Stockholm Research Reports in Demography | no 2020:2

Age Variations and Over-Coverage:

Is the Migrant Mortality Advantage Merely a Data Artefact?

Matthew Wallace and Ben Wilson

ISSN 2002-617X | Department of Sociology

Stockholm University
Demography Unit

Age Variations and Over-Coverage: Is the Migrant Mortality Advantage Merely a Data Artefact?

Matthew Wallace [a] | Ben Wilson* [a,b]

a: Stockholm University, b: London School of Economics

* Corresponding author: matthew.wallace@sociology.su.se

Abstract

The migrant mortality advantage has been extensively observed in the literature, but its veracity continues to be debated. In particular, concerns persist that the advantage is merely an artefact, generated by issues identifying, and capturing the mobility of, foreign-born populations. Here, we study this by working at the intersection of two recent developments; one uncovering age variation in the advantage (specifically a deep U-shape of advantage at peak migration ages) and another uncovering high levels of over-coverage – the main source of data artefact – at the same ages. Using survival analysis on Swedish population registers from 2010 to 12, we aim to determine whether age variation in the advantage persists after correcting over-coverage and, ultimately, whether the advantage is an artefact. We find the U-shaped age profile of advantage in Sweden and, crucially, discover that peak migration age differentials are not caused by over-coverage. We demonstrate the advantage to be real and, in all likelihood, generated by selection effects.

Keywords: *International immigration, Health, Mortality, Over-coverage, Data artefacts, Unregistered emigration, Sweden*

Stockholm Research Reports in Demography 2020:2
ISSN 2002-617X

Introduction

Over the past few decades, many studies have documented low all-cause mortality rates among the foreign-born, as compared with native-born populations in high-income receiving countries (Aldridge et al., 2018). This ‘migrant mortality advantage’ has attracted much attention in recent years due to the growing share, diversification and ageing of foreign-born populations in high-income destinations. The health of immigrants has profound implications for overall population health and present substantial challenges for national health and welfare systems to adapt to better understand and manage immigrant health (Rechel et al., 2011, Guillot et al., 2018).

This migrant mortality advantage exhibits several common features that transcend the country context. It is often found, paradoxically, among migrant populations that have a lower average socioeconomic position than the native-born population in the country that they move to (Ruiz et al., 2013, Deboosere and Gadeyne, 2005). Additionally, studies suggest that the advantage is strongest upon arrival and wears off over time spent in the host country (Hammar et al., 2002, Harding, 2003, Harding, 2004, Hajat et al., 2010, Vandenheede et al., 2015, Syse et al., 2016, Syse et al., 2018, Wallace et al., 2019) and appears to be weaker, or absent, among immigrants arriving as children (Guillot et al., 2018, Juárez et al., 2018, Mehta et al., 2019). These common threads have often been used to try to advance our understanding of what causes the migrant mortality advantage, but its explanations remain fiercely debated (Guillot et al., 2018).

Until recently, little was known about how the migrant mortality advantage varies over *age* – the most fundamental of demographic characteristics. Such a gap has limited our understanding of *how* and *why* mortality among immigrants differs from native-born. This is because any estimates produced net of age (rather than by age) assume that the size of the mortality differentials, and the causes working to generate such differentials, remain proportional over the entire life course (Guillot et al., 2018). Moreover, this means that policy recommendations regarding immigrant health and mortality are based upon age-adjusted mortality trends that can mask variation in the presence, scale and even direction of migrant mortality advantage over age.

Having recognised this issue, researchers have recently developed a theoretical framework that predicts how the advantage would vary over age (Guillot et al., 2018). Regarding the empirical evidence, systematic age variation has been documented in the United States, United Kingdom and France (Guillot et al., 2018). To elaborate, a relative excess mortality among immigrants at infant and child ages was swiftly followed by a pronounced 'U-shape' of advantage among immigrants at young adult ages, the tail of which gradually became smaller with age (Guillot et al., 2018). The authors surmised that these patterns were most consistent with in-selection effects. This implies that people who migrate are healthier than people who stay in the origin country because they self-select directly upon their good health and indirectly on measured and unmeasured factors associated with health (Riosmena et al., 2013). The age patterns that have been documented were also somewhat consistent with cultural explanations, which predict that some origin groups practice healthier behaviours than the destination average due to prevailing cultural norms derived from their country of origin (Abraido-Lanza et al., 2016). However, the analysis of age profiles of advantage was inconsistent with out-selection effects (or the salmon bias), which predicts that sick migrants will return to their country of origin (Wallace and Kulu, 2018).

Crucially, one explanation that the researchers were unable to dismiss when testing their new framework was data artefacts; the belief that mortality differences between foreign-born and native-born are generated by issues stemming from the (in)ability to identify, and capture the movements of, immigrant populations (Wallace, 2019). Although they offered a meticulous overview of the main issues associated with this explanation (principally, the under-coverage of deaths and the over-coverage of the population), they could not theorise with certainty how data artefacts might cause the advantage to vary across age (Guillot et al., 2018). Thus, while the article was crucial in extending the framework around the advantage and providing new empirical evidence concerning age, it could not dispel concerns as to whether the advantage is real or merely a data artefact (Wallace, 2019). This is concerning because if it *is* a data artefact, then this undermines most of what we think that we know about the health and mortality of immigrants.

This concern becomes even more pertinent in light of more recent research on population over-coverage in Sweden (Monti et al., 2019), the principal source of data artefact in foreign-born versus native-born mortality differentials. Testing three approaches for correcting this problem, the authors documented the largest levels of over-coverage (i.e. the greatest number of people considered to no longer be resident in Sweden), and thus the largest bias, in core demographic estimates for immigrants at peak migration ages (15-50). These were the *exact* ages at which the advantage was found to be most pronounced (Guillot et al., 2018). This calls into question whether this U-shape of advantage is attributable to data artefacts, as opposed to substantive causes.

In this paper, we address this concern with two aims. Initially, we examine whether the same U-shape pattern of migrant mortality advantage can be found among immigrants in Sweden, as has been documented in the United States, United Kingdom and France. Then, our main aim is to determine whether correcting for the main data artefact, over-coverage, can account for U-shaped advantage at young adult ages. In doing so, we set out to establish whether or not the advantage is wholly a data artefact. To achieve our aims, we use event history models to obtain the foreign-born versus native-born mortality differentials over age and sex (uncorrected and corrected for over-coverage). We do this for Sweden in the period 2010-2012, which allows us to compare our findings with previous research on the three high-income contexts mentioned above.

Sweden represents an ideal context in which to conduct this research. Sweden's foreign-born population has grown and diversified rapidly over the past several decades, rising from 10.6% in 1995 to 17.6% in 2017, the highest level of any country in the EU with a population above one million (Agafiței and Ivan, 2017). Additionally, at least as compared with the other two countries from Europe included in the prior research (Guillot et al., 2018), Sweden's migration history is unique, characterised by a much larger proportion of refugee migration, rather than migration from the former colonies (as is the case for France and the United Kingdom). Our analysis is also based on registers of the entire population (over 9 million people). This permits a thorough and detailed examination of how migrant mortality varies over age. Furthermore, we are able to build upon an existing body of research on over-coverage in Sweden, as several approaches for correcting for population over-coverage already exist (Weitoft et al., 1999, Aradhya et al., 2017, Monti et al., 2019). The true value of our findings is reflected in the ability to contribute to ongoing debate on the causes of the migrant mortality advantage and whether foreign-born versus native-born mortality differences are genuine, or merely an artefact of the data.

Background

Over-coverage refers to a situation in which people continue to be recorded as resident in a population, even though they have left the country (Monti et al., 2019). This almost always takes place because we lack any record or proof of their departure in national data sources. If such cases are not identified in studies of mortality, these individuals become statistically immortal as they continue to age in data sources, despite being unable to die in the host country, as their (eventual) death will be registered elsewhere (Kibele et al., 2008). This generates a downward bias in death rates because we overestimate risk-time and underestimate deaths. Hypothetically, if the level of over-coverage is large enough, then the advantage could simply be a consequence of the inaccurate presence of some immigrants in the analysis. While native-born individuals are also susceptible to over-coverage, immigrant populations are expected to be disproportionately affected due to their increased mobility, including recent diversification in forms of migration (e.g. repeat, onwards and circular) (Aradhya et al., 2017). Here, we provide an overview of previous studies that have explicitly studied over-coverage in migrant death rates, starting with Sweden. We focus on data, methods and the potential impact of over-coverage.

