Immigration and the Health of Children:
Lessons from the United States at the turn of the 20th Century

Martin Dribe
Center for Economic Demography
Department of Economic History
Lund University
Martin.Dribe@ekh.lu.se

J. David Hacker
Department of History
University of Minnesota
hacker@umn.edu

Francesco Scalone
Department of Statistical Sciences
University of Bologna
Francesco.scalone@unibo.it

Draft 8/15/2016

Abstract

The societal integration of immigrants is a great concern in many of today's Western societies, and has been so for a long time. Whether we look at Europe in 2015 or the United States at the turn of the 20th century, large flows of immigrants pose challenges to receiving societies. While much of the research have focused on the socioeconomic integration of immigrants there has been less attention to the demographic integration, even though this can tell us as much about the way immigrants fare in their new home country. In this paper we study the disparities in infant- and child mortality across ethnic groups and generations, using new census data of greater density than previously available. In addition to charting the main patterns we focus special attention on the importance of community-level factors in determining the health disparities across groups, and how these disparities changed over immigrant generations. The results are broadly consistent with assimilation theory and also points to contextual variables being important both for child mortality and to explain ethnic differentials in child mortality.

Introduction

The United States has long struggled with difficulties associated with immigration. During the "classic" period of largely unrestricted immigration between the American Civil War and the introduction of stringent numerical quotas in the 1920s, the foreign born population composed between 13 and 15 percent of the overall population, the highest percentages in U.S. history. The integration of these immigrants into American society has been a topic of social research for over a century (e.g. Walker 1891; Drashler 1920; Gordon 1964). Much of the focus has been on how immigrants fared in the labor market in terms of earnings and occupational career, and how immigration affected the economic realities of the native born, and the U.S. economy at large (see Abramitsky and Boustan 2016). There has also been considerable interest in the marriage patterns of immigrants and intermarriage as an indicator of assimilation (Alba and Golden 1986; Lieberson and Waters 1988; Pagnini and Morgan 1990). U.S. immigrants at the turn of the twentieth century from eastern and southern Europe were much less likely to intermarry with natives than were old immigrant groups from northern and western Europe (Sassler 2005; Wildsmith et al. 2003). Intermarriage was also positively associated with different assimilation indicators such as ability to speak English, socioeconomic status and belonging to the second generation. Moreover, there were important associations between contextual factors and different marital outcomes. The relative size and sex ratio of the origin group, ethnic diversity, the share of the native born white population, and the proportion of life time spent by immigrants in the United States were all associated with exogamy (Dribe, Hacker and Scalone forthcoming).

In the literature on the immigration of the late twentieth century there has also been a great interest in other aspects of the well-being of immigrants as manifested in their health and mortality. Especially large attention has been devoted to the unexpectedly low mortality of immigrants of Hispanic origin, and the extent to which this is associated with health-related behavior or more a matter of immigrants selection (e.g., Hummer et al. 1999; Singh and Hiatt 2006; Hummer, Melvin and He 2015). For the historic immigration to the United States there has not been as much focus on these aspects of immigrant assimilation, although there are exceptions to this (Preston and Haines 1991; Preston, Ewbank and Hereward 1994). With the release of more extensive historical data more detailed analyses are now possible. In this paper we estimate infant and child mortality among 16 immigrant groups in 1910, and the native born white and black populations of native parentage. We then model that mortality as a function a rich set of social, economic, and demographic variables. Our data come from new high density samples of the U.S. census, which included several questions designed to measure immigrants' social and economic integration (e.g., occupation, literacy, ability to speak English, year of immigration, mother tongue, and parental mother tongue, nativity, nativity of parents, and nativity of spouse). We supplement these data with new measures of neighborhood characteristics constructed from new complete-count census data collected by Ancestry.com (Ruggles et al. 2015), which allow us to test hypotheses related to the influence of couples' environment and neighbors on child mortality. In relation to previous research on child mortality of immigrants at the turn of the twentieth century our main contribution is to analyze the importance of contextual factors in explaining ethnic mortality differences. We do this by estimating county-level fixed-effects models and by studying the impact of county-level variables measuring the size of immigrant populations, presence of natives and overall child mortality. In addition, we also analyze how child mortality among immigrants is shaped by intermarriage, immigrant generation, and time in the United States.

Theory and previous research

Ethnic differences in mortality stem from a variety of causes related to income, education, place of residence, housing quality, health-related behavior, migrant selection, etc. Some of these risk factors are relatively easy to measure while others are more difficult. In this paper we focus our attention exclusively on infant and child mortality and how it is related to ethnicity and the various background explanations, just mentioned. There are several models of the determinants of infant and child mortality more general. One frequently used framework is the one devised by Mosely and Chen (1984) where child health is modeled using five sets of proximate determinants; maternal factors (e.g. mother's age, parity and birth interval); injuries; nutrition; environmental contamination (e.g. disease environment, poor hygiene); and illness control (e.g. preventive measures such as vaccination). Socioeconomic status, ethnicity, and immigrant's degree of assimilation into host society are all background factors that could have an impact on the proximate determinants (see Hummer et al. 1999). Socioeconomic status affects both nutrition and the living environment through place of residence (see Van Poppel, Jonker and Mandemakers 2005; Woods 2000).

Urban residence is a well-known risk factor for child health, especially in pre-industrial and early industrial times when hazardous emissions and diseases plagued cities (e.g. Condran and Crimmins 1980). An important reason for the high disease prevalence was deficient sanitation and impure water (see, e.g., Burström et al. 2005) as well as crowding, which for example is a risk factor for measles (Burström, Diderichsen and Smedman 1999). Also within the urban context, a higher income enabled families to avoid the worst areas, and to afford housing of good quality. Socioeconomic status, and especially education, also affects parental child-rearing capabilities, by increasing their responsiveness to information about healthy behavior and child-rearing practices (Caldwell 1979; Caldwell and McDonald 1984; Desai and Alva 1998). One such factor often stressed in the historical literature is breastfeeding. In the past breastfeeding played the double role of delaying new pregnancies, thus giving more time for the mother to care for her child, and providing the infant with uncontaminated nutrition (e.g. Preston and Haines 1991). In contexts where both cow milk and water was often impure, such as in urban areas, this could greatly reduce the risk of disease and mortality among young children, and since infant mortality forms a big part of total mortality under age five, breastfeeding can have a major impact on child mortality (see, e.g. Preston, Ewbank, and Hereward 1994; Woods, Williams, and Galley 1993). Woodbury's survey of early twentieth-century breastfeeding practices in Baltimore (1925), suggested that breastfeeding and socioeconomic status were offsetting factors in infant mortality: immigrant groups with lower socioeconomic status, such as Italian immigrants, were more likely to breastfeed and did so for longer durations.

