

MAX PLANCK INSTITUTE
FOR DEMOGRAPHIC RESEARCH

Konrad-Zuse-Strasse 1 · D-18057 Rostock · Germany · Tel +49 (0) 3 81 20 81 - 0 · Fax +49 (0) 3 81 20 81 - 202 · www.demogr.mpg.de

MPIDR Working Paper WP 2023-050 I December 2023
Revised December 2025
https://doi.org/10.4054/MPIDR-WP-2023-050

Leaving for Life: Using Online Crowd-Sourced Genealogies to Estimate the Migrant Mortality Advantage for the United Kingdom and Ireland Between the 18th and 20th Centuries

Elena Maria Pojman | pojman@demogr.mpg.de
Duke Mwedzi
Orlando Olaya Bucaro
Stephanie Zhang
Michael Y. C. Chong
Monica Alexander
Diego Alburez-Gutierrez

© Copyright is held by the authors.

Working papers of the Max Planck Institute for Demographic Research receive only limited review. Views or opinions expressed in working papers are attributable to the authors and do not necessarily reflect those of the Institute.

Leaving for Life: Using Online Crowd-Sourced Genealogies to Estimate the Migrant Mortality Advantage for the United Kingdom and Ireland Between the 18th and 20th Centuries

Elena Maria Pojman,[1,7] Duke Mwedzi,[2] Orlando Olaya Bucaro,[3] Stephanie Zhang,[4] Michael Y. C Chong,[5] Monica Alexander,[6] Diego Alburez-Gutierrez[7]

[1] Department of Sociology and Criminology, The Pennsylvania State University, University Park, PA, USA

[2] Department of Global Development, Cornell University, Ithaca, NY, USA

[3] International Institute for Applied Systems Analysis, Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/OEAW, University of Vienna), Vienna, AT

[4] Department of Sociology, University of California, Los Angeles, CA, USA

[5] Department of Sociology, Nuffield College, and Leverhulme Centre for Demographic Science, University of Oxford, Oxford, United Kingdom

[6] Departments of Statistical Sciences and Sociology, University of Toronto, Toronto, Canada

[7] Kinship Inequalities Research Group, Max Planck Institute for Demographic Research, Rostock, DE

E.M.P., pojman@demogr.mpg.de, https://orcid.org/0000-0001-8982-6324

D.M., dem327@cornell.edu, https://orcid.org/0000-0001-7394-2427

O.O.B., olayabucaro@iiasa.ac.at, https://orcid.org/0000-0002-9847-1374

S.Z., szhang19@g.ucla.edu, https://orcid.org/0000-0001-9804-6522

M.Y.C., michael.chong@sociology.ox.ac.uk, https://orcid.org/0000-0003-1158-3822

M.A., monica.alexander@utoronto.ca, https://orcid.org/0000-0002-8135-3435

D.A.G., alburezgutierrez@demogr.mpg.de, https://orcid.org/0000-0002-9823-5179

E.M.P., D.M., O.O.B., and S.Z. contributed equally to this work and should all be considered first authors.

Corresponding author: Elena Maria Pojman, pojman@demogr.mpg.de

Leaving for Life: Using Online Crowd-Sourced Genealogies to Estimate the Migrant Mortality Advantage for the United Kingdom and Ireland Between the 18th and 20th Centuries

Abstract

Demographic studies consistently find a mortality advantage among migrants, but a lack of longitudinal data tracking individuals across national borders has limited the study of historical international migration. To address this gap, we use the crowd-sourced online genealogical database Familinx to estimate the migrant mortality advantage for migrants from the United Kingdom and Ireland born between 1735-1894. We compare age at death for non-migrants and migrants to Canada, the United States, South Africa, New Zealand, and Australia using mixed-effects regression models that account for unobserved factors shared between siblings. Results suggest an overall expected migrant advantage of 5.9 years 95% CI [5.7, 6.2] even after accounting for between-family variation, with migrants estimated to live an additional 2.4 [0.8, 4.0] to 7.9 [6.1, 9.6] years depending on the country of destination. This study contributes to the understanding of the migrant mortality advantage in a historical context and shows the potential for online genealogies to contribute to demographic research.

Keywords: *crowd-sourced genealogies, migrant mortality advantage, United Kingdom, Ireland, sibling effects*

Acknowledgements and funding: We are grateful for helpful feedback from attendees at the 2023 Southern Demographic Association Annual Meeting. This project was funded by the 2023 Summer Incubator Program of the Max Planck Institute for Demographic Research. We thank all participants and MPIDR researchers for their suggestions.

INTRODUCTION

Across a range of contexts, migrants have been found to live longer than their non-migrant counterparts in both origin and destination countries (Razum, 2008; Wallace & Wilson, 2019). Scholars debate whether this migrant mortality advantage is driven by selection (the healthy migrant hypothesis, Guillot et al., 2023; Mourits & Puschmann, 2023; Wallace & Wilson, 2019) or contextual factors specific to the destination country (Galt, 2025; Hatton, 2021). Existing research observing a migrant mortality advantage has largely been in contemporary migration contexts (Abraído-Lanza et al., 1999; Palloni & Arias, 2004; Razum et al., 1998), but findings for a migrant mortality advantage in historical migrant flows are mixed and less studied (Bakhtiari, 2022; Mourits & Puschmann, 2023; Puschmann et al., 2016, 2017).

Much of the research on the migrant mortality advantage relies on data and comparisons with those from the destination country, prohibiting comparison of migrants to their origin country peers who did not migrate. The mortality outcomes of migrants are likely similar to those of their family members, particularly siblings; however, data limitations often prevent researchers from accounting for family ties. Accounting for sibling ties could thus be a useful contribution to our understanding of the migrant mortality advantage in historical populations, and particularly the mechanisms driving this advantage.

We study the migrant mortality advantage during a period of mass emigration out of the United Kingdom and Ireland, from the mid-18th century until the early 20th century to Canada, the United States, South Africa, New Zealand, and Australia. As British colonies, these countries facilitated virtually free emigration, mitigating the effects of selection bias in more recent migration flows due to legal restrictions on migration. The variation in these migrant destinations also allows us to understand how the receiving country context may contribute to differences in the migrant mortality advantage.

The current study is guided by the following questions: (1) What is the size of the migrant mortality advantage for migrants from the United Kingdom and Ireland between the 18th and 20th centuries? (2) How does this mortality advantage vary across the migration destination? To answer these questions, we use the online crowd-sourced genealogical dataset Familinx (Kaplanis et al., 2018) and mixed effect regression models where we account for between-family variation. Our results suggest a migrant advantage of 5.9 years, 95% CI [5.7, 6.2], though this advantage is heterogeneous by destination, ranging from 2.4 [0.8, 4.0] years in South Africa to 7.9 [6.1, 9.6] years in New Zealand.

Though historical in coverage, these findings offer a novel contribution to the migrant mortality literature by providing evidence for another context in which migrants outlived their non-migrant peers. First, by leveraging the genealogical structure of the data, we improve upon prior work by using sibling random effects to capture between-family variation, thereby better accounting for migrant selection and isolating family characteristics from the effect of migration. Second, we compare migrants to their non-migrant compatriots for a more direct comparison, rather than comparing migrants to non-migrants in the destination country. Finally, our study covers five countries to assess whether the advantage holds across different contexts, thereby illuminating contextual mechanisms driving mortality disparities. This study highlights the utility of crowd-sourced genealogical data sources in demographic research.

BACKGROUND

Mechanisms Driving the Migrant Mortality Advantage

The migrant mortality advantage describes how migrants often have better mortality outcomes compared to their native-born counterparts (Abraído-Lanza et al., 1999; Feliciano, 2020; Razum, 2008). In some contexts, this finding is paradoxical, as migrants tend to have lower

socioeconomic status and to be from less developed nations with poorer health relative to that of the destination country (Abraído-Lanza et al., 1999; Bakhtiari, 2022; Wallace & Wilson, 2019). Two common explanations for the migrant mortality advantage include migrant selectivity (the healthy migrant hypothesis) and the migrant destination context.

The healthy migrant hypothesis argues that migrants are positively selected relative to non-migrants in both the origin and destination countries (Guillot et al., 2023; Mehta et al., 2016; Razum, 2008). Critics of this hypothesis point to measurement issues and data limitations, such as migrants being undercounted in population statistics, leading to numerator-denominator bias and rendering them "statistically immortal" if their deaths are not recorded in the destination country (Pablos-Méndez, 1994; Puschmann et al., 2017). Additionally, population statistics in either the origin or destination country may suffer from migration censorship, that is, an overrepresentation of individuals who die before they can migrate, artificially decreasing the average age of death for non-migrants relative to migrants (Kasakoff & Adams, 1995; Ruggles, 1992). Despite these challenges, studies still find a mortality advantage after accounting for many of these potential biases (Abraído-Lanza et al., 1999; Puschmann et al., 2017).

The healthy migrant literature originally focused on the paradoxical mortality advantages found among Hispanic migrants living in the United States (e.g., Abraído-Lanza et al., 1999; Pablos-Méndez, 1994; Palloni & Arias, 2004), but over time has come to describe a range of the experiences of migrant groups in various destination countries, ranging from more recent migration flows in Europe (Andersson & Drefahl, 2017; Guillot et al., 2023; Helgesson et al., 2019; Mehta et al., 2016) to more limited evidence in historical flows (Mourits & Puschmann, 2023; Puschmann et al., 2016, 2017). Although impressive in terms of research design and

findings, these latter studies are more constrained to primarily domestic migration flows in the Netherlands and other Northwestern European cities.

Some scholars argue that migrant selectivity occurs between rather than within families (Mourits & Puschmann, 2023). In other words, migrants themselves are healthier, rather than coming from healthier families. If true, this may reflect a risk diversification strategy on the part of families, who send the healthiest and therefore most likely to succeed (Stark & Bloom, 1985). Other scholars note that family members have different abilities, economic opportunities, or familial responsibilities which would make them more or less likely to migrate (Abramitzky et al., 2012; Mourits & Puschmann, 2023). For example, individuals who do not expect to inherit their family's land, by virtue of gender or birth order, could have an increased propensity to migrate (Abramitzky et al., 2013; Wegge, 1999).

A smaller parallel literature argues that conditions in the destination country drive differences in migrant outcomes. This literature leverages a comparative approach to avoid problems of selection bias, either comparing between migrant groups that had the same origin country (e.g., Hatton, 2021), or between family members with mixed migration histories (e.g., Galt, 2025). For example, compared to Rotterdam-born individuals who never moved, their peers who left Rotterdam—regardless of their origin—saw lower mortality risks in the nineteenth century (Puschmann et al., 2017). Another study, also focused on nineteenth century migration flows, leveraged direct sibling comparisons between a British brother who did not migrate and his brother who moved to New Zealand and found a direct effect of migration on mortality, with migrant brothers living nearly five years longer on average (Galt, 2025). Similar studies comparing within families confirm this survival advantage (e.g., Mourits & Puschmann, 2023). Yet, this literature is geographically limited in scope and difficult to generalize to other migration flows.

Thus, the mechanisms driving the migrant mortality advantage are unclear, particularly in historical cases.

Evidence for a Historical Migrant Mortality Advantage

Though the migrant mortality advantage has been found in numerous *contemporary* migrant groups, such Mexicans in the United States (Palloni & Arias, 2004) and Turkish in Germany (Razum et al., 1998), there is less evidence in historical migrant flows. Much of this literature is limited to a few Northwestern European cities (Mourits & Puschmann, 2023; Puschmann et al., 2016, 2017), or to country-specific analyses, such as the case of British migrants to New Zealand (Galt, 2025). There is evidence that many European migrants during the period 1850-1914 held economic advantages over native-born Americans (Abramitzky et al., 2014), which could have translated to mortality advantages.

Instead, European migrants to the United States in the early 20th century saw a mortality *dis*advantage compared to the native-born population. Before immigration restrictions were passed in the 1920s, much of this disparity was driven by settlement patterns, as immigrants were likelier to live in cities which had greater exposure to infectious diseases (Bakhtiari, 2022). Lacking evidence on other immigration flows and a wider time scope, however, it is difficult to conclude whether a migrant mortality advantage extends to migrants in the past.

Historical Emigration from the United Kingdom and Ireland

We use historical emigration from the United Kingdom and Ireland to Canada, the United States, South Africa, Australia, and New Zealand as a case study of the migrant mortality advantage in a historical context. The sheer size of this flow, combined with the lack of legal restrictions on immigration to these countries, makes it an ideal case to study mortality differences

between migrants and non-migrants (Hatton, 2021). The United Kingdom's colonial relationship with its overseas territories facilitated virtually free international movement for European migrants on a scale that is no longer possible due to legal and immigration restrictions. As such, there was less selection of migrants on characteristics such as socioeconomic status or education level (Hatton, 2004, 2021). The varied destinations of the flow also make it possible to test whether the migrant mortality advantage holds across different contexts.

The Demographic Transition together with the Industrial Revolution set the stage for mass emigration from Europe starting in the 18th century (Hatton & Williamson, 1994; Richards, 2018). As the United Kingdom and Ireland moved through the second and third stages of the Demographic Transition, population growth soared (Friedlander & Okun, 2022), exceeding agricultural labor demands, despite increasing agricultural productivity (Richards, 2018). Consequently, economic circumstances worsened for many, with higher rent costs and declining access to land (Hatton & Williamson, 1994; Horn, 1998; Richards, 2018). Coupled with rapid urbanization during the Industrial Revolution from 1750–1850, urban areas were subject to overcrowding and poor health, resulting in decreased lifespan, particularly for low socioeconomic status individuals (Taylor, 1988). These poor conditions prompted the out-migration of an estimated 19 million people in total from the United Kingdom and Ireland between the 18th and 19th centuries (Richards, 2018) to colonies in North America, Australasia, and Africa, with the vast majority settling in present-day United States, Canada, Australia, and New Zealand (Hatton, 2021). In sum, emigration served as a "safety valve" relieving the pressure caused by overcrowding and few economic opportunities (Blanc & Wacziarg, 2025; Hatton & Williamson, 1994).