In the most recent Swedish research, the authors tested three approaches for correcting over-coverage. The first is the zero-income approach, which is based on the logic that those without economic activity in a welfare state such as Sweden in a given year, or years, can be assumed to have left the country (Aradhya et al., 2017). The other two approaches are the cross-sectional and longitudinal register trace approaches, which assume that people have emigrated if they are not correctly registered and they fail to show any 'traces' of activity (e.g. internal migration or enrolment in education) across different registers, the former at a specific point in time and the latter over time. The authors estimated age-specific foreign-born death rates, corrected and uncorrected for over-coverage. For each approach, they documented the largest impact of over-coverage at peak migration ages. The zero-income approach offered the most conservative correction (with immigrant mortality rates between 1.4 to 2.5 times higher, compared to between 1.2 to 1.6 times higher for the other two approaches). Between ages 40 and 75 the three approaches produced indistinguishable ratios. These results show the extent to which foreign-born death rates are downwardly biased by over-coverage. However, they say nothing about the extent to which over-coverage can explain the migrant mortality advantage. As detailed below (in the next section on methods), we draw upon these approaches in our own analysis.

To the best of our knowledge, the only study to examine the role of over-coverage in explaining mortality differences between immigrants and native-born in Sweden was conducted around two decades ago (Weitoft et al., 1999). Using an approach based upon income and the receipt of social benefits, the authors corrected mortality rates among immigrants aged 20-64 for the period 1987-1994. Initially, a migrant mortality advantage was found among immigrants from Southern Europe, ex-Yugoslavia and Turkey, Latin America, Africa and Asia, and the rest of Europe, Canada, US and Oceania. After excluding people who did not fulfil the income or social benefits criteria, the advantage decreased, but persisted in all these groups. However, further restricting based upon income only, the migrant mortality advantage was lost among most of these groups. The authors concluded that there was some underestimation of mortality, but its extent was difficult to assess (Weitoft et al., 1999). In a direct response to this study, an article from Germany calculated mortality among the foreign-born population aged 15+ and 15-64 in Germany using the German Socio-Economic Panel, based upon the logic that that cohort studies are less vulnerable to over-coverage than register data. They found advantages among immigrants of a similar size to those found in German register studies (Razum et al., 1998).

In a study of England and Wales, one study used life event indicators from civil registers (birth of children, deaths of respondents, migration, and other life events) and presence at decennial censuses from a linked, longitudinal data source to identify the over-covered. The authors then tested several different scenarios (examining the effect of people exiting 2, 4, and 7 years after their final census, assuming that no more life events were registered after that date). They found that over-coverage could explain some, but not all, of the migrant mortality advantage (Wallace and Kulu, 2014). In France, one paper (Khlat and Courbage, 1996) used an indirect approach initially developed in an earlier article (Courbage and Fargues, 1979) to show a 23% over-coverage of Moroccan men from 1980 to 1990. Nevertheless, adjusted life tables still gave a 2.4-year mortality advantage for Moroccan men compared to the native-born. In Belgium, a study of migrant mortality performed an over-coverage check by reducing the risk time of foreign-born by 194 days (the amount of days later, on average, that it took for those who did not record their emigration to be 'administratively removed' from the risk set). A large advantage persisted and the author argued over-coverage could not explain their results (Anson, 2004).

Rather than attempting to explicitly correct over-coverage, like Razum et al. (2000), several other studies have used data sources they reason are more suited to capturing the immigrations and emigrations of foreign-born individuals. One study has used the Federal German Statutory Pension Scheme (DRV) to estimate the mortality of pensioners, arguing that the data is more accurate because the survival of pensioners has to be tracked carefully to give correct pension payments (Kibele et al., 2008). Pensioners living abroad must provide annual confirmation of being alive to receive a pension, reducing the likelihood that pensioners continue to be included in this database after they have died. Among male immigrants aged 65+, large advantages were found in official data but not in the pension data. Further, growing disparities were documented with age, consistent with the accumulation of people who have become statistically immortal. Similarly, another study used the same German pension data and, additionally, the Central Register of Foreigners (*AZR*) (argued to be more accurate due to its explicit focus on the foreign-born population) to uncover disparities between a migrant mortality advantage in the official data and the lack of advantage when combining the pensions data with the AZR (Kohls, 2010).

Along similar lines, researchers in the United States have combined pension data with data from annual questionnaires about beneficiaries living in the United States and abroad. The authors calculated age-adjusted and age-specific death ratios (65-90+) comparing foreign-born and native-born subpopulations, both including and excluding emigration, as recorded by the questionnaires. In contrast with the German studies, however, the authors continued to observe large mortality advantages among Hispanic immigrants that could not be explained by over-coverage, even after factoring in that leavers were negatively selected on health (Turra and Elo, 2008).

In summary, researchers have used two different approaches to address over-coverage in studies of the migrant mortality advantage. The first approach explicitly corrects for over-coverage by using available data to attempt to identify unregistered emigrants. The second approach makes no such correction but uses alternative data sources that are assumed to more accurately capture the residency status of immigrant populations. In both cases, we have reasons to be cautious about the conclusions that have been reached. With the first approach, studies have used indirect information to identify the over-covered, which could lead to the wrongful *inclusion* of leavers and *exclusion* of stayers in the risk population. With the second approach, alternative data sources may be less susceptible to over-coverage, but they have so far been unable to study the migrant mortality advantage in enough detail to either examine its determinants or to isolate its likely extent. This is because alternative data sources are either underpowered or do not cover the ages essential for estimating the advantage. For example, the absence of an advantage at pensionable ages could be explained by the fact selection effects among immigrants have worn off and they have adapted to the host society. Thus, it is difficult to examine whether over-coverage can explain away the advantage without focussing on young adult ages, when immigrants have low average durations of stay. This is especially important given the U-shaped age-profile of the advantage that has been identified in recent research. In articles that *do* explicitly correct over-coverage, the bias rarely explains mortality differences between immigrants and native-born, but none of these studies disaggregate by age. Rather, they estimate the average impact of over-coverage bias across all age groups. Given this gap, we set out to use high-quality register data in order to establish the effect of over-coverage over ages, including the peak migration ages when the mortality advantage has been found to be the largest.

Data and method

Our study uses the collections of Swedish register data organised at Stockholm University that are accessible for research under ethical approval from the regional ethics board in Stockholm. We use longitudinal individual-level data from several administrative datasets. Available data cover the entire resident population of Sweden annually from 1961 until 2017. However, we focus on period 2010-12, which enables us to compare our estimates of the migrant mortality advantage, by age, with those from the most recent research on other high income destinations (Guillot et al., 2018). We take our information from four registers: (1) the total population register (Mikrodata för Registret över totalbefolkningen; RTB), which acts as the base register for the production of statistics on the size and the composition of Sweden's population; (2) the register for labour market studies and health insurance (Longitudinell integrationsdatabas för sjukförsäkrings- och arbetsmarknadsstudier; LISA), which contains annual data on education, employment and health and social benefits; (3) the migration register, which contains detailed data on registered immigrations and emigrations of Sweden's resident population; and (4) the death register, which is of high quality and covers all deaths in Sweden. Despite prior evidence of over-coverage in Sweden (Monti et al., 2019), the Swedish population data is of high quality, in particular because residents of Sweden are obligated to register their address in order to work in the country and have the opportunity to access all of the benefits and social services that are available (e.g. to access health care, to receive welfare benefits, or for their children to attend school).

Measuring population over-coverage

We follow studies that have calculated over-coverage based upon the absence of labour market activity and social welfare receipt (Weitoft et al., 1999, Aradhya et al., 2017). These studies have considered people to be living in Sweden if they have received an income in the year(s) prior to, or during, the study period. The assumption is that one must be resident in Sweden to receive an income from work or social benefits. This has been referred to as the zero-income approach and it has recently been shown to provide the most conservative estimate of over-coverage available (Monti et al., 2019). For the purpose of trying to explain away differences between foreign-born and native-born, it makes sense to use the most conservative approach. This is because, if mortality differences persist after the correction, it is very unlikely that the age variation in the migrant mortality advantage is generated solely by over-coverage. Further, the zero-income approach is attractive because the receipt of income (including social benefits) is most likely at the ages of interest (i.e. the U-shape of advantage among young adult foreign-born).