In the context of immigrants, assimilation is likely of prime importance as well. By learning the native language mothers will be more likely to receive information about child rearing behavior to the extent that they did not know about them when they arrived. It is important to remember that the period we are looking at was a time of dramatic change in mortality throughout the Western world, partly related to improved nutrition following the agricultural and industrial revolutions, but to a large extent thanks to scientific discoveries and better knowledge about what caused different diseases and how they could be prevented (e.g., Easterlin 1996). Thus, immigrants arriving from less developed countries than the United States were exposed to a flow of new information about how to properly take care of themselves and their children, and mastering English can be expected to have been important in accessing this information, as well as in getting access to basic health services, such as a family doctor. In a similar way, exposure to the native population, for example by not residing in ethnic enclaves, would be expected to have a similar effect. For similar reasons, we would expect second-generation immigrants to have a different mortality pattern than the first generation, and also that intermarriage, and time spent in the United States would have an impact on child mortality.

At the same time it is important to keep in mind that not all immigrants came from contexts with less knowledge and higher mortality than the United States. The low mortality of Jewish immigrants has, for example, often been noted, and has been related to their high rates of breastfeeding as well as their (religious) practice of washing hands before meals (e.g., Preston, Ewbank, and Hereward 1994). An important factor behind a lower mortality of some immigrants may also have been better physical conditions of the mothers as a result of better nutrition and less disease exposure in the home country, which would benefit the health of their children (see, e.g., Bhalotra and Rawlings 2013).

In the literature on the health of contemporary immigrants in the United States there has been an enormous attention to the so called "Hispanic Paradox", i.e. that immigrants from Latin America, and especially Mexico, have comparatively low mortality (Hummer, Melvin and He 2015), and that over time there seems to be a negative assimilation going on in the sense that their mortality is increasing with longer time spent in the United States (Landale, Oropesa and Gorman 2000). There has been an intense debate over the causes behind this paradox, and it seems that even though migrant selectivity, both in moving to the United States and in returning home, is part of the explanation it is not the whole explanation (Markides and Eschbach 2005; Hummer et al. 2007; Palloni and Arias 2004; Riosmena, Wong, and Palloni 2013). Instead lower smoking prevalence and better diet are important reasons for the rather good health and mortality situation of these immigrant groups (Hummer, Melvin and He 2015).

For the United States at the turn of the twentieth century, there were also some immigrant groups with higher mortality than the native whites, and some with distinctly lower mortality. Much of the difference can be explained by factors related to socioeconomic status and place of residence (see Preston, Ewbank and Hereward 1994), but there are also differences remaining after such controls. An interesting question is if accounting for assimilation always leads to a convergence, and hence that we have a positive association between mortality and assimilation (or acculturation) for immigrants from low-mortality origins and a negative association for immigrants from high-mortality origins.

Data

We use data from the 1910 IPUMS census sample (Ruggles et al. 2010) and the 1910 complete-count microdata collected by Ancestry.com, recently made available by the Minnesota Population Center. Conducted during the high point of European immigration, the 1910 census includes information on birthplace and parental birthplaces, duration of marriage, ability to speak English, literacy, year of immigration, employment and occupation. It also includes variables on the number of times married, language spoken, mother tongue, parental mother tongue, and for all ever married women, the number of children ever born and the number of children still surviving. As discussed below, the latter two variables allow us to construct a measure of infant and child mortality for currently-married couples in the dataset.

We analyze the experience of 18 different ethnic groups: Native whites of native parentage (NWNP), native blacks of native parentage (NBlNP), French Canadians, English Canadians, British (English, Scottish and Welch), Irish, German, Danish, Norwegian, Swedish, Dutch, French, Italian, Portuguese, Russian Jewish, Eastern and Central European Jewish, Polish and Mexican. These were the most numerous ethnic groups of European and American origin living in the United States in 1910.They were defined based on nativity, race and mother tongue. NWNP and NBlNP were both based on nativity and race, Canadian, German, Scandinavian, French, Dutch, Italian, Portuguese, Polish and the two Jewish groups were defined based on nativity and mother tongue, while British, Irish and Mexican were defined solely based on nativity (see Watkins 1994 for a similar approach).

A number of variables at the individual level are included to measure assimilation and place of residence. We distinguish three different immigrant generations: 1G (foreign born arriving in the United States after the age of 12), 1.5G (foreign born arriving in the United States at age 12 or younger), and 2G (U.S. born with at least one foreign born parent). For the second generation ethnicity was defined based on mother's origin unless she was U.S.-born in which cased we based it on father's origin.

We categorize marital outcomes into five different groups based on origin and immigrant generation: 1G Endogamy (married to a spouse from the same origin), 2G endogamy (married to a second-generation spouse from the same origin), NWNP Exogamy (married to a U.S.-born white spouse with two U.S.-born parents), 2G Exogamy (married to a second-generation spouse with at least one foreign born parent from a different ethnic group), Other Exogamy (married to any other spouse, including foreign born and U.S.-born blacks).

Place of residence distinguishes rural areas from urban areas of different sizes (2,500-9,999, 10,000-99,999, 100,000 or more). In addition we have an indicator for farm residence, as it can expected to be associated with better access to nutrition. Literacy and ability to speak English is included for both the wife and her husband. We also include the husband's occupation score, a measure of the median earnings of the occupation in 1950 and an assumed proxy for socioeconomic status in 1910 (Sobek 1995). In addition we have an indicator of whether or not the woman is gainfully employed.