The emigrants themselves often were those most poised to benefit from migration: typically young, single, unskilled men or young couples with small children (Hatton &

Williamson, 1994; Horn, 1998; Thompson, 2009; Tomlins, 2001). However, it is also important to note that many migrants moved involuntarily, such as convicts and indentured servants (Richards, 2018; Tomlins, 2001), although migration conditions changed over time. In the case of Australia, for example, what was originally the forced migration of convicts later became a flow of migrants seeking land and wealth in the Gold Rush (Murdoch, 2004).

Migrants' destinations varied depending on their background and reason for moving. During the 19th century, British migrants to Australia and New Zealand tended to be more skilled on average than those moving to Canada and the United States (Haines, 1997; Hatton, 2021; Murdoch, 2004). This may have been due to cost differences, with fare from London to New York costing a third of the fare to Australia. As such, migrants from professional and commercial backgrounds were more likely to migrate to Australia and New Zealand, while agricultural and manual workers tended to migrate to North America (Hatton, 2021). There is also evidence that British migrants to South Africa tended to be from poorer backgrounds and were motivated to move by the desire for greater economic security and independence (Tosh, 2017). Importantly, relative to the native-born, migrants from the United Kingdom and Ireland to the United States tended to earn more, even soon after arrival (Abramitzky et al., 2014).

Mortality disparities may also be driven by characteristics of the migration journey itself, particularly distance between the origin and destination countries. Individuals who survived a long arduous journey may have been predisposed to better mortality outcomes, whereas those vulnerable to disease did not survive the journey. Indeed, prior work finds that migrant survival advantages are stronger the further they move (Puschmann et al., 2016, 2017). It may be that mortality outcomes vary because the journey from the United Kingdom and Ireland to Australia and New Zealand was much longer than the journey to the United States, Canada, or South Africa.

Additionally, infectious diseases may have also contributed to differences in mortality rates among migrants. Although infectious diseases such as tuberculosis, measles, and smallpox were often brought by European settlers, there were some infectious diseases in destination countries such as yellow fever and malaria to which they would not have had resistance (Haines & Shlomowitz, 1998). Later groups of migrants, such as Europeans who moved to the United States in the early 20th century, were more vulnerable to infectious disease mortality because they were likelier to reside in densely populated urban areas (Bakhtiari, 2022). In short, the migration destination could have influenced disease susceptibility, resulting in disparate health outcomes by destination.

Another possible driver of mortality differences among migrants is variations in exposure to violent conflict between European migrants and indigenous populations in colonies, specifically regarding variations in the scale, type, and intensity of conflict. Contexts that saw more prolonged and large-scale violent conflict would have experienced more casualties for British and Irish migrants and thus, increased mortality. South Africa is particularly notable for its conflicts in the 18th and 19th centuries due to large-scale, centrally coordinated military resistance by indigenous groups over more than a century, causing heavy casualties to the British settlers (Beckett, 2019; Evans et al., 2018; Webb, 2015). Similar centrally organized resistance took place in New Zealand, but the scale of losses for colonial forces was smaller (Carpenter, 2015; Draper et al., 2023). There were multiple protracted wars between the 17th and 19th centuries between European settlers and various indigenous groups in the United States (Draper et al., 2023; Nester, 2004).

By contrast, where indigenous resistance was less intense or less confrontational, British and Irish migrants had less exposure to violence and thus, lower risk of mortality. In the Pacific, indigenous groups favored guerilla tactics and economic sabotage over open conflict (Duckett

White, 2023; Kerkhove, 2020). Where open conflict did take place, it was small in scale and resulted in few casualties for migrants (Carpenter, 2015). In Canada, conflict was also small in scale and localized, as settlers made greater use of diplomacy and treaties (Dixon, 2014).

DATA AND METHODS

Data and Sample

In this study, we ask: (1) What is the size of the migrant mortality advantage for migrants from the United Kingdom and Ireland between the 18th and 20th centuries? (2) How does this mortality advantage vary across the migration destination? Data come from Familinx, a genealogical dataset that was scraped from Geni.com with information on the timing and location of vital events such as birth and death for over 86 million individuals (Kaplanis et al., 2018). Familinx data represents the work of amateur genealogists who upload their family trees which were then matched with historical records (e.g., burials and baptisms) to create more complete genealogies spanning centuries and world regions. Genealogies were organized into directed acyclic graphs, and an automated pipeline was used to identify and correct invalid biological structures (such as cycles or individuals with more than two parents). Conflicts between overlapping profiles were algorithmically resolved through local merging and pruning steps, yielding biologically valid and deduplicated pedigrees that showed more than 90% concordance with expert genealogists' manual curation (Kaplanis et al., 2018). Individuals born in the United Kingdom and Ireland are well-represented in Familinx, and the data's genealogical structure is better suited for studying migration than similar data which suffer from migration censorship (e.g., administrative data or family reconstructions using parish records; Ruggles, 1992). This genealogical structure also allows us to identify transnational ties between kin and thereby account for unobserved similarities between siblings.

Because the data are user-generated, there were several instances of reporting errors and other issues which necessitated extensive data cleaning and treatment before analysis. For example, some profiles in the dataset had missing information for country of birth or death, some records had implausible values for year of birth and death leading to an impossible age at death, and the names of some birth and death locations varied by language and changed over time. In general, we use birth and death information from baptism and burial records when exact birth and death dates are missing, and discard profiles with incomplete information on year and location of birth and death. We describe this further in section A of the Online Supplement.

The initial sample consisted of profiles who were born in the United Kingdom or Ireland and died in either the United Kingdom, Ireland, Canada, the United States, South Africa, Australia, or New Zealand between the years 1735-1894. They were not missing information on either the year or location of their birth or death, and their birth was recorded as before their death (N = 147,651). Next, we drop profiles missing gender or information on both parents, and those who died younger than age 15 or older than 110. After these steps, our final sample consisted of n = 61,794 non-migrants, that is, individuals who died in either the United Kingdom or Ireland and n = 35,874 migrants, that is, individuals who died abroad (n = 5,042 in Canada, n = 18,332 in the United States, n = 1,181 in South Africa, n = 8,585 in Australia, and n = 2,734 in New Zealand), for a total of N = 97,668 observations. Due to data limitations, we cannot identify return migrants and therefore consider them to be non-migrants; given the period and destinations we analyze, we estimate there are few such cases. We describe in detail our data preparation procedure, including our treatment of missingness, in section A of the Online Supplement. Replication materials for the entire data cleaning process and analysis are available at https://osf.io/b87t6/?view_only=73ca840636424c9c9f78687103c3d90f.

Each individual has a unique identification number and, where applicable, the identification number of at least one parent. From this information we can derive sibling ties: individuals who share at least one parent are considered to be siblings. Sibling ties are determined first through the maternal line and if that is not available, the paternal line (see Table S2). Lacking full data on every individual's parents, it is difficult to determine if individuals are full or half siblings. As a result, half-siblings are treated as full siblings, and we are unable to identify stepsiblings. Further details on the sibling tie matching are available in section A of the Online Supplement. These sibling ties, recorded as a family identification number, are included in our models as a random effect.

Genealogical data have been used to investigate a variety of demographic outcomes (e.g., Chong et al., 2022; Corti et al., 2024; Cozzani et al., 2023; Gavrilov et al., 2002; Gay et al., 2023; Minardi et al., 2024; Piraino et al., 2014), and represent an opportunity for demographers to answer long-discussed questions about the nature of inter- and intragenerational demographic processes (Alburez-Gutierrez et al., 2019). Despite the biases inherent in crowdsourced genealogies (Calderón Bernal et al., 2025; Colasurdo & Omenti, 2024; Stelter & Alburez-Gutierrez, 2022), we believe that our analytical choices represent best practices in dealing with genealogical data.

Measures

We measure the outcome of age at death by subtracting one's birth year from their death year. While the data does not allow us to measure migration directly, individuals whose birth is recorded in the United Kingdom or Ireland and death is recorded in Australia, New Zealand, South Africa, the United States, or Canada are considered migrants for the purposes of this study, as we infer that they would have migrated to their eventual death location. Individuals whose birth and death are recorded in the United Kingdom or Ireland are considered non-migrants.

To account for potential confounding between age at death and migration status, we control for gender and birth cohort, the latter measured as a categorical variable in 10-year intervals (e.g., 1735, 1745). We also include controls for the number of siblings, coded as a categorical variable (0, 1, 2, 3-5, 6+), because siblings compete for scarce resources amongst themselves (Donrovich et al., 2014; Lam & Marteleto, 2008) which could diminish the magnitude of shared mortality advantages. Table 1 shows descriptive statistics of selected variables, broken down by the overall sample and by the country of death. On average, individuals in the sample died in their mid-60s, were more likely to be male, and were roughly one-third migrants. About half of the profiles had zero recorded siblings, though the other half were split between having one, two, three to five, or more than six siblings.

[INSERT TABLE 1 ABOUT HERE]

Analytic Strategy

Among the sample of individuals born in the United Kingdom and Ireland between 1735-1894, we examine the migrant mortality advantage by comparing those who stayed in the United Kingdom and Ireland with those who migrated to other countries. We use mixed effects models with age at death as the outcome variable to estimate the effect of migration, while accounting for gender, birth cohort, and heterogeneity between families. We present results from two sets of models. In the first set (Model A1; migrant effects), we examine general mortality differences between migrants and non-migrants, incorporating fixed effects for migrant status, gender, birth cohorts, and interaction terms between migrant status and each birth cohort. In the second set (Model B1; destination country effects), we estimate migrant mortality differences for each of the destination countries separately, incorporating a fixed effect for each of the destination countries, gender, birth cohort, and interaction terms between each of the destination countries and each birth

cohort. The birth cohort and birth cohort interaction effects account for changing mortality conditions over time, and that the change may not be the same between countries.

In all models, we also include sibling-level random effects to account for heterogeneity between families (see section A of the Online Supplement for more information). Each set of siblings receives one level of the random effect, thereby capturing mortality variation between families that may be due to unobserved factors shared by siblings. This allows for the comparison of migrant and non-migrant siblings, in the sense that the variation between siblings is left to be explained by the migrant, birth cohort, and gender fixed effects. This approach eliminates the need to construct specific comparisons between sibling pairs, and the fixed effects can be interpreted as resulting from a comparison between all migrants and non-migrants in the population accounting for the possibility that some families may experience higher or lower mortality on average. Additionally, sets of siblings where all are migrants or all are non-migrants (including individuals without siblings) can be used in the estimation. Information from these profiles still contribute to the estimation of the fixed effects and the variance in the outcome, while allowing some variance to be explained by the sibling random effect.

RESULTS

Migration Flows in the 18th and 19th Centuries

Figure 1 shows the migration flows identified in the sample. The United States is the largest destination country for migrants across all cohorts followed by Australia. South Africa has the fewest migrants, with migration beginning in earnest for the cohorts born between 1765-1775. Across all destinations, flows are relatively low for earlier cohorts then begin to surge, particularly in the cohorts born between 1775-1855, before declining among later birth cohorts. A large majority of migrants across destinations are male, though the gender composition shifts between

cohorts. Female migrants make up a larger share of the flow between the 1770s and 1850s, but when the overall migration flows are smaller, the share of women is smaller.

[INSERT FIGURE 1 ABOUT HERE]

Migrant Mortality Advantage

We estimate the relationship between migration and age at death using mixed effect regression models. Results are presented as average marginal effects (AMEs), allowing us to interpret our findings as the average effect of being a migrant (versus a non-migrant) on age at death in our sample, while holding controls at their means (Mize, 2019). Positive values indicate that migrants have a mortality advantage whereas negative values indicate a mortality disadvantage.

When comparing migrants to non-migrants overall in Model A1, we find an AME of 5.96 years, 95% CI [5.7, 6.2] for individuals who migrate, suggesting a nearly six-year mortality advantage of migrants over non-migrants (see Table S3 of the Online Supplement for the full regression estimates). We present differences in the migrant mortality advantage by birth cohort in Figure 2, with the overall AME overlaid in green. While there is fluctuation in the magnitude of the AME, it remains positive and statistically significant for every cohort in our sample, indicating that migrants had a mortality advantage over their non-migrant peers across all birth cohorts. For migrants in cohorts born between 1815-1865, this advantage is particularly large, at over six extra years lived compared to non-migrants. Further, the sibling random effect shows a variance of 40.17 (see Table S3), indicating strong clustering of longevity among siblings.

[INSERT FIGURE 2 ABOUT HERE]

Heterogeneity by Migrant Destination

Next, we examine heterogeneity by migrant destination (Model B1; see Figure 3). Here, the AMEs reflect the country-specific mortality advantage over non-migrants. To highlight the contrasts between the overall migrant advantage and country-specific advantages, we include the overall AME and 95% confidence intervals from Model A1 in green. When disaggregating by destination country in Model B1, we find that the AME ranges from 2.4 years, 95% CI [0.8, 4.0] in South Africa to 7.9 years, 95% CI [6.1, 9.6] in New Zealand. The AMEs for all destinations are positive and highly statistically significant ($p < 0.001$), suggesting a clear mortality advantage for migrants. The regression estimates (see Table S3) also suggest that the mortality advantage is particularly strong for migrants to the U.S. and New Zealand who were born after 1815, with similar, though relatively weaker, evidence for migrants to Canada. Here, the sibling random effect shows a variance of 39.90, again showing that longevity is strongly clustered within sibling groups.

Comparing models A1 and B1 shows that the overall migrant mortality advantage in this context is similar to the mortality advantage among migrants to Australia and the United States, while migrants to Canada and New Zealand have an even greater advantage, and migrants to South Africa have less of an advantage than the average migrant.

[INSERT FIGURE 3 ABOUT HERE]

Benchmarking and Supplementary Models

In the absence of administrative or demographic data against which we could benchmark our estimates, we compare our findings to those of Galt (2025) who used the self-constructed RA*/Ryan dataset based on a sample of men with surnames beginning with either “Ra” or “Ryan” in New Zealand before 1913 (Galt, 2023). These migrant men were linked to their non-migrant

brothers living in the United Kingdom, among other countries. Galt finds that migrants to New Zealand lived on average 4.6 years longer than their non-migrant brothers, with the largest advantages for those born between 1830 and 1850. This estimate is smaller than ours by 4.1 years, which likely reflects the direct comparison employed in her analysis—we compare migrants to all non-migrants, while accounting for similarities within sibling groups.

