Within Families and Between Countries
Investigating the Lifespan of US Immigrants Compared to Origin, Destination and Non-Migrant Siblings in Cohorts from 1850 to 1890

Saverio Minardi, University of Bologna
Paul Puschmann, Radboud University Nijmegen
Nicola Barban, University of Bologna

Abstract

In this research note, we explore the relationship between migration and lifespan for individuals born between 1850 and 1890 who migrated to the U.S from anywhere in the world. Using micro-level genealogical data from Familinx, our study compares the lifespans of migrants to those of the U.S.-born population, the non-migrant population in their countries of origin, and their non-migrant siblings. We employ sibling fixed effect models to investigate whether migrants enjoy survival advantages compared to their non-migrant siblings, considering that health and mortality are often clustered in families due to shared environments, behaviors, resources, and household dynamics. We also examine differences in the migrant mortality nexus across the main migrant groups of the time. Consistent with previous studies, we find a negative healthy migrant effect compared to U.S.-born individuals and a positive migrant effect compared to the origin populations and non-migrant siblings. Country-specific analysis reveals cross-country differences in these effects. The general negative effect compared to the U.S. population is predominantly driven by historically disadvantaged groups such as Irish and Eastern European migrants. All groups except the Irish experience a mortality advantage compared to their origin populations and siblings.

Key words: Healthy migrant; immigrant paradox; unhealthy migrant; lifespan; migrant selection; USA

Words: 3981

Introduction

The healthy migrant effect (HME), or the immigrant paradox, is the counterintuitive observation that migrants are in better health than non-migrants despite worse average socio-economic conditions and lower educational status. Migrants have lower risks of dying even if they grew up in countries with a higher mortality regime and live under worse circumstances than the native population in the destination society.

While several reasons, such as statistical artefacts or the selective return of frail migrants (the Salmon Bias hypothesis) have been invoked to explain this phenomenon, increasing evidence points to the critical role of positive selection into migration as a key explanation (Palloni and Arias, 2004; Feliciano, 2020). This mechanism implies that individuals who migrate are positively selected from their population of origin, based on a set of characteristics related to better health. As a result, migrants exhibit a health advantage not only compared to individuals in the destination country, but also to those in the origin country.

Despite evidence of a healthy migrant effect in numerous contexts, our understanding of the actual characteristics driving the selection process and how they vary across time and place is still incomplete. Most studies have compared the health of migrants to the population in the destination country, and when compared to the country of origin, they more often focused on a single national group. This has led to two underexplored areas.

First, most studies assume that the HME is a universal process related to the migration experience. However, migrants come from very different contexts and migrate for very different reasons, meaning that their health advantage (or disadvantage) may differ with regard to both origins, destinations, and the interaction of the two. Very few studies have documented how migrants' selection differs across origin groups, comparing migrants of multiple nationalities to the respective origin and destination populations.

Second, the factors linking migration to better health are not well understood. Migration and health may both be influenced by individual attributes — like skills, ambition, and health — or by group and family-level characteristics—such as religious beliefs, social background, or regional origin. However, most studies lack detailed information about these individual traits before migration and cannot always identify an appropriate comparison group from the same pre-migration background. Therefore, existing studies tend to compare migrants with the general population of their home countries, making it difficult to pinpoint the exact mechanisms linking migration to health.

In order to shed light on these unexplored areas we investigate the lifespan advantage of migrants to the United States from seven major nationalities in the late 19th and early 20th centuries, comparing their lifespan to the U.S.-born population and to that of their country of origin. Moreover, we compare migrants to their non-migrant siblings in the origin country. Using siblings fixed effects models we explore whether selection into migration is based on family-constant characteristics or within-family-individual characteristics.

We overcame the complex data requirements to address these questions by using online crowdsourced genealogies. Genealogical data offer a unique opportunity for the study of migration because they record individual and family histories across borders, making it possible to identify migrants and connect them to their country and family of origin (Kaplanis, 2018).

We focus on the US during the turn of the twentieth century for several reasons. First, it was a period of mass and relatively uncontrolled migration until the first strong migration laws for European immigrants were introduced in the 1920s, the Emergency Quota Act of 1921 and the Immigration Act of 1924. Second, previous studies have shown that during those years migrants displayed a disadvantage compared to the U.S.-born population making it a rare exception, since the healthy migrant effect is observed in a multitude of contexts (Bakhtiari, 2023). Finally, despite being an historical period, the US-vital statistics recorded aggregate mortality rates disaggregated by migration status, offering a benchmark and an opportunity for comparison for our micro-level analysis. These are important reference points considering the possible biases of online genealogies.

Results indicate that the reversed HME effect observed during those years was not uniform among all origins groups. Historically disadvantaged groups, outside the White Anglo-Saxon Protestant (WASP) norm, such as Irish and Eastern European were the most penalized. Irish migrants, in particular, exhibited a lifespan penalty, not only compared to the U.S.-born population, but also compared to non-migrant Irish. We also highlight that the migrant effect on lifespan is robust after the inclusion of sibling fixed-effects, suggesting that selection into migration results from individual characteristics rather than family-constant ones, i.e., migrants do not come from families selected on the health and mortality status.

Our findings are consistent with recent research confirming a heterogeneous migrant effect, contingent on migrant origin and destination, as well as the presence of migrant advantages (or disadvantages) when comparing to siblings and relatives. Mourits and Puschmann (2023) find a mortality advantage for migrants but not for their siblings or offspring within the Dutch province of Zeeland. Using Familinx data and focusing on migrants' from the the U.K., Pojman et al. (2023) find that degree of the migration effect on longevity varies depending on the migrant destination, even when compared to non-migrant siblings. Feliciano (2005) finds that migrants to the U.S. vary substantially in their degree of selectivity on education, depending upon the origin country. Similarly, Ro et al (2016) find different degrees of health selection comparing migrants to the U.S. to various origin countries.

Data and methods

The investigation of lifespan differences between migrants, their families in the country of origin, and the U.S.-born population requires numerous and complex information that is not collected by any administrative data source. It requires not only information on migration status, birth and death years and locations, but also on family links, meaning information on at least one parent for each reference individual. Data of this kind would require to follow individuals and families across multiple borders for their entire lifespan, which is an enormous task for any centralized institution.

To overcome these challenges, we use Familinx, a dataset built from online crowdsourced genealogies downloaded from the website Geni.com (Kaplanis et al., 2018). This is a bottom-up data source produced by individuals and amateur genealogists who have reconstructed their own family trees and, as such, it is not bound to any national borders. Individuals record their family members and their basic demographic information regardless of their place of birth and death. Online genealogies have recently been used to study several historical demographic processes, and to investigate the nature and direction of their biases, suggesting that, despite certain biases, they are consistent with official data sources in terms of long-term trends (Hsu et al., 2021; Stelter and Alburez-Gutierrez, 2022; Minardi et al., 2023; Cozzani et al., 2023; Blanc, 2023a, 2023b; Calderón-Bernal et al. 2023).

Overall, existing literature suggests that genealogical data tends to underestimate mortality and especially among younger individuals and females. This underestimation arises from structural reasons, such as the omission of childless individuals and the omission of lineages that did not result in descent (Minardi et al., 2023). This is less of a problem for our current study as long as the direction and magnitude of bias are comparable for migrants and natives. None of the existing studies has investigated the representativeness of the data in terms of migration background and mortality differentials. We therefore present in the Supplement Materials to this article a detailed comparison to official data sources.

