated by the birth year coefficient from the main regression on the vertical axis and the corresponding age (in the year 2000) on the horizontal axis.

The negative and significant coefficient on age of arrival in column one of Table 5 is consistent with previous studies, but this coefficient is no longer significant once the child mortality controls are introduced. This, along with the fact that the coefficient on the interaction between child mortality and age of arrival is negative and significant, indicates that the negative significance of age of arrival in the regression that does not control for child mortality is driven by the large negative effect among people emigrating from countries with high child mortality. Likewise, the coefficient on the interaction term between age of arrival and the dummy variable indicating if the country of origin is an English speaking country is positive and significant in the regression that does not control for child mortality, but becomes insignificant once child mortality controls are introduced. A possible explanation for this is that countries that are English speaking are more likely to be countries with lower child mortality rates (correlation coefficient of -.348 between the English speaking indicator and child mortality[12]), and so the positive coefficient on the English-speaking interaction term is capturing the effect of the lower child mortality rates in these countries.

The main coefficient of interest is the coefficient on the interaction term between child mortality and age of arrival. As expected, this coefficient is negative and significant at the 5% level (and nearly significant at the 1% level with a p-value of .012), which indicates that the effect of age of immigration on adult wages becomes more negative as child mortality in the country of origin increases. This gives the expected result that an immigrant is worse off arriving later from a country with a high child mortality rate than from a country with a low child mortality rate. The coefficient on the interaction between the dummy variable indicating whether the country of origin is an English-speaking country and age of arrival was positive, but not significant. As conjectured previously, this coefficient is most likely not significant due to the sample being composed of people who immigrated as young children. The

[12] The complete correlation table can be found in appendix B.

coefficient on child mortality is not significant, but this coefficient does not capture the total effect of child mortality. The child mortality coefficient would have to be added to the coefficient of the interaction term between age of arrival and child mortality for a given age of arrival in order to determine the entire magnitude of the effect of child mortality. The sum of the coefficient on child mortality and the coefficient on the interaction term between age of arrival and child mortality is negative for any age of arrival greater than one.

To give meaning to the magnitude of the coefficient of interest in the main regression, I calculated the predicted log income of the average immigrant arriving to the United States in 1969 (the mean year of immigration in the sample) at age 0 and age 10. I calculated this predicted value for immigrants arriving from both the country with the child mortality rate at the 25th percentile of the 1969 distribution, and the 75th percentile of the 1969 distribution. Italy was the country at the 25th percentile of the child mortality distribution in 1969 (with a child mortality rate of 35.1 deaths per 1000 live births), and Guatemala was the country at the 75th percentile (with a child mortality rate of 179.5 deaths per 1000 live births). I made these calculations holding sex, race, birth year, and language in the country of origin at the sample mean values. The predicted log income for a person who emigrated from Italy at age 0 is 10.5218, and the predicted value for someone who emigrated from Italy at age 10 is 10.4704. The predicted log income for a person who emigrated from Guatemala at age 0 is 10.4010, and the predicted log income for a person who emigrated from Guatemala at age 10 is 10.2734. The difference in log income for those arriving at age 0 is 10.4010-10.5218=-.1208, and the difference in log income for those arriving at age 10 is 10.2734-10.4704=-.1970 Thus, the difference-in-differences for these two countries is -.1970-(-.1208)=-.0762. This value of -.0762 indicates that the log wages of a person immigrating from Guatemala at age 10 are .0762 below what they would have been if the log wages vs. age of immigration plot (see figure 2) followed the parallel trends. To convert this number into dollar terms, the predicted value of 10.2734 for someone arriving from Guatemala at age 10 corresponds to an income of $28,951.31. If you add .0762 to the predicted value, you get 10.3496, which corresponds to an income of $31,244.98. Thus, this difference-in-differences of -.0762 corresponds to $2,293.67 lower income, and this is the effect of the interaction term between age of arrival and child mortality.[13]

I also ran two modified versions of the main regression. The first is the main regression with educational attainment, measured as the highest year of school completed, as the dependent variable. The results of this regression can be found in Table 1 in appendix A. The sign and significance of the coefficients of interest is quite similar to the coefficients from the regression with income as the dependent variable. The main difference is that the coefficient on the interaction term between age of arrival and child mortality is negative and significant at the .1% level, compared to at the 10% level in the regression with income as the dependent variable. The second modified version of the main regression is a probit regression with all the same independent variables as the main regression, but with a dummy variable indicating