We rely partly on the complete count data (which includes all the individuals in the census) to investigate the influence of contextual characteristics on immigrant behavior. Although the 1910 complete-count data contains a limited number of variables and does not identify neighborhood or census tract, it does identify individuals' residence location by county and census enumeration district (ED), birthplace, parental birthplace, marital status, age, number of children ever born, and number of children surviving. In this study we analyze contextual effects at the county level.

We constructed three different contextual variables. Relative group size is defined as the proportion of foreign born from the country group of origin in relation to the total population. Because we lack information on mother tongue in the full-count dataset we calculated this variable from the IPUMS sample, while the other two contextual variables were based on the full-count data. The proportion NWNP is the share of the population in the county that is white and born in the United States with two native born parents. Finally, we estimate the background disease environment from the mortality of white children born to white women of native parentage. Specifically, we construct a mean index of infant and child mortality for currently married white women of native parentage age 20-49 in each county with the "age model" described in the United Nations' *Manual X* for indirect estimation methods (United Nations 1983:76-81; see also Haines and Preston 1997) standardized to Model West life table level 14.0 (Coale and Demeny 1966). American data were used in the construction of Model West and previous studies have found that it closely approximates mortality in the early twentieth-century United States (Preston and Haines 1991: ch. 2). As shown below, level 14 was a very close fit for the mortality of children born to women in our sample.

Table 1 shows the descriptive statistics of the sample. In total we look at a sample of 139,229 women, corresponding to a total population of children born of about 35 million (based on population weights and number of children ever born). About 55 percent of the sample are native whites (NWNP) and about 11 percent are native blacks (NBlNP)(other native born with native parentage ethnic groups are not included in the analysis). Among the immigrant groups German, Irish and British are the three largest. A majority of the sample live in rural areas (63 percent), less than one fifth live in urban areas with a population of more than 100,000. Almost 60 percent live on a farm. Over 90 percent speak English and are literate, while only 9 percent of the women are employed. Looking only at the non-native population (i.e., excluding NWNP and NBlNP) 32 percent are first-generation immigrants arriving as adults (1G), 13 percent are foreign born arriving as children, and 55 percent are belong to the second generation (native born with at least one foreign-born parent). About 42 percent of the immigrants are married to a foreign-born spouse from the same ethnic origin and an additional 20 percent are married to a second-generation immigrant from the same origin, indicating the great importance of endogamy among early twentieth-century immigrants (see Dribe, Hacker and Scalone forthcoming). Limiting the sample to only the foreign born (1G and 1.5G) a majority of the sampled women have spent less than 10 years in the United States before getting married.

Table 1 here

As shown in Figure 1, the background mortality environment, as proxied by infant and child mortality among children born to NWNP women, varied substantially by county. Some of this heterogeneity, of course, reflects higher infant and child mortality in urban areas and higher population density areas, such as that in Chicago and New York, and lower mortality in rural areas. But there were many pockets of unexpected high infant mortality in rural areas, such as in the Arkansas counties bordering the Mississippi River and in the Anthracite coal mining counties in eastern Pennsylvania. The map strongly suggests that the inclusion of urban-rural and regional dummy variables would be insufficient in controlling for the geographic differentials in the disease environment. In the subsequent analysis we apply county-level fixed effects to account for this diversity in experience.

Figure 1 here

Methods

We rely on indirect estimates of infant and child mortality constructed from census questions on the number of live births that an ever-married woman had in her life (i.e. parity or children ever born) and how many of those children were still living (i.e. children surviving). The question was asked to all ever-married women in the 1910 census. The method has been detailed elsewhere (United Nations 1983, Preston and Haines 1991; Haines and Preston 1997). Briefly, a mortality index was constructed for each woman by dividing the number of actual child deaths she experienced by the expected number of deaths. The latter was based on her parity, the children's length of exposure to the risk of dying (proxied by the mother's duration of marriage or age), the overall age pattern of fertility, and the model life table standard. For women in the IPUMS sample data, we relied on the "marital duration model," which provides the best estimates for populations with little childbearing outside of marriage. We limited our universe to currently-married white and black women with spouses present, in their first marriage, and married less than 35 years. Unfortunately, the times married variable was not collected by Ancestry.com and is not available in the complete count dataset. For our contextual background mortality estimate we were forced to rely on the "age model," with the universe limited to currently-married white women age 20-49.

The index is readily interpretable. A mortality index of 1.0 means that the woman (or group of women) was experiencing child mortality equivalent to West Model 14.0 (Coale and Demeny 1966), which indicates an infant mortality probability (${}_5q_0$) of 0.122 (both sexes combined) and expectation of life of about 51 years. Values above 1.0 mean that the woman (or a group of women) was experiencing child mortality above the life table standard, while values below 1.0 means that the woman was experience mortality below the standard. Any mortality parameter desired can be obtained by multiplying the index by the ${}_5q_0$ value in the standard table – in our case 0.17972 – and then finding the appropriate table in the life table system (Hacker and Haines 2005).

In the multivariate analysis we use linear regression (OLS) with the mortality index at the individual level as the dependent variable. Prior research has shown the mortality index to be robust and econometrically well-behaved when used as a dependent variable in a regression model (Trussell and Preston 1982). Because infant and child mortality was declining in the early twentieth-century United States, however, the index will be biased by differences in women's (or differences among groups of women's) marital duration or age. To account for this bias, we constructed estimates of the midpoint of the period to which the mortality estimates refer to for each woman (also detailed in United Nations 1983; Haines and Preston 1997) and included this mortality reference date (MRD) in our models (the MRD is expressed in years before 1910). The regressions are weighted by a combination of the sample probability weights and the number of children born in order to reflect the population of children at risk of mortality, even though the unit of analysis is women.

We start by estimating a basic model only controlling for age and ethnic origin. This gives an estimate of the gross ethnic differentials in child mortality. We then add variables measuring place of residence and assimilation which gives an idea about how much of the raw ethnic differences that can be explained by these factors. In this model we also include the mortality reference date. Subsequently we add county-level fixed effects to the model to account for unobserved variables at this level, and finally we compare with a model including three contextual variables at the county level: relative group size, proportion NWNP and mortality index.