We first compared trends in the share of migrants by nationality to the U.S. censuses from 1850 to 1940 (Figure A1). Results show that both levels and trends are remarkably similar, with the exception of Irish migrants, who are underreported[1]. Second, we compare age-specific mortality rates for white natives and migrants by sex and period to U.S.-vital statistics (Figure A2). We observe that for both groups, Familinx underestimates mortality rates for individuals below age 30. To reduce this source of bias, we limit our sample to individuals who lived at least to the age of 45. Third, we compare the magnitude of bias for migrants and natives. Figure A3 shows the ratio between the mortality rates in the U.S.-vital statistics and Familinx by age, nativity, period, and sex. A ratio of one would indicate that the mortality rate is identical in the two data sources, and a higher ratio would indicate an underestimation of mortality in Familinx. More than the actual underestimation we are interested in differences between migrants and non-migrants. We do see that Familinx data tend to underestimate mortality more for migrants than natives and more so at younger ages, but the difference is contained above age 45. As discussed below, differences in underestimation between natives and migrants may be partially due to the underrepresentation of Irish who are the most disadvantaged group. However, by benchmarking our microlevel analysis to aggregate official statistics we are able to ascertain the consistency of our results in

---

[1] This underestimation could result from two factors. First an under-recording of Irish ancestors due to lower fertility, lower rate of surviving descent, lower available documentation, or lower rate of genealogists among Irish descendants. Second, different patterns of return migration. Migration in the censuses is observed in the census year, while migration in Familinx is inferred by comparison between place of birth and place of death, meaning that if migrants returned and died in the origin country they would not be considered as migrants at any point.

terms of directionality of the association. Therefore, it is important to keep in mind that our estimates of an un-healthy migrant effect are likely conservative, while any evidence of a healthy migrant effect is likely overestimated.

The raw Familinx data are composed of approximately 86 million records without any geographical or time limitations. However, the majority of records do not have information on all variables of interest. We therefore first restrict our sample to individuals with complete information on the country of birth and death[2], as well as their birth and death year. We further limit our sample to individuals who die after reaching 45 years of age due to representativity reasons and to ensure that all individuals have the chance to migrate. Our focus is on migration to the U.S. and we restrict the sample to individuals born between 1850 and 1890. This means that all individuals in our sample were at least 34 years old when the Immigration Act of 1924 was introduced and all individuals included died before 2000.

We built a set of different data to explore the migrant advantages in regards to both origin and destination population. First, we built the destination data, made up of U.S.-born individuals and migrants. Second, we built nine origins data, one for each country of origin, each including individuals, who were born and died in the origin country and those who were born in the origin country and died in the U.S. We investigate nine specific origins groups which are the most numerous in the 1850 to 1940 U.S. censuses. The origin groups consist of migrants from Germany, Ireland, The UK[3], Italy, Russia, Scandinavia, Eastern Europe, and Western Europe[4]. A stylized depiction of the sample structures is reported in Figure 1 and the sample distribution is reported in Table A1 of the online appendix. Sample sizes range from a maximum of 679,918 records for the destination sample and a minimum of 9,942 for the Irish origin one.

For each of the original samples, we also constructed a second version which includes only individuals with information on at least one parent in order to identify same-sex siblings and to estimate siblings fixed-effects.

We first estimate the origin and destination effects of migration on lifespan without distinguishing by country of origin. For these models, the dependent variable is *lifespan* (calculated as the year of death minus the year of birth) and the main independent variable is *migration*, defined as having a country of

---

[2] Information on location is recorded both as unstructured strings and coordinates (see Kaplanis et al., 2018 for more details). We obtain structured categorical information from unstructured strings by matching each entry to a list of possible country names and abbreviations from different languages and periods through the Damerau-Levenshtein distance algorithm. We extract information from coordinates through reversed geoparsing. The two procedures give almost identical results, but, when both information sources are available, we give priority to unstructured text as it was directly recorded by users.

[3] North Ireland is considered as Ireland. We refer to the UK instead of Great Britain because some records are broadly located by users in "UK" and therefore we cannot define with certainty their precise location. Some could be from one of the colonies during the time the recorded individual was alive. As a robustness check, we ran the analysis including only those individuals that can be located with certainty in England, Wales, or Scotland as their location matches one of the three names of a British country. Results from this check are almost identical.

[4] Scandinavia includes Sweden, Finland, Norway, and Denmark; Eastern Europe includes Poland, Ukraine, Belarus, Hungary, Lithuania, Czech Republic, Romania, Slovakia, Croatia, Latvia, Estonia, Slovenia; Western Europe includes Austria, the Netherlands, Switzerland, France, Belgium, Luxemburg.

death different from the country of birth. We estimate baseline models separately by sex, and control for birth year. In the second step, we estimate the origin effects, including siblings' fixed effects for the siblings subsample.

After assessing the general effect of migration on lifespan, we further investigated differences based on nationality and sex, considering both countries of origin and destinations. For the destination model we included nationalities-specific migration dummies; for the origin models we estimated separate migration effects for each country. All models included birth-year fixed effects. Finally, we estimated the origins models, separately by origin group, including sibling fixed-effects.

Results

Figure 2 descriptively reports the full depiction of the relationship of interest. For each nationality it shows birth cohort rolling averages of average lifespan (conditional of living until age 45) for migrants, non-migrants in the country of origin, and non-migrants in the U.S. First, we observe an increase in average lifespan for all nationalities considered. Second, Figure 2 highlights the temporal and national specifics of migrants' lifespan in relation to non-migrants. As documented elsewhere, we also find a migrant disadvantage for specific cohorts and nationalities, most pronounced for Irish migrants throughout the full period. The Irish are the most disadvantaged migrant groups and their underrepresentation likely influences our pulled estimates. Striking is also the fact that migrants from Eastern Europe and Russia enjoy a huge health advantage compared to their origin population.

Figure 3 reports the average estimate for all nationalities and cohorts, by origin and destination, disaggregated by sex (Full results in Table A3, Online Appendix). It presents different specifications, a baseline association including fixed effects for birth cohort, and, for the origins sample, a model including siblings fixed effects.

The first coefficient from Figure 3 confirms the presence of a migrant mortality penalty compared to the U.S.-born population. This negative effect is likely conservative due to the biases in our data and the undercounting of Irish. The largest group of migrants, the Irish, consisted of people who were negatively selected, fleeing harsh conditions at home following the Great Famine of the 1840s and early 1850s, which resulted in an estimated one million deaths (Kinealy 2006). From 1847 to 1854 1.2 million Irish fled to the U.S. Upon arrival, Irish migrants were highly disadvantaged in the destination country. They faced strong stigma and prejudice, because of their sheer numbers, but also because of their different religion - Catholicism -, their poverty and their different perceived race. Of all immigrant groups the Irish were hated the most among nativists who tried to keep them at the bottom of the society (Hardin 2016). Due to discrimination and a lack of skills, the Irish occupied the most dangerous, unhealthy and worst paying jobs, such as mining and the construction of railroads and canals. Last but not least, the

Irish lived predominantly in overcrowded urban ethnic enclaves in the North East (Gurpegui Palacios 2020).

When looking at the origin model we observe a reversal of the effect of migration on health. This result suggests that, despite not exhibiting an advantage compared to the destination population, migrants are nevertheless positively selected compared to their country of origin. Thus, the reversed immigrant paradox observed at the end of the 19th and the first half of the twentieth century is not explained by the absence of a migrant selection process, at least not for the general migrant population, but rather in the conditions at destination, i.e. the higher living and working standards of U.S. natives compared to migrants.

Finally, an important result which emerges from Figure 3 is that a positive migrant effect is robust after including siblings fixed-effects, which means that the healthy migrant effect (in relation to the origin population) remains after controlling for all family constant characteristics, which include important factors such as country and region of origin, parental resources, neighbors and religion. Instead, the results point in the direction of individual-level characteristics as the main drivers, factors that vary between siblings, which may include, personal drive, health, and skills.