[13] I also did this calculation with the two countries that had the most immigrants in the sample. Mexico and England were the countries with the most immigrants in the sample, with 23% and 5% of the sample respectively. The child mortality was 112.1 in Mexico in 1969, and 21.4 in England in 1969. The difference-in-differences for these two countries is -.0479, indicating that the log wages of a person immigrating from Mexico at age 10 are .0479 below what they would have been if the log wages vs. age of immigration plot followed the parallel trends.

if a person was employed as the dependent variable. This dummy variable was given a value of one if the person was employed, and zero if the person was unemployed but still in the labor force (I removed people not in the labor force from the sample). This probit model aims to test if the variables of interest from the main regression have a significant impact on the probability of being employed. The results of this regression can be found in Table 2 in appendix A[14]. The results indicate that none of the variables of interest had a statistically significant effect on the probability of being employed.

The dataset used in the main regression contains people who immigrated to the United States from 1936 to 1985, and during this time there were major changes in U.S immigration policy. The most drastic shift occurred with the passing of the Immigration and Nationality Act of 1965 which abolished the national origins quota system that had been in place since 1921.[15] The national origins quota system had restricted the number of immigrants from a country to a percent of the foreign-born population from that country who were residing in the United States in a specified base year (both the percent and base year were adjusted over time). The Immigration and Nationality Act replaced the quota system with an annual maximum of 170,000 immigration visas for immigrants from countries outside of the Western Hemisphere, with at most 20,000 visas for immigrants from a particular country. Likewise, an annual maximum of 120,000 immigration visas was established for immigrants from countries in the Western Hemisphere. This marked a large change as Western

[14] Table 2 in appendix A reports the coefficients from the probit regression, not marginal effects.
[15] Keely, pages 158-162

Hemisphere countries had been exempt from the quota system, and thus this Act established the first maximum on the number of immigrants from the Western Hemisphere. The Act also changed the preference system for immigrants so that family relationships played a more prominent role in the process of immigrant selection.

The results of the main regression restricted to people who immigrated in or after 1970 and prior to 1965 are presented in Table 6 on page 22 in order to examine the potential impact of this shift in U.S immigration policy. Column 1 of Table 6 shows the results for immigrants who arrived in or after 1970, and column 2 shows the results for immigrants who arrived prior to 1965. The Act was phased in between 1965 and 1968, so I choose 1970 as the starting point to examine the impact of the Act in order to give ample time for the changes in policy to take effect. The coefficient on the interaction between age of arrival and child mortality is negative and significant, and similar in magnitude in both periods. This indicates that the passing of the Immigration and Nationality Act of 1965 did not significantly change the impact that the childhood environment in the country of origin has on the effect of age of immigration on adult wages. One interesting result in this table is that the coefficient on age of arrival is positive and significant for immigrants who arrived prior to 1965. A positive coefficient on age of arrival appears inconsistent with the findings of previous studies that have examined age of arrival, but this inconsistency is due to this coefficient not capturing the total effect of age of arrival. The total effect of age of arrival requires adding the age of arrival coefficient and the coefficient on the interaction term between age of arrival and child mortality for a specified child

Table 6. Results pre-1965 and post 1970.

	(1) Post 1970	(2) Pre 1965
Female	-0.314***	-0.608***
	(0.0287)	(0.0183)
Black/Negro	-0.0109	-0.0600
	(0.0243)	(0.0648)
Chinese	0.198***	0.340*
	(0.0415)	(0.130)
Japanese	0.130***	0.101***
	(0.0214)	(0.0196)
Other Asian or Pacific Islander	0.167***	0.0393
	(0.0268)	(0.0586)
Other race, nec	-0.000906	-0.126***
	(0.0160)	(0.0293)
Two major races	-0.0609**	-0.0361
	(0.0222)	(0.0365)
Three or more major races	-0.136*	-0.0259
	(0.0639)	(0.147)
Age of Arrival	-0.00380	0.0127*
	(0.00424)	(0.00582)
Child Mortality	0.000347	-0.000177
	(0.00131)	(0.000818)
Child Mortality*Age of Arrival	-0.0000836*	-0.0000830**
	(0.0000335)	(0.0000316)
English Speaking Country*Age of Arrival	0.00555	-0.00631
	(0.00484)	(0.00460)
Birthplace Controls	Yes	Yes
Birth Year Controls	Yes	Yes
Observations	47964	25708

Adjusted R^2	0.090	0.113

Clustered standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

mortality. The average child mortality over all the countries and all the years prior to 1965 was 202.95 deaths per 1000 live births, and at this value of child mortality the marginal effect of arriving one year later is .0127+(-.0000830*202.95)=-.004145. Thus, while the coefficient on age of arrival is positive and significant, the marginal effect of age of arrival for an immigrant arriving from a country with the average child mortality rate is negative once the total effect of age of arrival is accounted for.