Specifically, our analysis assumes that people are over-covered (i.e. have left Sweden) if one is present in the total population register at the end of 2009 and aged 16 or above, but does not receive income for *at least* two consecutive years from 2010 to 2012 from employment, social benefits, sick-pay, or pensions, and have no recorded emigration date between 1st January 2010 and 31st December 2012. Conversely, individuals are assumed to be resident if they are: (i) younger than 16, (ii) older than age 16 and receiving *at least two* consecutive years of income, and/or (iii) have a death registered between 1st January 2010 and 31st December 2012. These criteria are illustrated in Figure 1 (for native-born) and Figure 2 (for foreign-born), which also provides detailed information on the absolute and relative proportions of over-coverage in each subgroup. We use two years of consecutive income in line with recommendations from the researchers who have used the zero-income approach. They suggest that a single year of zero-income is more likely to indicate an aberration that is unrelated to presence or not in the country (Aradhya et al., 2017). Nonetheless, we conducted additional sensitivity analysis conditional upon having a source of income in all three years of the study period (see Figure A3 in online supplementary materials). We discuss this additional set of analyses in the final section of the paper.

Statistical methods and study parameters

We estimate uncorrected and corrected age-specific mortality hazard ratios (HRs) for all-cause mortality using Cox Proportional Hazards (PH) models, comparing all foreign-born women and men versus native-born women and men. For individuals, entry into the study is at the beginning of 2010 (1st January) and we follow the resident population for 3-years until the end of 2012 (31st December). Exit from the population-at-risk takes place when people die, emigrate (in the cases where the emigrations are registered) or reach the end of the three-year risk window alive. Age acts as the 'clock' in our event history models, specified using the age at entry and age at exit of subjects when setting the data. We only adjust for a single covariate: our foreign-born versus native-born dummy, with the latter acting as the reference group in all models. We stratify our models by sex and into 5-year age groups, ranging from 5-9 to 85+. We begin with 5-9, rather than 0-4, so as to maintain consistency with previous research that documented age variation in the advantage (Guillot et al., 2018). The choice of beginning at age 5 has been justified by the fact that there are very few foreign-born present in the age group 0-4.

Initially, we estimated uncorrected HRs. Then we remove the over-covered (we simply drop those with a '1' in our over-coverage identifier variable) and refit our models without these individuals to obtain HRs 'corrected' for over-coverage. As such, we identify over-coverage using the most conservative method and then simply remove the over-covered people from our analysis (rather than, say, projecting potential exit scenarios, as has been the case in previous studies (Wallace and Kulu, 2014)). Consequently, we interpret our uncorrected HRs as lower bound estimates for the relative mortality differentials and our corrected HRs as upper bound estimates. In effect, we propose that the true foreign-born versus native-born differential lies somewhere between the two. We estimate all-cause mortality for the entire resident population of Sweden, not a sample, but include 95% confidence intervals as some measure of population variability.

Results

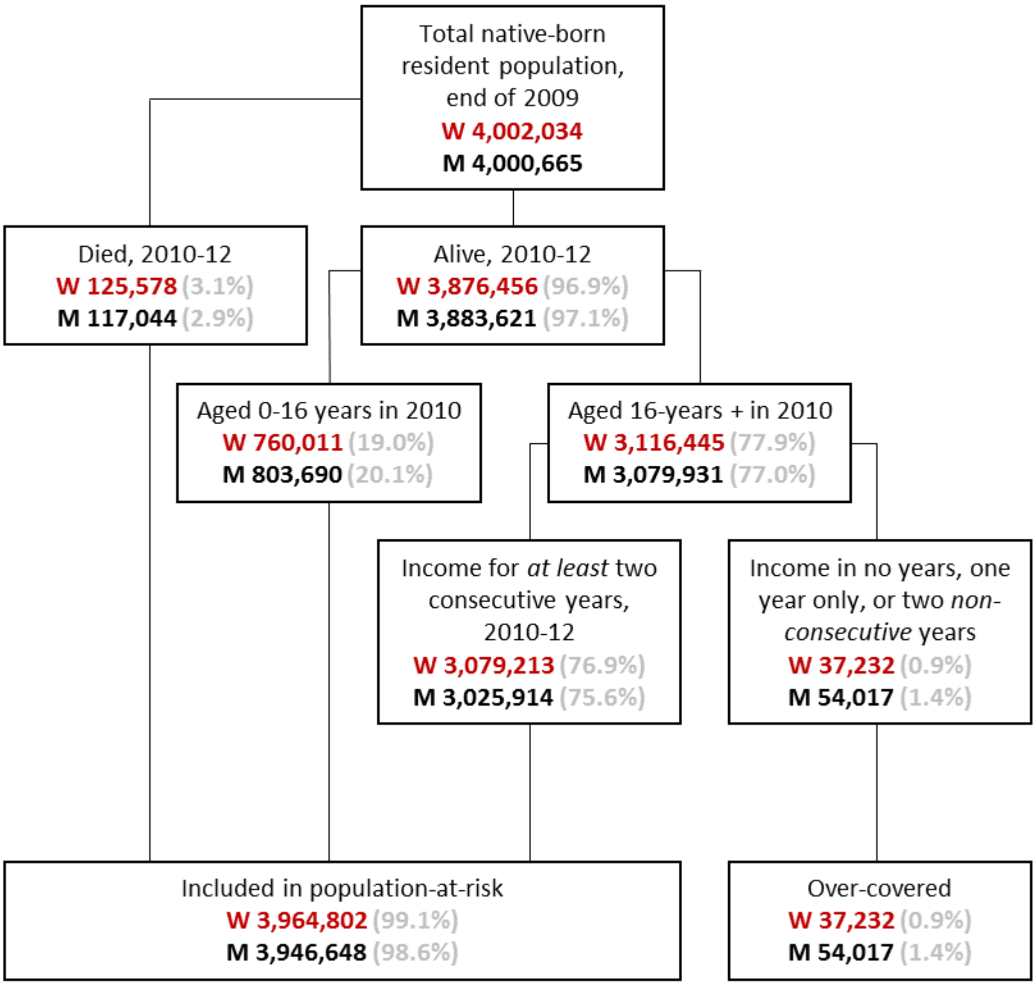

Figure 1. Flow chart documenting number of resident and over-covered native-born.
Notes: (1) % always relates to "total resident population end of 2009" (2) W=women, M=men.

Figures 1 and 2 provide flow charts for native-born and foreign-born individuals, documenting the absolute and relative proportions of individuals included in our initial analysis, as well as those excluded due to over-coverage. These figures highlight three general trends. The absolute numbers of over-covered people are quite similar when comparing the foreign-born and native-born (e.g. 54,059 foreign-born men versus 54,017 native-born men). Accordingly, the relative proportions of over-coverage are higher among the foreign-born as compared to native-born (e.g. 8.3% among foreign-born men compared to 1.4% among native-born men). Additionally, irrespective of their nativity, the absolute and relative proportions of over-coverage are always lower among women (e.g. 43,243 among foreign-born women as compared to 54, 059 among foreign-born men, which translates to relative levels of 6.3% in the former and 8.3% in the latter).