As shown in Figure 2, pooled relationships do not necessarily correspond to the experience of migrants of each nationality. In Figure 4 we present the results of origins and destination models, separately for each origin group (full results in tables from A4 to A6, Oline Appendix). Here, national differences are clearly portrayed. The reverse HME at the destination appears to be limited to specific origin groups: Irish, Russians, and Eastern Europeans exhibit strong negative effects, and to a lesser extent also British males and Italian females. Many of these groups not only came from more disadvantaged countries but were also overrepresented among the lower and manual classes and predominantly resided in urban centers (Table A2, online appendix). Other groups show either no advantage or a positive effect compared to the US population. These less disadvantaged groups more often lived in rural contexts and were independent farmers and landowners (Table A2, online appendix). It is striking that most of the migrant groups who faced a health disadvantage compared to the U.S.-born population were at the time considered as a threat to US society, as they were deemed incompatible with mainstream society that was dominated by White Anglo-Saxon Protestants. These migrants were perceived as ‘racially inferior’ and faced xenophobia, which was voiced by nativist political movements, such as the Know Nothings (Lee 2021). Canadians do not show major differences, which is not surprising, given the geographical proximity and migration from Canada did not require major selection. One exception are male Italians, who show a positive effect with respect to the US population regardless of being a discriminated group, living in urban centres, and being highly represented in lower classes. This may be explained by several factors that cannot currently be investigated, such as specific social and behavioral protective factors, specific patterns of return migration, or selective recording of more longevous ancestors.

With respect to origin countries, all groups show a positive selection, confirming the result estimated for the general population, with the exception of Irish. Irish migrants emerge once again as the most disadvantaged group regardless of gender, and are the only group to be negatively selected from the origin population, underlying different migration motives predominantly driven by pushing factors due to harsh living conditions in Ireland. This result is in line with the existing literature documenting that Irish migrants to England were in worse health compared with both the English-born and the Irish-born who remained in Ireland (Delaney et al., 2013). Russian and Eastern European migrants show a very strong positive effect compared to their origin population despite a negative effect compare to U.S.-born, confirming results on later periods documenting the Russian immigrants report lower levels of disability compared with Russians in Russia but poor health relative to U.S.-born whites (Mehta and Elo, 2012).

We finally tested the origin models controlling for siblings fixed effects, which is accounting for selection processes at the family level for each specific nationality (Figure 5, full results in tables A7 and A8, Online Appendix). The fixed effects models require restricting the sample to individuals with information on at least one parent. Furthermore, only observations with within-siblings variation on migration contribute to the estimate, meaning individuals with at least one recorded sibling with a different migration status. As a result of this strict data requirement, the sample size for each nationality estimate is quite low. Nevertheless, controlling for sibling fixed effects does not overturn results. Quite the contrary: the estimates from Figure 5 are consistent with previous estimates, suggesting that individual characteristics rather than group-level characteristics drive migrant selection processes.

Discussion

The presence of an HME is considered nearly a universal phenomenon. Despite considerable evidence that migrants are, on average, healthier and longer-lived than non-migrants in both origin and destination countries, the precise mechanisms behind this association remain unclear. Moreover, existing studies have mostly focused on single nationalities, comparing migrants to the general population in the origin or destination country, and extending this positive association to migrants in general, regardless of the origin and specifics of each group.

Here, we leveraged recent online crowdsourced genealogies to investigate the association between migration and lifespan in the U.S. for cohorts born between 1850 and 1890. We compared migrants' lifespan to the destination population, the origin population, and to their siblings in the country of origin. Moreover, we distinguished by nationality of origin in order to investigate heterogeneity across migration groups.

In line with existing evidence, we show that during the research period the general migrant population in the U.S. had a lifespan disadvantage compared to the native population, but not their origin population (except for the Irish until the 1880's). This can be attributed to a combination of higher lifespan among U.S. born and lower lifespans among origin populations. However, upon checking for the origin heterogeneity, only a few but numerous groups exhibited an actual penalty, suggesting that the reverse in migrant effects was likely driven by a few disadvantaged origins groups, who faced a particularly hard time in the U.S., mostly likely related to xenophobia as the majority who faced a health penalty were non-WASP.

Concerning positive selection from the country of origin, we show that a positive association is present even when comparing migrants to their non-migrant siblings. First, this is evidence that the positive selection commonly observed is not the result of group level characteristics or different comparison groups at the subnational level. Second, it suggests that the selection into migration is not driven by family constant characteristics, but rather by individual ones that vary within families and between siblings.

Despite the unique contribution of the data used here, they are not without limitation. As shown by previous literature, reported records are selected based on a number of characteristics and individuals registered are usually longer lived and likely in higher status. This bias is acceptable in our study as long as it is constant across migrant and non-migrant groups. We contribute to the recent debate on the representatives of genealogical data by showing their portrayal of migrant characteristics and migration processes. By comparing nationalities to the census data we show that genealogy fairly represents migrants' composition by nationality and trends in their variation.

Second, we compare age-specific death rates for migrants and natives to the U.S. vital statistics. We show that Familinx generally underestimates mortality but this underestimation declines with age. We also show that the underestimation of mortality is stronger for migrants than natives, but once again the statistics based on the genealogies and the data from the statistical office tend to converge with higher ages. While we are partially underestimating the negative migrant effect compared to destination and overestimating the positive effect compared to origin, the bias should be contained given our age cut-off at 45 years.

References

Bakhtiari, E. (2022). The missing mortality advantage for European immigrants to the United States in the early twentieth century. Demography, 59(4), 1517-1539.

Blanc, G. (2023a). The Cultural Origins of the Demographic Transition in France.

Blanc, G. (2023b). Crowdsourced Genealogies. https://www.guillaumeblanc.com/files/theme/Blanc_crowdsourced.pdf

Bostean G. 2013. Does selective migration explain the Hispanic paradox? A comparative analysis of Mexicans in the U.S. and Mexico. J. Immigr. Minor. Health 15:624–35

Calderón-Bernal, L. P., Alburez-Gutierrez, D., & Zagheni, E. (2023). *Analyzing Biases in Genealogies Using Demographic Microsimulation*. Max Planck Institute for Demographic Research.

Cozzani, M., Minardi, S., Corti, G., & Barban, N. (2023). Birth month and adult lifespan. *Demographic Research*, *49*, 201-218.

Delaney, L., Fernihough, A., & Smith, J. P. (2013). Exporting poor health: the Irish in England. Demography, 50(6), 2013-2035.

Dribe, M., Hacker, J. D., & Scalone, F. (2020). Immigration and child mortality: Lessons from the United States at the turn of the twentieth century. *Social Science History*, *44*(1), 57-89.

Feliciano C. (2005). Educational selectivity in U.S. immigration: How do immigrants compare to those left behind? *Demography* 42:131–52

Feliciano, C. (2020). Immigrant selectivity effects on health, labor market, and educational outcomes. *Annual Review of Sociology*, *46*, 315-334.

Gurpegui Palacios, J.A. (2020). So Far So Close: Irish and Mexican Migrant Experience in the United States, Oceánide 13, 111-119

Hardin, C (2016). The Irish Experience in America: Nativism, Argumentation, and Responses. Dissertation from University of Amsterdam.

Hsu, C. H., Posegga, O., Fischbach, K., & Engelhardt, H. (2021). Examining the trade-offs between human fertility and longevity over three centuries using crowdsourced genealogy data. *Plos one*, *16*(8), e0255528.

Kaplanis J. et al. 2018 Quantitative analysis of population-scale family trees with millions of relatives.Science360,171-175.DOI:10.1126/science.aam9309

Kinealy, C. (2006). *This Great Calamity: The Irish Famine 1845-1852*. Dublin: Gill and Macmillan

Lee, E. (2021). *America for Americans: A History of Xenophobia in the United States*. New York: Basic Books.

Mehta, N. K., & Elo, I. T. (2012). Migrant selection and the health of US immigrants from the former Soviet Union. Demography, 49(2), 425-447.

Minardi, S., Corti, G., & Barban, N. (2023). Historical patterns in the intergenerational transmission of lifespan and longevity: Evidence from the United States, 1700-1900. *SocArXiv. doi*, *10*.