One final effect to consider is the potential that region-specific events temporarily change the type of people who immigrate to the United States. For example, it may be that immigrants arriving from Europe in the 1940s are more likely to be refugees from World War II than economic migrants. I added a term to control for yearly continent fixed effects to account for this possibility. To do this I created a variable to indicate the continent from which an immigrant arrived. This variable took on integer values from 1 to 6 (there were no immigrants from Antarctica), and I added an interaction term between the set of continent indicator factor variables and the vector of year of immigration factor variables to the main regression. The addition of this yearly continent fixed effect control did not significantly alter the magnitude or significance of the coefficients of interest. The results of this regression can be found in Table 3 in appendix A.

As stated in the empirical design section, the main identifying assumption implicit in this model is that the trends would be parallel if the childhood environment in the country of origin has no impact on the effect of age of immigration on adult wages. One of the main ways that this assumption would be violated is if parental characteristics that predict a child's adult wages, such as parental education or income, were correlated with the interaction term between child mortality and age of immigration after controlling for all the other variables in the model. For example, the parallel trend assumption would be violated if the difference in education between parents who immigrate with young children from countries with low and high child mortality is different from the difference in parental education between parents who immigrate with older children from countries with low and high child mortality rates (non-zero difference-in-differences in parental education across child's age of arrival).

I examined a sample of parents from the 1980 Census who immigrated with young children in order to test if the parallel trend assumption is valid. I used a sample from the 1980 Census for this robustness check as all of the necessary variables are available in the 5% IPUMS sample of the 1980 Census, but are only available in the 1% sample of earlier Censuses. One difficulty in this approach is that the Census started precisely recording how long immigrants had been in the U.S in the year 2000. Prior to the year 2000, the Census recorded how long an immigrant had been in the U.S in five-year intervals (0-5 years, 6-10 years, etc). This makes it impossible to precisely determine the age of arrival of an immigrant's children prior to the year 2000. In order to get around this difficulty, I limited the sample to parents whose eldest child was 9, 10, or 11 in 1980. I further restricted the sample to parents who were listed as the head of the household, were age 25 or older, and all of whose children were immigrants. I created a dummy variable to indicate if the eldest child arrived young (approximately age 0-5) or old (approximately age 6-10). It would be necessary to restrict the sample to parents whose eldest child was age 10 in 1980 in order to precisely construct this variable, but I included parents whose eldest child was ages 9 or 11 to increase the sample size. Sample statistics for this dataset are presented in Tables 7, 8, 9, and 10. The high percentage of males in the sample is likely because this sample is composed of household heads.

Table 7. Child's Age of Arrival Summary Statistics.

Distribution of Child's Age of Arrival	Percent
Arrived Young	56.02
Arrived Old	43.98
Total	100
N	4209

Table 8. Race Distribution Summary Statistics.

Race Distribution	Percent
White	59.45
Black/Negro	7.29
Chinese	4.04
Japanese	2.45
Other Asian or Pacific Islander	24.61
Other race, nec	2.16
Total	100
N	4209

Table 9. English vs. Non-English Speaking Distribution.

Distribution of Country Languages	Percent
Non-English	88.52
English	11.48
Total	100
N	4209

Table 10. Sex Distribution Summary Statistics.

Sex Distribution	Percent
Male	86.22
Female	13.78
Total	100
N	4209

The model used for this robustness check takes the following form:

$$educ_i = \alpha + \beta_1[sex_i] + \beta_2[race_i] + \beta_3[age_i] + \beta_4[bpl_i] + \beta_5[marst_i] + \beta_6[nchild_i] + \beta_7[cmort_{it}] + \beta_8[childarrive_i] + \beta_9[childarrive_i * cmort_{it}] + \beta_{10}[English_i * childarrive_i] + u_i \quad (2)$$

Where:

- *educ* is the education of person i measured as highest year of school completed.
- *sex* is a dummy variable indicating the sex of person i.
- *race* is the vector of race dummy variables of person i.
- *age* is the vector of age dummy variables of person i.
- *bpl* is the vector of birthplace dummy variables for person i.
- *marst* is the vector of marital status dummy variables for person i.
- *nchild* is the vector of number of children dummy variables for person i.
- *cmort* is the child mortality in the birthplace of person i in the year of birth of their eldest child.
- *childarrive* is a dummy variable indicating if the eldest child of person i was young (~age 0-5) or old (~age 6-10) when they arrived to the United States. This variable is assigned a value of one if the parent arrived when their eldest child was age 6-10.