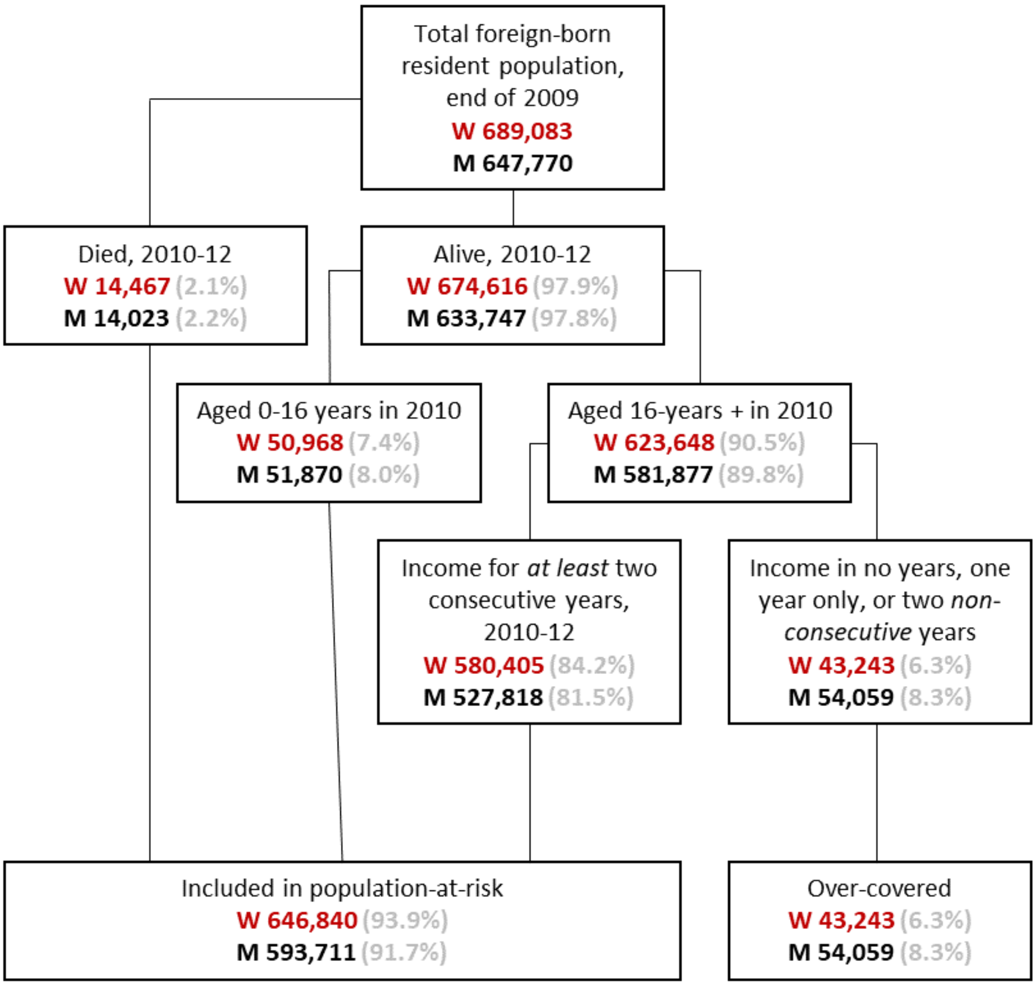

Figure 2. Flow chart documenting number of resident and over-covered foreign-born.
Notes: (1) % always relates to “total resident population end of 2009” (2) W=women, M=men.

Figure 3 illustrates the population age-structure for native-born and foreign-born, highlighting the proportion in each subgroup who are over-covered (starting from ages 15-19). This serves to emphasise the differences between the age-structure of the two groups alongside age-specific variation in levels of over-coverage. As expected, there is a younger age-structure for foreign-born, reflective of the fact that they are a younger, more mobile population. Regarding over-coverage, the general pattern for foreign-born is that levels are highest at peak migration ages (15-39) especially in the age groups 20-24 and 25-29 (in which the share is near double that of the average levels shown in Figure 2) and then fall over age. These patterns are very consistent with previous work testing this approach (Monti et al., 2019) and suggest that the greatest bias will be induced in the foreign-born versus native-born mortality differentials at peak migration ages.

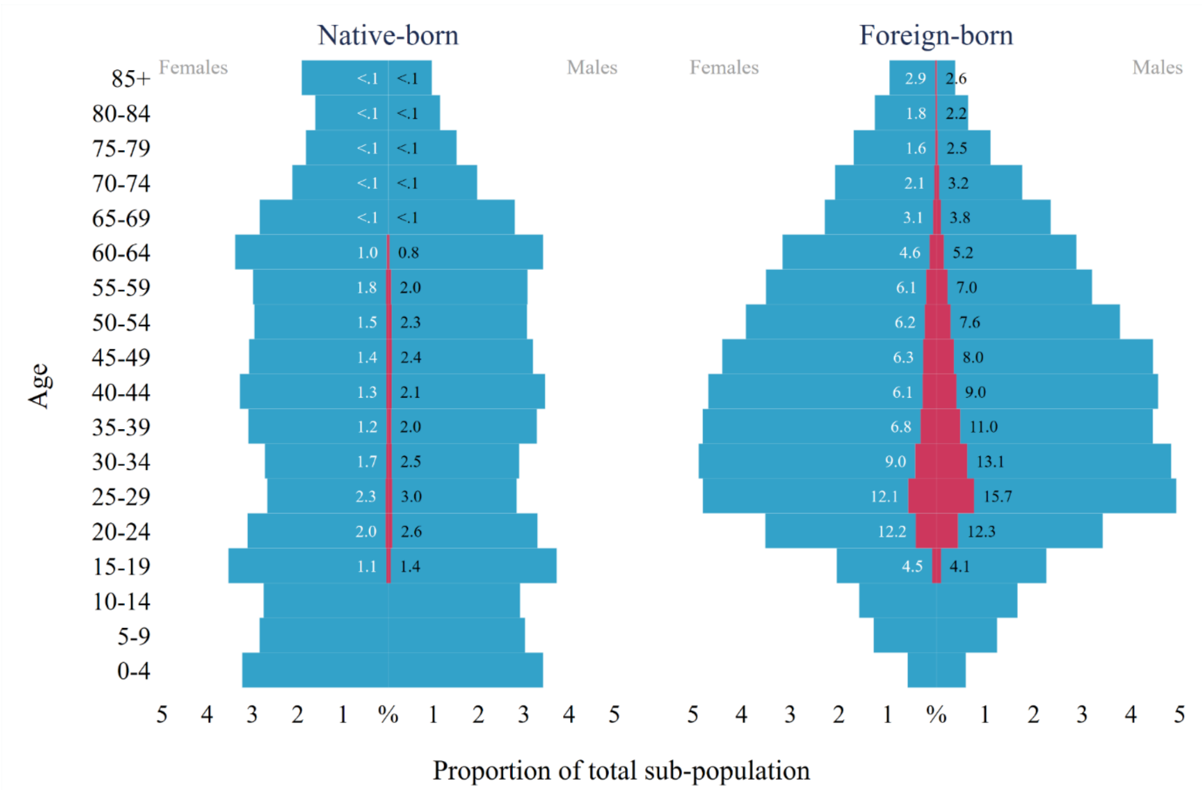

Figure 3. Age-sex structures and levels of over-coverage for native-born and foreign-born. Notes: (1) values next to red/darker bars detail the proportion over-covered in that age group.

Figure 4 relates directly to the first aim of our paper – to determine whether the same U-shape of migrant mortality advantage observed in the United States, United Kingdom and France by

Guillot et al. (2018) can also be observed in Sweden. Specifically, Figure 4 shows age variation in the migrant mortality advantage among foreign-born living in Sweden in period 2010-12, plotted with permissions alongside the age variation documented in the three aforementioned countries (Guillot et al., 2018). For women and men, we document the same age profile for Sweden. In short, we document excess all-cause mortality for the youngest age group (5-9), followed by a deep 'U-shaped' advantage starting at ages 15-19 and ending at 35-39, then a gradual wearing off of the advantage among women and the reversal of the advantage among men.