Mourits, R. J., & Puschmann, P. (2023). Exploring familial factors in the migrant mortality advantage among domestic migrants in later life: Zeeland, the Netherlands, 1812–1962. *SSM-Population Health*, *22*, 101359.

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

Ro A, Fleischer NL, Blebu B. 2016. An examination of health selection among U.S. immigrants using multinational data. Soc. Sci. Med. 158:114–21

Stelter, R., & Alburez-Gutierrez, D. (2022). Representativeness is crucial for inferring demographic processes from online genealogies: Evidence from lifespan dynamics. *Proceedings of the National Academy of Sciences*, *119*(10), e2120455119.

Figures

Figure 1 Conceptual representation of comparison populations and empirical samples

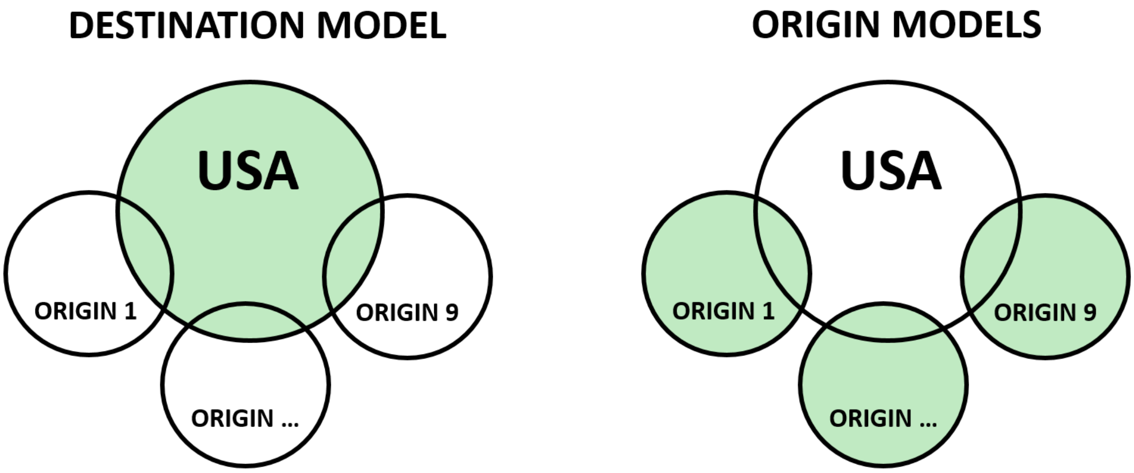


Figure 2 Trends in lifespan (rolling averages) by nativity, country of origin, and cohorts

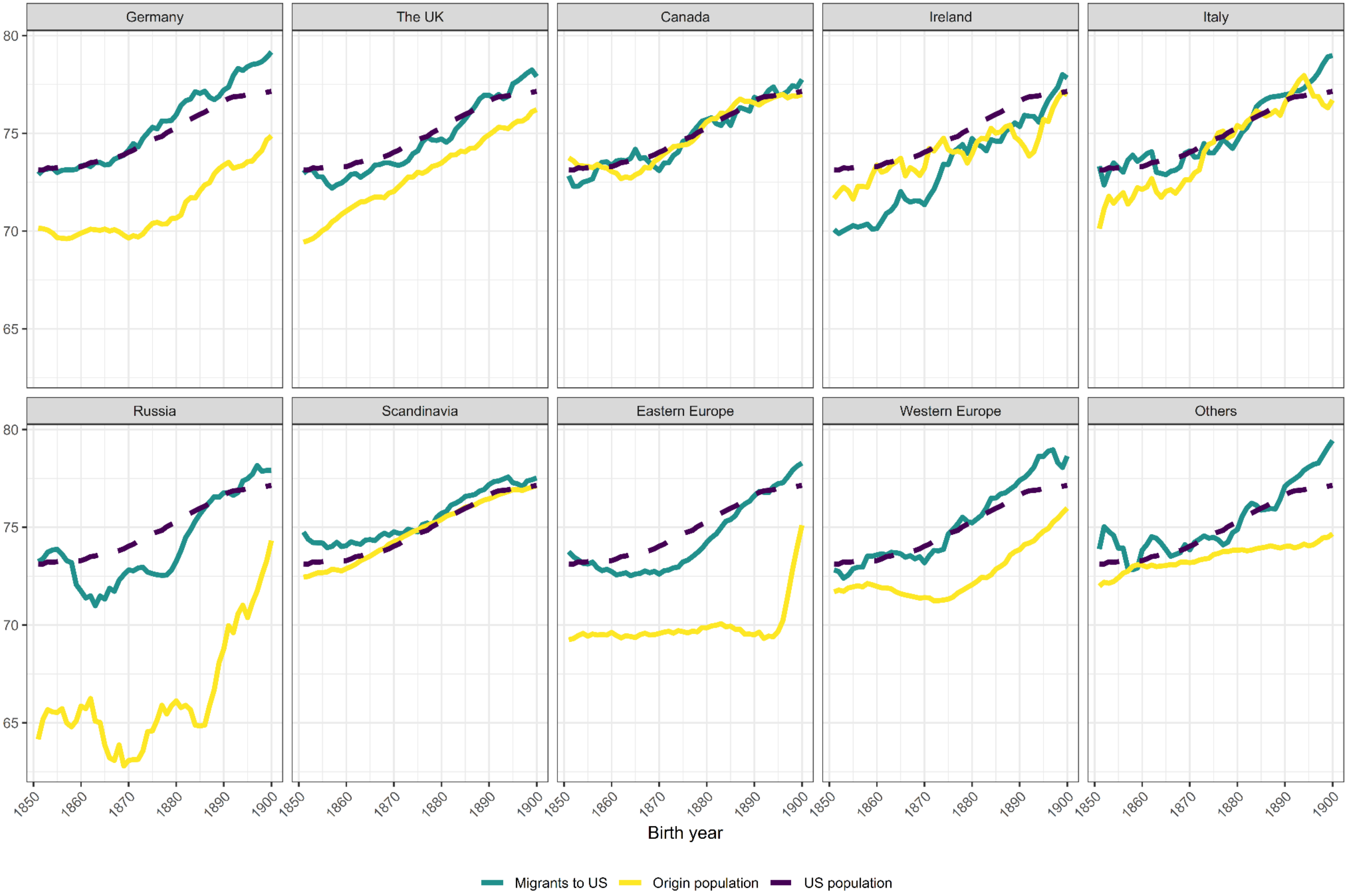


Figure 3 Average marginal effect of migration on lifespan by origin and destination and including siblings fixed effects

