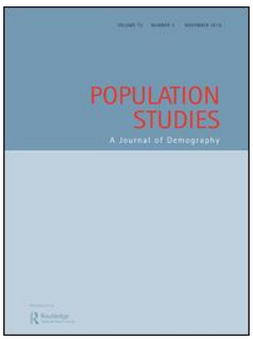

Routledge
Taylor & Francis Group

Population Studies
A Journal of Demography

ISSN: 0032-4728 (Print) 1477-4747 (Online) Journal homepage: https://www.tandfonline.com/loi/rpst20

Over-coverage in population registers leads to bias in demographic estimates

Andrea Monti, Sven Drefahl, Eleonora Mussino & Juho Härkönen

To cite this article: Andrea Monti, Sven Drefahl, Eleonora Mussino & Juho Härkönen (2019): Over-coverage in population registers leads to bias in demographic estimates, Population Studies, DOI: 10.1080/00324728.2019.1683219

To link to this article: https://doi.org/10.1080/00324728.2019.1683219

Published online: 14 Nov 2019.

Population Studies, 2019
https://doi.org/10.1080/00324728.2019.1683219

Over-coverage in population registers leads to bias in demographic estimates

Andrea Monti [1], Sven Drefahl[1], Eleonora Mussino[1] and Juho Härkönen[2]
[1]Stockholm University, [2]European University Institute

Estimating the number of individuals living in a country is an essential task for demographers. This study assesses the potential bias in estimating the size of different migrant populations due to over-coverage in population registers. Over-coverage—individuals registered but not living in a country—is an increasingly pressing phenomenon; however, there is no common understanding of how to deal with over-coverage in demographic research. This study examines different approaches to and improvements in over-coverage estimation using Swedish total population register data. We assess over-coverage levels across migrant groups, test how estimates of age-specific death and fertility rates are affected when adjusting for over-coverage, and examine whether over-coverage can explain part of the healthy migrant paradox. Our results confirm the existence of over-coverage and we find substantial changes in mortality and fertility rates, when adjusted, for people of migrating age. Accounting for over-coverage is particularly important for correctly estimating migrant fertility.

Keywords: over-coverage; fertility; mortality; foreign-born; Sweden; register-based; register bias; population estimates; healthy migrant paradox

[Submitted January 2019; Final version accepted August 2019]

Introduction

Estimating population size is a fundamental task of demographers. Accurate estimation of population size is particularly challenging for migrant populations, where population size may be under- or over-estimated. Underestimation (under-coverage) of migrant populations is a problem particularly for unregistered immigrants (e.g., Woodrow and Passel 1990; Strozza 2004; van der Heijden et al. 2006). At the same time, many population registration systems lack accurate documentation of emigrations due to lack of knowledge of the need to register an emigration or low incentives to do so, thus low compliance, leading to over-coverage in population registers.

In addition to leading to inaccurate representations of both the stocks and characteristics of migrant populations, problems with estimating the size of these populations can contribute to biased estimates of core demographic phenomena including demographic rates. Even when immigrant populations are not accurately documented, their vital events (such as births and deaths) are often well recorded, leading to inflation of the respective demographic rates in the case of under-coverage. Over-coverage leads to the opposite problem: emigrated individuals continue to be erroneously regarded as being at risk of vital events and are less likely to have their vital events documented in the country which they have left but where they are still registered (e.g., Qvist 1999; Weitoft et al. 1999; Loeb et al. 2013). These problems can contribute to apparent demographic paradoxes, such as the unexpectedly low mortality and fertility rates in some migrant populations (e.g., Qvist 1999; Palloni and Arias 2004).

In this paper, we compare procedures for identifying over-coverage in population registers, focusing on Sweden. Previous research has focused more on issues of under-coverage, and the problems of estimating the size and characteristics of undocumented populations, than on over-coverage. However, the problems associated with over-coverage are becoming more pertinent due to ongoing changes in demographic data collection, in which an increasing number of countries have moved to register-based systems and register-based censuses (Poulain and

Herm 2013; Skinner 2018), as well as due to increases in re-migration (such as return or onward migration) among immigrants (Castles et al. 2009; Jeffery and Murison 2011) and in circular migration (Aradhya et al. 2017). Consequently, over-coverage has been identified as a potential source of bias in register-based censuses, official statistics, population forecasts, and academic research (e.g., Cortese and Greco 1993; Fortini et al. 2007; Crescenzi et al. 2008, 2009; Statistics Sweden 2015a), and also in survey sampling (Commissione per la Garanzia dell'Informazione Statistica 2002; Salentin 2014) among migrants in particular (Maehler et al. 2017).

Despite general acknowledgment of the problems of over-coverage, there is to date no praxis for identifying the prevalence of over-coverage or for assessing its consequences for demographic research. We compare estimates of over-coverage of foreign-born individuals based on two different approaches suggested by researchers (Qvist 1999; Aradhya et al. 2017) and by Statistics Sweden (SCB 2015a), both of which rely on traces of activity or the lack of such traces as reported in the register. More specifically, we estimate the prevalence of over-coverage and its trend, and compare the prevalence of over-coverage in different immigrant groups. We also assess the socio-demographic predictors of over-coverage and the potential bias from over-coverage on age-specific fertility and death rates.

Sweden provides a good setting for this research, given its large and heterogeneous migrant population (which in relative terms is bigger than that of the United States (US)) and its relatively high rates of re-migration. At the end of 2016, 17.85 per cent of residents registered in Sweden's total population were foreign born (Statistics Sweden 2018b); the equivalent share for the US in the same year was 13.25 per cent (United States Census Bureau 2018). Also, among those immigrating to Sweden between 1990 and 1995, almost 27 per cent had emigrated from Sweden within 10–15 years (Monti 2018). Together with the other Nordic countries, Sweden is a global forerunner in terms of its comprehensive and widely used population registration system. As such, our findings can offer important lessons for estimating over-coverage and its consequences in other national settings as well. Today, most European countries have centralized national register systems that either replace or complement national censuses and are used as survey sampling frames. A majority of these registers are kept for administrative purposes and do not include continuously updated copies controlled by the national statistical bureaus, as is the case in the Nordic countries, Belgium, and the Netherlands (Poulain and Herm 2013). As a consequence, the administrative registers are bound to definitions matching local policies, which are not always in line with Eurostat's recommendations for terms such as 'usual residence' or 'migrant'.

Nevertheless, the method presented here should also be applicable to most other European countries after adapting the definitions to each national context—providing they more or less correspond to Eurostat's recommendation of the definitions regarding usual residency and migration (EUR-Lex—L 2007, Regulation No 862, article 2):

(a) ' "usual residence" means the place at which a person normally spends the daily period of rest, regardless of temporary absences for purposes of recreation, holiday, visits to friends and relatives, business, medical treatment or religious pilgrimage, or, in default, the place of legal or registered residence';
(b) ' "immigration" means the action by which a person establishes his or her usual residence in the territory of a Member State for a period that is, or is expected to be, of at least 12 months, having previously been usually resident in another Member State or a third country';
(c) ' "emigration" means the action by which a person, having previously been usually resident in the territory of a Member State, ceases to have his or her usual residence in that Member State for a period that is, or is expected to be, of at least 12 months.'