Results

We start by looking at the mortality index by ethnic group in Table 2. The mean for the whole sample is 0.99 (based on the Model West 14.0 life table). The native born whites have lower child mortality than the overall mean. Blacks and Mexicans have the highest child mortality

(1.47 and 1.58 respectively). French Canadians, Irish, Italians, and Poles also have higher than average child mortality while the other ethnicities have lower than average mortality levels, with the Russian Jews, the Dutch and the Danes having the lowest levels. It is of course premature to put too much interpretation into these differences before analyzing more in depth how they depend on residential patterns, or the degree of assimilation into U.S. society.

Table 2 here

Before turning to these questions, however, something should be said about the actual mortality levels. As we standardized the mortality index on the Model West 14 life table (both sexes combined), it is straightforward to get estimates of the probability of dying at various ages by simply multiplying the mortality index with the ${}_nq_x$ in the life table. For example, the value of 0.91 for NWNP implies that the probability of death before age five (${}_5q_0$) is 0.163. This also makes it possible to make at least a rough comparison with the mortality levels in the countries of origin, at least for the countries where data is readily available. In Table 3 the implied ${}_5q_0$s for the origin groups calculated from the mortality index and the Model West 14 life table are compared with ${}_5q_0$s in 1900-1910 taken from the Human Mortality Database (www.humanmortality.org). Except among Swedes and Norwegians mortality levels are lower among ethnic groups in the U.S. than in the countries of origin. As could be expected differences are larger between the countries of origin than between U.S. ethnic groups, but overall the ranking of countries are similar, with the exception of the Dutch who show much lower mortality in the United States than in the country comparison.

Table 3 here

Turning to the multivariate analysis, Table 4 shows regression estimates for four different models: a basic model controlling only for age and age squared (M1); the full model with all individual-level variables (M2), the full model plus county-level fixed effects (M3) and the full model plus the contextual variables at the county level (M4). By comparing the estimates for ethnic origin in the basic model and the full model we get an idea of how much the observed variables related to assimilation and place of residence contribute in explaining the ethnic differences. Similarly, comparing the full model and the fixed-effects model shows the contribution of unobserved county-level factors in explaining the ethnic differentials. The comparison of the fixed-effects model and the full model including contextual variables indicates the importance of the measured contextual effects in relation to the total influence of the county level on ethnic differences in child mortality.

Table 4 here

We begin by comparing the estimates for ethnic origin across the different models. In most cases the mortality differences to the native whites are smaller in the full and fixed effects models, implying that some of the gross difference can be explained by place of residence and the assimilation variables included in the models. This is for example clearly evident for native blacks, Mexicans, Irish, Italians and Polish. For Scandinavian and Dutch immigrants there are also considerable differences between the full model and the fixed-effects model, indicating that unobserved factors at the county level are important explanations for the mortality differentials. For the other origins unobserved county-level factors does not seem to explain much of the mortality differences in addition to what is explained by the individual-level factors capturing place of residence and assimilation. For most origins the differentials are also highly similar between the fixed-effects model and the model including the three contextual variables, indicating that there are no vital determinants for ethnic-group-specific mortality missing at the county level.

Most notably for the Jewish immigrants, the gross mortality differences are much smaller than the ones we get when controlling for assimilation and place of residence. When taking these factors into account the child mortality of Jewish immigrants, whether from

Russian or Eastern-Central European origins, have less than half the mortality level of native whites.

Turning to the control variables reflecting place of residence and assimilation, living in a larger urban area is related to higher mortality, which shows the impact of adverse sanitary conditions and housing standards in the large cities explaining the so called urban penalty, which was the normal pattern in preindustrial and early-industrial societies (Preston and Haines 1991; Woods 2000). Non-English speakers show substantially higher mortality than English speakers, but there is no added impact of spouse being a non-English speaker. As expected, being literate is associated with lower child mortality, and in this case there is also an additional association with spouse being literate. Women who are employed experience higher mortality of their children than non-employed women, which could both be related to neglect of children and to a selection effect because families with lower capabilities caring for their children were also more likely to be dependent on women's work for their subsistence. Finally living on a farm is associated with lower child mortality over and above simply living in rural areas. This could be related to better access to food and higher-quality housing on farms than in rural areas in general.

Looking at the contextual variables, relative group size shows a positive association with mortality, implying that a higher proportion of one's own ethnic group in the county of residence is associated with higher mortality. The proportion of native whites does not seem to be related to child mortality when looking at all ethnic groups, while the county-level mortality index is positively associated with individual-level child mortality. A one unit higher mortality index in the county of residence increases the individual-level mortality index by about 0.8, which is a sizeable magnitude.

To look more in detail into how these assimilation indicators are associated with mortality we divided the sample into four groups: the native whites, the native blacks, those with an average mortality index above one (High mortality: French Canadian, Irish, Italian, Polish, and Mexican) and those with a mortality index lower than one (Low Mortality: German, Scandinavian, Dutch, and Jewish). We excluded four origin groups with levels close to the average (British, French, Portuguese, and English Canadian). Table 5 shows the net estimates from interaction models looking at the impact if the contextual variables on child mortality. Relative group size shows positive association with mortality for native whites and for immigrants from both high-mortality and low-mortality origins. Immigrants from high-mortality origins show the strongest association. For native blacks the association is negative, implying that the higher the proportion native blacks in the county, the lower the child mortality, which runs contrary to expectations. A higher proportion of native whites in the county is associated with lower child mortality for both native blacks and immigrants from high-mortality origins, while there is almost no association for immigrants from low-mortality origins. A higher mortality index is associated with higher child mortality in all groups, as expected, but the association is significantly weaker for immigrants from low-mortality regions. Thus, it seems clear that the context in which immigrants live is important for their mortality experience, but that this is more so for immigrants originating from high-mortality countries, while those from low-mortality countries seem less affected by the contextual setting.