We also estimate several supplementary models to be sure of the robustness of our results; these models are further described in section B of the Online Supplement. These models reveal that our findings are robust to alternate sample selections (i.e., only profiles with siblings, survival to different ages) and modeling strategies.

DISCUSSION AND CONCLUSION

Using the case of mass emigration from the United Kingdom and Ireland to the United States, Australia, New Zealand, Canada and South Africa during the 18th and 19th centuries, this study contributes to our understanding of the migrant mortality advantage using the crowd-sourced genealogical dataset Familinx. In mixed effect regression models, we compare migrants to their origin country compatriots who did not migrate. We find that migrants live 5.9 years, 95% CI [5.7, 6.2] longer on average than non-migrants, and this advantage is heterogeneous by destination country, ranging from 2.4 years [0.8, 4.0] in South Africa to 7.9 years [6.1, 9.6] in New Zealand. Our findings provide suggestive evidence of a migrant mortality advantage in this historical context and highlight the possibility for online crowdsourced genealogical data to contribute to demographic research.

This study offers several contributions to the literature on the migrant mortality advantage, particularly in historical migration flows (Galt, 2025; Mourits & Puschmann, 2023; Puschmann et

al., 2016, 2017). Our data encompass roughly two centuries worth of individuals born in the present-day United Kingdom and Ireland, over one-third of whom migrated abroad. Further, the five destination countries provide the opportunity to understand heterogeneity in the migrant mortality advantage across different contexts. We leverage the genealogical structure of the data and the multiple migrant destinations to understand two potential mechanisms driving this mortality advantage: the healthy migrant hypothesis and the migration destination context.

Our findings suggest that migrants lived longer than non-migrants, providing support for the healthy migrant hypothesis. We also find evidence of strong clustering of mortality behavior within sibling groups, contradicting Mourits and Puschmann (2023) who argued that migrants are healthier rather than coming from healthier families. Further, we find that the mortality advantage is particularly concentrated among migrants born in the mid-nineteenth century.

Lacking information on the age at which individuals migrated, their occupation, socioeconomic status, or other factors that could shed greater light on the conditions they met after migration, we see our work as offering limited evidence to the latter debate regarding the role of migration destination. However, the use of sibling random effects is our best effort to account for unobserved family characteristics that could shape migration and mortality outcomes. That the United States and South Africa—the two countries that saw the most violent conflict with natives and settlers—had relatively less of a mortality advantage compared to migrants to Australia, Canada, or New Zealand could suggest that conditions in the destination country are a key component driving mortality differences. We are unable to fully disentangle this mechanism but see it as an important direction for future research.

Alternatively, the lifespan discrepancy by migrant status may be a result of poor conditions in the United Kingdom and Ireland (e.g., exposure to diseases such as smallpox, persistent poverty,

overcrowding in cities) that migrants were able to avoid. This framing runs contrary to typical explanations of the healthy migrant hypothesis because it posits that mortality advantages are due to poorer conditions experienced by non-migrants which depress their mortality. In reality, the observed variations in the migrant mortality advantage are likely a combination of all the above mechanisms: selection into migration, differing mortality conditions by destination, and avoiding deleterious conditions in the origin country. Disentangling these mechanisms is not possible in this study but may prove a promising avenue for future work.

There are some limitations to our study. Crowdsourced genealogical data have several biases with respect to the completeness of family trees (Calderón Bernal et al., 2025): access to information about relatives is unequal, meaning that records may exclude childless people, particularly women, and infant and child mortality may be underreported. While we partially address the latter bias by conditioning our analyses on those who survived to age 15, we identify a gender bias in recording such that men are overrepresented. Minardi et al. (2023) notes, however, that both the median and mean age at death for men and women were similar during much of our study period, and the female advantage in lifespan likely did not occur until after the 1850s, suggesting minimal impact of on our findings.

Moreover, though algorithmically cleaned and deduplicated by Kaplanis et al. (2018), these data are subject to issues of data completeness and accuracy. Missing or inaccurate information on vital events, especially for death, can affect lifespan measurement and the reconstruction of kinship and migration histories. Such issues are not random: data coverage varies across time and geography, with earlier historical periods or less digitally active regions especially underrepresented (Koylu et al., 2021; Koylu & Kasakoff, 2025). As recent comparative studies have shown, online family trees disproportionately represent individuals of European descent,

older age groups, and socioeconomically advantaged populations, while underrepresenting Indigenous, Black, and migrant populations in the U.S. (Koylu et al., 2021; Koylu & Kasakoff, 2025), and likely over-represent higher socioeconomic strata in Europe (Stelter & Alburez-Gutierrez, 2022). These biases should be kept in mind when interpreting the results of our study. Genealogical data offer unprecedented temporal depth but should be interpreted as describing the subset of populations with documented and connected family histories, rather than a statistically representative sample of all individuals (Colasurdo & Omenti, 2024; Otterstrom et al., 2024; Spencer & Otterstrom, 2024). As such, we see our findings as likely representative of individuals whose descendants had the free time, money, or other resources to trace their ancestry; such bias would affect the non-migrants as much as the migrants. Even among this more selective group, we find a significant mortality advantage.

Though Familinx is rich in terms of sample size, it suffers from a lack of detail at the individual level. As a result, we are unable to control for socioeconomic status (i.e., literacy, financial resources, etc.) directly, which may influence not only one's propensity to migrate but also their mortality. Migrant outcomes may be partially determined by one's country of residence during early life (Alexander & Ward, 2018), but we are unable to capture one's "exposure" to different epidemiological or socioeconomic conditions at specific points in their lifetime. Though it is possible to identify some "salmon" migrants who migrated abroad but returned to die in their home country using the birthplace of their children, we are unable to do so for childless individuals, making this a promising avenue for future work. Yet, given the length of the journey from the United Kingdom and Ireland to our destinations of interest and the costs involved, return migrants are likely a small portion of migrants who would not substantially change our results. Future

studies should investigate the impact of exposure to poor conditions in the origin country as a mechanism driving the migrant mortality advantage.

Further, we classify individuals who moved between Ireland and the United Kingdom as “non-migrants" rather than “migrants" because our interest was primarily in migration to Canada, the United States, South Africa, New Zealand, and Australia. As such, we do not investigate the possibility of a migrant mortality advantage between the United Kingdom and Ireland. We also do not capture migration within the United Kingdom (e.g., domestic rural-urban migration) during this period. These limitations are in part due to the ambiguous usage of “United Kingdom” to describe locations within the dataset and the difficulty in extracting more granular location data from free text entries. Future work should certainly investigate this possibility, as migration within the United Kingdom and Ireland occurred at non-trivial levels, often driven by many of the same factors that encouraged migration to the countries in this study (Nicholas & Shergold, 1987).

Though our setting is historical, our findings are relevant for present-day migration flows. On average, we find a clear mortality advantage for migrants leaving the United Kingdom and Ireland but the magnitude of this difference varied by birth cohort and destination, suggesting that the healthy migrant effect also varies by the migration context. Our sample of migrants faced an extremely different set of legal, social, and environmental circumstances compared to those faced by contemporary migrants, but the underlying mechanisms driving mortality differences are likely similar. In sum, our study provides evidence that the migrant mortality advantage is driven by both positive selection and the destination context.

REFERENCES

Abraído-Lanza, A. F., Dohrenwend, B. P., Ng-Mak, D. S., & Turner, J. B. (1999). The Latino mortality paradox: A test of the "salmon bias" and healthy migrant hypotheses. *American Journal of Public Health, 89*(10), 1543–1548. https://doi.org/10.2105/AJPH.89.10.1543

Abramitzky, R., Boustan, L. P., & Eriksson, K. (2012). Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration. *American Economic Review, 102*(5), 1832–1856. https://doi.org/10.1257/aer.102.5.1832

Abramitzky, R., Boustan, L. P., & Eriksson, K. (2013). Have the poor always been less likely to migrate? Evidence from inheritance practices during the age of mass migration. *Journal of Development Economics, 102*, 2–14. https://doi.org/10.1016/j.jdeveco.2012.08.004

Abramitzky, R., Boustan, L. P., & Eriksson, K. (2014). A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration. *Journal of Political Economy, 122*(3), 467–506. https://doi.org/10.1086/675805

Alburez-Gutierrez, D., Aref, S., Gil-Clavel, B. S., Grow, A., Negraia, D. V., & Zagheni, E. (2019). Demography in the digital era: New data sources for population research. In *Smart statistics for smart applications: Book of short papers SIS2019* (pp. 23–30). Pearson. https://doi.org/10.31235/osf.io/24jp7

Alexander, R., & Ward, Z. (2018). Age at Arrival and Assimilation During the Age of Mass Migration. *The Journal of Economic History, 78*(3), 904–937. https://doi.org/10.1017/S0022050718000335

Andersson, G., & Drefahl, S. (2017). Long-Distance Migration and Mortality in Sweden: Testing the Salmon Bias and Healthy Migrant Hypotheses. *Population, Space and Place, 23*(4), e2032. https://doi.org/10.1002/psp.2032

Bakhtiari, E. (2022). The Missing Mortality Advantage for European Immigrants to the United States in the Early Twentieth Century. *Demography, 59*(4), 1517–1539. https://doi.org/10.1215/00703370-10111916

Beckett, I. F. W. (2019). *Rorke's Drift and Isandlwana*. Oxford University Press.

Blanc, G., & Wacziarg, R. (2025). *Malthusian Migrations* (Working Paper No. 33542; NBER Working Paper Series). National Bureau of Economic Research. https://doi.org/10.3386/w33542

Calderón Bernal, L., Alburez-Gutierrez, D., & Zagheni, E. (2025). *Analysing biases in genealogies using demographic microsimulation* (No. WP-2025-034; MPIDR Working Paper). Max Planck Institute for Demographic Research. https://doi.org/10.4054/MPIDR-WP-2023-034

Carpenter, L. (2015). Māori chivalry during a British defeat: 150 years of remembering Gate Pā. In K. Reeves, G. R. Bird, L. James, B. Stichelbaut, & J. Bourgeois, *Battlefield Events: Landscape, commemoration and heritage*. Routledge.

Chong, M., Alburez-Gutierrez, D., Del Fava, E., Alexander, M., & Zagheni, E. (2022). *Identifying and correcting bias in big crowd-sourced online genealogies* (No. WP-2022-005). Max Planck Institute for Demographic Research. https://doi.org/10.4054/MPIDR-WP-2022-005

Colasurdo, A., & Omenti, R. (2024). Using online genealogical data for demographic research: An empirical examination of the FamiLinx database. *Demographic Research, 51,* 1299–1350. https://doi.org/10.4054/DemRes.2024.51.41

Corti, G., Minardi, S., & Barban, N. (2024). Trends in assortative mating in the United States, 1700–1910. Evidence from FamiLinx data. *The History of the Family, 29*(4), 461–481. https://doi.org/10.1080/1081602X.2024.2352539

Cozzani, M., Minardi, S., Corti, G., & Barban, N. (2023). Birth month and adult lifespan: A within-family, cohort, and spatial examination using FamiLinx data in the United States (1700–1899). *Demographic Research, 49,* 201–218. https://doi.org/10.4054/DemRes.2023.49.9

Dixon, D. (2014). *Never Come to Peace Again: Pontiac's Uprising and the Fate of the British Empire in North America*. University of Oklahoma Press.

Donrovich, R., Puschmann, P., & Matthijs, K. (2014). Rivalry, solidarity, and longevity among siblings: A life course approach to the impact of sibship composition and birth order on later life mortality risk, Antwerp (1846-1920). *Demographic Research, 31,* 1167–1198. https://doi.org/10.4054/DemRes.2014.31.38

Draper, M., Kerby, M., & Baguley, M. (2023). War on the frontier. *Historical Encounters: A Journal of Historical Consciousness, Historical Cultures, and History Education*, *10*(2), 1–11. https://doi.org/10.52289/hej10.201

Duckett White, S. C. (2023). (Flint)lock, stock and two smoking barrels: A modern military interpretation of frontier economic warfare. *Historical Encounters: A Journal of Historical Consciousness, Historical Cultures, and History Education*, *10*(2), 52–62. https://doi.org/10.52289/hej10.205

Evans, J., Grimshaw, P., Philips, D., & Swain, S. (2018). *South Africa*. https://www.manchesterhive.com/display/9781526137333/9781526137333.00013.xml

Feliciano, C. (2020). Immigrant Selectivity Effects on Health, Labor Market, and Educational Outcomes. *Annual Review of Sociology*, *46*(1), 315–334. https://doi.org/10.1146/annurev-soc-121919-054639

Friedlander, D., & Okun, B. S. (2022). Demographic Transition and the Industrial Revolution in England: Inverse Rural and Urban Processes. *Journal of Family History*, *47*(4), 401–412. https://doi.org/10.1177/03631990221114065

Galt, M. (2023). Following through: The value of tracing British settlers across time and space. *The Journal of New Zealand Studies*, *NS36*. https://doi.org/10.26686/jnzs.iNS36.8330

Galt, M. (2025). The Impact of Migrant Selection and Destination on the Wellbeing Outcomes of British Migrants to Nineteenth-Century New Zealand. *Journal of Migration History*, *11*(3), 290–320. https://doi.org/10.1163/23519924-11030003

Gavrilov, L. A., Gavrilova, N. S., Olshansky, S. J., & Carnes, B. A. (2002). Genealogical data and the biodemography of human longevity. *Biodemography and Social Biology*, *49*(3–4), 160–173. https://doi.org/10.1080/19485565.2002.9989056

Gay, V., Gobbi, P. E., & Goñi, M. (2023). *Revolutionary Transition: Inheritance Change and Fertility Decline*. CEPR. https://hal.science/hal-04285818

Guillot, M., Khlat, M., Gansey, R., Solignac, M., & Elo, I. (2023). Return Migration Selection and Its Impact on the Migrant Mortality Advantage: New Evidence Using French Pension Data. *Demography*, 10938784. https://doi.org/10.1215/00703370-10938784

Haines, R. (1997). ‘The idle and the drunken won’t do there’: Poverty, the new poor law and nineteenth-century government-assisted emigration to Australia from the United Kingdom∗. *Australian Historical Studies*, *27*(108), 1–21. https://doi.org/10.1080/10314619708596025

Haines, R., & Shlomowitz, R. (1998). Explaining the Modern Mortality Decline: What can we Learn from Sea Voyages? *Social History of Medicine, 11*(1), 15–48. https://doi.org/10.1093/shm/11.1.15

Hatton, T. J. (2004). Emigration from the UK, 1870-1913 and 1950-1998. *European Review of Economic History*, *8*(2), 149–171. https://doi.org/10.1017/S1361491604001121

Hatton, T. J. (2021). Emigration from the United Kingdom to the United States, Canada and Australia/New Zealand, 1870–1913: Quantity and quality. *Australian Economic History Review*, *61*(2), 136–158. https://doi.org/10.1111/aehr.12218

Hatton, T. J., & Williamson, J. G. (1994). What Drove the Mass Migrations from Europe in the Late Nineteenth Century? *Population and Development Review*, *20*(3), 533. https://doi.org/10.2307/2137600

Helgesson, M., Johansson, B., Nordquist, T., Vingård, E., & Svartengren, M. (2019). Healthy migrant effect in the Swedish context: A register-based, longitudinal cohort study. *BMJ Open*, *9*(3), e026972. https://doi.org/10.1136/bmjopen-2018-026972

Horn, J. (1998). British Diaspora: Emigration from Britain, 1680–1815. In P. J. Marshall, A. Low, & Wm. R. Louis (Eds.), *The Oxford History of the British Empire: Volume II: The Eighteenth Century* (p. 0). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198205630.003.0002

Kaplanis, J., Gordon, A., Shor, T., Weissbrod, O., Geiger, D., Wahl, M., Gershovits, M., Markus, B., Sheikh, M., Gymrek, M., Bhatia, G., MacArthur, D. G., Price, A. L., & Erlich, Y. (2018). Quantitative analysis of population-scale family trees with millions of relatives. *Science*, *360*(6385), 171–175. https://doi.org/10.1126/science.aam9309

Kasakoff, A. B., & Adams, J. W. (1995). The Effect of Migration on Ages at Vital Events: A Critique of Family Reconstitution in Historical Demography. *European Journal of Population*, *11*(3), 199–242.