Previous approaches

Over-coverage is an ongoing issue for all countries with a sizable immigrant resident population. However, the extent of the problem and its variation across countries is largely unknown and directly related to the different definitions of international long-term migrants used in different countries. In the Swedish case, some earlier studies have attempted to address this issue for different migrant populations. Kirwan and Harrigan (1986) found an over-coverage rate of about 2.5 per cent for Finnish migrants in Sweden, and concluded that an error of that magnitude is unlikely to bias conclusions for their studied outcomes. They, however, had the advantage of accessing both Swedish and Finnish data, and they caution that emigrations to non-Nordic destinations may be more problematic to address (Kirwan and Harrigan 1986). Statistics

Sweden addressed the same issue in a report in the late 1990s, concluding that over-coverage of immigrants recorded as residing in Sweden is about 1 per cent for migrants from Nordic countries and 2.8 per cent for migrants from other countries (Qvist 1999). In a recent paper, Ludvigsson et al. (2016) concluded from personal communications with Statistics Sweden that their estimate for over-coverage is equal to 0.25–0.50 per cent of the entire Swedish population. For Nordic immigrants, over-coverage may be about 0.1 per cent, but it is substantially higher for individuals born outside the Nordic countries (potentially 4–8 per cent). Statistics Sweden furthermore argued that the often-low mortality of foreign-born individuals suggests that a significant proportion of them no longer reside in Sweden, with substantial variation by age and country of origin (Statistics Sweden 2015a). The impact of over-coverage for the estimation of demographic rates is thought to be largest at the very highest ages, where the number of surviving individuals becomes smaller and registration errors tend to accumulate (Statistics Sweden 2015a). Consequently, the Swedish Tax Agency performs routine checks on individuals aged 100 and above.

Previous studies (e.g., Turra and Elo 2008 for the US; Wallace and Kulu 2014 for England and Wales, and Syse et al. 2016 for Norway) have applied different 'correction' methods to explain the lower mortality among migrants vs. their host populations (the healthy migrant paradox). However, none have found a reliable and repeatable measure. Recently, Aradhya et al. (2017) suggested income-based exclusion as a method for dealing with over-coverage in register-based research. The suggestion is based on the idea that all individuals without any economic activity in a welfare state like Sweden in a given year can be assumed to not live in the country and should thus be excluded from the study population. This criterion, which we call the *zero personal income approach* in the remainder of this paper, provides a relatively straightforward rule for excluding individuals who are thought of as no longer belonging to the population counts. Although this solution is appealing because zero-income individuals can easily be identified in most register-based research, very little is known of its validity.

A second approach, which we will call the *register-trace approach* in the remainder of this paper, was proposed by Statistics Sweden in their efforts to evaluate the quality of the population registers. Similar approaches have been used by other national statistical bureaus (e.g., Tiit and Maasing 2016). This approach tracks a larger number of activities in different linked Swedish registers (Statistics Sweden 2015a). In 2015, Statistics Sweden developed the register-trace approach further by considering not only cross-sectional but also longitudinal information. A central focus of our study is to compare and examine these different methods of over-coverage estimation and to show the impact of these over-coverage measures on demographic estimates of fertility and mortality.

Data and method

The data used for this study are Swedish administrative register data for 1990–2012, on foreign-born residents aged 18–75 who have been registered in the official national population register of the total population in Sweden. Detailed annual data are derived from several administrative registers (Registers on the Total Population, Social Insurance, Emigration and Immigration, Domestic migration, Cause of Death, Civil Status Changes, and Education), and enable us to create different measures of over-coverage.

In Sweden, individuals whose main place of actual or planned residence is within the country for at least one year are registered in the official national population registers. 'Residing' in Sweden requires spending your daily rest in the country on a regular basis, corresponding to at least 52 days a year (SFS 1991, p. 481). Likewise, emigrants are those whose actual or planned residence is outside the country for at least one year. The definitions are in line with the European Commission's regulation on the definitions of usual residence, immigration, and emigration (EUR-Lex—L 2007). Registration is a requirement for obtaining a personal identification number, which is needed for all formal contact with the authorities as well as for everyday life (e.g., employment, housing, and banking). Individuals who leave the country for at least one year are obliged to report their move, and by doing so become deregistered. However, the incentives to comply are low and knowledge about this obligation is limited. The under-reporting of emigration leads to over-coverage in population registers.

Attempts to detect and correct for over-coverage have focused on different ways of assessing presence in the Swedish population by looking at officially recognized, and thus registered, activities: if an individual indeed resides in Sweden, this should be somehow visible in the national registers. Following previous studies, we replicate and compare three different ways of validating presence in Sweden by

searching for activity in the registers. Our register-trace variables are similar but not identical to the ones used by Statistics Sweden (2015a); differences are due to our not accessing exactly the same register variables and also the fact that we are looking at a specific age span of 18–75 years. For example, 'being born' is therefore not part of our measure. Individuals not found active by any of these approaches are considered as not belonging to the Swedish population and thus contributing to over-coverage. Each approach is described next.

The zero personal income approach

One way to ensure correct coverage of study populations in empirical studies has been to exclude people with no personal income (i.e., Weitoft et al. 1999; Aradhya et al. 2017). The argument is that with no economic means to secure one's livelihood, it is unlikely that a person is regularly active in the country. In this approach, individuals are classified as not residing in Sweden (referred to as 'over-covered' from now on) in a given year if they have no reported personal income from earnings, social allowances, parental leave, sick leave, student finance, unemployment benefits, labour market programmes, elderly pensions, home care allowances, or other pensions or social benefits.

This intuitively appealing approach requires access to a dozen or so variables that are routinely available in register-based data. Although it is likely to classify over-coverage correctly in a large share of cases, a limitation of this approach is that it does not apply to children and youth. Additionally, there is a risk of excluding residents who for some reason do not have any registered income, for example due to black-market employment or family support. On the other hand, it is still possible to have a recorded income while living outside the country, for example from pensions or Swedish employment located abroad. In this study we include the approach in our comparisons because earlier studies have used it; however, we think that its practical limitations are too significant for it to be used as a standard correction for over-coverage.

The register-trace approaches

Statistics Sweden uses a broader approach, based on a similar logic, that should overcome the limitations of the zero personal income approach. Register-trace approaches are based on the idea that regularly resident individuals should show some type of activity in the national registers. Personal income is only one of several 'traces' that vouch for individual presence; other such traces include vital events, household income, and educational changes.

We use two versions of the register-trace approach, a cross-sectional version that is only based on information during a single calendar year, and a longitudinal version based on information from three consecutive years. In the cross-sectional register-trace approach, over-coverage is assumed when a person is not found active in any of the following domains in a given year:

- Immigration
- Emigration
- Change of civil status (though not due to the death of spouse)
- Change of citizenship
- Domestic move within the country
- Graduation from the gymnasium (upper secondary education at ages 16–19)
- Enrolment in any higher education (above gymnasium level), measured both from information on student allowance and latest year of obtaining course credits
- Employment (including self-employment if reaching a certain level of income)
- Unemployment or Labour market activation programme, as registered by the Public Employment Service
- Being linked to any household income, measured as the sum of the personal incomes of all members in a household
- Death.

The longitudinal register-trace approach extends this idea, and has been used since 2015 by Statistics Sweden as an additional check to classify individuals initially identified as part of the over-covered population by the cross-sectional register-trace approach (Statistics Sweden 2015a). It complements information for the specific year for which over-coverage is estimated with information from the previous year and subsequent year. It is based on a weighted sum of indicators that consider activity one year before and one year after the given year in relation to individual characteristics. The different register-based indicators are classified into two groups, one indicating correctly registered residence and another indicating over-coverage. The specific indicators and their weights used in this study all originate from the method proposed by Statistics Sweden (2015a). Small developments to the register approach were

suggested by Statistics Sweden in 2018 (Statistics Sweden 2018a); however, the indicators and weights presented here were not affected. This means that in 2018 the register-based approach is still the most developed method used by Swedish authorities to control for over-coverage. The indicators and weights are summarized in Table A1 in the Appendix. If the weighted number of indicators signifying correct registration (indicators 1–6 in Table A1) exceeds the weighted number of indicators signifying over-coverage (indicators 7–18), the non-active individuals from the cross-sectional approach are removed from the over-covered population (i.e., they are considered to be resident). Because the longitudinal register-trace approach is only applied to those classified by the cross-sectional approach as belonging to the over-covered population, the estimates of the longitudinal approach will always be lower than those of the cross-sectional one.