- *childarrive*cmort* is an interaction term between a dummy variable indicating if the eldest child of person i was young (~age 0-5) or old (~age 6-10) when they arrived to the United States and the child mortality in the birthplace of person i in the year of birth of their eldest child.
- *English*childarrive* is an interaction term between a dummy variable indicating if person i is from a country where English is the most commonly spoken language and a dummy variable indicating if the eldest child of person i was young (~age 0-5) or old (~age 6-10) when they arrived to the United States. *English* is assigned a value of one if the country of origin was an English speaking country.
- u_i is the error term for person i.

I also ran this regression with the log of 1980 parental income as the dependent variable, and the results of these regressions are presented in Table 11 on page 28. The coefficient of interest in this robustness check is the coefficient on the interaction term between age of arrival (which in this case is a dummy variable) and child mortality. This coefficient is negative, but not significant, in both of the regressions, with a p-value of .171 in the education regression and .066 in the income regression. Given these results, there is no statistically significant evidence to suggest that the parallel trend assumption is violated.

Table 11. 1980 robustness check.

	(1) Education	(2) Income
Female	-0.957***	-0.860***
	(0.195)	(0.0774)
Black/Negro	0.410	0.142
	(0.347)	(0.123)
Chinese	-0.956	-0.746
	(0.666)	(0.387)
Japanese	1.597**	0.593*
	(0.590)	(0.243)
Other Asian or Pacific Islander	0.0173	-0.0279
	(0.532)	(0.114)
Age of Arrival	0.327	-0.148
	(0.323)	(0.0908)
Child Mortality	-0.0444**	-0.00649
	(0.0142)	(0.00421)
Child Mortality *Age of Arrival	-0.00335	-0.00127
	(0.00243)	(0.000680)
English Speaking Country*Age of Arrival	0.318	-0.132
	(0.284)	(0.128)
Birthplace Controls	Yes	Yes
Age Controls	Yes	Yes
Number of Children Controls	Yes	Yes
Marital Status Controls	Yes	Yes
Observations	4209	3586
Adjusted R^2	0.499	0.221

Clustered Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

VI. Conclusion

The results presented in this paper suggest that the childhood environment in the country of origin has a significant impact on the effect of age of immigration on adult labor market outcomes. If the parallel trend assumption is accepted, then exposure to the higher child mortality rate in Guatemala (the country at the 75th percentile of child mortality distribution in the mean year of immigration in the sample) results in approximately 7.5% lower adult wages for someone who emigrated from Guatemala at age 10 compared to someone who emigrated from Italy at age 10 (the country at the 25th percentile of the child mortality distribution). These results suggest that prolonged exposure to the factors that contribute to higher child mortality rates as a young child, such as poor healthcare and lack of access to clean water, results in lower adult wages after immigrating to the United States. Controlling for whether immigrants arrived from an English speaking country allows me to conclude that the negative impact of a poor childhood environment is not due to the potential effect of language skills that would result if countries with high child mortality rates also tend to be non-English speaking countries.

The key assumption necessary to accept these conclusions is that the parallel trend assumption is indeed valid. The fact that the coefficient on the interaction term between age of arrival and child mortality is not significant in either of the robustness check regressions provides some credibility for accepting this assumption. That being said, the low p-values on the coefficient (.171 and .066) prevent this robustness check from emphatically validating the parallel trend assumption. Future research could be done to remove this potential endogeneity concern by implementing an instrumental variable for the interaction between age of arrival and child mortality. One potential source of an instrumental variable would be the change in immigration policy that occurred with the passing of the Immigration and Nationality Act of 1965. To conclude, this is the first study which has shown that it is important to consider the environment in the country of origin when examining the effect of age of immigration on adult labor market outcomes.

References

Schaafsma, Joseph, and Arthur Sweetman. “Immigrant Earnings: Age at Immigration Matters”. *Canadian Journal of Economics* 34.4 (2001): 1066-1099. Print.