Kerkhove, R. (2020). Tribal alliances with broader agendas?: Aboriginal resistance in southern Queensland's "Black war." *Cosmopolitan Civil Societies: An Interdisciplinary Journal*, *6*(3), 38–62. https://doi.org/10.3316/informit.931199815359739

Koylu, C., Guo, D., Huang, Y., Kasakoff, A., & Grieve, J. (2021). Connecting family trees to construct a population-scale and longitudinal geo-social network for the U.S. *International Journal of Geographical Information Science*, *35*(12), 2380–2423. https://doi.org/10.1080/13658816.2020.1821885

Koylu, C., & Kasakoff, A. B. (2025). Ethical Challenges in Analyzing and Mapping Historical Demographic Changes and Migration Using Population-Scale Family Trees. *Cartographic Perspectives*. https://doi.org/10.14714/CP105.1945

Lam, D., & Marteleto, L. (2008). Stages of the Demographic Transition from a Child's Perspective: Family Size, Cohort Size, and Children's Resources. *Population and Development Review*, *34*(2), 225–252. https://doi.org/10.1111/j.1728-4457.2008.00218.x

Mehta, N. K., Elo, I. T., Engelman, M., Lauderdale, D. S., & Kestenbaum, B. M. (2016). Life Expectancy Among U.S.-born and Foreign-born Older Adults in the United States: Estimates From Linked Social Security and Medicare Data. *Demography*, *53*(4), 1109–1134. https://doi.org/10.1007/s13524-016-0488-4

Minardi, S., Corti, G., & Barban, N. (2023). *Historical Patterns in the Intergenerational Transmission of Lifespan and Longevity: Evidence from the United States, 1700-1900* [Preprint]. SocArXiv. https://doi.org/10.31235/osf.io/gfxcm

Minardi, S., Corti, G., & Barban, N. (2024). Historical Patterns in the Intergenerational Transmission of Lifespan and Longevity: A Research Note on U.S. Cohorts Born Between 1700 and 1900. *Demography*, *61*(4), 979–994. https://doi.org/10.1215/00703370-11458359

Mize, T. (2019). Best Practices for Estimating, Interpreting, and Presenting Nonlinear Interaction Effects. *Sociological Science*, *6*, 81–117. https://doi.org/10.15195/v6.a4

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. https://doi.org/10.1016/j.ssmph.2023.101359

Murdoch, A. (2004). *British emigration, 1603-1914*. Palgrave Macmillan.

Nester, W. R. (2004). *Frontier War for American Independence* (First Edition). Stackpole Books.

Nicholas, S., & Shergold, P. R. (1987). Internal migration in England, 1818–1839. *Journal of Historical Geography*, *13*(2), 155–168. https://doi.org/10.1016/S0305-7488(87)80144-5

Otterstrom, S., Selander, J., & Deo, R. (2024). Analyzing Multi-Generational Gathering and Dispersion to and from Ten 19th-Century American Cities. *Genealogy*, *8*(4), 148. https://doi.org/10.3390/genealogy8040148

Pablos-Méndez, A. (1994). Mortality Among Hispanics. *JAMA: The Journal of the American Medical Association*, *271*(16), 1237. https://doi.org/10.1001/jama.1994.03510400023017

Palloni, A., & Arias, E. (2004). Paradox lost: Explaining the Hispanic adult mortality advantage. *Demography*, *41*(3), 385–415. https://doi.org/10.1353/dem.2004.0024

Piraino, P., Muller, S., Cilliers, J., & Fourie, J. (2014). The transmission of longevity across generations: The case of the settler Cape Colony. *Research in Social Stratification and Mobility, 35*, 105–119. https://doi.org/10.1016/j.rssm.2013.08.005

Puschmann, P., Donrovich, R., Grönberg, P.-O., Dekeyser, G., & Matthijs, K. (2016). Disfavored in Life, Favored in Death? Later-Life Mortality Differences (Ages 30+) between Migrants and Natives in Antwerp, Rotterdam and Stockholm, 1850-1930. *Historical Social Research, 41*(4). https://doi.org/10.12759/HSR.41.2016.4.257-290

Puschmann, P., Donrovich, R., & Matthijs, K. (2017). Salmon Bias or Red Herring?: Comparing Adult Mortality Risks (Ages 30–90) between Natives and Internal Migrants: Stayers, Returnees and Movers in Rotterdam, the Netherlands, 1850–1940. *Human Nature, 28*(4), 481–499. https://doi.org/10.1007/s12110-017-9303-1

Razum, O. (2008). Migrant Mortality, Healthy Migrant Effect. In W. Kirch (Ed.), *Encyclopedia of Public Health* (pp. 932–935). Springer Netherlands. https://doi.org/10.1007/978-1-4020-5614-7_2188

Razum, O., Zeeb, H., Akgün, H. S., & Yilmaz, S. (1998). Low overall mortality of Turkish residents in Germany persists and extends into a second generation: Merely a healthy migrant effect?*. *Tropical Medicine & International Health*, *3*(4), 297–303. https://doi.org/10.1046/j.1365-3156.1998.00233.x

Richards, E. (2018). *The genesis of international mass migration: The British case, 1750-1900*. Manchester University Press.

Ruggles, S. (1992). Migration, Marriage, and Mortality: Correcting Sources of Bias in English Family Reconstitutions. *Population Studies, 46*(3), 507–522. https://doi.org/10.1080/0032472031000146486

Spencer, R. W., & Otterstrom, S. M. (2024). *Massive Genealogies Distinguish Frontier from Steady-State Internal Migration* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2410.18235

Stark, O., & Bloom, D. E. (1985). The New Economics of Labor Migration. *The American Economic Review*, *75*(2), 173–178.

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. https://doi.org/10.1073/pnas.2120455119

Taylor, D. (1988). Public Health in Britain 1750–1900. In D. Taylor, *Mastering Economic and Social History* (pp. 301–331). Macmillan Education UK. https://doi.org/10.1007/978-1-349-19377-6_19

Thompson, R. (2009). *Mobility and Migration: East Anglian Founders of New England, 1629-1640* (Reprint edition). University of Massachusetts Press.

Tomlins, C. (2001). Reconsidering Indentured Servitude: European Migration and the Early American Labor Force, 1600–1775. *Labor History*, *42*(1), 5–43. https://doi.org/10.1080/00236560123269

Tosh, J. (2017). From the ‘cape of despair’ to the Cape of Good Hope: Letters of the emigrant poor in early nineteenth-century England. *Social History*, *42*(4), 480–500. https://doi.org/10.1080/03071022.2017.1368233

Wallace, M., & Wilson, B. (2019). Migrant Mortality Advantage Versus Origin and the Selection Hypothesis. *Population and Development Review*, *45*(4), 767–794. https://doi.org/10.1111/padr.12298

Webb, D. A. (2015). WAR, RACISM, AND THE TAKING OF HEADS: REVISITING MILITARY CONFLICT IN THE CAPE COLONY AND WESTERN XHOSALAND IN THE NINETEENTH CENTURY. *The Journal of African History*, *56*(1), 37–55. https://doi.org/10.1017/S0021853714000693

Wegge, S. A. (1999). To Part or Not to Part: Emigration and Inheritance Institutions in Nineteenth-Century Hesse–Cassel. *Explorations in Economic History*, *36*(1), 30–55. https://doi.org/10.1006/exeh.1998.0703

Tables

Table 1 Sample descriptive statistics

	Whole Sample Mean/Prop. (SD)	**Non-Migrants** Mean/Prop. (SD)	**United States of America** Mean/Prop. (SD)	**Australia** Mean/Prop. (SD)	**New Zealand** Mean/Prop. (SD)	**Canada** Mean/Prop. (SD)	**South Africa** Mean/Prop. (SD)
Age at death	65.85 (18.27)	63.73 (19)	69.4 (16.3)	68.48 (16.39)	70.8 (15.36)	71.63 (15.97)	66.72 (17.55)
Male	0.61	0.61	0.61	0.59	0.59	0.63	0.65
Number of siblings							
0	0.51	0.48	0.58	0.53	0.53	0.54	0.49
1	0.15	0.16	0.13	0.14	0.15	0.14	0.15
2	0.10	0.11	0.08	0.09	0.10	0.10	0.12
3-5	0.17	0.19	0.14	0.17	0.16	0.17	0.16
6+	0.07	0.07	0.07	0.07	0.07	0.06	0.07
Migrant	0.37	-	-	-	-	-	-
N	98,057	61,794	18,332	8585	2734	5042	1181
Number of sibling groups	67,530	43,228	14,275	6495	2106	3959	884

Note: Authors' calculation using Familinx data. "Prop." = proportion, "SD" = standard deviation.

Figures

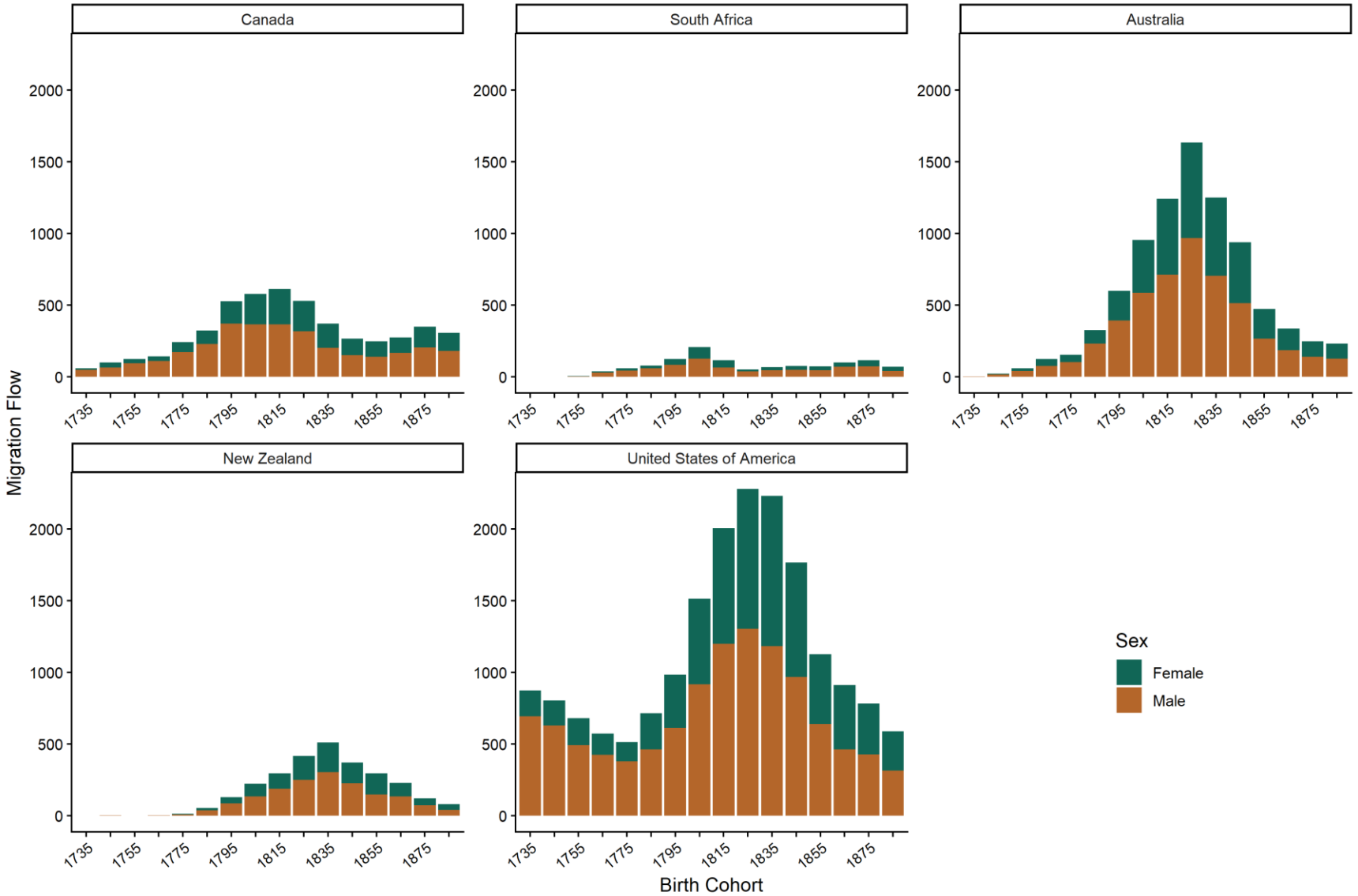


Figure 1 Observed migration flows in Familinx sample by sex, birth cohort, and destination.