Over-coverage is defined by all three approaches as where a person is registered in the national population register but not active during the same year. This means that a person living in Sweden in January, leaving the country in February and registered as part of the Swedish population in November, will not be considered as contributing to the over-coverage. Our measures of over-coverage are thereby probably slightly underestimated compared with an approach where the same calculations are made using monthly data.

The advantage of the register-trace approaches is that they cover a larger number of life domains and thus should also be able to identify activity for those individuals without any economic activity of their own. That said, the data requirements for this approach are clearly much higher, particularly if longitudinal information is used.

Analysis

Our analysis follows five steps. First, we estimate trends in over-coverage according to the three approaches, for each calendar year 1990–2012. Second, we assess the potential consequences of overestimation of the foreign-born population in demographic research by estimating their age-specific fertility and death rates (ASFRs and ASDRs) and total fertility rate (TFR). In this part of the analysis, we use data from 2010, the latest year for which we have all the information needed for the longitudinal register-trace estimation. The adjustment for over-coverage in a particular year can be made by excluding all individuals who according to the different approaches are characterized as being part of the over-coverage in the studied population. Evidently, most adjustments relate to denominator data. Thus, this can often be done by solely excluding the over-covered individuals from contributing to 'at risk' time (i.e., from the denominator). Similar to Aradhya et al. (2017), for the zero personal income approach we choose to exclude over-covered individuals both in any numerator and denominator data when calculating corresponding demographic rates. The difference is minor since one of our observed events (death) is already part of the two register-trace approaches.

Third, we assess the population groups most likely to be over-covered, by analyzing the individual-level socio-demographic predictors of belonging to the over-covered population in 2010 according to the three different approaches, using logistic regressions. Our predictor variables are sex, age, citizenship, reason for residence permit, and the latest year of immigration. Fourth, we present trends in over-coverage by country of birth.

Fifth, and finally, we assess which of the 18 indicators used in the register-trace approaches can best complement zero personal income as an adjustment of over-coverage. The aim is to derive an estimation that is similar to existing register-trace approaches, but with a more parsimonious combination of indicators and consequentially lower data requirements.

Results

Comparing the three indicators of over-coverage

Over-coverage in Sweden increased between 1990 and 2012 according to all three indicators (Figure 1). A comparison of the different approaches to measuring over-coverage shows how the two register-trace approaches provide lower (similar) and more stable estimates than the zero-income approach, which produces higher and more volatile estimates of over-coverage over the whole period.

When using the zero-income approach, estimates of the prevalence of over-coverage range from around 4 per cent of the foreign-born population according to the register-trace measures to over 12 per cent. The general increase over time in estimated over-coverage can be linked to the overall rise in registered (and non-registered) emigration during the same period, a trend that has been especially noted among the foreign-born but is also prevalent among the Swedish-born population (Statistics Sweden 2015b).

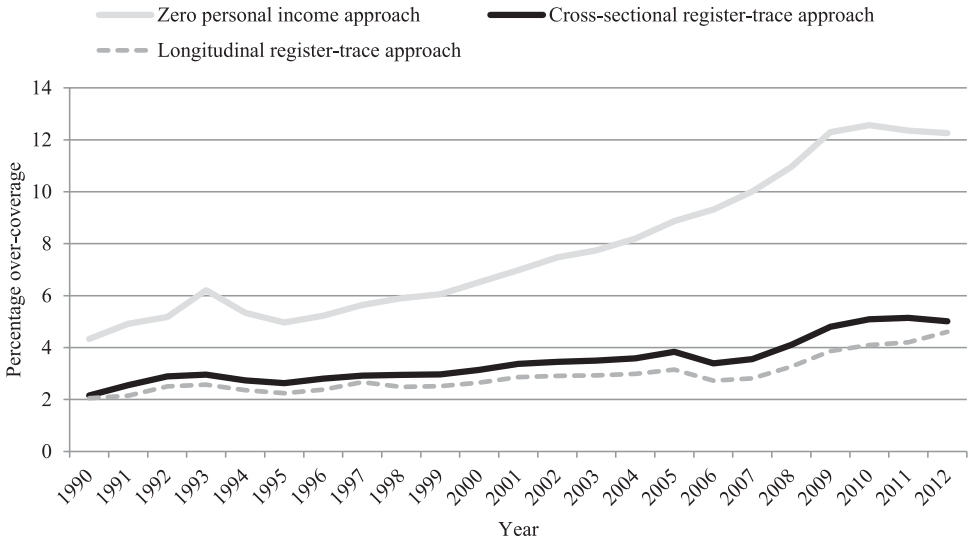

Figure 1 Over-coverage among the foreign-born population aged 18–75 according to three different measurement approaches, Sweden 1990–2012
Source: Authors' calculations based on Swedish register data.

In 2005 Swedish Tax Authorities made an extensive effort to deregister over-covered individuals, which led to a temporal drop in over-coverage (Swedish Tax Authorities 2006). The decrease can be noted in the two register-trace approaches. The increase starting around 2006–07 should thus partly be interpreted as a continued increase from earlier years. Additionally, this increase could be interpreted as a consequence of the re-migration of those who moved to Sweden following the EU expansions in 2004 and 2007.

These results show substantial differences between the zero personal income and register-trace approaches. These differences have increased over time, while few differences are observed between the cross-sectional and longitudinal register-trace approaches. The zero personal income approach thus risks overestimating over-coverage compared with the register-trace approaches which take into account traces of activity other than income alone. In 2010, when this difference is at its peak, almost 84,000 individuals estimated to be over-covered using the zero personal income approach were misclassified due to the oversight of other aspects of activity in the registers (representing 67 per cent of over-coverage).

Over-coverage bias in mortality and fertility rates

In the next step, we examine consequences of over-coverage for the estimation of demographic rates among the foreign-born population. To show the possible impact of over-coverage, we estimate fertility rates (for women) and death rates (for women and men) for 2010, before and after adjusting for over-coverage. We expect the bias to vary by age and that it will be sensitive to the type of process we study, the age-specific intensity of that process, and the age-specific intensity of migration.

Figure 2 shows, as rate ratios, the relative differences in ASDRs for the foreign-born population aged 18–75, with and without adjusting for over-coverage using the three approaches. A value of '1' would indicate that there is no difference in the mortality rate before and after adjustment for over-coverage, suggesting that over-coverage does not bias mortality rates. However, if over-coverage biases the results, the adjusted rates will be higher than the non-adjusted ones, as individuals are removed from the denominators of the rate calculations. In other words, over-coverage bias would lead to underestimation of the respective rates if not adjusted for.

In terms of mortality we find that the over-coverage adjustment has a large impact at ages with high migration intensity, up to around age 40, and a low to very low impact for ages 40 and above. When using the zero personal income approach, mortality rates are up to 2.5 times higher after adjustment at ages 20–30. When using the longitudinal and cross-sectional register-trace approaches, we find mortality rate differences typically of about 25–50 per cent at those ages. With our data we are not able to address the impact of over-coverage at ages of high mortality intensity, that is, ages above 75. The observed patterns do not suggest that the impact of over-coverage increases again at higher ages.