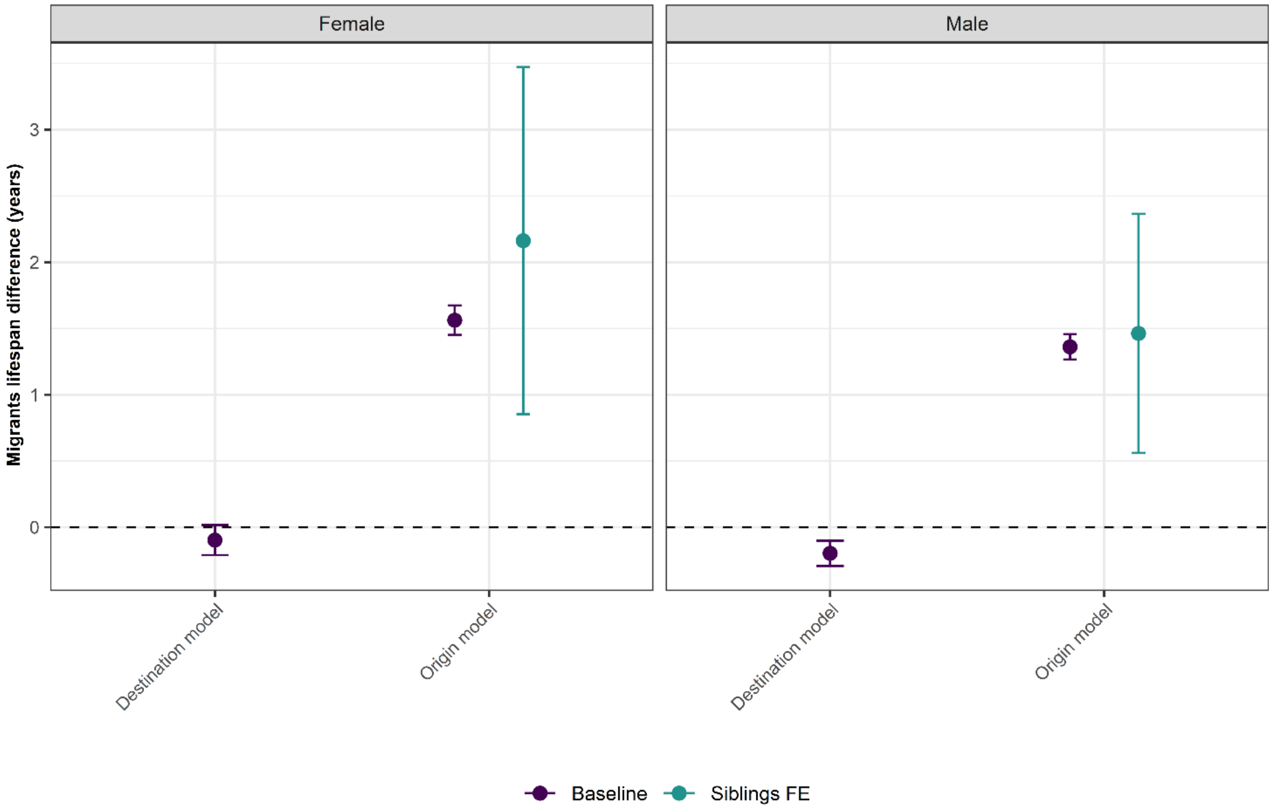


Figure 4 Average marginal effects of origin country on lifespan compared to origin and destination population by sex

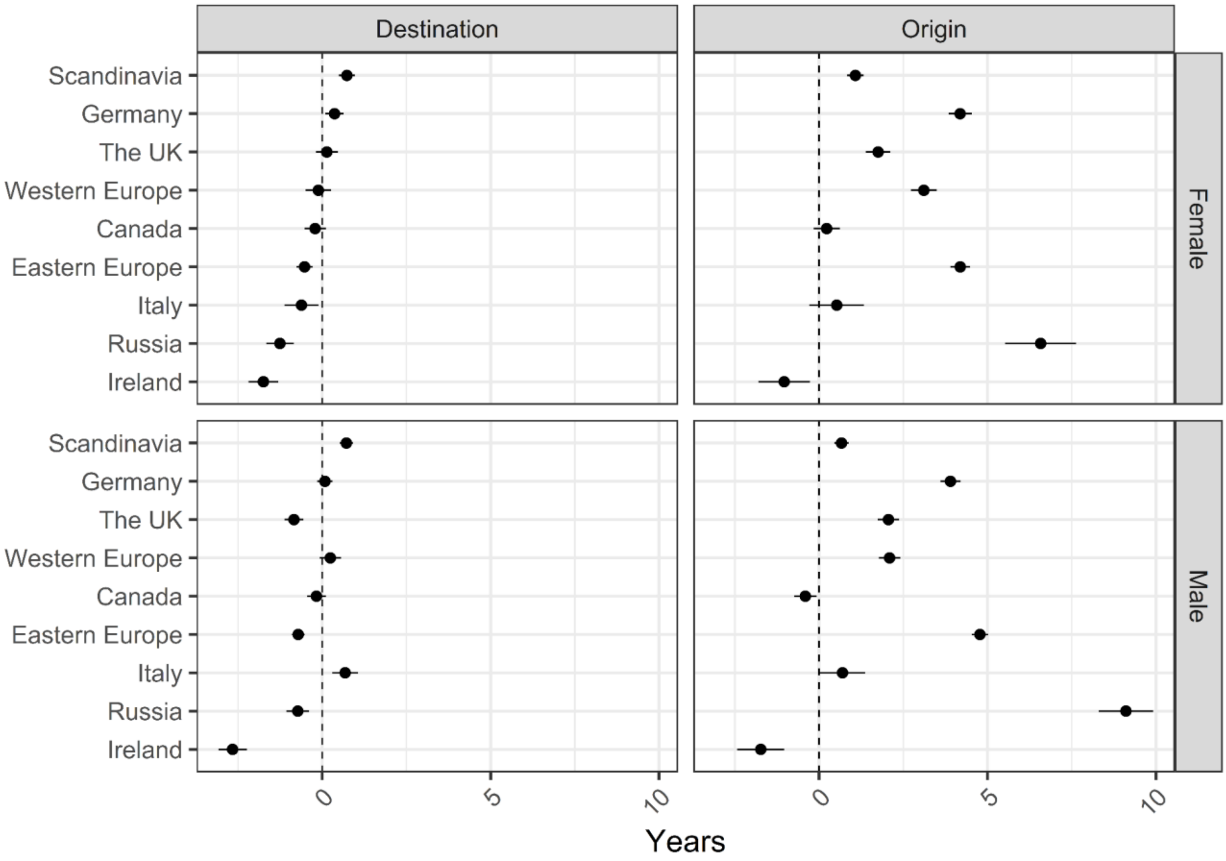
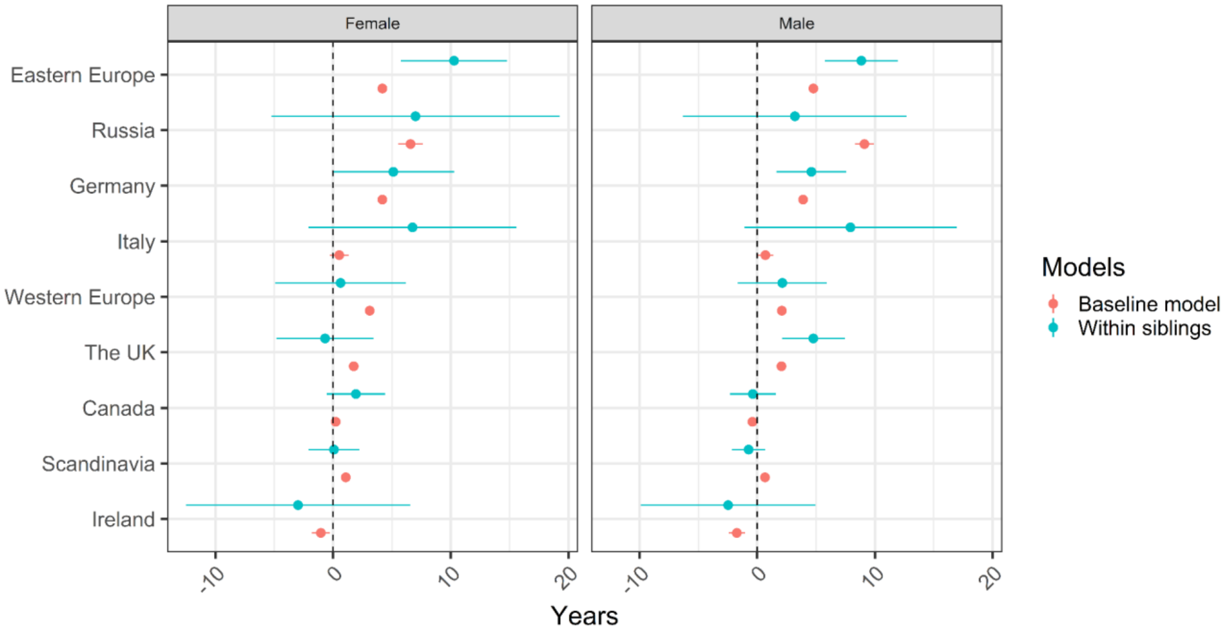