Table 5 here

Table 6 shows regression estimates for immigrant generation and intermarriage looking only at the sample of immigrants (1G, 1.5G and 2G). Compared to first-generation immigrants (1G), the 1.5G (foreign-born who came as children) show no different mortality pattern, while there is dramatically lower child mortality among the second-generation immigrants in the basic model. When adding controls for assimilation and place of residence in the full model, however, most of this mortality advantage disappears, and when also adding the county-level fixed effects the association is no longer statistically significant. In other words, what appeared to be a clear mortality advantage for the second generation is largely explained by where they live, and by their greater degree of integration into U.S. society in terms of language and socioeconomic status. When all of this is taken into account we cannot demonstrate any differences in child mortality between the generations.

Table 6 here

Looking at the association with exogamy, women married endogamously with a second-generation immigrant have a clear mortality advantage compared to those endogamously married with a first-generation immigrant. Even though the advantage declines substantially in the full model, and somewhat further in the fixed effects model there is still a statistically significant difference. The patterns are similar for intermarriage with native whites, but in this case the advantage is not statistically significant in the fixed-effects model. For exogamy with other 2G immigrants there is an advantage only in the basic model, and for other exogamy there are no visible differences in mortality compared to 1G endogamy.

Table 7 displays the net estimates from interaction models with High mortality as the reference category (full models including county-level fixed effects). Looking first at immigrant generation associations are in opposite directions in the two groups, consistent with assimilation theory. In the high-mortality group the second generation show lower mortality than the first generation, while the opposite holds true for the low mortality group. In other words, immigrants arriving to the United States with higher mortality that the country average converges to the mainstream in the next generation, by lowering their child mortality. Conversely, immigrants from low-mortality origins, experience the opposite development with increasing child mortality, also converging to the average levels in the United States. The differences in the development between the two groups across generations are also statistically significant.

Table 7 here

If we instead look at intermarriage, in panel B of Table 7, the pattern is similar. Even though not all coefficients are statistically significant the overall pattern is clear: assimilation implies a convergence in mortality for immigrants from both high-mortality origins and low-mortality origins. Marrying someone from the second generation, for example, lowers child mortality in the high-mortality group compared to marrying a first-generation immigrant, while it increases mortality in the low-mortality group. The associations are similar for intermarriage with the native white population, and even though the base estimate for in the high-mortality group is not statistically significant, the difference in the estimate between the low-mortality and high-mortality groups is statistically significant.

In Table 8 we study the extent to which the time spent in the United States before getting married is associated with child mortality. This analysis is made on a sample only including foreign-born. In panel A both 1G and 1.5G are included and in panel B the sample is further restricted to only include the 1G. Few of the coefficients are statistically significant, and it is difficult to identify a consistent pattern. When removing the 1.5 G, no associations are statistically significant. Thus, there is no support for the hypothesis that a longer time in the United States would be linked to mortality assimilation. However, one reason for this result is likely the negative selection of women spending a long time in the United States (10-20 years) before getting married. If we assume many adults who arrived were in the twenties, these women would have been in their late thirties or even early forties before getting married and having children, and we would expect child mortality to be quite different for these women than for those forming a family in more typical ages.

Table 8 here

Conclusions

Immigrant assimilation into host societies is a multifaceted process involving not only learning the language and getting employment and equal pay, but also involves demographic aspects such as life expectancy, marriage and fertility. A crucial indicator of living standards is the health of children. In this study we look at mortality differentials across ethnic groups in early twentieth-century America. Besides looking at the overall differences in mortality under age five across a number of immigrant groups and compare with native-born whites and blacks, we also study in more detail how different assimilation indicators and place of residence contribute to explain these differences. We also look at how childhood mortality differs across immigrant generations and by exogamy, to assess the importance of assimilation for health and mortality.

Our analysis demonstrates large mortality differentials across ethnic groups in 1910. Native-born blacks show much higher child mortality than native-born whites, and among immigrants Mexicans, French Canadians and Irish have higher than average child mortality, while Russian Jews, Dutch and Scandinavians are among the groups with lower than average levels. However, once we control for different individual-level variables aimed to measure the degree of assimilation and urbanity the ethnic differentials often becomes smaller, an exception being the Jewish population whose mortality advantage grows considerably larger when adjusting for their adverse living conditions. Similarly, when we control for unobserved county-level variables, mortality differentials sometimes further diminish. These results show that while there were large ethnic differences in the United States in the early twentieth century, they could to a large extent be explained by other factors, such as where different groups tended to live, and their socioeconomic and cultural integration into U.S. society. Not all differences could be explained this way. Also after controlling for assimilation and place of residence, blacks and Mexicans have 40 percent higher child mortality than the native-born whites, and French Canadians have almost 30 percent higher mortality. At the same time the

Jewish populations have almost 40 percent lower child mortality than the native whites. For most other ethnic groups the differentials are 5 percent or less.

Our analysis also showed that the context in which people live is important for their mortality, as measured by the proportion co-ethnics, the proportion native whites, and the average mortality index. The contextual variables help to explain some of the ethnic differentials in mortality, and are also important in their own right. Their impact also differs considerably depending on the origin, with the more disadvantaged groups – blacks and immigrants from high-mortality origins – being more affected by the contextual setting.

We find substantial mortality differences between immigrant generations as well as in relation to marital exogamy. Overall second generation immigrants, as well as immigrants intermarried with natives or second-generation immigrants, have a clear mortality advantage. Much of this advantage disappears when adding controls for assimilation and place of residence, even though some of the advantage for intermarried immigrants remains. More interestingly, the associations are opposite for immigrants from low-mortality and high-mortality origins. For immigrants from low-mortality origins assimilation (being second generation or married to the native-born) is associated with higher child mortality, while it is associated with lower child mortality for immigrants from high-mortality origins. This is consistent with assimilation theory and suggests that demographic behavior – in this case child mortality – is converging to the native white pattern. A similar pattern of negative acculturation has been found for Hispanic immigrants in contemporary United States, who begin with lower mortality than the national average and then converge to the higher level in subsequent generations or with a longer time in the country (Landale, Oropesa and Gorman 2000; Hummer, Melvin and He 2015).

References

Abramitzky, R. and L. Boustan. Forthcoming. Immigration in American economic history. *Journal of Economic Literature*.