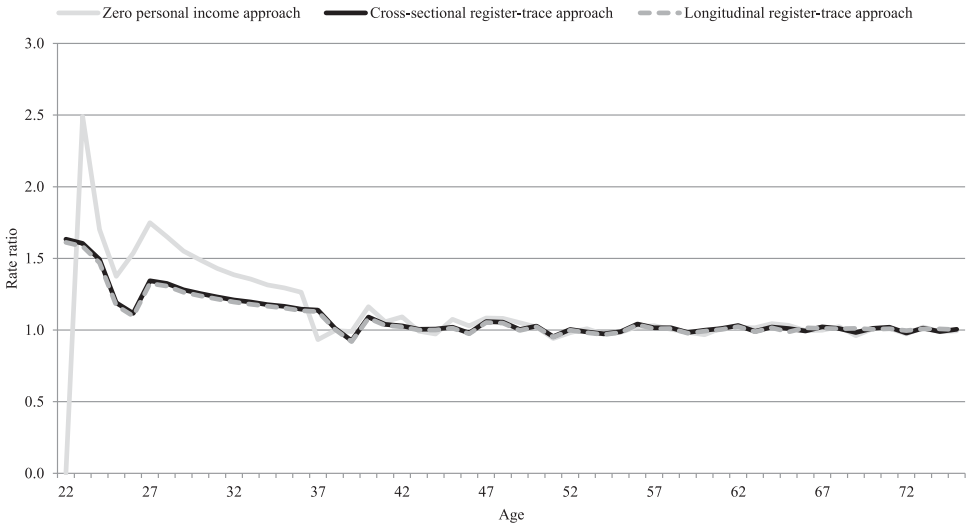

Figure 2 Rate ratio between adjusted and observed age-specific death rates among the foreign-born population aged 22–75, Sweden 2010

Notes: The rate ratio is the adjusted ASDR divided by the observed ASDR (a value of '1' indicates no difference; values greater than '1' indicate that the adjusted rates are higher). Due to zero registered deaths at ages 18–21, and no registered deaths at ages 18–22 if using a zero personal income approach, these ages are omitted from the figure.

Source: As for Figure 1.

Figure 3 shows the corresponding results for ASFR estimates. Similar to mortality, adjusting for over-coverage increases fertility rates, particularly when we use the zero personal income approach. Using this indicator, we find a peak in the rate difference at around age 24, with fertility rates about 50 per cent higher after adjustment. Using the longitudinal and cross-sectional register-trace approaches, we observe less dramatic but still substantive rate ratios for many ages with high fertility intensity,

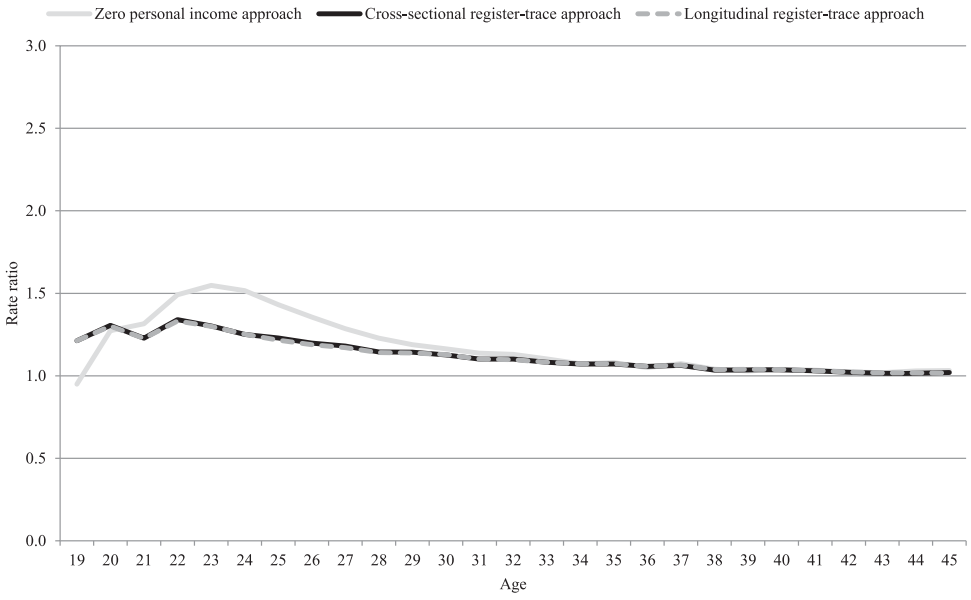

Figure 3 Rate ratio between adjusted and observed age-specific fertility rates among foreign-born women aged 19–45, Sweden 2010

Notes: The rate ratio is the adjusted ASFR divided by the observed ASFR (a value of '1' indicates no difference; values greater than '1' indicate that the adjusted rates are higher). Due to zero registered births at age 18 if adjusting for over-coverage (using any approach), this age is omitted from the figure.

Source: As for Figure 1.

Table 1 Total Fertility Rate (TFR) of foreign-born women, observed and adjusted for over-coverage, Sweden 2010

TFR			
Observed	Adjusted, by approach		
	Zero personal income	Cross-sectional register-trace	Longitudinal register-trace
2.23	2.63	2.51	2.51
	Rate ratio		
1.00	1.18	1.12	1.12

Note: The rate ratio is the adjusted TFR divided by the observed TFR (values greater than '1' indicate that the adjusted rates are higher).
Source: Authors' calculations based on Swedish register data.

with an emphasis towards the younger part of the fertility schedule. The differences between the indicators are highest for ages where women are more likely to be outside the labour market—for example, because of being in education—and thus to have a higher risk of zero observed personal income.

For both mortality and fertility, over-coverage does not seem to lead to a substantive bias at ages with low migration intensity, that is, at ages after the mid-30 s. For younger ages, the bias is important for both processes. However, its real-world impact is likely to be more substantive for fertility estimates, as the bias is concentrated at ages of high fertility but low mortality. If the bias is high at ages of high process intensity, both relative and absolute differences become high, meaning that projected estimates of the number of births would be more biased than those of the number of deaths.

Table 1 shows the unadjusted TFR calculated from the observed ASFRs of foreign-born women in Sweden in 2010 (first column). This value is contrasted with the adjusted measures where we control for different indicators of over-coverage. We can see that the adjusted TFRs are higher. When using the zero personal income approach, the TFR increases by 18 per cent; when using the register-trace approaches, the increase is about 12 per cent. Not adjusting for over-coverage can thus lead to a rather substantial underestimation of immigrant fertility in register-based research.

Over-coverage by socio-demographic characteristics

Over-coverage is caused by the non-random process of failing to register emigration. In Sweden this is the duty of the emigrating individual, who may be weakly informed and incentivized to do so, and the process is not well monitored. Although we can only speculate why individuals do not deregister when emigrating, we believe there to be two main reasons: (1) individuals are not aware of the requirement to deregister, do not care, or simply forget to do so; and (2) individuals do not want to deregister due to fear of losing, for example, their resident permit or the right to return. In Table 2, we present some socio-demographic predictors of the probability of belonging to the over-covered population in 2010. Because our aim is to compare different over-coverage measurements, we present our results for all three approaches to assessing over-coverage.

Socio-demographic predictors of over-coverage confirm the links to registered emigration probabilities and the sensitivity to the type of over-coverage approach. We find that characteristics known to be associated with higher registered emigration are also linked to higher over-coverage likelihoods, for example being male, not having Swedish citizenship, or being a student migrant (Table 2, columns a–c). Additionally, over-coverage probabilities increase with less time spent in Sweden (Table 2, columns a–c). It is plausible that this reflects unawareness of the obligation to deregister. Non-Swedish citizens might also be more afraid of losing any formal connection to Sweden that might hinder a return.