Figure 5 Average marginal effect of migration on lifespan compared to origin population including siblings fixed effects



Supplementary materials

Figure A1 Percentage of migrants over the total population in Familinx (yellow) and Census (purple)

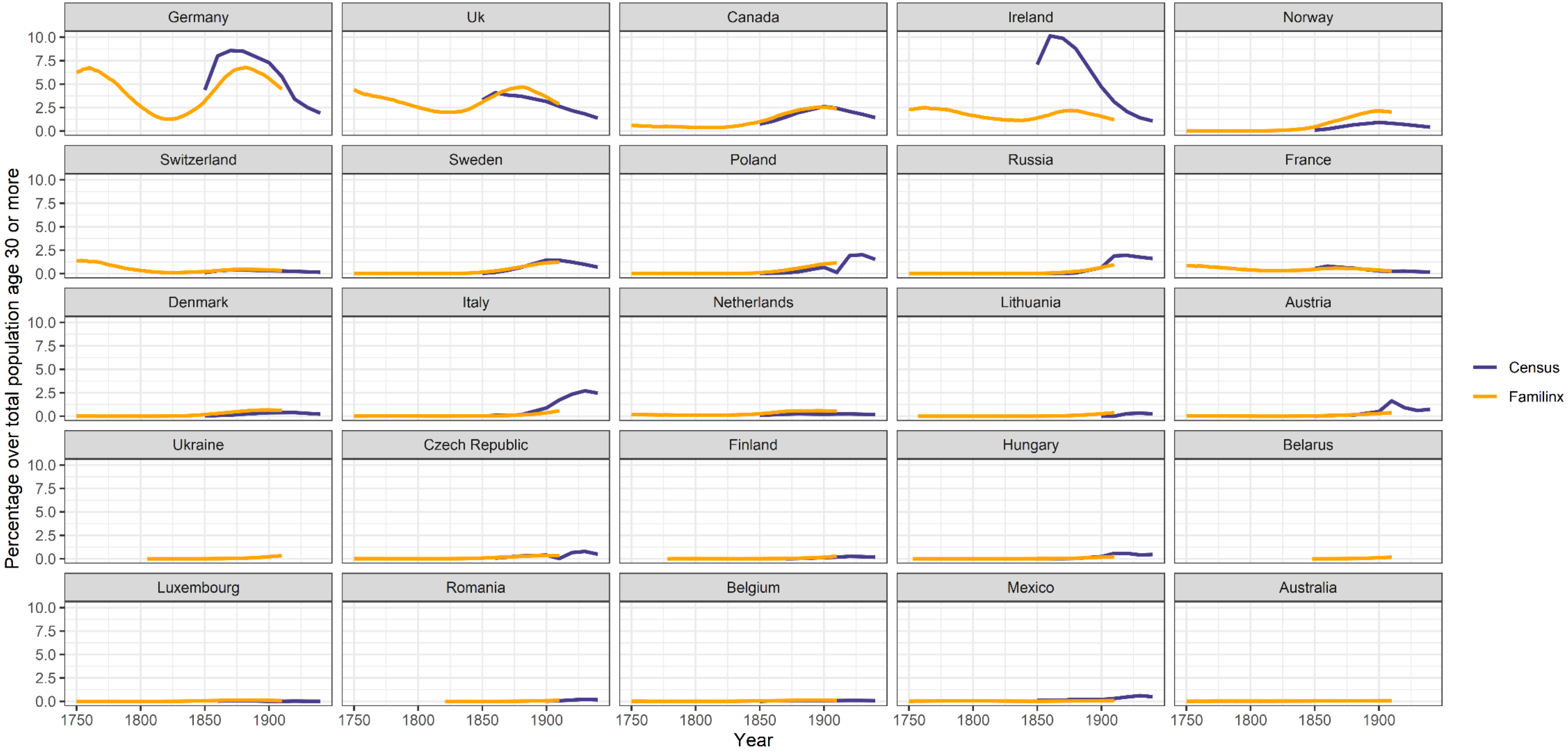
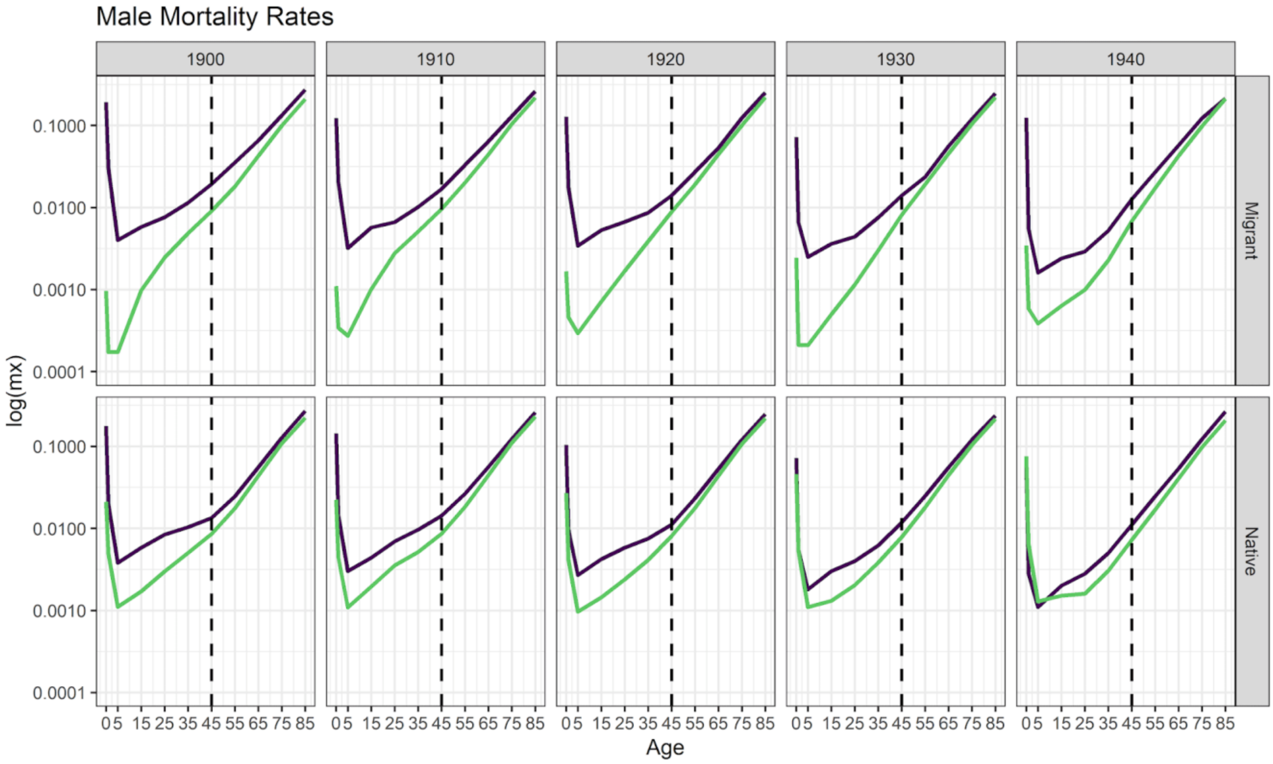


Figure A2 Age specific morality rates for migrants and natives by period and sex (1900-1940)