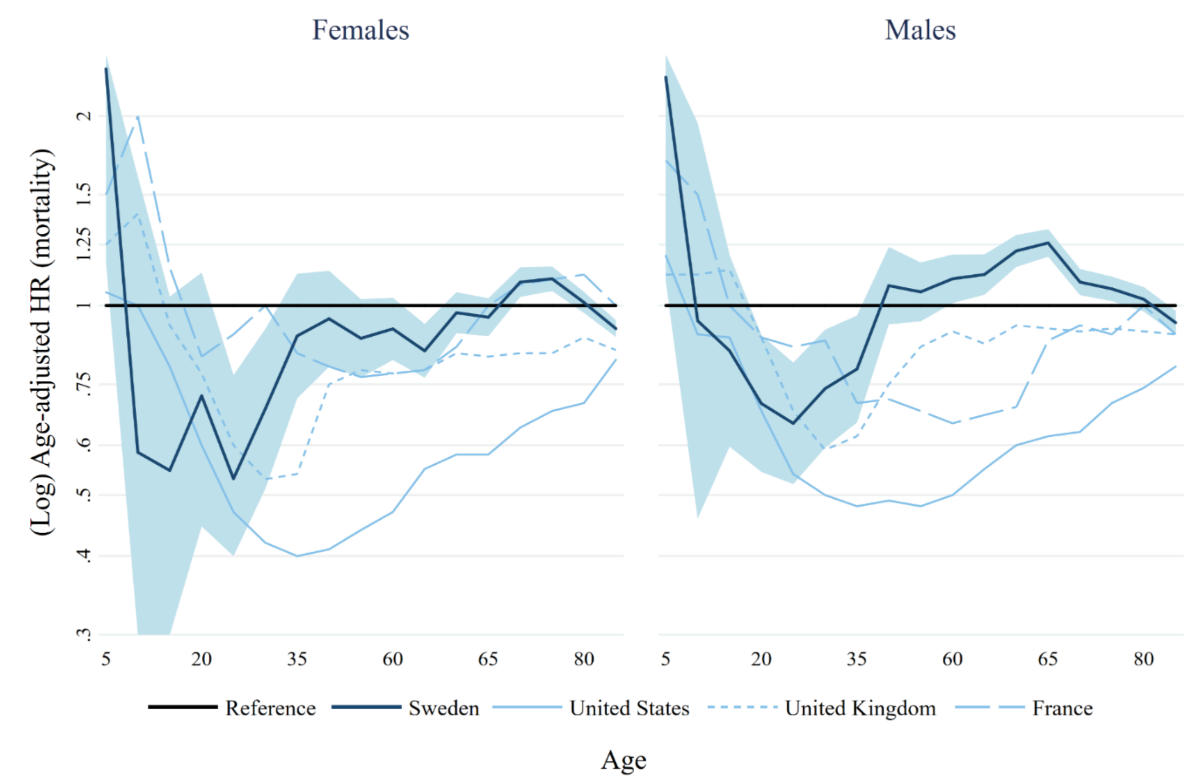

Figure 4. Age-specific relative hazard ratios for mortality, foreign-born versus native-born in Sweden and three other countries, 2010-12.
Notes: permissions to use estimates for the United States, United Kingdom and France provided by authors of the paper "Understanding age variations in the migrant mortality advantage: An international comparative perspective" (https://doi.org/10.1371/journal.pone.0199669) in line with the journal's requirements.

Nevertheless, there are some differences between countries, even if the overall shape of the age variation is similar. First, we note the different scale of advantage between the foreign-born and native-born across the four countries. For example, at its peak, a relative advantage of 0.4

among foreign-born women and 0.5 among foreign-born men living in the United States is considerably larger than it is in other countries. Second, while mortality differentials have fully attenuated, and even reversed, among foreign-born living in the United Kingdom, France and Sweden, a large advantage remains among foreign-born in the United States, even at the oldest ages. Third, while the U-shape in Sweden falls near perfectly within peak migration ages (as it also does in the United Kingdom), the position of the U-shape among the foreign-born living in France is at later ages for both women and men, starting at 35-39 and finishing at 60-64. Fourth, the reversal in the migrant mortality advantage that we document among foreign-born men living in Sweden represents a considerable departure from the patterns of the other three countries.

This excess male hump in Sweden from 35 to 74 merited closer examination. Consequently, we conducted further analysis of these ages (available in the appendix). We first examined the composition of foreign-born by country of origin in each age group (Tables A1 and A2). For both men and women, the ages of peak migration are characterised by higher proportions of migrants from Asia and particularly the Middle East. Between age groups 15-19 and 35-39, these two regions of origin consistently accounted for approximately 40% of all foreign-born individuals. From ages 40-44, however, we begin to see a shift in the declining share of these two origin groups, combined with a growing share of foreign-born from other Nordic countries (particularly Finland). By the age group 50-54, this group alone accounted for one quarter of all foreign-born, a proportion that rises to approximately 40% by ages 60-64 (over 50% among women).

Thus, one explanation for excess foreign-born male mortality at older working ages in Sweden may relate to the specific mortality pattern of immigrants from other Nordic countries. So, we plotted the age variation in migrant mortality of the specific origin regions. Most of the origin-specific patterns were unstable, due to a low number of death events and low exposure among specific origin groups. Nonetheless, we documented a persistent and pervasive excess mortality among immigrants from other Nordic countries living in Sweden (Figure A1). Additionally, we documented the presence of a strong U-shaped migrant mortality advantage among the foreign-born men and women from the Middle East (also Figure A1). As one final step, we removed foreign-born from the other Nordic countries from the subpopulation of all foreign-born and recalculated relative mortality differentials for foreign-born versus native-born. We plotted this alongside the initial age variation from Figure 3 (see Figure A2). As we can see, this 'hump' at older ages disappears and the age variation among foreign-born men in Sweden now closely resembles the age pattern documented among foreign-born men in the three other contexts.

In addition to Figure 4, we also calculated age-adjusted HRs for the entire adult age range (20-85+), young working ages (15-39), older working ages (40-64) and also retirement ages (65+). Respectively, for women we observed HRs of 0.98, 0.74**[1], 0.92**, and 0.99; for men we observed HRs of 1.06, 0.73**, 1.14**, and 1.06**. These estimates, along with the age patterns from Figure 4, highlight the importance of incorporating age when calculating foreign-born versus native-born mortality differentials. Analyses that ignore age mask great complexity and may also lead to wrong conclusions about the size and direction of the advantage, including specific ages in which the health and health care needs of migrants differ from native-born. In particular, the 20-85+ HRs for men suggest a small overall excess, which hides the substantial advantage among foreign-born men at peak migration ages (HR=0.73 for ages 15-39, and even lower in specific age groups within this range). It also hides the large excess that emerges at older working ages (HR=1.14 for ages 40-64, and even higher in specific age groups in this range).

[1] Value significant to $p<0.01$

Having established that the migrant mortality advantage varies over age in Sweden in a similar way as the United States, United Kingdom and France using comparable data and methods to that used for the three other countries, our final analysis addresses our main aim, to see whether correcting for over-coverage can explain the U-shape of advantage at peak migration ages. Figure 5 plots the age variation for foreign-born illustrated in Figure 4 (which we now refer to as the 'uncorrected' HRs) alongside the age variation for foreign-born after having removed those individuals that we identify as over-covered (which we refer to as being the 'corrected' HRs).

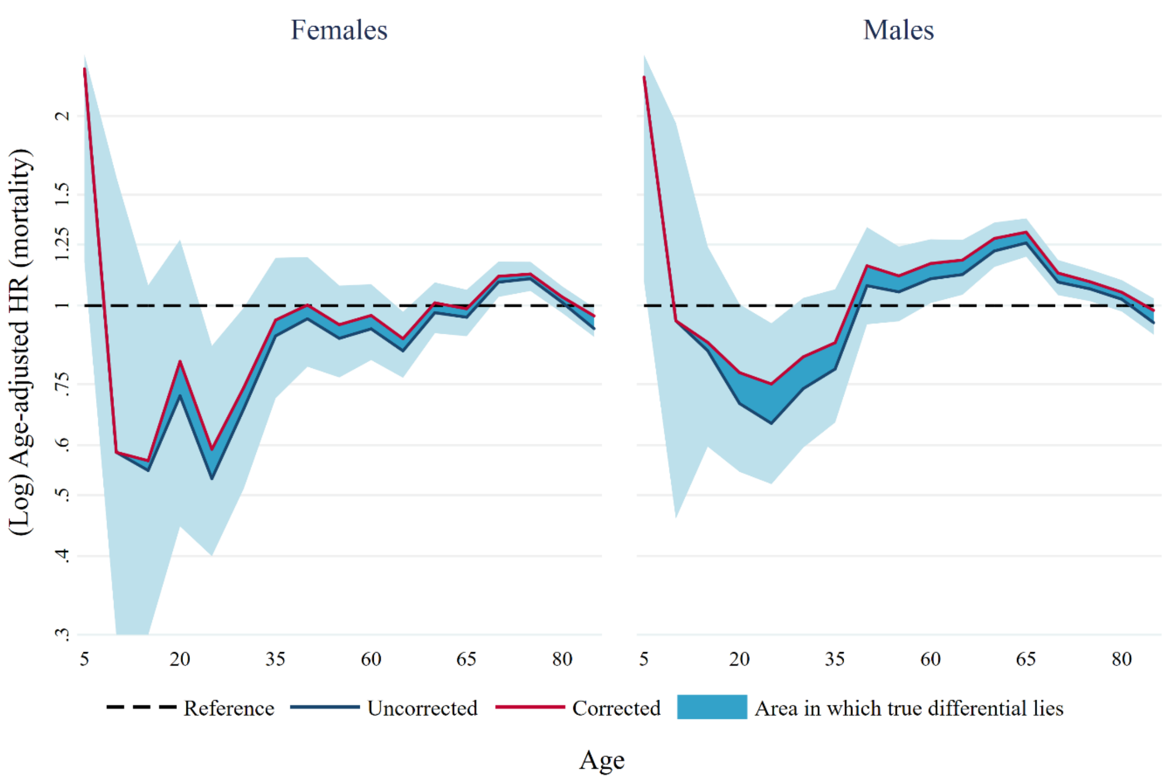

Figure 5. Uncorrected and corrected age-specific relative hazard ratios for mortality, foreign-born versus native-born in Sweden, 2010-12.