Alba, R. D. and R. M. Golden. 1986. Patterns of ethnic marriage in the United States. *Social Forces* 65: 202–23.

Bhalotra, S. and Rawlings, S. 2013. Gradients of the intergenerational transmission of health in developing countries. *Review of Economics and Statistics* 95: 660–672.

Burström B., F. Diderichsen, and L. Smedman. 1999. Child mortality in Stockholm during 1885-1910: the impact of household size and number of children in the family on the risk of death from Measles. *American Journal of Epidemiology* 149(12): 1134-1141.

Burström B., G. Macassa, L. Öberg, E. Bernhardt, and L. Smedman. 2005. Equitable child health interventions: the impact of improved water and sanitation on inequalities in child mortality in Stockholm, 1878 to 1925. *American Journal of Public Health* 95(2): 208-216.

Caldwell, J. C. 1979. Education as a factor in mortality decline: an examination of Nigerian data. *Population Studies* 33(3): 395-413.

Caldwell J. C. and P. McDonald. 1982. Influence of maternal education on infant and child mortality: levels and causes. *Health Policy and Education* 2:251-267.

Coale, A. J. and P. Demeney. 1966. *Regional Model Life Tables and Stable Populations*. Princeton: Princeton University Press.

Condran, G. A. and E. Crimmins. 1980. Mortality differentials between rural and urban areas of states in the northeastern United States 1890-1900. *Journal of Historical Geography* 6 (2) 179-202.

Desai S. and S. Alva. 1998. Maternal education and child health: is there a strong causal relationship? *Demography* 35(1): 1-81.

Drachsler, J. 1920. *Democracy and Assimilation: The Blending of Immigrant Heritages in America*. New York: Macmillan.

Dribe, M., J. D. Hacker and F. Scalone. 2016. Becoming American: How context shaped intermarriage during the great migration to the United States at the turn of the twentieth century. Manuscript.

Easterlin, R. A. 1996. How beneficent is the market? A look at the modern history of mortality. *European Review of Economic History* 3: 257-294.

Gordon, M. M. 1964. *Assimilation in American Life: The Role of Race, Religion, and National Origins*. New York: Oxford University Life.

Hacker, J. D and M. R. Haines. 2005. American Indian mortality in the late nineteenth century: the impact of federal assimilation policies on a vulnerable population. *Annales de Démographie Historique* 110: 17-45.

Haines, M. R., and S. H. Preston. 1997. The use of the census to estimate childhood mortality: comparisons from the 1900 and 1910 United States Public Use Samples. *Historical Methods* 30: 77-96.

Hummer, R. A. et al. 1999. Race/ethnicity, nativity, and infant mortality in the United States. *Social Forces* 77: 1083-1118.

Hummer, R. A., J. E. Melvin and J. E. He. 2015. Immigration: Health and mortality. *International Encyclopedia of the Social & Behavioral Sciences*, Vol. 11.

Landale, N. S., R. S. Oropesa, and Bridget K. Gorman. 2000. Migration and infant death: Assimilation or selective migration among Puerto Ricans? *American Sociological Review* 65: 888-909.

Lieberson, S. and M. C. Waters. 1988. *From Many Strands. Ethnic and Racial Groups in Contemporary America*. New York: Russel Sage Foundation.

Markides, K.S. and K. Eschbach. 2005. Aging, migration and mortality: Current status of research on the Hispanic paradox. *Journals of Gerontology: Series B*: 60B (Special Issue II):68–75.

Mosley, W. H. and L. C. Chen. 1984. An analytical framework for the study of child survival in developing countries. *Population and Development Review* 10: S25-S45.

Pagnini, D. L. and S. P. Morgan. 1990. Intermarriage and social distance among U.S. immigrants at the turn of the century. *American Journal of Sociology* 96: 405-432.

Palloni, A. and E. Arias. 2004. Paradox lost: explaining the Hispanic adult mortality advantage. *Demography* 41:385–415.

Preston, S. H. and M. R. Haines. 1991. *Child Mortality in Late Nineteenth-Century America*. Princeton: Princeton University Press.

Preston, S. H., D Ewbank and M. Hereward. 1994. Child mortality differences by ethnicity and race in the United States: 1900-1910. In: S. C. Watkins (ed.), *After Ellis Island. Newcomers and Natives in 1910 Census*. New York: Russel Sage Foundation.

Riosmena, F., R. Wong, and A. Palloni. 2013. Migration selection, protection, and acculturation in health: A binational perspective on older adults. *Demography* 50:1039–1064.

Ruggles, S., K. Genadek, R. Goeken, J. Grover, and M. Sobek. 2015. Integrated Public Use Microdata Series: Version 6.0 [Machine-readable database]. Minneapolis: University of Minnesota.

Sassler, S. 2005. Gender and ethnic differences in marital assimilation in the early twentieth century. *International Migration Review* 39: 608-636.

Singh, G. K. and R. A. Hiatt 2006. Trends and disparities in socioeconomic and behavioral characteristics, life expectancy, and cause-specific mortality of native-born and foreign-born populations in the United States, 1979-2003. *International Journal of Epidemiology* 35: 903-919.

Trussell, J. and S. Preston. 1982. Estimating covariates of childhood mortality from retrospective reports of mothers. *Health Policy and Education* 3: 1-36.

United Nations. 1983. *Indirect Techniques for Demographic Estimation. Manual X*. New York. United Nations.

Van Poppel, F., M Jonker, and K. Mandemakers. 2005. Differential infant and child mortality in three Dutch regions, 1812-1909. *Economic History Review* 58 (2): 272-309.

Walker, F. A. 1891. Immigration and degradation. *Forum* 11: 634–44.

Watkins, S. C. (ed.). 1994. *After Ellis Island. Newcomers and Natives in 1910 Census*. New York: Russel Sage Foundation.

Wildsmith, E., M. P. Gutmann, and B. Gratton. 2009. Assimilation and intermarriage for U.S. immigrant groups, 1880–1990. *History of the Family* 8: 563-584.

Woods, R. 2000. *The Demography of Victorian England and Wales*. Cambridge: Cambridge University Press.