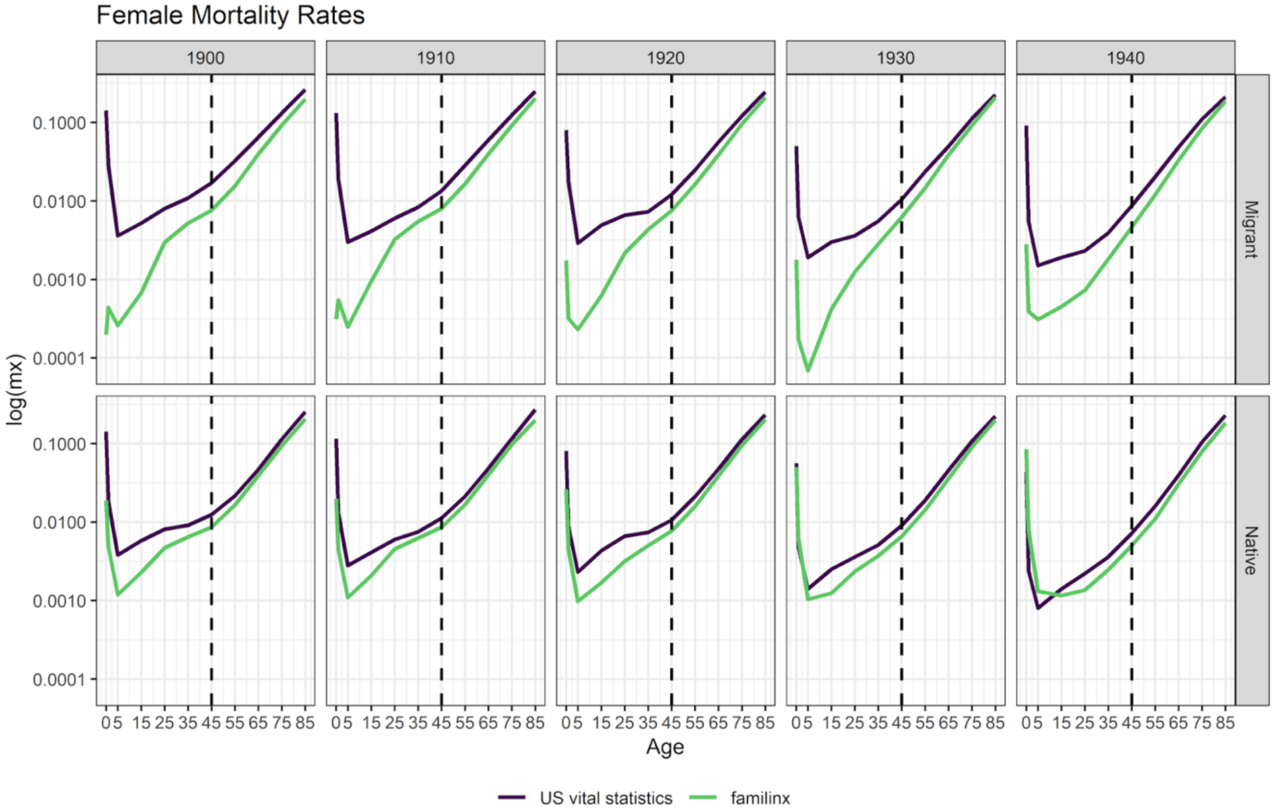
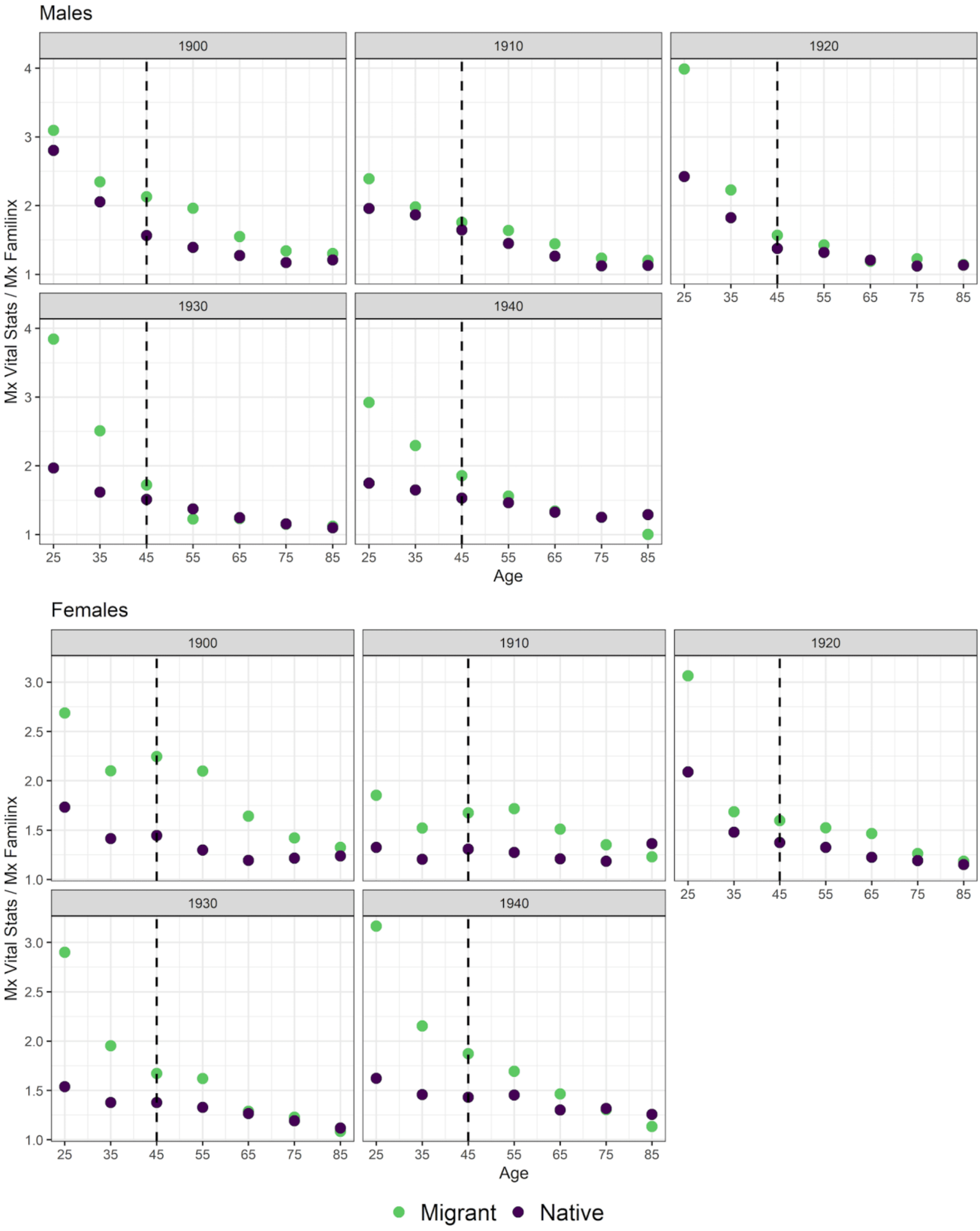


Figure A3 Ratio of US Vital Statistics mortality rates to Familinx mortality rates by age, sex, period and nativity (1900-1940)



Tables

Table A1 Sample distribution by origin, destination and migration status

Country of origin	Non-migrant	Migrant to the US	Total	% of migrants in the US	% of migrants in origin country
Destination					
USA	542137	137781	679918	20%	
Origin					
Canada	27783	13698	41481	2%	33%
Eastern Europe	43015	27493	70508	4%	39%
Germany	23525	20182	43707	3%	46%
Ireland	3614	6328	9942	1%	64%
Italy	3822	6337	10159	1%	62%
Other	116132	5742	121874	1%	5%
Russia	1899	8737	10636	1%	82%
Scandinavia	162097	26199	188296	4%	14%
The UK	48813	13034	61847	2%	21%
Western Europe	114860	10031	124891	1%	8%

Table A2 Distribution of occupational class and urban status by country of origin for males in the US age 30 to 50 born between 1850 and 1890 from pooled US censuses 1850 to 1950