As we can see from Figure 5, correcting for population over-coverage has a greater impact for men than women, and the most material impact at the peak migration ages (signified by the higher distance between the two lines at these ages). Nonetheless, the characteristic U-shape persists among both foreign-born men and women in Sweden, even after accounting for over-coverage. We note that the upper confidence also shifts upward but emphasise the fact that we are working with entire population data, so we place greater trust in the point estimates of the HRs.

Table 1. Proportion of migrant mortality advantage attributable to data artefact or substantive causes for foreign-born in Sweden, 2010-12.

Age	HR mortality		Absolute difference in estimates	% advantage attributable to:	
	Uncorrected	Corrected		Data artefact	Substantive causes
Females					
15-19	0.55	0.57	0.02	4%	96%
20-24	0.72	0.82	0.10	34%	66%
25-29	0.53	0.59	0.06	13%	87%
30-34	0.69	0.74	0.06	18%	82%
35-39	0.89	0.95	0.05	51%	49%
15-39	**0.74**	**0.79**	**0.06**	**22%**	**78%**
Males					
15-19	0.85	0.87	0.03	17%	83%
20-24	0.70	0.78	0.08	28%	72%
25-29	0.65	0.75	0.10	29%	71%
30-34	0.74	0.83	0.09	35%	65%
35-39	0.79	0.87	0.08	39%	61%
15-39	**0.73**	**0.82**	**0.08**	**31%**	**69%**

Notes: (1) 15-39 represents a 'peak migration' ages model, averaging over ages 15-19 to 35-39 (2) Calculated by (uncorrected-corrected ratio) / (1 – uncorrected ratio) (3) Calculated by 1 – calculation for data artefact column

In Table 1, we provide the corrected and uncorrected HRs in Table 1 for peak migration ages (that are used to plot Figure 4), alongside estimates of the amount of the migrant mortality advantage that is attributable to over-coverage. First, we show the absolute difference between uncorrected and corrected estimates for the peak migration age groups. Then, we calculate the percentage of the advantage within each age group attributable to data artefact. In absolute terms, correcting for over-coverage explains a similar amount of the advantage in each age group; usually from 0.04 to 0.10, with an average across the peak migration age range of 0.08 for men and 0.06 for women. For men in all the five-year age groups between 15-19 and 35-39, around one third of the advantage can be attributed to over-coverage, with the remaining two thirds (i.e. the residual, after removing unregistered emigrants) explained by substantive causes. For women, the share attributed to data artefacts varies more across age groups, perhaps because mortality is lower among women at these ages, and estimates are therefore inherently more variable. Nevertheless, the average across this age ranges suggests only one fifth (22%) of the migrant mortality advantage among foreign-born women aged 15-39 is caused by data artefact.

Discussion

Despite being documented extensively over the last few decades, scholars continue to debate what causes the migrant mortality advantage and, of greater importance, whether or not it is real. Such debates have persisted due to reservations over the accuracy of demographic data on international immigration and a lack of large-scale datasets permitting the study of native-born versus foreign-born mortality differentials in detail. Indeed, we have only recently begun to understand how these differentials vary over age, the most fundamental of all demographic characteristics. Here, we have contributed extensively to these two debates by building upon recent work documenting major age variation in the migrant mortality advantage in three high-income countries (Guillot et al., 2018) and levels of population over-coverage among migrants in Sweden (Monti et al., 2019). Our study is the first examination of the intersection between these two developments. Our initial aim was to see whether the age profile of advantage varied in a similar way in Sweden as in the United States, United Kingdom, and France (Guillot et al., 2018). Our second aim was to determine whether the documented age variation in Sweden, and particularly the U-shape of advantage at peak migration ages, could be explained by over-coverage.

Regarding our first aim, we found the same systematic age pattern in native-born and foreign-born mortality differentials in Sweden as observed in the United States, United Kingdom, and France (Guillot et al., 2018). In short, for both women and men, we documented a large and initial relative excess mortality in the youngest age group, followed by the emergence of a pronounced 'U-shape' of advantage between age groups 15-19 and 35-39. From age 40 onwards, an attenuation of the advantage towards the mortality of native-born was observed among women. In contrast, among men, an excess hump emerged between ages 40-44 and 75-79. This hump, which did not present among men in any of the countries included by Guillot et al. (2018) was investigated further. Our analysis revealed that this excess was driven by men from other Nordic countries (notably Finland), who comprise a sizeable proportion of all foreign-born men at older ages and experience systematic and considerable excess mortality from age 35 onwards that elevates the overall foreign-born mortality rates at these ages. With reference to our second aim, we found that correcting for over-coverage could explain *some*, but not *all*, of the migrant mortality advantage that was observed at peak migration ages. Indeed, a substantial part of the advantage – around four fifths among foreign-born women and two thirds among men – remained after having corrected for population over-coverage in Sweden.

Our findings advance our understanding of the migrant mortality advantage in two significant aspects. First, we add to the small body of evidence that has examined the role of over-coverage in the size and presence of the migrant mortality advantage. Here, in line with previous work that has corrected foreign-born versus native-born mortality differentials for population over-coverage, we determine that this particular data artefact can explain *some*, but not *all*, of the advantage (when it is observed). Importantly, we go beyond this existing body of research by revealing that over-coverage introduces differential amounts of bias into the migrant mortality advantage at different ages. We recommend that future studies, where feasible, adopt an age-specific perspective when investigating the impact of over-coverage on the size and presence of the migrant mortality advantage. Additionally, our results question whether studies that use alternative data sources (such as pension data) to overcome over-coverage problems are well-suited to investigating the migrant mortality advantage. While such data may better capture the resident population, we found little to no residual advantage among foreign-born women and a sizeable excess among men in post-retirement ages. As such, there was no advantage to be explained. Conclusions based on these data may give the false impression that the advantage is an artefact of other data sources, whereas we demonstrate that this is not the case, at least for Sweden.

Second, our findings contribute to ongoing debate regarding what causes the migrant mortality advantage. By documenting similar age variation as observed in other high-income countries (Guillot et al., 2018), but showing that foreign-born versus native-born mortality differentials are not caused by over-coverage, we go an important step further than the existing evidence by ruling out this data artefact as a main cause of the migrant mortality advantage. This, in turn, implies that the residual mortality differentials must be generated by other, substantive causes. So, what causes the migrant mortality advantage? In line with conclusions from prior work on age variation, our findings suggest that selection plays a prominent role. At the youngest ages, excess relative mortality is consistent with children arriving as dependants who do not self-select into migration. Then, the emergence of a deep, U-shaped advantage at peak migration ages speaks to the idea of large numbers of mobile, healthy and young adults self-selecting into migration. Their arrival rapidly alters the composition of the foreign-born population at these ages, so the U-shape represents compositional changes, rather than genuine age effects. The narrowing of the advantage after age 40 reflects that the composition of the foreign-born population remains largely unchanged, as fewer new immigrants arrive in the host country. Consequently, average duration of residence rises with age, which may signify the wearing off selection effects, perhaps accelerated by a gradual adaptation to the host society (Guillot et al., 2018).