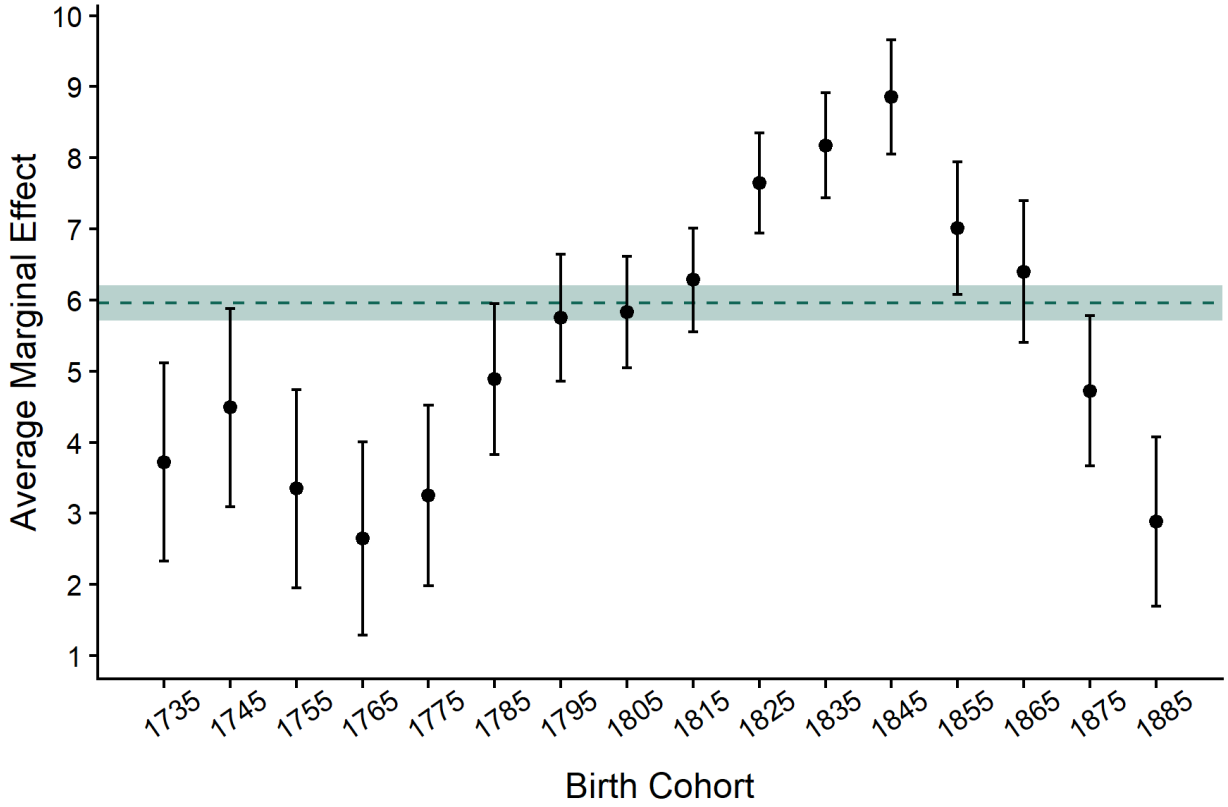


Figure 2 Average marginal effect of migration on age at death, across birth cohort (Model A1). Overall AME (migrant versus non-migrant) is shown by the dashed line, with 95% CIs shaded in green.

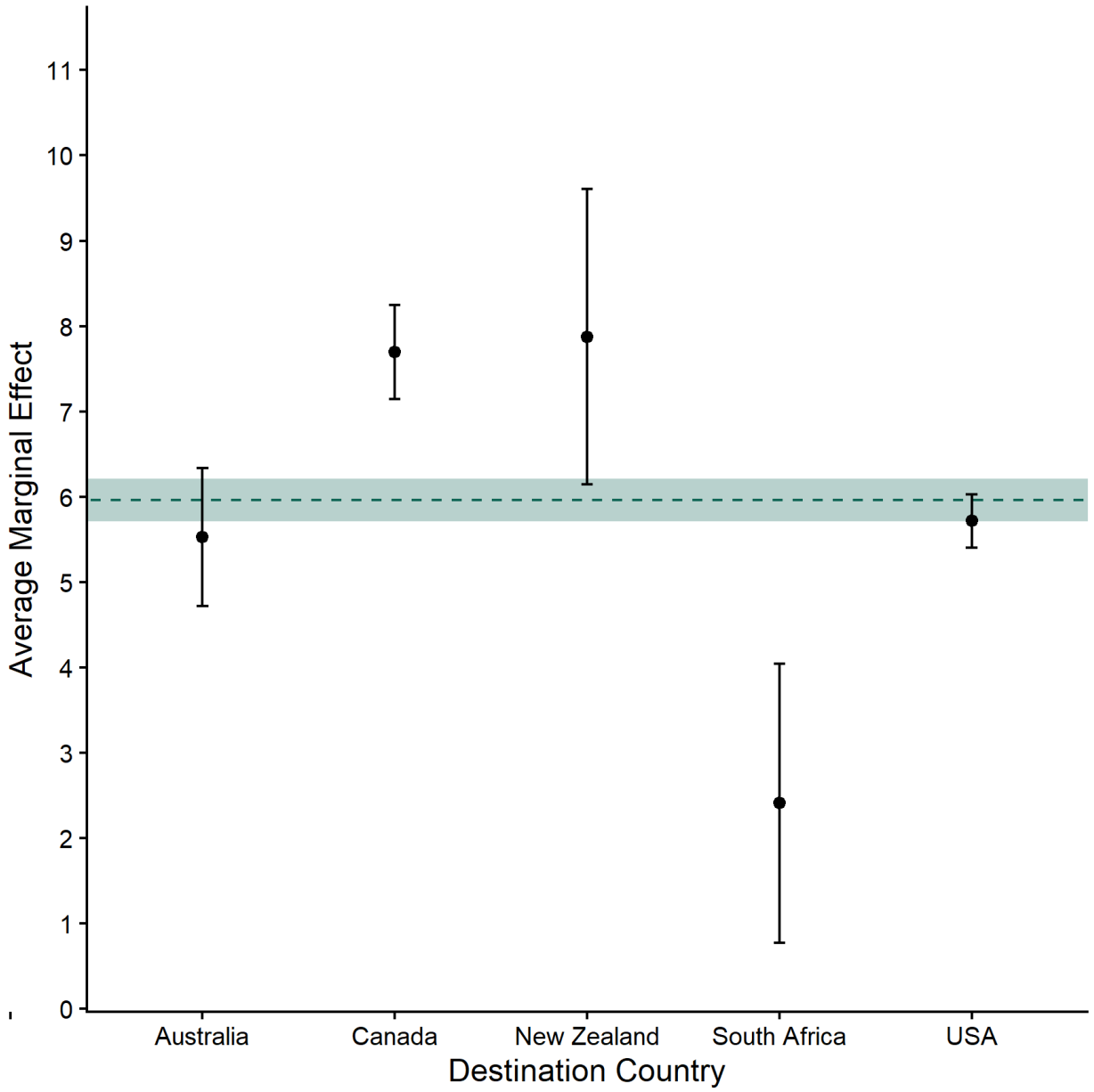


Figure 3 AME of migrating on age at death by destination country (Model B1) in black with 95% CIs. AME for Model A1 (migrant versus non-migrant) is shown by the dashed line, with 95% CIs shaded in green.

Online Supplement

Leaving for Life: Using Online Crowd-Sourced Genealogies to Estimate the Migrant Mortality Advantage for the United Kingdom and Ireland Between the 18th and 20th Centuries

Table of Contents

A. Data Cleaning & Imputation 38
Imputation of missingness 38
Sibling Ties 42
B. Results and Robustness Checks 43
Supplementary models 49
C. References 66

A. Data Cleaning & Imputation

Imputation of missingness

Due to the messiness of the raw data, we take steps, such as documenting and publishing the data preparation procedure, to ensure transparency and reliability of our data and results. We began with a dataset of over 86 million individuals, many of whom had missing values across most variables. In preparing the data for imputations, only individuals with observed kinship ties were kept from the full data set. To address the high degree of missingness in key variables among the remaining observations, we imputed several types of missingness: (1) missingness in the birth and death year, (2) missingness in the birth and death location, and (3) missing gender. Figure S1 offers a visual representation of our data cleaning process to obtain our initial sample.

For type (1), we imputed missing birth and death years with baptism and burial years, respectively, if they were also non-missing in the Familinx dataset. Similarly for type (2), we used coordinate information on birth, death, baptism, and burial to extract countries of interest for each event. Missing birth and death locations were imputed using baptism and burial locations, respectively. This included baptism and burial locations that were derived from coordinate information. Finally, entries containing abnormal birth years (negative years and years occurring after 2020) that likely contained errors were removed.

The coordinate information matching process was conducted as follows. Using the *countrycode* package (Arel-Bundock et al., 2018), columns containing country code information were then used to fill in entries that were still missing in these four columns, with imputed baptism and burial information from their respective country codes being used to fill in country level information. The *countrycode* package was also used to fill in birth and death countries with “USA” from the state-based location columns if they matched names or abbreviations of U.S. states. Finally, the free text columns were used to extract country level information through regex for rows that were still missing country information. This was repeated on the baptism and burial free text columns which were then used to impute birth and death country information that was still missing.

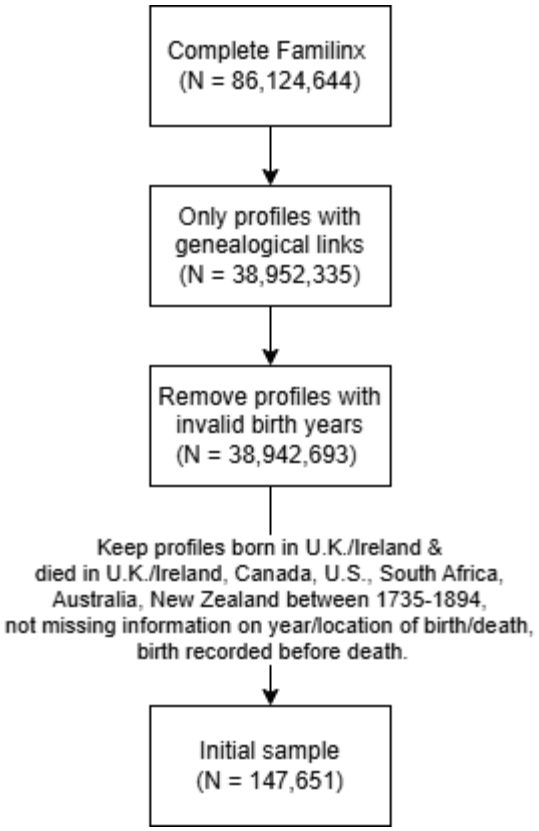


Figure S1 Flow of data cutting process to obtain initial sample

To fill in type (3) information for individuals with a missing gender, kinship ties were used. When two individuals were both parents of another individual in the data set and the gender information for one parent was known, this was used to fill in the gender information for the parent missing gender information.

To create the final country categories, a combination of term matching and regex approaches were used. For Ireland and Northern Ireland, regex pattern matching was used to detect references to Ireland or "eire". For England, Scotland, Wales, a term matching approach was adopted to match various spellings, languages and cities referenced. Finally, for the United States, Australia, South Africa, New Zealand and Canada, terms matching in the country column included references to territories, regions, states (that had not already been picked up previously), alternate spellings and names in other languages. The full list of terms used for term matching is presented in Table S1.

To produce the final data set, only rows that were complete (i.e., had no missing information regarding birth/death year, gender, or birth/death location) were kept. All individuals born in the U.K./Ireland and dying in the U.K./Ireland, United States, Australia, Canada, New Zealand, and South Africa were kept.