The differences between over-coverage indicators are not restricted to differences between income-based and register-trace measures. For example, the longitudinal register-trace approach shows remarkably low probabilities for the oldest age group (Table 2, column c). Accounting for several years when searching for activity in registers is thus more important for older individuals, as they do not appear as often in the registers as people of working ages, although still residing in the country.

The higher probability of over-coverage among students is especially elevated according to the zero personal income measure (Table 2, column a), which may be because many foreign students finance their studies with personal savings or foreign grants and allowances. Although elevated

Table 2 Socio-demographic predictors (odds ratios) of over-coverage in the foreign-born population aged 18–75, Sweden 2010

	Adjustment approach		
	a Zero personal income	b Cross-sectional register-trace	c Longitudinal register-trace
Sex			
Male	ref.	ref.	ref.
Female	0.83***	0.71***	0.69***
Age			
18–25	1.15***	0.92***	1.02
26–30	0.94***	1.01	1.11***
31–35	0.89***	0.98	1.04
36–40	ref.	ref.	ref.
41–50	1.28***	1.16***	1.08***
51–60	1.72***	1.59***	1.35***
61–75	1.17***	1.74***	0.54***
Citizenship			
Swedish	ref.	ref.	ref.
European	2.62***	3.36***	3.14***
Non-European	1.93***	2.66***	3.07***
Unknown	1.46***	2.04***	2.15***
Resident permit			
Asylum	ref.	ref.	ref.
Work	1.94***	2.6***	3.01***
Family	2.51***	1.94***	1.84***
Student	12.12***	6.54***	6.83***
Other	7.61***	7.75***	3.47***
No need/missing	7.57***	8.98***	9.67***
Latest year of immigration			
Earlier than 1990	ref.	ref.	ref.
1990–99	5.03***	7.2***	5.94***
2000–04	9.89***	16.65***	11.61***
2005–06	11.84***	20.22***	12.01***
2007–08	16.42***	24.65***	14.69***
2009	24.05***	21.83***	10.29***
Constant	0***	0***	0***
Number of observations	865,728	865,728	865,728
Log likelihood	−214,905	−132,448	−112,899

***<0.0001; ** < 0.001; * < 0.01

Notes: The table shows the result from three individual logistic regressions with the results from each over-coverage estimation method as dependent variables. Ref. stands for the reference category.

Source: As for Table 1.

during the first year after arrival, the highest risks of over-coverage are found a couple of years after immigration according to the register-trace approaches (Table 2, columns b and c), which is consistent with the probabilities of registered emigration from Sweden (Monti 2018). On the other hand, the zero personal income approach shows a linear relationship, with the most recently arrived migrants having the highest probabilities of over-coverage (Table 2, column a), although this could be explained by them just not having had the time to declare any income.

Following on from the socio-demographic characteristics of over-covered individuals, we can detect the following mechanisms and contexts behind over-coverage:

(1) Not registering emigrations. This is the main and most basic source of over-coverage and is true regardless of the measure used. For this reason, over-coverage levels are linked to emigration levels as well as the characteristics of the people emigrating from Sweden (i.e., being male, being a student, not having Swedish citizenship, having spent one to three years in Sweden). Over-covered individuals like these are not resident in the country (unless they have re-entered at a later time).

(2) Leading a 'quiet life'. This refers to individuals outside the labour market and outside any official bureaucracy that results in traces in the population registers. These individuals might actually still reside in the country, but are not active in a way that is noticeable on a yearly basis in the official records. To verify the presence of these individuals a longitudinal approach is required, looking at several years in order to capture any registered activity. If this is done, lower over-coverage rates will be found, as in our example with older individuals.

(3) Not receiving any officially registered income. This mechanism by itself results in over-coverage only if using the zero personal income approach. Many of the individuals marked as over-covered are likely to still live in the country, which makes this approach inappropriate, especially when studying groups outside any official labour market, such as newly arrived migrants.

Trends in over-coverage by country of birth

Results from Table 2 stress how over-coverage probabilities are linked to previous immigration (reason and timing) and how the zero personal income approach is therefore less appropriate for some groups, those who are likely to lack personal income according to the registers. To get a fuller picture of the associations between over-coverage and immigration, we now turn to over-coverage levels by country of birth. To give an idea of differences in over-coverage by country of birth and how they change over time, we present trends using the cross-sectional register-trace approach, as it has the advantages of including more information than the zero-income approach but being less data intensive than the longitudinal register-trace approach (Figure 4, panels A and B).

Between 1990 and 2012 the overall differences in over-coverage levels by country of birth have increased. To a large extent these differences can be explained by the different levels and trends in registered emigration across the groups (Appendix Figure A1, panels A and B). For example, migrants born in the US, Canada, Australia, or New Zealand exhibit the highest proportion of over-coverage during the overall period, which corresponds to their high proportions emigrating. As previously mentioned, in 2005 the Swedish Tax Authorities initiated a greater check than usual on over-coverage which, after further investigation of each individual case, led to a corrected number of people assumed to be living in Sweden (Swedish Tax Authorities 2006). This should explain the sharp decline in over-coverage for migrants from some countries around 2006 (Figure 4, panel A). People not found active in the registers were registered as having emigrated, which led to a sharp increase in emigration numbers for these country groups during this period (Appendix Figure A1, panel A).

In the late 2000s we observe a notable increase in over-coverage rates, especially for Western, Southern, and Eastern European countries, such as Poland. Parts of the increase, at least for Eastern European countries, might be explained by the expansion of EU member states in 2007 when more countries joined the Schengen Agreement, making it easier to move between EU member states. A similar but smaller increase is seen around the time of the previous EU expansion in 2004, which confirms this interpretation. However, as the increase in the late 2000s is also found among migrants from Asia, other factors, such as irregular administrative register checks, should not be disregarded.

Due to a Nordic agreement, all intra-Nordic immigration is automatically reported to the sending country. As soon as an individual registers their presence in a Nordic country—for example a Finnish-born immigrant to Sweden who is now returning to Finland—a message is sent to the sending country, in this case Sweden, for deregistration. This explains the low over-coverage rates for Finnish-born migrants. However, in regions with close geographic proximity and relatively dense populations, this system still fails to detect a large number of false registrations. This is because individuals can work and live simultaneously in two countries and often register their presence in the country that maximizes their own economic advantages. Such false registrations provide a plausible explanation behind the increasing over-coverage rates among Danish and Norwegian migrants. The opening of the Öresund bridge between Sweden and Denmark in 2000 connected the densely populated areas of Malmö and Copenhagen and made it even easier to move across the national borders of Denmark and Sweden. In 2006 the Swedish Tax Authorities (2006) voiced their concern over such practices of 'false immigration', as it contributes to over-coverage and has led to a continued increase in over-coverage of Danish nationals in Sweden.