Finally, we cannot *definitively* state that age variations in the migrant mortality advantage in the United States, United Kingdom and France by Guillot et al. (2018) are not caused by over-coverage. However, given that available data sources in these countries are of a comparable quality to those in Sweden and the size of the U-shaped advantages were of a similar size, we think it is reasonable to suggest that over-coverage probably could not explain the differentials in these countries either. We recommend that future work looks to correct age variations in the advantage in these three countries for over-coverage to see whether these mortality differentials persist.

The main limitation of our study relates to the approach used to correct for population over-coverage. The zero-income approach is one of three competing methods available. As yet, no consensus has been reached as to which one is the most accurate. We used the zero-income approach because previous work has shown it to be the most conservative one, particularly for the ages in which we were most interested (i.e. 15-39). That it identifies more individuals as over-covered than the other two approaches worked well within our narrative of determining whether or not age variations in the migrant mortality advantage were a data artefact. Of course, we could have been even more conservative by specifying that people had to have income in every year of the study period. We took the decision to use two consecutive years based upon the recommendations from previous users of this zero-income approach. We conducted a sensitivity analysis, which can be found in Figure A3. This shows the age profile of migrant mortality advantage when we restrict to three consecutive years of income, rather than two. We observe a minor upshift at peak migration ages for foreign-born women and a more moderate shift for men, but still not enough to explain away the large differentials in peak migration age groups.

Regarding the zero-income approach itself, and not only how to implement it, the use of income and social benefits as indicators of residence is clearly more relevant at stages of the life course where many people are in work. As such, it is perhaps less effective at older ages. This approach also likely excludes those at the fringes of society. People may be incorrectly recorded as over-covered if they have been marginalised from the workforce, do not have a consistent source of income, or have vast amounts of wealth, so they do not need to work or receive benefits. It is unclear how (if at all) such individuals would influence the age profile of migrant mortality advantage.

Additionally, over-coverage is also only an example of one data artefact, albeit the most salient one. Thus, it is possible – albeit unlikely – that other data artefacts continue to bias the residual mortality differentials. Other artefact includes ethnic misclassification (irrelevant here because we use country of birth), under-registration of deaths (though the number of unrecorded deaths among foreign-born residents abroad would have to be substantial to fully explain the U-shaped advantage, at ages where death is a particularly rare event), age misreporting (an issue specific to foreign-born from developing countries, of which there are many in Sweden; however, such an issue should generate an *increasing* bias with age and not the pattern that we observe here), and population under-coverage (which would work in the opposite direction to over-coverage). Nonetheless, it is still worth conducting future research on these other types of data artefact in Sweden.

Conclusions

Overall, we find that the migrant mortality advantage, and in particular its characteristic age profile, cannot be explained by population over-coverage; the principal source of data artefact. This tells us that the advantage is real and generated by substantive processes. We should now redirect focus to understanding what explains the advantage and what combination of processes cause it to narrow, or reverse, across age. Alongside the suggestions for future research already given, we think that further insight could be provided by continuing to replicate this age profile in other countries and seeing whether the age profile remains stable over time in the countries in which it has already been found. Moreover, investigating foreign-born versus native-born mortality differentials over duration of residence (paying close attention to age) would speak more directly to a selection and adaptation narrative. Decision makers can be reassured that the advantage is genuine, but should remain wary that foreign-born versus native-born mortality differentials uncorrected for over-coverage may overestimate the size of the advantage to some degree.

Acknowledgements

We thank the colleagues who provided helpful comments in the development of this research.

Funding

This work was supported by the Swedish Research Council (Vetenskapsrådet) via the Swedish Initiative for Research on Microdata in the Social and Medical Sciences (SIMSAM), grant 340-2013-5164.

References

ABRAIDO-LANZA, A. F., ECHEVERRIA, S. E. & FLOREZ, K. R. 2016. Latino Immigrants, Acculturation, and Health: Promising New Directions in Research. *Annu Rev Public Health,* 37**,** 219-36.

AGAFIŢEI, M. & IVAN, G. 2017. *First and second-generation immigrants - statistics on main characteristics.* [Online]. Available: http://ec.europa.eu/eurostat/statistics-explained/index.php/First_and_secondgeneration_immigrants_-_statistics_on_main_characteristics [Accessed].

ALDRIDGE, R. W., NELLUMS, L. B., BARTLETT, S., BARR, A. L., PATEL, P., BURNS, R., HARGREAVES, S., MIRANDA, J. J., TOLLMAN, S., FRIEDLAND, J. S. & ABUBAKAR, I. 2018. Global patterns of mortality in international migrants: a systematic review and meta-analysis. *Lancet,* 392**,** 2553-2566.

ANSON, J. 2004. The Migrant Mortality Advantage: A 70 Month Follow-up of the Brussels Population. *Eur J Popul,* 20**,** 191-198.

ARADHYA, S., SCOTT, K. & SMITH, C. D. 2017. Repeat immigration: A previously unobserved source of heterogeneity? *Scand J Public Health,* 45**,** 25-29.

COURBAGE, Y. & FARGUES, P. 1979. A method for deriving mortality estimates from incomplete vital Statistics. *Population Studies,* 33**,** 165-180.

DEBOOSERE, P. & GADEYNE, S. 2005. Adult Migrant Mortality Advantage in Belgium: Evidence Using Census and Register Data. *Population (English Edition),* 60**,** 655-698.

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**,** e0199669.

HAJAT, A., BLAKELY, T., DAYAL, S. & JATRANA, S. 2010. Do New Zealand's immigrants have a mortality advantage? Evidence from the New Zealand Census-Mortality Study. *Ethn Health,* 15**,** 531-47.

HAMMAR, N., KAPRIO, J., HAGSTRÖM, U., ALFREDSSON, L., KOSKENVUO, M. & HAMMAR, T. 2002. Migration and mortality: a 20 year follow up of Finnish twin pairs with migrant co-twins in Sweden. *J Epidemiol Community Health,* 56**,** 362-366.

HARDING, S. 2003. Mortality of migrants from the Indian subcontinent to England and Wales: effect of duration of residence. *Epidemiology,* 14**,** 287-292.

HARDING, S. 2004. Mortality of migrants from the Caribbean to England and Wales: effect of duration of residence. *Int J Epidemiol,* 33**,** 382-6.

JUÁREZ, S. P., DREFAHL, S., DUNLAVYA, A. & ROSTILAA, M. 2018. All-cause mortality, age at arrival, and duration of residence among adult migrants in Sweden: A population-based longitudinal study. *Soc Sci Med (Pop Health),* 6**,** 16-25.

KHLAT, M. & COURBAGE, Y. 1996. Mortality and Causes of Death of Moroccans in France, 1979-91. *Population (English Edition),* 8**,** 59-94.

KIBELE, E., SCHOLZ, R. & SHKOLNIKOV, V. M. 2008. Low migrant mortality in Germany for men aged 65 and older: fact or artifact? *Eur J Epidemiol,* 23**,** 389-93.

KOHLS, M. 2010. Selection, Social Status or Data Artefact - What Determines the Mortality of Migrants in Germany? *In:* SALZMANN, T., EDMONSTON, B. & RAYMER, J. (eds.) *Demographic Aspects of Migration.* Germany: Springer VS.

MEHTA, N. K., MARTIKAINEN, P. & CEDERSTROM, A. 2019. Age at Immigration, Generational Status, and Mortality among Children of Immigrant Mothers: a Longitudinal Analysis of Siblings. *Am J Epidemiol*.

MONTI, A., DREFAHL, S., MUSSINO, E. & HARKONEN, J. 2019. Over-coverage in population registers leads to bias in demographic estimates. *Popul Stud (Camb)***,** 1-19.

RAZUM, O., ZEEB, H., AKGUN, 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? *Trop Med Int Health,* 3**,** 297-303.

RAZUM, O., ZEEB, H. & ROHRMANN, S. 2000. The 'healthy migrant effect'—not merely a fallacy of inaccurate denominator figures. *International Journal of Epidemiology,* 29**,** 189-192.

RECHEL, B., MLADOVSKY, P., DEVILLE, W., RIJKS, B., PETROVA-BENEDICT, R. & MCKEE, M. 2011. *Migration and health in the European Union,* Berkshire, England, Open University Press.

RIOSMENA, F., WONG, R. & PALLONI, A. 2013. Migration selection, protection, and acculturation in health: a binational perspective on older adults. *Demography,* 50**,** 1039-64.