Table S1 Term matching terms

Country	Term
United Kingdom	'united kingdom', 'x-england', 'england', 'x-scotland', 'x-united-kingdom', 'scotland', '(present uk)', 'x-great-britain', 'x-wales', 'gb', 'x-northern-ireland', 'uk', 'northern ireland', 'england, uk', 'u.k.', 'wales', 'england, united kingdom', 'great britain', 'england uk', 'uk:great britain', 'uk:northern ireland', 'england (present uk)', 'scotland, united kingdom', 'uk:northern ireland', 'uk.', 'scotland, uk', 'uk:isle of wight', 'ireland (present northern ireland)', 'north ireland', 'scotland, uk', 'scotland, united kingdom', 'scotland uk', 'south wales', 'britain', 'england/ uk', 'n.ireland', 'engand', 'englnd', 'northern ireland, uk', 'middlesex', 'n. ireland', 'huntingdonshire', 'lancashire', 'london', 'uk:scotland:shetland islands:mainland', 'u k', 'united kingdom of great britain and ireland', 'northern-ireland', 'nothern ireland, uk', 'ireland (northern)', 'n. ireland', 'ireland or scotland', 'scotland or ireland', 'enfland', 'storbritannia', 'bonhill, dunbartonshire, scotland', 'cambridgeshire', 'chatham', 'crickdale, wiltshire, uk', 'eicester, leicestershire, uk', 'endgland', 'england.', 'englanmd', 'englnad', 'enland', 'essex', 'fifeshire', 'gloucester', 'great briatin', 'great britai', 'herfordshire', 'lower bebington', 'newcastle upon tyne', 'north wales', 'reino unido', 'royaume uni', 'scotlans', 'swindon, wiltshire, england', 'tyrone', 'uk /england', 'uk/wales', 'uk:wales:anglesey', 'umited kingdom', 'united kinbgdom', 'united kinbgdom', 'winwick', 'woolwich, kent, uk', 'yorkshire', 'xengland', "yhdistynyt kuningaskunta", "(present u.k.)", "(present day united kingdom)", "-england", "(now united kingdom)", "present united kingdom)", ","המלכה המאוחדת "aberdeen city", "englang", "england,uk", "england or pa", "wales or england", "verenigd koninkrijk", "unitedkingdom", "united-kingdom", "uk:scotland:barra", "uk, england", "u. k.", "heathfield, sussex, england", "united kingdrom"
Ireland	'ireland', 'ie', 'republic of ireland', 'eire', 'bydoney,tyrone ,ireland', 'ireland, uk', 'uk (ireland)', 'ireland ???', 'ireland (eire)', 'ireland.', 'or ireland', 'kilkenny', 'tipperary', 'waterford', 'ulster', 'galway', 'down', 'carlow'
United States of America	us", 'usa', 'united states', 'united states of america', 'america', '(present usa)', 'colonial america', 'province of new york', 'new netherland colony', 'new england colonies', 'new england', 'present united states', 'american colonies', "british america", "new netherlands", 'new york', 'american colonies [present united states]', 'british north america', 'u.s.a.', 'new netherlands (usa)', 'nieuw netherlands', 'british colonies', 'nouvelle france', "british colonies of north america", 'american colonies (present usa)', 'usa:new york:long island', 'usa:49', 'usa:massachusettes:nantucket island', 'usa:massachusettes:martha\'s vineyard', 'usa:44', 'the united states of america', 'massachusetts colony', 'british amercia', 'now usa', 'martin county, indiana, usa', 'richland county', 'u,s.a.', 'u.sa.', 'united sates', 'unitet states', 'usa.', 'usa:hawaii:kauai', "ee.uu.", "(currently) united states", "(present usa)", "(present (usa)", "(present) usa", "amerikas forente stater", "cleveland", "estados unidos", "usa:california:santa catalina island", "usa (all present day)", "us virgin islands", "kittery, york, maine", "in what will be america"

Country	Term
Australia	'australia', 'au', 'australien', 'australia:tasmania', 'new south wales', 'australia [green slopes hospital]', 'aust', 'western australia', 'austrailia', 'australis', 'nsw', 'port melbourne', 'portsea', 'sydney nsw', 'tasmania'
New Zealand	'new zealand', 'nz', 'new zealand:north island', 'new zealand:south island', 'new zealand.', ', new zealand', 'christchurch', 'king street, sydenham, christchurch, nz', 'new zeaand', 'new zealand of senile decay', 'new, zealand', 'tennyson street, sydenham, christchurch, nz', ''king street, sydenham, christchurch, nz''
Canada	'canada', 'ca', 'kanadas', 'canada:27', 'united province of canada', 'british north america (present canada)', 'canada:cape breton island', 'province of canada', 'canada:11', 'canada:newfoundland', 'upper canada', '(present canada)', 'acadie', 'canad', 'can'', 'canadá', 'canada:15', 'canada:vancouver island'
South Africa	'za', 'south africa', 'suid afrika', 'cape of good hope', 'cape colony', 's africa', 'cape colony (south africa)', 'rep south africa', 'south africa', 'south afica', 'south africa.', 'union of south africa'

Sibling Ties

We define siblings as individuals who share at least one parent. Siblings are first determined through the maternal line and second through the paternal. Using the unique ID of one's mother, or of the father if the mother's ID is missing, we construct a family ID that allows us to identify groups of siblings. Individuals without recorded siblings are given an individual ID. Table S2 displays the percent of observations, by country of death, that were assigned their mother's, father's, or personal unique ID, respectively. Roughly 99 percent of profiles in each country were assigned their mother's id as their sibling tie.

Table S2 Proportion assigned each parent ID

	Assigned mother ID	Assigned father ID
United Kingdom and Ireland	0.993	0.007
United States of America	0.987	0.013
Australia	0.994	0.006
New Zealand	0.989	0.011
Canada	0.993	0.007
South Africa	0.990	0.010

Note: Authors' calculation using Familinx data.

B. Results and Robustness Checks

This section is organized as follows. We first present regression estimates from the main models (Table S3). Next, we estimate supplementary models on a reduced sample of individuals with at least one sibling (Figures S4-S5, Table S4). To compare models A1 and B1 with A2 and B2, we compare the AMEs directly (Table S5). Then, we show that our models are robust to alternate minimum age cutoffs (Figures S6-S7). Finally, we present estimates from Cox mixed effects models (Table S6).

Table S3 Results from Main Models

Dependent variable: Age at Death	Model A1: Migrant	Model B1: Country of Death
Migrant (*reference: Non-Migrant)*	4.535**	
	(1.421)	
Destination country *(reference: United Kingdom/Ireland)*		
Canada		3.844
		(2.385)
South Africa		0.513
		(2.194)
Australia		6.244
		(10.229)
New Zealand		2.292
		(2.067)
United States of America		3.701***
		(0.728)
Birth cohort (*reference: 1735)*		
1745	0.101	0.072
	(0.950)	(0.550)
1755	0.436	0.464
	(0.934)	(0.532)
1765	3.200***	2.374***
	(0.900)	(0.516)
1775	2.462**	2.246***
	(0.863)	(0.494)
1785	1.540†	1.157*
	(0.836)	(0.481)
1795	0.711	0.174
	(0.810)	(0.473)
1805	0.385	-0.764
	(0.797)	(0.470)
1815	-0.991	-1.775***
	(0.786)	(0.464)
1825	-2.521**	-3.511***
	(0.783)	(0.466)
1835	-3.552***	-4.838***

Dependent variable: Age at Death	Model A1: Migrant	Model B1: Country of Death
	(0.790)	(0.471)
1845	-4.521***	-5.533***
	(0.794)	(0.475)
1855	-2.537**	-4.255***
	(0.803)	(0.484)
1865	-0.862	-2.585***
	(0.807)	(0.487)
1875	2.296**	0.272
	(0.819)	(0.493)
1885	5.178***	3.564***
	(0.845)	(0.517)
Male *(reference: female)*	0.305†	0.028
	(0.178)	(0.117)
Number of siblings *(reference: none)*		
1	-	-2.060***
	-	(0.175)
2	-0.784**	-2.824***
	(0.271)	(0.213)
3-5	-0.801**	-2.816***
	(0.244)	(0.182)
6+	-0.166	-2.084***
	(0.365)	(0.304)
Migrant X Birth cohort interaction *(reference: 1735)*		
1745	1.829	
	(1.940)	
1755	0.670	
	(1.935)	
1765	-1.878	
	(2.000)	
1775	-0.942	
	(1.893)	
1785	2.258	
	(1.751)	
1795	2.907†	
	(1.627)	
1805	2.417	
	(1.565)	
1815	3.247*	
	(1.538)	
1825	4.353**	
	(1.519)	

Dependent variable: Age at Death	Model A1: Migrant	Model B1: Country of Death
1835	4.607**	
	(1.524)	
1845	5.363***	
	(1.544)	
1855	3.556*	
	(1.585)	
1865	2.121	
	(1.606)	
1875	0.793	
	(1.631)	
1885	-0.832	
	(1.672)	
Destination country X Birth cohort interaction *(reference: U.K./Ireland & 1735)*		
Canada:1745		1.935
		(3.016)
South Africa:1745		-7.775
		(17.975)
Australia:1745		-3.687
		(10.950)
New Zealand:1745		-3.453
		(10.484)
United States of America:1745		0.709
		(1.036)
Canada:1755		3.765
		(2.905)
South Africa:1755		12.443
		(8.239)
Australia:1755		-1.335
		(10.495)
United States of America:1755		-1.336
		(1.063)
Canada:1765		3.330
		(2.835)
South Africa:1765		-0.765
		(3.664)
Australia:1765		-3.291
		(10.359)
New Zealand:1765		5.116
		(17.959)
United States of America:1765		-2.093†
		(1.098)

Dependent variable: Age at Death	Model A1: Migrant	Model B1: Country of Death
Canada:1775		1.828
		(2.665)
South Africa:1775		1.455
		(3.226)
Australia:1775		-2.808
		(10.336)
New Zealand:1775		10.824*
		(5.357)
United States of America:1775		-1.749
		(1.117)
Canada:1785		4.548†
		(2.598)
South Africa:1785		3.441
		(2.990)
Australia:1785		-1.234
		(10.281)
New Zealand:1785		8.114*
		(3.217)
United States of America:1785		-0.750
		(1.027)
Canada:1795		3.689
		(2.524)
South Africa:1795		1.271
		(2.751)
Australia:1795		-0.973
		(10.259)
New Zealand:1795		3.933
		(2.617)
United States of America:1795		1.839†
		(0.962)
Canada:1805		3.870
		(2.514)
South Africa:1805		2.356
		(2.559)
Australia:1805		-1.218
		(10.249)
New Zealand:1805		5.526*
		(2.401)
United States of America:1805		2.028*
		(0.899)
Canada:1815		5.188*

Dependent variable: Age at Death	Model A1: Migrant	Model B1: Country of Death
		(2.506)
South Africa:1815		2.245
		(2.766)
Australia:1815		-1.664
		(10.245)
New Zealand:1815		6.703**
		(2.328)
United States of America:1815		2.610**
		(0.867)
Canada:1825		5.891*
		(2.523)
South Africa:1825		2.703
		(3.360)
Australia:1825		-0.547
		(10.242)
New Zealand:1825		6.320**
		(2.261)
United States of America:1825		4.775***
		(0.857)
Canada:1835		5.870*
		(2.575)
South Africa:1835		0.167
		(3.140)
Australia:1835		0.736
		(10.245)
New Zealand:1835		5.771**
		(2.232)
United States of America:1835		5.123***
		(0.863)
Canada:1845		5.000†
		(2.639)
South Africa:1845		4.509
		(3.029)
Australia:1845		2.419
		(10.250)
New Zealand:1845		8.519***
		(2.285)
United States of America:1845		5.017***
		(0.887)
Canada:1855		3.213
		(2.660)

Dependent variable: Age at Death	Model A1: Migrant	Model B1: Country of Death
South Africa:1855		-0.359
		(3.036)
Australia:1855		1.562
		(10.267)
New Zealand:1855		7.349**
		(2.340)
United States of America:1855		2.769**
		(0.947)
Canada:1865		3.577
		(2.634)
South Africa:1865		0.927
		(2.864)
Australia:1865		0.049
		(10.281)
New Zealand:1865		6.996**
		(2.411)
United States of America:1865		2.274*
		(0.985)
Canada:1875		1.637
		(2.589)
South Africa:1875		4.632†
		(2.752)
Australia:1875		-1.030
		(10.297)
New Zealand:1875		5.403*
		(2.625)
United States of America:1875		0.022
		(1.015)
Canada:1885		-0.426
		(2.619)
Australia:1885		-3.437
		(10.303)
United States of America:1885		-0.698
		(1.092)
Constant	62.574***	66.190***
	(0.718)	(0.402)
Sibling random effect, variance	40.17	39.90
Residual variance	278.52	278.23
Observations	97,668	97,668

Note: † p<0.1 * p<0.05 ** p<0.01

Supplementary models

We estimate several supplementary models. We first limit our sample to only individuals with at least one sibling in the data (Models A2 and B2; N = 48,224). The results are substantively similar to the previous models, both for the migrant model (Figure S2) and by destination country (Figure S3). Table S4 presents regression estimates from these models.

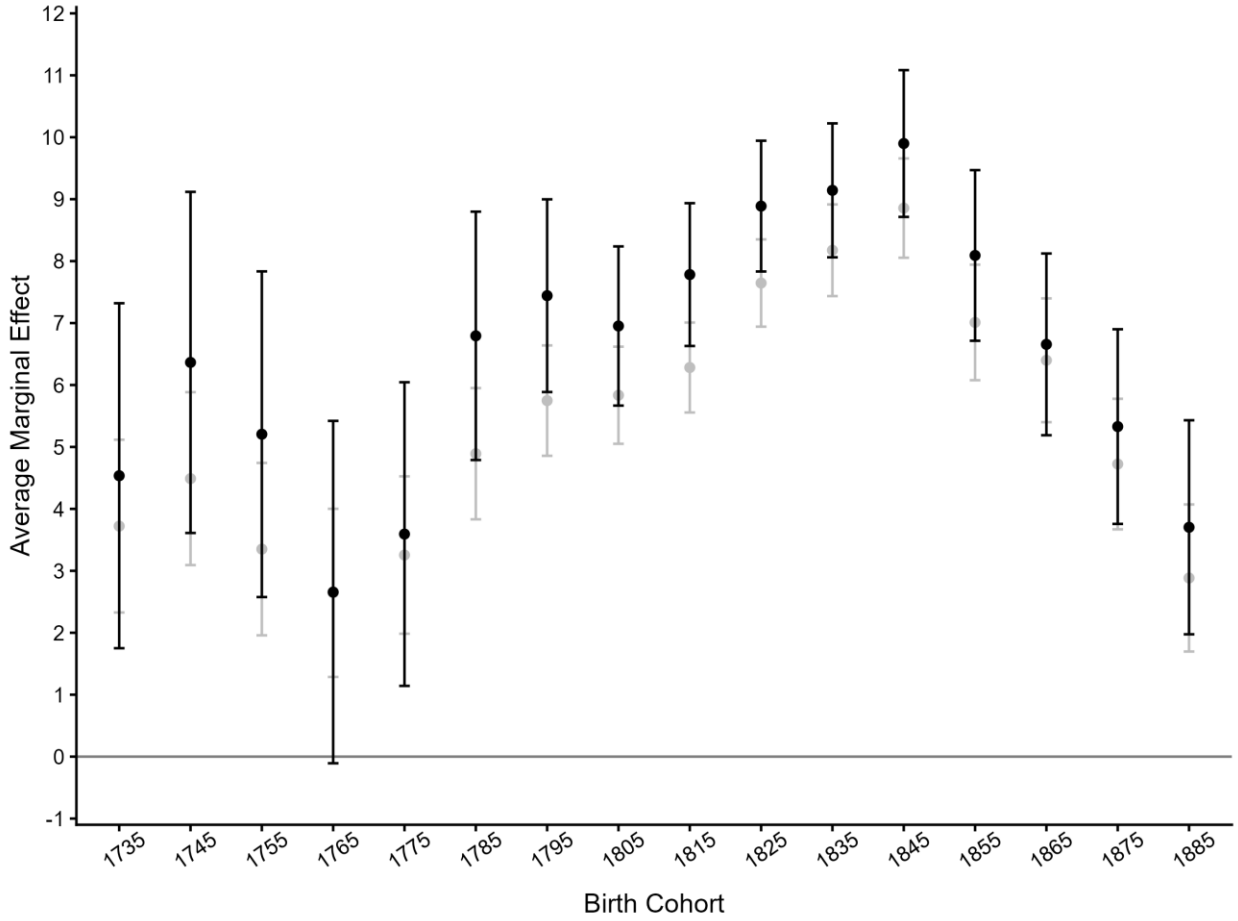


Figure S2 Average marginal effect of migration on age at death, across birth cohort (Model A2) shown in black. AMEs for full sample (Model A1) in grey.