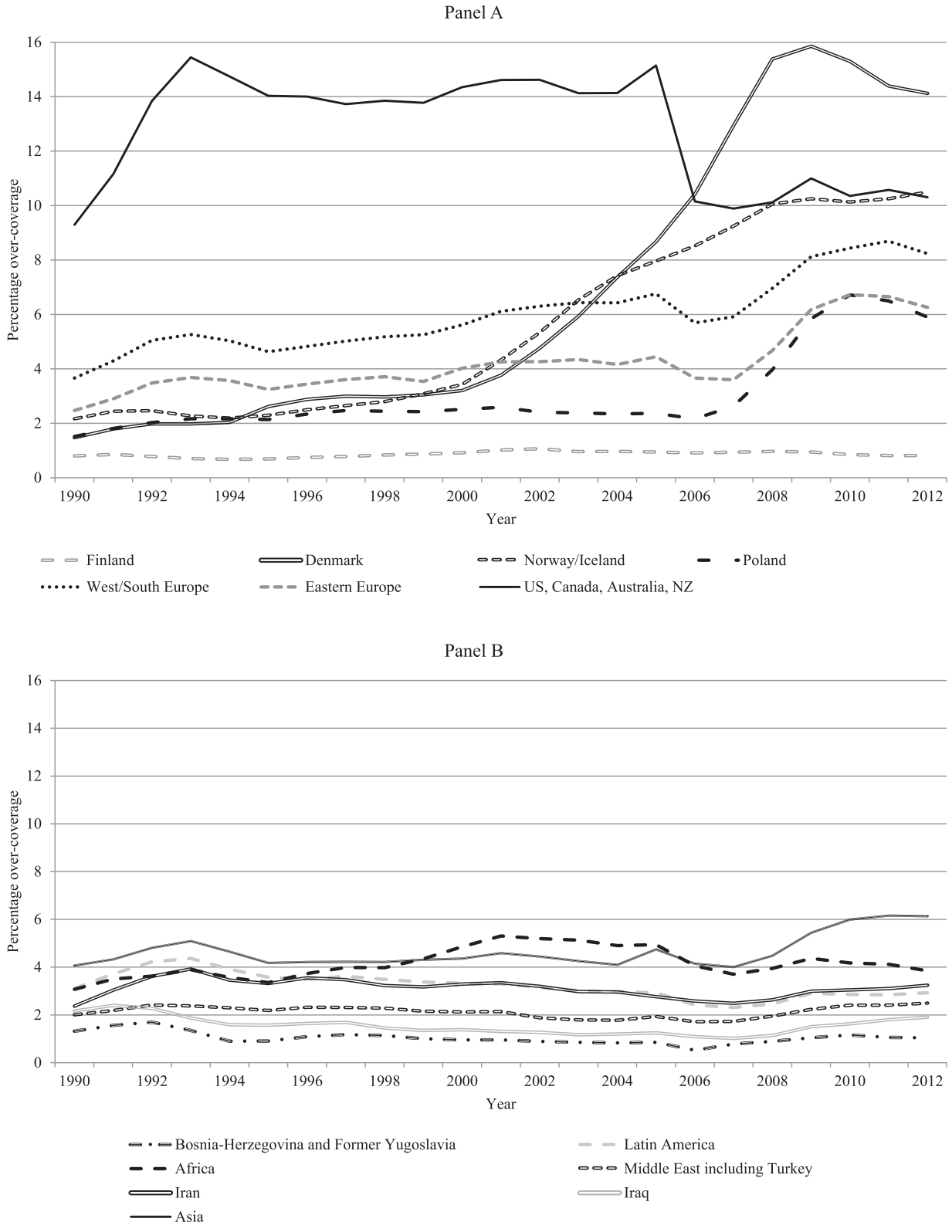

Figure 4 Over-coverage (measured through cross-sectional register traces) among the foreign-born population aged 18–75, by country (group) of birth, Sweden 1990–2012
Source: As for Figure 1.

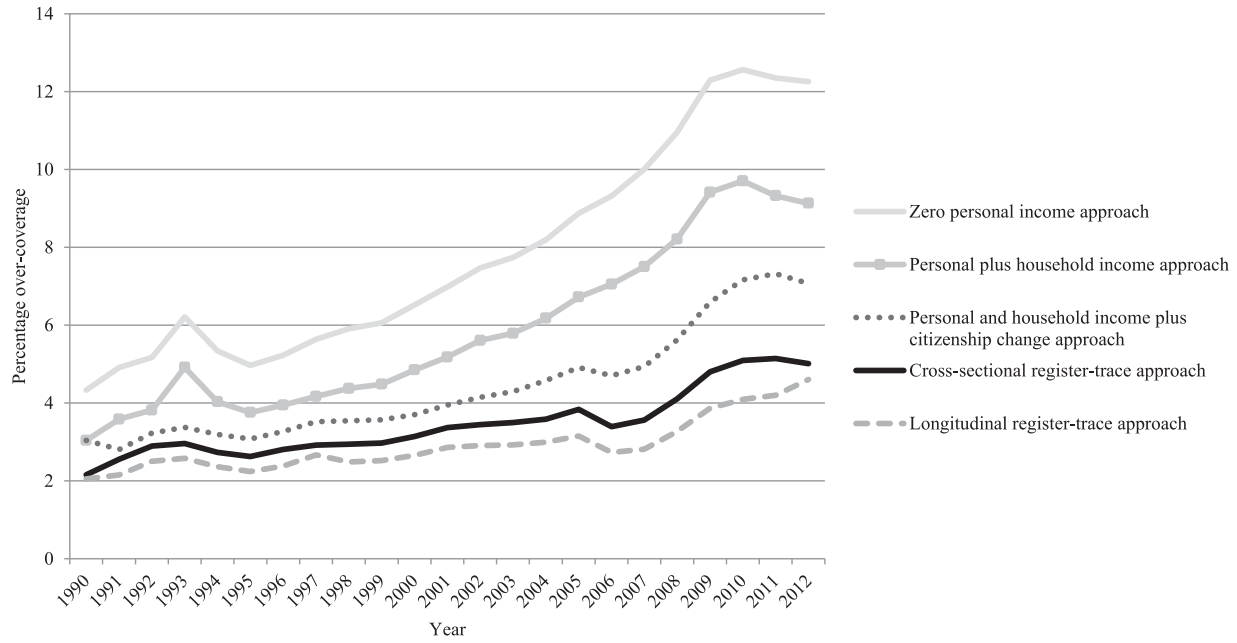

Figure 5 Changes in over-coverage among the foreign-born population aged 18–75 when adding single variables to the zero personal income approach, Sweden 1990–2012
Source: As for Figure 1.

Improving indicators of over-coverage

The zero personal income approach has been the only approach used in previous academic research. In this paper, we have argued that its limitations are too substantive for it to be used as a standard correction for over-coverage. However, the full register-tracing of individuals requires a lot of information that the analyst seldom has easily to hand, while personal income is a more easily accessible variable. We argue that the zero personal income approach should not therefore be rejected as a basis for over-coverage estimation but should be complemented with additional information. Thus, in the next step we aim to find an accommodating way of improving current estimation methods, by adding one indicator at a time from the cross-sectional register-trace approach to the zero personal income approach. The purpose is to derive a skimmed-down version of the register-trace approach that is easier to apply than the full version with 18 indicators but that still addresses the limitations of the zero personal income approach. For practical reasons we do not report all the steps in this procedure here, but in Figure 5 we show the most parsimonious combinations of variables added to the zero personal income approach. The variable that seems to discriminate most, together with personal income, is household income. Adding a variable on citizenship change decreases the difference between the two approaches even further (Figure 5). This citizenship change variable also modifies the socio-demographic composition of the people marked as over-covered (in terms of sex, birth cohort, citizenship, resident permit, and latest year of migration) in the direction of the register-trace approaches (Table A2 in the Appendix). This combination of personal and household income variables plus a third variable not related to income but rather to migration history should therefore be considered as the most preferable option.