Woods, R., N. Williams, and C. Galley. 1993. Infant mortality in England, 1550-1950. In: C. A. Corsini and P. P. Viazzo (ed.) *The Decline of Infant Mortality in Europe, 1800-1950*. Four National Case Studies. Florence: UNICEF.

Table 1. Descriptive statistics of the sample (weighted means and proportions).

Ethnic origin (%)	
Native White Native Parentage	55.1
Native Black Native Parentage	10.8
French Canadian	1.4
English Canadian	1.8
British	4.1
Irish	6.8
German	12.0
Danish	0.5
Norwegian	1.2
Swedish	1.4
Dutch	0.4
French	0.4
Italian	1.4
Portuguese	0.1
Russian Jewish	0.8
Eastern & Central Eur. Jewish	0.3
Polish	1.5
Mexican	0.3
Place of residence (%)	
Rural	62.9
Urban pop. 2,500-10,000	7.7
Urban pop. 10,000-100,000	12.3
Urban pop. <100,000	17.2
English speaker (%)	
No	2.6
Yes/NA	97.4
Spouse English speaker (%)	
No	1.6
Yes/NA	98.4
Literate (%)	
No/NA	9.9
Yes (read and write)	90.2
Spouse literate (%)	
No/NA	9.5
Yes (read and write)	90.5
Employed (%)	
No	91.0
Yes	9.0
Farm residence (%)	
No	59.0
Yes	41.0
Immigrant generation (%)*	

1G	32.2
1.5G	12.9
2G	54.9
Intermarriage (%)*	
1G Endogamy	42.0
2G Endogamy	19.5
NWNP Exogamy	17.9
2G Exogamy	12.9
Other Exogamy	7.8
Time in US before marriage (%)**	
0-4	44.5
5-9	25.0
10-14	15.1
15-19	10.4
20-24	3.9
25-29	0.8
30-34	0.3
Age (mean)	38.3
Mortality Reference Date (mean)	9.1
Occupation score (mean)	21.7
Contextual variables (county)	
Mortality index	0.89
Relative group size	0.46
Proportion NWNP	0.55
Sample N	139 229
Population	34 592 202

* 1G, 1.5G and 2G (N=47 558)
** 1G and 2G (N=19 663)
Source: IPUMS, Ruggles et al. 2015.

Table 2. Mortality index by ethnic origin (weighted means).

	Mean	s.e.
Native White Native Parentage	0.91	0.005
Native Black Native Parentage	1.47	0.014
French Canadian	1.30	0.038
English Canadian	0.93	0.029
British	0.97	0.019
Irish	1.10	0.016
German	0.88	0.010
Danish	0.77	0.052
Norwegian	0.82	0.032
Swedish	0.81	0.031
Dutch	0.75	0.055
French	0.93	0.066
Italian	1.15	0.038
Portuguese	0.99	0.113
Russian Jewish	0.74	0.038
Eastern & Central Eur. Jewish	0.80	0.068
Polish	1.11	0.035
Mexican	1.58	0.072
Total	0.99	0.004
N	139 229	

Note: Mortality index calculated based on the Model West 14 life table (Coale and Demeny 1966).

Table 3. Under 5 mortality (${}_5q_0$) comparisons. US ethnic groups and country of origin in 1900/1910.

	US ethnic	**Country of origin**
Britain	0.17	0.19
Denmark	0.14	0.15
Norway	0.15	0.15
Sweden	0.15	0.13
Netherlands	0.13	0.19
France	0.17	0.20
Italy	0.21	0.27

Note: Country of origin Britain is England and Wales. ${}_5q_0$ for US ethnic groups is calculated by multiplying the mortality index in table 2 with ${}_5q_0$ from Model West 14 life table (0.17972).
Source: ${}_5q_0$ for countries of origin from the Human Mortality Database (www.humanmortality.org).

Table 4. Regression estimates, mortality index (weighted OLS).

	M1		M2		M3 (FE)		M4 (Contextual)	
	Coef	P>t	Coef	P>t	Coef	P>t	Coef	P>t
Ethnic origin								
Native White Native Parentage	ref	ref	ref	ref	ref	ref	ref	ref
Native Black Native Parentage	0.571	0.000	0.423	0.000	0.385	0.000	0.394	0.000
French Canadian	0.391	0.000	0.245	0.000	0.307	0.000	0.274	0.000
English Canadian	0.026	0.376	-0.033	0.252	0.051	0.092	0.043	0.165
British	0.053	0.007	-0.009	0.649	0.030	0.144	0.049	0.032
Irish	0.180	0.000	0.060	0.000	0.082	0.000	0.086	0.000
German	-0.046	0.000	-0.116	0.000	-0.057	0.000	-0.030	0.048
Danish	-0.151	0.003	-0.155	0.002	-0.052	0.325	-0.012	0.814
Norwegian	-0.105	0.001	-0.112	0.000	0.018	0.638	0.064	0.062
Swedish	-0.114	0.000	-0.178	0.000	-0.069	0.036	-0.054	0.098
Dutch	-0.161	0.004	-0.215	0.000	-0.044	0.489	-0.080	0.141
French	0.006	0.927	-0.077	0.238	-0.048	0.464	-0.019	0.773
Italian	0.270	0.000	-0.044	0.281	-0.025	0.553	-0.013	0.761
Portuguese	0.099	0.376	-0.134	0.215	0.018	0.879	-0.041	0.709
Russian Jewish	-0.143	0.000	-0.379	0.000	-0.408	0.000	-0.390	0.000
Eastern & Central Eur. Jewish	-0.083	0.218	-0.330	0.000	-0.410	0.000	-0.381	0.000
Polish	0.230	0.000	-0.022	0.543	0.072	0.062	0.079	0.039
Mexican	0.690	0.000	0.385	0.000	0.417	0.000	0.415	0.000
Age	-0.006	0.037	0.005	0.110	0.008	0.012	0.007	0.030
Age squared	0.000	0.000	0.000	0.030	0.000	0.047	0.000	0.026
Mortality Reference Date			-0.011	0.000	-0.014	0.000	-0.014	0.000
Place of residence								
Rural			ref	ref	ref	ref	ref	ref
Urban pop. 2500-10,000			-0.016	0.309	-0.009	0.612	-0.010	0.511
Urban pop. 10,000-100,000			0.036	0.010	0.032	0.057	0.017	0.243
Urban pop. <100,000			0.157	0.000	0.089	0.000	0.046	0.002
English speaker								
No			0.207	0.000	0.159	0.000	0.164	0.000
Yes/NA			ref	ref	ref	ref	ref	ref
Spouse English speaker								
No			0.003	0.936	-0.051	0.237	-0.040	0.347
Yes/NA			ref	ref	ref	ref	ref	ref
Literate								
No/NA			ref	ref	ref	ref	ref	ref
Yes (read and write)			-0.145	0.000	-0.100	0.000	-0.106	0.000
Spouse literate								
No/NA			ref	ref	ref	ref	ref	ref
Yes (read and write)			-0.100	0.000	-0.087	0.000	-0.084	0.000