RUIZ, J. M., STEFFEN, P. & SMITH, T. B. 2013. Hispanic mortality paradox: a systematic review and meta-analysis of the longitudinal literature. *Am J Public Health,* 103**,** e52-60.

SYSE, A., DZAMARIJA, M. T., KUMAR, B. N. & DIAZ, E. 2018. An observational study of immigrant mortality differences in Norway by reason for migration, length of stay and characteristics of sending countries. *BMC Public Health,* 18.

SYSE, A., STRAND, B. H., NAESS, O., STEINGRÍMSDÓTTIR, O. A. & KUMAR, B. N. 2016. Differences in all-cause mortality: A comparison between immigrants and the host population in Norway 1990–2012. *Demographic Research,* 34**,** 615-656.

TURRA, C. M. & ELO, I. T. 2008. The Impact of Salmon Bias on the Hispanic Mortality Advantage: New Evidence from Social Security Data. *Popul Res Policy Rev,* 27**,** 515-530.

WALLACE, M. 2019. *Low migrant death rates: actual or artefactual?* [Online]. Available: http://www.niussp.org/article/low-migrant-death-rates-actual-or-artefactualfaible-mortalite-des-immigres-realite-ou-artefact/ [Accessed 19/07/2019].

WALLACE, M., KHLAT, M. & GUILLOT, M. 2019. Mortality advantage among migrants according to duration of stay in France, 2004-2014. *BMC Public Health (accepted, forthcoming).*

WALLACE, M. & KULU, H. 2014. Low immigrant mortality in England and Wales: a data artefact? *Soc Sci Med,* 120**,** 100-9.

WALLACE, M. & KULU, H. 2018. Can the salmon bias effect explain the migrant mortality advantage in England and Wales? *Population, Space and Place,* e2146**,** e2146.

VANDENHEEDE, H., WILLAERT, D., DE GRANDE, H., SIMOENS, S. & VANROELEN, C. 2015. Mortality in adult immigrants in the 2000s in Belgium: a test of the 'healthy-migrant' and the 'migration-as-rapid-health-transition' hypotheses. *Trop Med Int Health,* 20**,** 1832-45.

WEITOFT, G. R., GULLBERG, A., HJERN, A. & ROSEN, M. 1999. Mortality statistics in immigrant research: method for adjusting underestimation of mortality. *Int J Epidemiol,* 28**,** 756-63.

Appendix

Table A1. Origin composition (%) of immigrant population at specific ages, women.

Age	Nordic	Other EU	CE Europe	Fr Yugo	CS Amer	North Africa	SS Africa	Asia	Mid East	Rest World
5	9	11	13	3	4	1	12	23	20	4
10	7	10	13	5	5	1	12	19	26	3
15	5	6	12	15	5	1	11	16	27	2
20	6	5	13	16	6	1	10	16	26	2
25	6	6	15	11	7	1	9	19	24	2
30	7	7	16	10	7	2	9	19	22	2
35	9	7	14	10	7	2	8	19	22	2
40	15	7	13	12	6	2	7	15	21	2
45	20	7	13	13	6	1	5	12	21	2
50	27	7	16	13	6	1	3	9	17	1
55	36	7	17	12	6	1	2	6	12	1
60	49	9	14	10	4	0	1	4	8	1
65	48	12	14	11	3	0	1	3	8	1
70	50	14	12	11	2	0	1	2	7	1
75	49	17	11	9	2	0	1	2	7	2
80	49	16	15	6	2	0	1	2	6	3
85	48	14	21	4	2	0	0	2	5	2

Table A2. Origin composition (%) of immigrant population at specific ages, men.

Age	Nordic	Other EU	CE Europe	Fr Yugo	CS Amer	North Africa	SS Africa	Asia	Mid East	Rest World
5	9	13	15	3	5	1	13	14	22	5
10	7	10	14	6	6	1	12	16	27	3
15	4	6	12	15	6	1	11	16	28	2
20	5	6	12	16	7	1	9	16	26	2
25	5	8	12	11	7	2	8	18	27	2
30	8	9	14	10	8	2	9	12	26	3
35	10	11	12	11	7	2	9	10	26	3
40	16	11	9	13	5	3	8	7	26	3
45	18	10	8	14	5	2	6	6	28	2
50	23	9	9	14	6	2	5	5	26	2
55	29	11	10	13	6	2	4	5	20	2
60	39	14	10	11	5	2	2	4	12	1
65	38	20	11	12	3	1	2	2	9	1
70	43	18	12	13	2	1	1	2	7	1
75	43	18	12	11	2	1	1	2	8	2
80	42	18	18	7	2	0	1	2	7	4
85	42	14	25	4	2	0	0	2	6	3

Figure A1. Age-specific relative hazard ratios for mortality, foreign-born from Middle East and other Nordic countries versus native-born in Sweden, 2010-12.

Figure A2. Age-specific relative hazard ratios for mortality, foreign-born (with and without immigrants from other Nordic countries) versus native-born in Sweden, 2010-12.

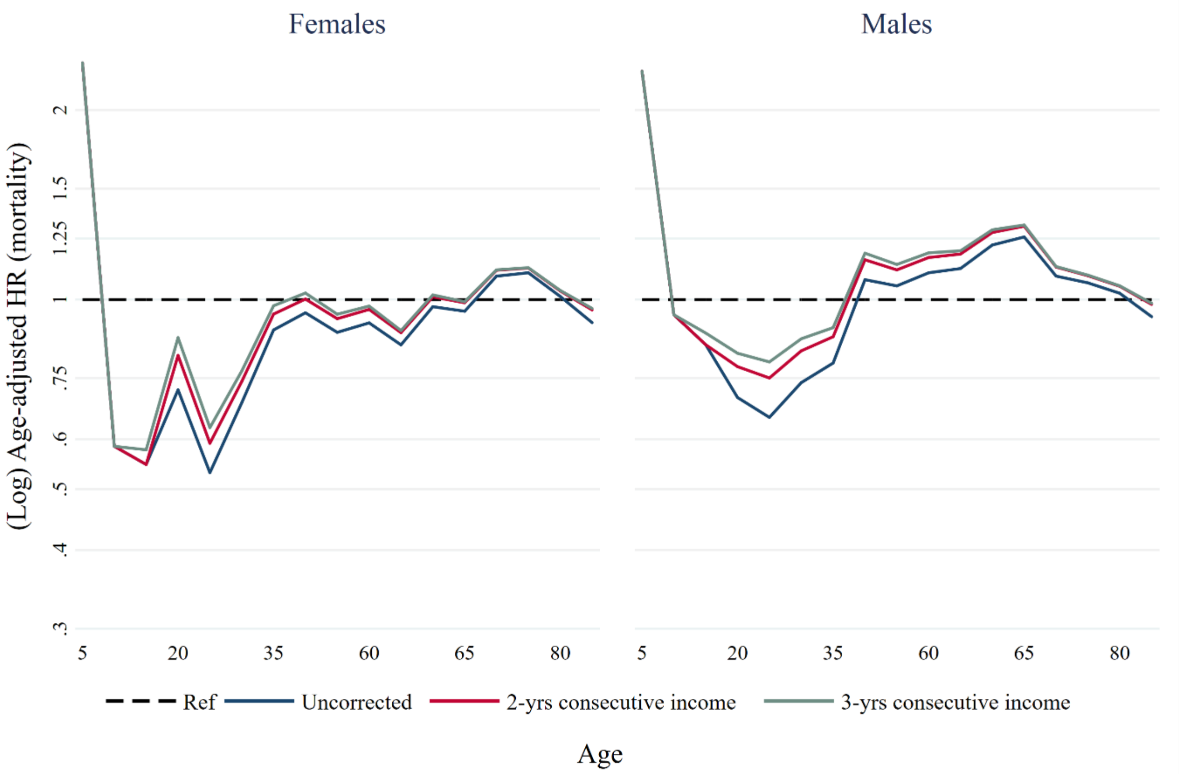

Figure A3. Age-specific relative hazard ratios for mortality, foreign-born (uncorrected vs. 2-years consecutive income vs. 3-years consecutive income).

Stockholm University

Demography Unit

Stockholm Research Reports in Demography
Stockholm University, 106 91 Stockholm, Sweden
www.su.se | info@su.se | ISSN 2002-617X