Discussion

A number of apparent paradoxes in research on migrant populations stem from two sources of potential bias: (1) when migration movements are associated with the outcomes of interest; and (2) when migration movements are not recorded perfectly. In the latter case, paradoxes can occur due to the fact that some people are incorrectly classified as being resident in a country (or incorrectly classified as not resident). For example, administrative registers may exclude some people currently residing in a country, such as unregistered immigrants whose immigration events were not recorded. This may lead to an underestimation of the size of the foreign-born population, labelled as under-coverage. Much research has focused on issues related to

under-coverage and the problem of estimating the size and characteristics of undocumented populations in a given host country. In this paper we have addressed a closely associated bias, that of over-coverage, which may occur when emigration movements are not recorded correctly. Over-coverage has become more important over time because it is tightly linked to: (1) migration processes, such as increases in re-migration among immigrants (Castles et al. 2009; Jeffery and Murison 2011); and (2) ongoing changes in demographic data collection, such as the number of countries that have moved to register-based data collection systems and register-based censuses. In this study, we compared different indicators to assess over-coverage, examined the variation in over-coverage across migrant groups, and calculated the bias in demographic rates that is produced by over-coverage. Our focus was on Sweden, a country with comprehensive population registration and a large and heterogeneous migrant population.

First, we evaluated the zero personal income approach used in previous academic research and addressed how to improve its accuracy. Then, we compared different approaches to estimating over-coverage in terms of its estimated prevalence. By constructing two different versions of a register-trace approach, we compared these with the zero personal income approach over time. We found that using zero personal income alone will likely overestimate over-coverage to a very large extent as compared with the register-trace approaches and that the differences between the approaches have increased over time.

Using different ways of estimating over-coverage, we analysed the extent to which over-coverage may bias estimates of mortality and fertility rates among immigrants in Sweden. Our results have shown that there is an upward adjustment to mortality rates among immigrant groups at ages of high migration when controlling for over-coverage, independent of which measure is used. This suggests that over-coverage could explain parts of the healthy migrant paradox, at least at the ages with high migration intensities. The impact of over-coverage may be even more important for the correct estimation of fertility. According to our measures, any potential over-coverage bias is largest at ages with relatively high fertility intensities, meaning that ASFRs for women in their 20 s are potentially underestimated by 30 per cent; adjusting for this bias increases the TFR by 12 per cent. This suggests that accounting for over-coverage is essential for correctly estimating fertility in migrant populations.

Errors in the registration system are known to accumulate at older ages, when population sizes tend to become so small that even a small number of errors in the vital registration system can lead to substantive bias in population-level estimates. Our study was limited in that we could not access the impact of over-coverage at ages above 75 because our data on many socio-economic characteristics were limited to ages up to and including 75. Future research needs to address this further. A similar limitation relates to the youngest age, 18, where income information is lacking in our data. This data limitation increases the over-coverage levels at this age. Previous reports (Statistics Sweden 2015a) have concluded that there are differences in over-coverage across different migration origins. We confirmed large variations in over-coverage depending on the origin of migrant groups, as defined by their country of birth, and also found that elevated over-coverage rates among specific migrant groups are associated with their high proportions emigrating and the possibilities of free mobility. Therefore, it is particularly important to adjust for over-coverage in demographic studies of migrant populations with known high emigration intensities. However, at the same time, especially for groups such as students and newly arrived migrants, only using zero personal income as an exclusion criterion should be used with caution.

As mentioned in the introduction, over-coverage is a different phenomenon from under-coverage, and only applies to migrants registered as part of the resident population. Migrants not yet registered (e.g., undocumented migrants or asylum seekers) constitute a selective and different group from the over-covered population. These two biases may therefore not cancel each other out.

In order to improve on currently available estimation methods, we aimed to take a parsimonious approach to see how adding selected variables to the zero personal income approach attenuated the estimated prevalence of over-coverage. Based on our analysis, we advise future users of the zero personal income approach to combine that variable with at least one additional measure of activity, preferably that of household income, in order to increase its accuracy. In order to get an even more stable measure that includes genuine people with no income of their own, an indicator of citizenship change could be added as a third variable.

Our results have shown that using individual and household income in combination increases the chances of more accurate over-coverage estimation at the national level in Sweden. In specific study

populations, for example those with larger shares of one-person households, the improvements may be less substantial and it might be advisable to combine the zero personal income approach with additional measures. On the same note, the correction using only zero personal income might introduce even more bias in countries other than the Nordic ones, where labour force participation is much lower and social transfers are also less all-encompassing.

In this paper we have analysed the total foreign-born population registered as usual residents in Sweden. Similar procedures for accounting for over-coverage are applicable, but not restricted, to countries with similar register-based population statistics. Adding variables to the zero personal income approach requires less detailed information than the register-trace approaches and should also be more feasible in other countries with restricted data access.

Regardless of data structure, this study has stressed the importance of the definition of the resident population and who is defined as a 'usual resident'. Additionally, it has shown the need for knowledge of the population studied and the administrative processes of data collection. Definitions and administrative processes become particularly important if we consider residents in regions close to national borders, sailors, long-distance commuters, or other transnationally mobile people, as in the case of Swedish individuals living in the Öresund region close to Denmark. Some of these people may be registered in Sweden and their income may be based in Sweden, although they spend most of their days and nights on the other side of the bridge. Whether they are classified as part of the Swedish population or not, depends on: (1) the definition of a usual resident; (2) what information enters the registers of Sweden and Denmark; and (3) the choice of estimation method. Some countries may use other data, such as border control systems for distinguishing commuters who spend their daily activities in one country but overnight in another. However, these types of data do not exist for Sweden and the idea of collecting such data is very politicized for ethical reasons.

In sum, our results have shown that the impact of over-coverage can be substantial and that there may be biases in estimates of different measures on migrant populations even in countries with the highest-quality registration systems. As such, our results offer lessons for estimating over-coverage and its consequences elsewhere as well. Research needs to acknowledge that any demographic estimates based on migrant populations are likely to be biased at ages of high migration intensity and that currently available correction methods need to be improved further.

Assessing error in register-based civil registration systems is becoming increasingly important now that numerous countries are moving from a traditional census to a register-based census. One of the primary goals of the traditional census was correcting population counts in terms of both under- and over-coverage of the resident population. For Sweden, a country characterized by a long history of the highest-quality registers, we have shown that the register-trace approach might help to reduce estimation errors, at least in terms of over-coverage.

Based on this study, we advise scholars to account for over-coverage in demographic analysis of populations where the overall proportions emigrating are high. For countries with access to less rich sources of register data, adding a few variables to the zero personal income approach can lead to results similar to the register-trace approach. This is especially the case if the variables added besides income relate to individuals' immigration history.

This study has addressed the bias in demographic estimates for migrant populations that stems from errors in data collection, in this case the under-recording of emigrations. However, there is at least one further and related source of bias that has been put forward in migration research: that occurring when migration movements are associated with the outcomes that are studied. This may happen when demographic events occur either more or less frequently in periods that cannot be observed, for example, shortly before immigration or after emigration. Conclusions that are drawn from time periods with available data are likely to be systematically biased. The fertility literature points out one common example. Studies have often found fertility levels to be high shortly after arrival in a destination country, which results from individuals postponing their fertility in anticipation of migration (e.g., Andersson 2004; Milewski 2010; Mussino and Strozza 2012). In turn, the mortality literature discusses the role of 'salmon bias' effects in mortality, which are produced when individuals emigrate or return to their country of origin due to poor health and in anticipation of death. This yields reduced mortality rates for immigrants in a given destination country (Abraído-Lanza et al. 1999; Andersson and Drefahl 2017). These examples are not based on actual problems with data coverage; they are rather manifestations of the endogeneity of migration and other demographic events, which can produce apparent paradoxes in the study of demographic processes.

We argue that in order to improve demographic and other estimates for migrant populations, researchers need to systematically distinguish and address all sources of potential bias.