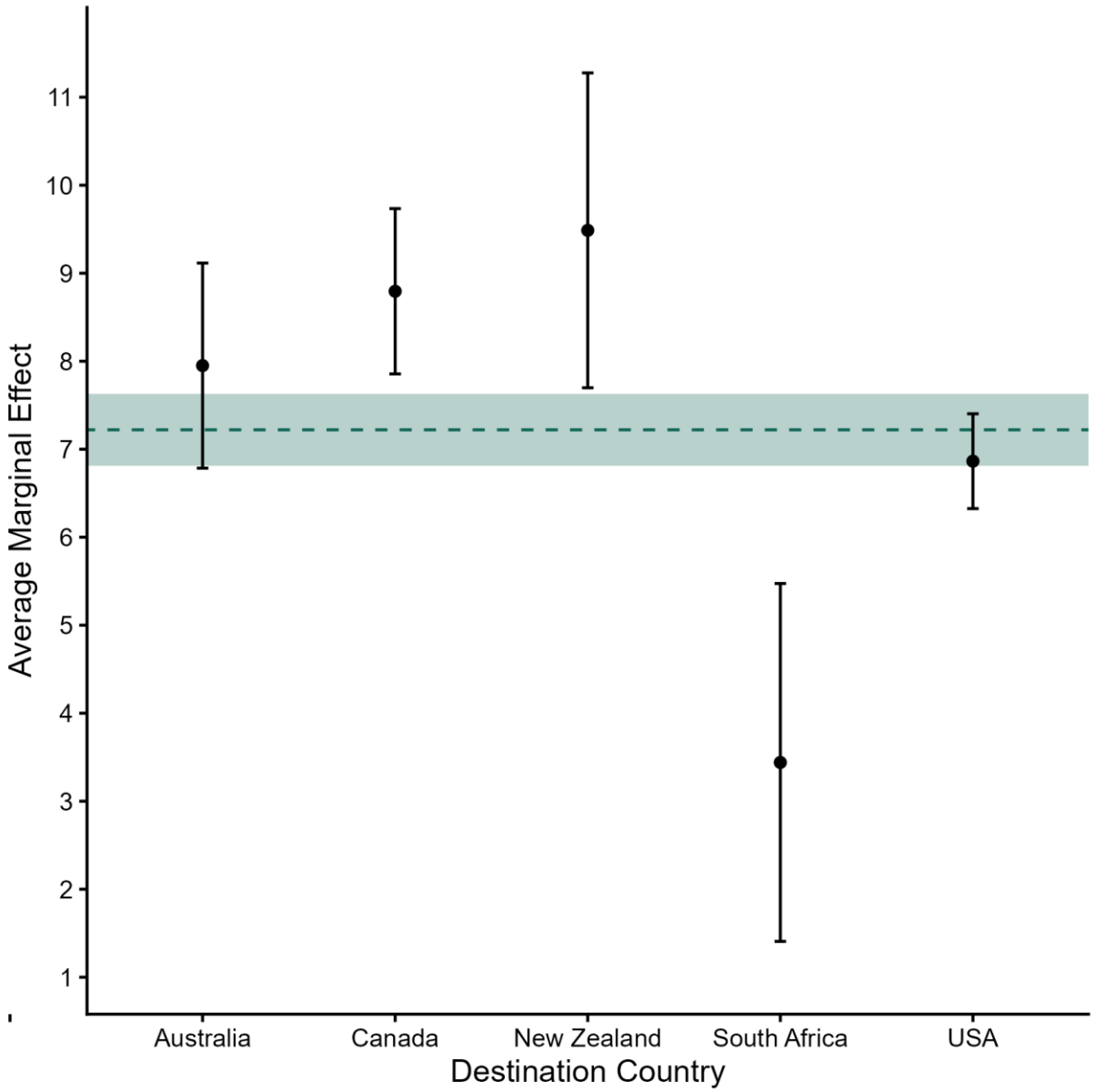


Figure S3 AME of migrating on age at death by destination country (Model B2) in black with 95% CIs. AME for Model A2 (migrant versus non-migrant) is shown by the dashed green line, with 95% CIs in green shade.

Table S4 Regression Estimates from Supplementary Models, Only Profiles with Siblings

Dependent variable: Age at Death	Model A1: Migrant	Model B1: Country of Death
Migrant (*reference: Non-Migrant)*	4.535**	
	(1.421)	
Destination country *(reference: United Kingdom/Ireland)*		
Canada		1.361
		(4.768)
South Africa		1.274
		(3.497)
Australia		20.977
		(13.359)
New Zealand		3.488
		(3.019)
United States of America		4.578**
		(1.457)
Birth cohort (*reference: 1735)*		
1745	0.101	0.102
	(0.950)	(0.949)
1755	0.436	0.444
	(0.934)	(0.933)
1765	3.200***	3.209***
	(0.900)	(0.900)
1775	2.462**	2.462**
	(0.863)	(0.863)
1785	1.540†	1.535†
	(0.836)	(0.836)
1795	0.711	0.712
	(0.810)	(0.809)
1805	0.385	0.384
	(0.797)	(0.797)
1815	-0.991	-0.997
	(0.786)	(0.786)
1825	-2.521**	-2.528**
	(0.783)	(0.783)
1835	-3.552***	-3.559***
	(0.790)	(0.790)
1845	-4.521***	-4.527***
	(0.794)	(0.794)
1855	-2.537**	-2.546**
	(0.803)	(0.803)
1865	-0.862	-0.866
	(0.807)	(0.807)
1875	2.296**	2.289**

Dependent variable: Age at Death	Model A1: Migrant	Model B1: Country of Death
	(0.819)	(0.818)
1885	5.178***	5.179***
	(0.845)	(0.845)
Male *(reference: female)*	0.305†	0.309†
	(0.178)	(0.178)
Number of siblings *(reference: 1)*		
2	-0.784**	-0.764**
	(0.271)	(0.271)
3-5	-0.801**	-0.764**
	(0.244)	(0.244)
6+	-0.166	-0.122
	(0.365)	(0.365)
Migrant X Birth cohort interaction *(reference: 1735)*		
1745	1.829	
	(1.940)	
1755	0.670	
	(1.935)	
1765	-1.878	
	(2.000)	
1775	-0.942	
	(1.893)	
1785	2.258	
	(1.751)	
1795	2.907†	
	(1.627)	
1805	2.417	
	(1.565)	
1815	3.247*	
	(1.538)	
1825	4.353**	
	(1.519)	
1835	4.607**	
	(1.524)	
1845	5.363***	
	(1.544)	
1855	3.556*	
	(1.585)	
1865	2.121	
	(1.606)	
1875	0.793	
	(1.631)	
1885	-0.832	
	(1.672)	

Dependent variable: Age at Death	Model A1: Migrant	Model B1: Country of Death
Destination country X Birth cohort interaction *(reference: U.K./Ireland & 1735)*		
Canada:1745		8.705
		(6.236)
Australia:1745		0.369
		(18.891)
New Zealand:1745		24.275
		(19.199)
United States of America:1745		1.144
		(2.005)
Canada:1755		7.050
		(5.625)
South Africa:1755		9.088
		(13.772)
Australia:1755		-14.694
		(14.791)
United States of America:1755		-0.165
		(2.047)
Canada:1765		10.357†
		(6.009)
South Africa:1765		-7.391
		(7.880)
Australia:1765		-12.531
		(14.013)
United States of America:1765		-3.833†
		(2.160)
Canada:1775		6.031
		(5.276)
South Africa:1775		-2.079
		(5.767)
Australia:1775		-17.020
		(13.798)
New Zealand:1775		1.517
		(11.314)
United States of America:1775		-2.474
		(2.183)
Canada:1785		9.169†
		(5.104)
South Africa:1785		5.437
		(5.074)
Australia:1785		-13.876
		(13.522)
New Zealand:1785		10.404†
		(6.302)

Dependent variable: Age at Death	Model A1: Migrant	Model B1: Country of Death
United States of America:1785		-0.610
		(2.058)
Canada:1795		7.524
		(4.952)
South Africa:1795		4.541
		(4.280)
Australia:1795		-13.802
		(13.441)
New Zealand:1795		7.677†
		(4.387)
United States of America:1795		1.802
		(1.890)
Canada:1805		6.572
		(4.909)
South Africa:1805		3.283
		(3.898)
Australia:1805		-14.477
		(13.407)
New Zealand:1805		6.849†
		(3.935)
United States of America:1805		2.329
		(1.738)
Canada:1815		8.603†
		(4.916)
South Africa:1815		4.593
		(4.282)
Australia:1815		-14.724
		(13.390)
New Zealand:1815		7.168*
		(3.534)
United States of America:1815		3.187†
		(1.664)
Canada:1825		8.587†
		(4.931)
South Africa:1825		7.225
		(5.232)
Australia:1825		-13.660
		(13.381)
New Zealand:1825		6.404†
		(3.352)
United States of America:1825		5.172**
		(1.620)
Canada:1835		10.822*

Dependent variable: Age at Death	Model A1: Migrant	Model B1: Country of Death
		(5.009)
South Africa:1835		-1.090
		(4.824)
Australia:1835		-12.413
		(13.385)
New Zealand:1835		3.949
		(3.280)
United States of America:1835		5.119**
		(1.613)
Canada:1845		7.406
		(5.121)
South Africa:1845		0.295
		(4.998)
Australia:1845		-10.963
		(13.394)
New Zealand:1845		9.426**
		(3.364)
United States of America:1845		5.065**
		(1.639)
Canada:1855		8.042
		(5.133)
South Africa:1855		-0.431
		(4.929)
Australia:1855		-12.833
		(13.420)
New Zealand:1855		6.152†
		(3.397)
United States of America:1855		3.222†
		(1.722)
Canada:1865		7.035
		(5.055)
South Africa:1865		-0.540
		(4.391)
Australia:1865		-13.946
		(13.447)
New Zealand:1865		6.255†
		(3.476)
United States of America:1865		1.216
		(1.767)
Canada:1875		3.694
		(4.989)
South Africa:1875		4.842
		(4.406)

Dependent variable: Age at Death	Model A1: Migrant	Model B1: Country of Death
Australia:1875		-13.344
		(13.470)
New Zealand:1875		3.649
		(3.814)
United States of America:1875		-0.334
		(1.814)
Canada:1885		1.546
		(5.033)
Australia:1885		-16.229
		(13.473)
United States of America:1885		-0.648
		(1.883)
Constant	62.574***	62.553***
	(0.718)	(0.718)
Sibling random effect, variance	53.52	53.39
Residual variance	313.58	313.35
Observations	48,224	48,224

*Notes: †p<0.1, *p<0.05, **p < 0.01, ***p<0.001.*

Table S5 compares the average marginal effects (AMEs) between the main models and these models. We interpret the similarities between the set of models as an indication of model robustness, as the profiles with more siblings are likely higher quality, which could be suggestive of higher socioeconomic status.

Table S5 AMEs of Models A1, B1, A2, B2

Dependent variable: Age at Death	Full Sample		Profiles with siblings	
	A1: Migrant	B1: Country of Death	A2: Migrant	B2: Country of Death
Migrant	5.96***		7.22***	
	(0.13)		(0.21)	
Australia		5.53***		7.95***
		(0.41)		(0.59)
Canada		7.70***		8.79***
		(0.28)		(0.5)
New Zealand		7.89***		9.49***
		(0.88)		(0.91)
South Africa		2.41**		3.44***
		(0.83)		(1.04)
United States of America		5.72***		6.86***
		(0.16)		(0.27)
Observations	97,668	97,668	48,224	48,224

*Notes: †p<0.1, *p<0.05, **p < 0.01, ***p<0.001.*

In the next set of supplementary models (Models A3 and B3) we repeat our original analyses, conditioning on survival to alternative ages (i.e., 0, 5, 10, 20, 25, …, 50). In doing so, we attempt to correct for the fact that the migrant mortality advantage may vary across age (Guillot et al., 2018) or may be sensitive to alternative cut-off points.

Figure S4 shows the AMEs from Model A3, demonstrating a consistent migrant mortality advantage when conditioning on survival through age 15. As the minimum age increases, the extent of the mortality advantage decreases, though it remains positive and statistically significant ($p < 0.001$). The size of the AME decreases with higher age cut-offs because samples with higher cutoffs have fewer individuals who die young and are biased towards those who live longer. As such, with higher age cut-offs there is less variability in age at death, so the difference between migrants and non-migrants diminishes.