Åslund, Olfaf, and Anders Bohlmark, and Oskar Nordstöm Skans. “Age at Migration and Social Integration”. *Institute for Labour Market Policy Evaluation* (2009). Print.

Borjas, George J., “Assimilation and Changes in Cohort Quality Revisited: What Happened to Immigrant Earnings in the 1980s?”. *Journal of Labor Economics* 13.2 (1995): 201-245. Print.

Smith, James P., “The Impact of Childhood Health on Adult Labor Market Outcomes”. *The Review of Economics and Statistics* 91-3 (2009). 478-489. Print.

Bleakley, Hoyt, and Aimee Chin. “Language Skills and Earnings: Evidence from Childhood Immigrants”. *The Review of Economics and Statistics* 86.2 (2004). 481-496. Print.

Keely, Charles. “Effects of the Immigration Act of 1965 on Selected Population Characteristics of Immigrants to the United States”. *Demography* 8.2 (1971). 157-169. Print.

Steven Ruggles, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]. Minneapolis, MN: Minnesota Population Center [producer and distributor], 2010.

The World Factbook 2014-15. Washington, DC: Central Intelligence Agency, 2015. https://www.cia.gov/library/publications/the-world-factbook/index.html

Kiara Johansson, Mattias Lindgren, Claes Johansson, Ola Rosling. "Under-Five Mortality Rate (per 1000 live births)" Gapminder. Version 8. 2015. http://www.gapminder.org/data/

Appendix A

Table 1. Main regression with Education as the dependent variable.

	(1) Education
Female	2.168**
	(0.721)
Black/Negro	-0.672
	(0.529)
Chinese	9.015***
	(2.180)
Japanese	5.934***
	(0.753)
Other Asian or Pacific Islander	6.050***
	(1.377)
Other race, nec	-2.283***
	(0.494)
Two major races	-2.375***
	(0.598)
Three or more major races	0.304
	(2.150)
Age of Arrival	-0.324
	(0.219)
English Speaking Country *Age of Arrival	0.409
	(0.236)
Child Mortality	0.0253
	(0.0155)
Child Mortality*Age of Arrival	-0.00308***
	(0.000856)
Birthplace Controls	Yes

Birth Year Controls	Yes
Observations	87508
Adjusted R^2	0.211

Clustered standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2. Probit Regression to determine effect on employment probability.

	(1) Employment Probability
Female	-0.0240
	(0.0333)
Black/Negro	-0.147**
	(0.0535)
Chinese	0.0973
	(0.0975)
Japanese	-0.0596
	(0.0445)
Other Asian or Pacific Islander	0.0777
	(0.0564)
Other race, nec	-0.0502***
	(0.0150)
Two major races	-0.0520
	(0.0270)
Three or more major races	-0.440***
	(0.0923)
Age of Arrival	-0.00657
	(0.00469)
Child Mortality	0.000356
	(0.000429)
Child Mortality*Age of Arrival	-0.0000195
	(0.0000383)

English Speaking Country*Age of Arrival	-0.00215
	(0.00441)
Observations	84636
Adjusted R^2	

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3. Main regression with yearly continent fixed effect controls.

	(1) With Child Mortality
Female	-0.427^{***}
	(0.0259)
Black/Negro	-0.00482
	(0.0270)
Chinese	0.235^{***}
	(0.0485)
Japanese	0.116^{***}
	(0.0146)
Other Asian or Pacific Islander	0.163^{***}
	(0.0275)
Other race, nec	-0.0290
	(0.0198)
Two major races	-0.0530^{**}
	(0.0197)
Three or more major races	-0.131
	(0.0750)
Age of Arrival	0.165
	(0.272)
Child Mortality	0.000288
	(0.000306)
Child Mortality*Age of Arrival	-0.0000677^{*}
	(0.0000277)

English Speaking Country*Age of Arrival	0.00266
	(0.00359)
Yearly Continent Fixed-Effects Controls	Yes
Birthplace Controls	Yes
Birth Year Controls	Yes
Observations	87508
Adjusted R^2	0.117

Clustered standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix B

Table 1. Correlation Table

	lnincwage	ageimmig	cmortality	english	birthyr
lnincwage	1				
ageimmig	-0.0285	1			
cmortality	-0.0318	0.234	1		
english	0.0624	-0.0714	-0.348	1	
birthyr	-0.163	-0.00714	-0.154	-0.221	1

Notes; the variables included are log income, age of immigration, child mortality, birth year, and an indicator variable for whether a country is English speaking