Occupation score			-0.004	0.000	-0.004	0.000	-0.004	0.000
Employed								
No			ref	ref	ref	ref	ref	ref
Yes			0.168	0.000	0.152	0.000	0.152	0.000
Farm residence								
No			ref	ref	ref	ref	ref	ref
Yes			-0.210	0.000	-0.208	0.000	-0.202	0.000
Contextual variables								
Relative origin group size							0.054	0.030
Proportion NWNP							-0.006	0.809
Mortality index							0.831	0.000
Constant	0.897	0.000	1.080	0.000	0.967	0.000	0.223	0.001
Number of obs	139 229		139 229		139 061		139 229	
F	111.07		108.18		72.33		135.97	
Prob > F	0.0000		0.0000		0.0000		0.0000	
R-squared	0.0242		0.0357		0.0715		0.0457	

Table 5. Net estimates from interaction models (weighted OLS). Contextual variables.

	Relative group size		Proportion NWNP		Mortality index	
	Coef.	P>t	Coef.	P>t	Coef.	P>t
NWNP	0.114	0.003	0.234	0.000	0.857	0.000
NBlNP	-0.188	0.000	-0.399	0.000	0.974	0.251
High mortality	0.580	0.000	-0.282	0.000	0.883	0.751
Low Mortality	0.263	0.081	0.019	0.004	0.493	0.000

Note: Models controls for the same variables as in M4, Table 4 except ethnic origin.
NWNP is the reference category and p-values in this category refer to base estimates for the contextual variables. Coefficients for other groups are net estimates (base estimate + interaction estimate) and p-values refer to the interaction estimates, testing if the difference to the base estimate is statistically significant.

Table 6. Regression estimates, immigrant generations and intermarriage (weighted OLS). Sample: 1G, 1.5G, 2G.

	M1		M1		M2 (Full)		M3 (FE)	
	Coef	P>t	Coef	P>t	Coef	P>t	Coef	P>t
Immigrant generation								
1G	ref	ref			ref	ref	ref	ref
1.5G	-0.032	0.156			0.009	0.697	0.044	0.060
2G	-0.147	0.000			-0.056	0.003	-0.031	0.115
Intermarriage								
1G Endogamy			ref	ref	ref	ref	ref	ref
2G Endogamy			-0.126	0.000	-0.050	0.014	-0.046	0.027
NWNP Exogamy			-0.129	0.000	-0.043	0.045	-0.028	0.222
2G Exogamy			-0.089	0.000	-0.030	0.201	-0.018	0.451
Other Exogamy			-0.010	0.696	0.006	0.809	0.007	0.788

Controls:
M1: Age, age squared, Mortality reference date
M2: M1 + Place of residence, English speaker, Spouse English speaker, Literate, Spouse Literate, Occupation score, Employed, Farm residence
M4: M2 + county fixed effects

Table 7. Net estimates from interaction models (weighted OLS). Immigrant generation and intermarriage.

A. Immigrant generation

	Low Mortality		High Mortality	
	Coef.	P>t	Coef.	P>t
1G	ref	ref	ref	ref
1.5G	0.092	0.257	0.031	0.487
2G	0.028	0.008	-0.067	0.038

B. Intermarriage

	Low Mortality		High Mortality	
	Coef.	P>t	Coef.	P>t
1G Endogamy	ref	ref	ref	ref
2G Endogamy	0.019	0.004	-0.102	0.005
NWNP Exogamy	0.039	0.049	-0.058	0.174
2G Exogamy	0.020	0.363	-0.027	0.529
Other Exogamy	0.037	0.538	-0.001	0.978

Note: High mortality is the reference category, p-values refer to base coefficients in the regression models.
P-values for low mortality refer to interaction terms, and tests if the estimate for high mortality and low mortality are different.
Based on full models with county-level fixed effects.

Table 8. Regression estimates, time in the US before marriage (weighted OLS).

A. 1G, 1.5G

	M1		M2 (Full)		M3 (FE)	
	Coef	P>t	Coef	P>t	Coef	P>t
0-4 years	ref	ref	ref	ref	ref	ref
5-9 years	-0.009	0.738	-0.001	0.980	-0.003	0.906
10-14 years	0.011	0.710	0.040	0.179	0.072	0.026
15-19 years	-0.047	0.179	-0.013	0.712	0.007	0.861
20-24 years	-0.133	0.013	-0.089	0.095	-0.072	0.206

B. 1G

	M1		M2 (Full)		M3 (FE)	
	Coef	P>t	Coef	P>t	Coef	P>t
0-4	ref	ref	ref	ref	ref	ref
5-9	-0.019	0.491	-0.012	0.665	-0.017	0.580
10-14	-0.004	0.943	0.017	0.734	0.039	0.468
15-19	0.029	0.826	0.066	0.605	0.006	0.962
20-24	0.029	0.889	0.042	0.839	-0.039	0.853

Controls:
M1: Age, age squared, Mortality reference date
M2: M1 + Place of residence, English speaker, Spouse English speaker, Literate, Spouse Literate, Occupation score, Employed, Farm, Intermarriage
M4: M2 + county fixed effects

Figure 1. Mortality index in U.S. counties in 1910.