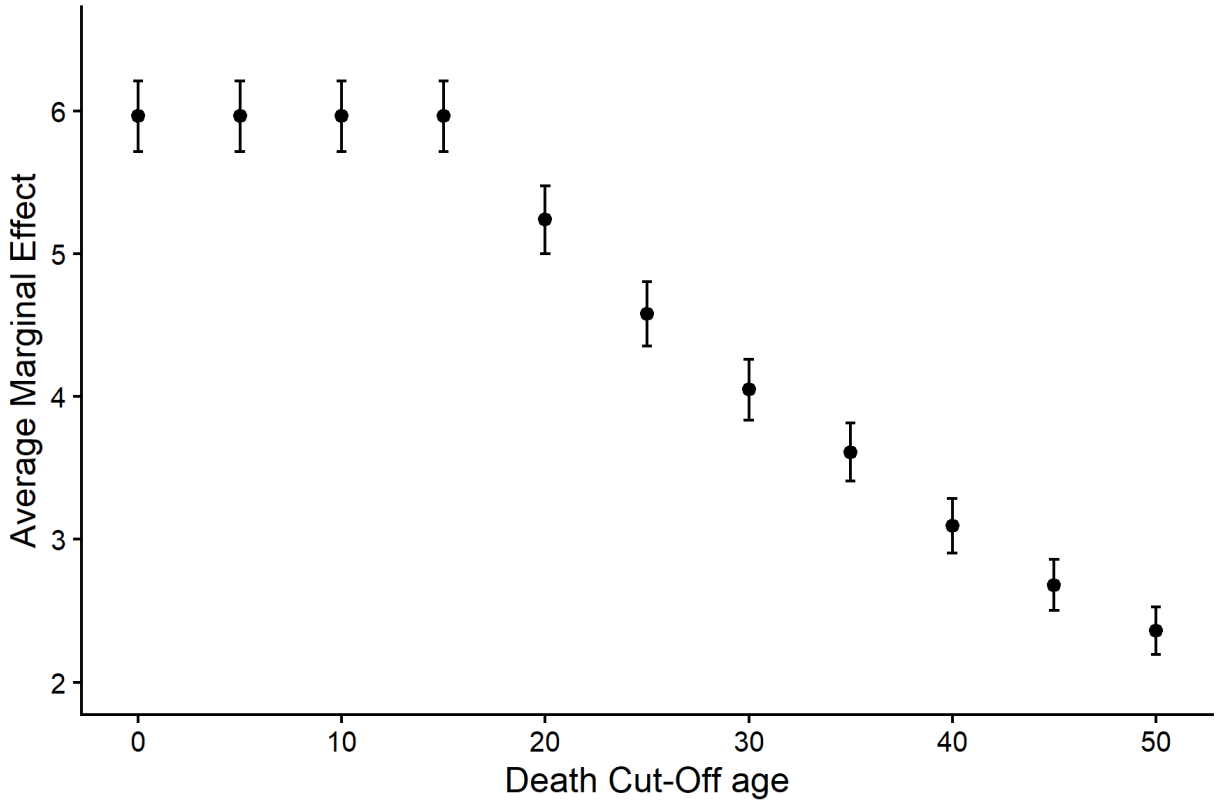


Figure S4 AMEs of migration on age at death across alternate age cutoffs (Model A3).

We find a similar pattern when looking at differences by destination country in Model B3 (see Figure S5): the extent of the migrant mortality advantage is constant through a cut-off of age 15, though this advantage decreases as the cut-off age increases. The AMEs are positive and statistically significant for most estimates except for cut-offs over age 25 in South Africa. The former point indicates that our findings are generally robust across age cut-off, though the estimated extent of the migrant mortality advantage varies. The latter point indicates that when the age cut-off is above 25, we do not find statistically significant differences in age at death between migrants to South Africa and non-migrants.

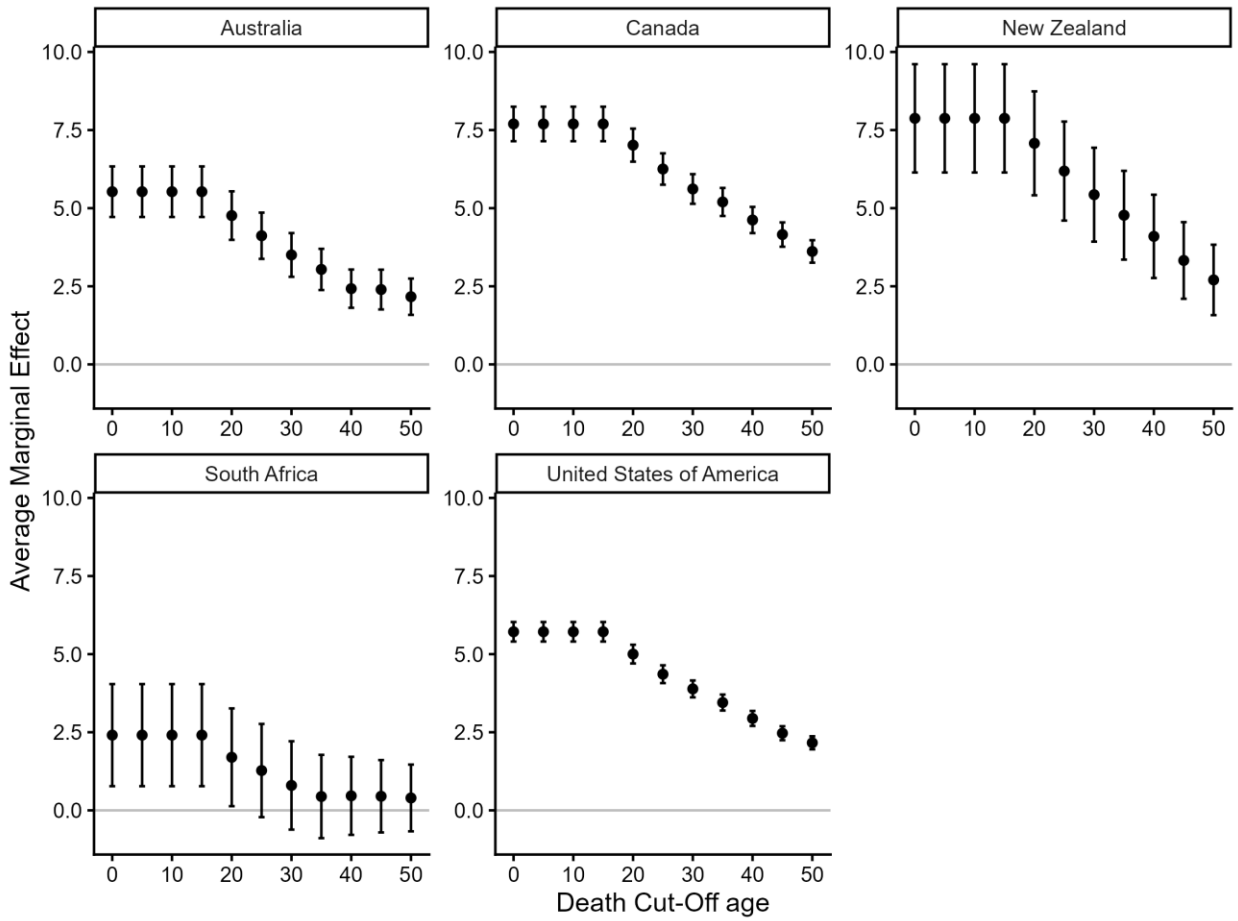


Figure S5 AMEs of migration on age at death across alternate age cutoffs and disaggregated by destination country (Model B3).

While survival models are common in mortality research, we did not present them in our main analysis due to concerns about how well our data fit the assumptions of hazard models. In addition, since hazard models are designed for incomplete data with right censoring issues, they were not necessary for our main analyses as all observations within our dataset are deceased individuals who have completed the life course. For transparency, we present regression estimates from Cox mixed effects models in Table S6. For the model to converge, we excluded the sibling random effects, and thus we caution over-interpretation of the results. Findings confirm that migrants experience significantly lower mortality compared to non-migrants.

Table S6 Cox mixed effects model

Dependent variable: Age at Death	Model A: Migrant HR (SE)	Model B: Country of Death HR (SE)
Migrant (*reference: Non-Migrant)*	0.743***	
	(0.029)	
Destination country *(reference: United Kingdom/Ireland)*		
Canada		0.820
		(0.109)
South Africa		1.131
		(0.138)
Australia		0.527
		(0.304)
New Zealand		1.100
		(0.125)
United States of America		0.739***
		(0.030)
Birth cohort (*reference: 1735)*		
1745	1.017	1.017
	(0.031)	(0.031)
1755	1.004	1.004
	(0.030)	(0.030)
1765	0.942*	0.942*
	(0.027)	(0.027)
1775	0.940*	0.941*
	(0.026)	(0.026)
1785	1.024	1.025
	(0.027)	(0.027)
1795	1.038	1.039
	(0.027)	(0.027)
1805	1.116***	1.117***
	(0.029)	(0.029)

Dependent variable: Age at Death	Model A: Migrant HR (SE)	Model B: Country of Death HR (SE)
1815	1.172***	1.172***
	(0.030)	(0.030)
1825	1.304***	1.305***
	(0.034)	(0.034)
1835	1.329***	1.329***
	(0.035)	(0.035)
1845	1.267***	1.267***
	(0.033)	(0.033)
1855	1.165***	1.166***
	(0.031)	(0.031)
1865	1.037	1.037
	(0.028)	(0.028)
1875	0.905***	0.905***
	(0.025)	(0.025)
1885	0.754***	0.754***
	(0.022)	(0.022)
Male *(reference: female)*	1.119***	1.120***
	(0.007)	(0.007)
Number of siblings *(reference: none)*		
1	1.045***	1.046***
	(0.010)	(0.010)
2	1.068***	1.068***
	(0.012)	(0.012)
3-5	1.066***	1.067***
	(0.010)	(0.010)
6+	1.012	1.010
	(0.014)	(0.014)
Migrant X Birth cohort interaction *(reference: 1735)*		
1745	1.056	
	(0.059)	
1755	1.085	
	(0.060)	
1765	1.087	
	(0.060)	
1775	1.052	
	(0.056)	
1785	1.019	

Dependent variable: Age at Death	Model A: Migrant HR (SE)	Model B: Country of Death HR (SE)
	(0.050)	
1795	1.013	
	(0.047)	
1805	1.015	
	(0.046)	
1815	0.977	
	(0.043)	
1825	0.880**	
	(0.039)	
1835	0.919†	
	(0.041)	
1845	0.942	
	(0.043)	
1855	1.034	
	(0.049)	
1865	1.080	
	(0.052)	
1875	1.096†	
	(0.054)	
1885	1.263***	
	(0.065)	
Destination country X Birth cohort interaction *(reference: U.K./Ireland & 1735)*		
Canada:1745		0.970
		(0.164)
South Africa:1745		2.685
		(2.705)
Australia:1745		2.070
		(1.279)
New Zealand:1745		0.673
		(0.396)
United States of America:1745		1.051
		(0.061)
Canada:1755		0.772
		(0.125)
South Africa:1755		0.477
		(0.221)
Australia:1755		1.697

Dependent variable: Age at Death	Model A: Migrant HR (SE)	Model B: Country of Death HR (SE)
		(1.005)
United States of America:1755		1.142*
		(0.067)
Canada:1765		0.728*
		(0.115)
South Africa:1765		0.899
		(0.183)
Australia:1765		1.633
		(0.955)
New Zealand: 1765		0.862
		(0.868)
United States of America:1765		1.162*
		(0.071)
Canada:1775		0.791
		(0.118)
South Africa:1775		0.799
		(0.143)
Australia:1775		1.453
		(0.848)
New Zealand:1775		0.385**
		(0.116)
United States of America:1775		1.186**
		(0.074)
Canada:1785		0.777†
		(0.113)
South Africa:1785		0.815
		(0.136)
Australia:1785		1.458
		(0.847)
New Zealand:1785		0.560**
		(0.100)
United States of America:1785		1.109†
		(0.063)
Canada:1795		0.802
		(0.113)
South Africa:1795		0.796
		(0.121)

Dependent variable: Age at Death	Model A: Migrant HR (SE)	Model B: Country of Death HR (SE)
Australia:1795		1.569
		(0.909)
New Zealand:1795		0.679**
		(0.099)
United States of America:1795		1.023
		(0.055)
Canada:1805		0.809
		(0.113)
South Africa:1805		0.758*
		(0.107)
Australia:1805		1.540
		(0.892)
New Zealand:1805		0.661**
		(0.088)
United States of America:1805		1.018
		(0.051)
Canada:1815		0.762†
		(0.106)
South Africa:1815		0.777
		(0.120)
Australia:1815		1.555
		(0.900)
New Zealand:1815		0.602***
		(0.077)
United States of America:1815		0.965
		(0.046)
Canada:1825		0.719*
		(0.101)
South Africa:1825		0.704†
		(0.130)
Australia:1825		1.323
		(0.765)
New Zealand:1825		0.593***
		(0.074)
United States of America:1825		0.866**
		(0.041)
Canada:1835		0.787†

Dependent variable: Age at Death	Model A: Migrant HR (SE)	Model B: Country of Death HR (SE)
		(0.113)
South Africa:1835		0.858
		(0.149)
Australia:1835		1.335
		(0.773)
New Zealand:1835		0.611***
		(0.075)
United States of America:1835		0.913†
		(0.044)
Canada:1845		0.878
		(0.129)
South Africa:1845		0.692*
		(0.117)
Australia:1845		1.327
		(0.768)
New Zealand:1845		0.578***
		(0.073)
United States of America:1845		0.959
		(0.047)
Canada:1855		0.959
		(0.142)
South Africa:1855		0.856
		(0.145)
Australia:1855		1.415
		(0.820)
New Zealand:1855		0.621***
		(0.080)
United States of America:1855		1.068
		(0.056)
Canada:1865		0.879
		(0.129)
South Africa:1865		0.736†
		(0.117)
Australia:1865		1.591
		(0.924)
New Zealand:1865		0.669**
		(0.089)

Dependent variable: Age at Death	Model A: Migrant HR (SE)	Model B: Country of Death HR (SE)
United States of America:1865		1.129*
		(0.061)
Canada:1875		0.896
		(0.129)
South Africa:1875		0.697*
		(0.107)
Australia:1875		1.568
		(0.912)
New Zealand:1875		0.672**
		(0.099)
United States of America:1875		1.178**
		(0.066)
Canada:1885		1.088
		(0.159)
Australia:1885		1.816
		(1.056)
United States of America:1885		1.246***
		(0.075)
Observations	97,668	97,668

Note: "HR" = hazard ratio, "SE" = standard error. $^{\dagger}p<0.1$, *$p<0.05$, **$p < 0.01$, ***$p<0.001$.

C. References

Arel-Bundock, V., Enevoldsen, N., & Yetman, C. (2018). countrycode: An R package to convert country names and country codes. *Journal of Open Source Software*, *3*(28), 848. https://doi.org/10.21105/joss.00848

Guillot, M., Khlat, M., Elo, I., Solignac, M., & Wallace, M. (2018). Understanding age variations in the migrant mortality advantage: An international comparative perspective. *PLOS ONE*, *13*(6), e0199669. https://doi.org/10.1371/journal.pone.0199669