	Professionals	Farmers (owners and managers)	Managers, officials and proprietors	Clerical and Kindred	Sales workers	Craftsmen	Operatives	Service Workers	Farm laborers	Laborers	Not in emp	Total	% urban
USA	5	26.43	9.83	4.24	5.19	16.78	11.14	3.73	5.53	11.14	0.98	100	46.88
Canada	5.03	10.07	10.13	3.98	4.91	27.43	17.38	3.99	2.58	13.23	1.28	100	66.49
Scandinavia	2.43	20.98	6.39	1.65	1.97	27.88	14.54	2.91	4.86	15.57	0.82	100	53.58
Great Britain	5.19	6.57	9.73	4.82	4.92	30.15	23.2	4.26	2.12	7.99	1.04	100	71.36
Ireland	2.06	3.26	5.73	3.52	3.21	20.47	21.37	8.03	2.35	28.32	1.68	100	81.68
Italy	1.54	2.39	9.62	0.75	2.71	17.32	20.6	7.77	2.26	34.55	0.5	100	81.46
Germany	2.98	17.31	10.29	2.1	3.86	26.02	15.69	5.13	3.03	12.16	1.43	100	66.46
Russia	3.64	5.5	23.68	1.15	7.63	24.28	19.84	2.57	1.08	10.14	0.49	100	87.16
Eastern Europe	1.44	5.61	8	0.9	2.35	20.63	28.44	3.13	1.65	27.36	0.51	100	78.61
Western Europe	3.35	10.95	10.13	1.34	3.33	19.32	23.23	5.37	3.65	18.5	0.85	100	66.8
Others	2.24	6.47	9.71	1.25	3.67	8.32	18.46	10.13	12.12	26.98	0.64	100	63.46
Total	4.44	22.06	9.89	3.67	4.81	18.06	13.5	4.12	5.04	13.45	0.95	100	53.13

Table A3 Baseline origins, destination, and siblings fixed-effects models by gender

	Male destination	Male origin	Male siblings	Female destination	Female orign	Female siblings
Migrant	-0.197***	1.361***	1.462**	-0.097+	1.562***	2.162**
	(0.048)	(0.049)	(0.460)	(0.058)	(0.057)	(0.668)
Birth year FE	X	X	X	X	X	X
Siblings FE			X			X
Constant	72.508***	70.734***		73.568***	72.340***	
	(0.141)	(0.135)		(0.168)	(0.154)	
Num.Obs.	372552	370605	76888	307366	312736	61045
R2	0.003	0.006	0.823	0.020	0.014	0.846
R2 Adj.	0.003	0.006	0.211	0.020	0.014	0.221
R2 Within			0.001			0.001
R2 Within Adj.			0.001			0.001

*p < .05; ** p < .01; *** p < .001; For fixed-effects models standard errors are clustered at the parent level

Table A4 Destination models by country of origin and gender

	Males	Females
Reference: US-born		
Canada	-0.171	-0.211
	(0.141)	(0.162)
Eastern Europe	-0.712***	-0.524***
	(0.098)	(0.121)
Germany	0.082	0.366**
	(0.114)	(0.138)
Ireland	-2.661***	-1.746***
	(0.215)	(0.225)
Italy	0.682***	-0.616*
	(0.194)	(0.256)
Russia	-0.725***	-1.255***
	(0.171)	(0.208)
Scandinavia	0.718***	0.734***
	(0.100)	(0.122)
The UK	-0.839***	0.137
	(0.145)	(0.166)
Western Europe	0.242	-0.114
	(0.160)	(0.194)
Birth year FE	x	x
(Intercept)	72.512***	73.531***
Num.Obs.	369199	304977
R2	0.004	0.020
R2 Adj.	0.004	0.020

*p < .05; ** p < .01; *** p < .001

Table A5 Models by origin country for males

	Scandinavia	Western Europe	The UK	Russia	Canada	Italy	Germany	Eastern Europe	Ireland
migrant	0.666***	2.093***	2.059***	9.110***	-0.409*	0.696*	3.900***	4.776***	-1.731***
	(0.107)	(0.164)	(0.161)	(0.410)	(0.170)	(0.345)	(0.153)	(0.126)	(0.355)
Birth year FE	X	X	X	X	X	X	X	X	X
(Intercept)	72.183***	71.169***	69.076***	63.020***	73.097***	68.959***	68.134***	67.745***	70.125***
	(0.267)	(0.294)	(0.389)	(1.352)	(0.528)	(1.350)	(0.406)	(0.418)	(0.972)
Num.Obs.	100178	64959	33312	6221	22975	6048	24134	40080	5211
R2	0.008	0.005	0.013	0.098	0.006	0.021	0.030	0.039	0.023
R2 Adj.	0.008	0.004	0.012	0.092	0.004	0.014	0.028	0.038	0.015

*p < .05; ** p < .01; *** p < .001

Table A6 Models by origin country for females

	Scandinavia	Western Europe	The UK	Russia	Canada	Italy	Germany	Eastern Europe	Ireland
migrant	1.077***	3.110***	1.752***	6.579***	0.225	0.524	4.187***	4.191***	-1.036**
	(0.124)	(0.193)	(0.184)	(0.537)	(0.198)	(0.413)	(0.175)	(0.150)	(0.389)
Birth year FE	X	X	X	X	X	X	X	X	X
(Intercept)	73.189***	71.731***	71.621***	66.589***	75.370***	71.329***	70.647***	70.154***	72.159***
	(0.295)	(0.312)	(0.448)	(1.636)	(0.642)	(1.711)	(0.499)	(0.533)	(1.108)
Num.Obs.	88118	59932	28535	4415	18506	4111	19573	30428	4731
R2	0.015	0.012	0.026	0.081	0.022	0.030	0.043	0.034	0.031
R2 Adj.	0.015	0.011	0.025	0.073	0.019	0.021	0.041	0.033	0.022

*p < .05; ** p < .01; *** p < .001

Table A7 Siblings fixed effects models for males

	Western Europe	Scandinavia	The UK	Canada	Eastern Europe	Germany	Ireland	Italy	Russia
migrant	2.132	-0.729	4.769***	-0.372	8.849***	4.606**	-2.472	7.921+	3.200
	(1.930)	(0.716)	(1.360)	(0.997)	(1.587)	(1.515)	(3.787)	(4.602)	(4.849)
Siblings FE	X	X	X	X	X	X	X	X	X
Num.Obs.	12328	20882	8101	5985	8221	4504	634	727	999
R2	0.824	0.827	0.851	0.774	0.889	0.838	0.872	0.871	0.892
R2 Adj.	0.209	0.203	0.238	0.093	0.274	0.194	0.116	0.234	0.165
R2 Within	0.001	0.000	0.007	0.000	0.029	0.011	0.004	0.042	0.001
R2 Within Adj.	0.000	0.000	0.006	-0.001	0.028	0.010	-0.006	0.034	-0.006

*p < .05; ** p < .01; *** p < .001

Table A8 Siblings fixed effects models for females

	Western Europe	Scandinavia	The UK	Canada	Eastern Europe	Germany	Ireland	Italy	Russia
migrant	0.635	0.077	-0.682	1.939	10.281***	5.118+	-2.980	6.750	7.000
	(2.838)	(1.099)	(2.100)	(1.258)	(2.304)	(2.646)	(4.865)	(4.501)	(6.246)
Siblings FE	X	X	X	X	X	X	X	X	X
Num.Obs.	11175	17174	6204	4471	6090	3206	547	482	686
R2	0.845	0.842	0.878	0.839	0.907	0.884	0.895	0.915	0.902
R2 Adj.	0.202	0.203	0.247	0.240	0.315	0.280	0.246	0.398	0.069
R2 Within	0.000	0.000	0.000	0.003	0.034	0.010	0.008	0.013	0.006
R2 Within Adj.	0.000	0.000	-0.001	0.001	0.033	0.008	-0.005	-0.001	-0.007

*p < .05; ** p < .01; *** p < .001