Notes and acknowledgements

1 Andrea Monti, Sven Drefahl, and Eleonora Mussino are based in the Demography Unit, Department of Sociology, Stockholm University, Sweden; Juho Härkonen is at the Department of Political and Social Sciences, European University Institute, Florence, Italy. Please direct all correspondence to Andrea Monti, Sociologiska institutionen, Stockholms universitet, SE—106 91 Stockholm, Sweden; or by E-mail: andrea.monti@sociology.su.se

2 We thank Gunnar Andersson for his valuable and constructive comments on the first draft of this manuscript. This research was supported by the Strategic Research Council of the Academy of Finland (Decision number: 293103); the Swedish Research Council (grant number: 340-2013-5164); and the Swedish Research Council for Health, Working life and Welfare.

ORCID

Andrea Monti http://orcid.org/0000-0002-9396-1446

References

Abraído-Lanza, A. F., B. P. Dohrenwend, D. S. Ng-Mak, and J. B. Turner 1999. The Latino mortality paradox: a test of the 'Salmon Bias' and healthy migrant Hypotheses, *American Journal of Public Health* 89 (10): 1543–1548.

Andersson, G. 2004. Childbearing after migration: fertility patterns of foreign-born women in Sweden, *International Migration Review* 38(2): 747–774.

Andersson, G. and S. Drefahl. 2017. Long-distance migration and mortality in Sweden: testing the salmon bias and healthy migrant hypotheses, *Population Space and Place* 23(4): e2032.

Aradhya, S., K. Scott, and C. Smith. 2017. Repeat immigration: a previously unobserved source of heterogeneity? *Scandinavian Journal of Public Health* 45(Suppl 17): 25–29.

Castles, S., H. de Haas, and M. J. Miller. 2009. *The Age of Migration: International Population Movements in the Modern World*. Basingstoke: Palgrave Macmillan.

Commissione per la Garanzia dell'Informazione Statistica. 2002. *Il campionamento da liste anagrafiche: analisi degli effetti della qualità della base di campionamento sui risultati delle indagini* [Sampling from registry lists: analysis of the effects of the quality of the sampling base on the results of the surveys], Rapporto di ricerca CGIS, n. 02.12, Roma Dicembre 2002.

Cortese, A. and M. Greco. 1993. *Il grado di copertura del censimento demografico 1991: considerazioni sulla base del confronto con le risultanze anagrafiche* [The degree of coverage of the 1991 demographic census: considerations based on the comparison with the registry data]. Quaderni di Ricerca Istat, Roma 1993, Serie Interventi e Relazioni.

Crescenzi, F., M. Fortini, G. Gallo, and A. Mancini. 2008. *Nota per il Presidente e il Consiglio dell'Istat, Linee generali di impostazione metodologica, tecnica e organizzativa del 15° Censimento generale della popolazione* [Note for the President and Council of Istat, General guidelines on methodological, technical and organizational set-up of the 15th General Population Census]. Roma, settembre 2008.

Crescenzi, F., M. Fortini, G. Gallo, and A. Mancini. 2009. *La progettazione dei censimenti generali 2010–2011 in Linee generali di impostazione metodologica, tecnica e organizzativa del 15° Censimento generale della popolazione* [The design of the general censuses 2010–2011 in General guidelines of methodological, technical and organizational set-up of the 15th General Census of the population]. Documenti ISTAT, n. n. 6/2009.

EUR-Lex—L:2007:199:TOC - EN - EUR-Lex. (n.d.). 2007. Regulation (EC) No 862/2007 of the European Parliament and of the Council of 11 July 2007 on Community statistics on migration and international protection and repealing Council Regulation (EEC) No 311/76 on the compilation of statistics on foreign workers. Official *Journal of the European Union*, L 199, 31 July 2007. Available: http://data.europa.eu/eli/reg/2007/862/oj (accessed: 24 April 2019).

Fortini, M., G. Gallo, E. Paluzzi, A. Reale, and A. Silvestrini. 2007. *Criticità di processo e di prodotto nel 14° Censimento generale della popolazione e delle abitazioni: aspetti rilevanti per la progettazione del 15° Censimento* [Process and product criticality in the 14th General Census of Population and Housing: Relevant Aspects for the Planning of the 15th Census]. In La progettazione dei censimenti generali 2010–2011, Documenti ISTAT, n. 10/2007.

Jeffery, L. and J. Murison. 2011. The temporal, social, spatial, and legal dimensions of return and onward migration, *Population, Space and Place* 17(2): 131–139.

Kirwan, F. and F. Harrigan. 1986. Swedish-Finnish return migration, extent, timing, and information flows, *Demography* 23(3): 313.

Ludvigsson, J. F., C. Almqvist, A. K. Bonamy, R. Ljung, K. Michaelsson, M. Neovius, M. Stephansson, and W. Ye. 2016. Registers of the Swedish total population and their use in medical research, *European Journal of Epidemiology* 31(2): 125–136.

Loeb, S., L. Drevin, D. Robinson, E. Holmberg, S. Carlsson, M. Lambe, and P. Stattin. 2013. Risk of localized and advanced prostate cancer among immigrants versus native-born Swedish men: a nation-wide population-based study, *Cancer Causes & Control* 24(2): 383–390.

Maehler, D. B., S. Martin, and B. Rammstedt. 2017. Coverage of the migrant population in large- scale assessment surveys. Experiences from PIAAC in Germany. Large-scale assessments in education, *IEA-ETS Research Institute Journal* 2017(5): 9.

Milewski, N. 2010. *Fertility of Immigrants: A Two-Generational Approach in Germany*. Heidelberg, Dordrecht, London, New York: Springer.

Monti, A. 2018. Re-emigration of foreign-born residents from Sweden: 1990–2015 working paper, *Stockholm Research Reports in Demography*, 2018: 15.

Mussino, E. and S. Strozza. 2012. The fertility of foreign immigrants after their arrival: the Italian case, *Demographic Research* 26(4): 99–130.

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

Poulain, M. and A. Herm. 2013. *Le registre de population centralisé, source de statistiques démographiques en Europe* [The Centralized Population Register, Source of Demographic Statistics in Europe], *Population* 68 (2): 183–212.

Qvist, J. 1999. *Täckningsproblem i Registret over totalbefolkningen RTB - Skattning av övertäckning med en indirekt metod* [Problems of Coverage in the Register of Total Population (RTB)—Estimation of Overcoverage by an Indirect Method]. R & D Report. Stockholm, Sweden: Statistics Sweden, 1999.

Salentin, K. 2014. Sampling the ethnic minority population in Germany. The background to 'migration background', *Methods, Data, Analyses* 8(1): 25–52.

SFS 1991:481. 1991. Folkbokföringslag (1991:481) [National Registration Law 1991:481]. Available: http://rkrattsbaser.gov.se/sfst?bet=1991:481 (accessed: 2 August 2019).

Skinner, C. 2018. Issues and challenges in census taking, *Annual Review of Statistics and its Application* 5: 49–63.

Statistics Sweden. 2015a. *Övertäckning i Registret över totalbefolkningen – en registerstudie* [Overcoverage in the Total Population Register—a register study], Bakgrundsfakta. Befolkning och Välfärd 2015: 1.

Statistics Sweden. 2015b. Invandringar och utvandringar efter födelseland och kön. År 2000–2018 [Immigration and emigration by country of birth and sex 2000–2018]. Available: http://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__BE__BE0101__BE0101J/ImmiEmiFod/ (accessed: 1 August 2019).

Statistics Sweden. 2018a. *Folkbokföringsfelet – En metodrapport om skattning av övertäckning, undertäckning och folkbokförda på fel adress* [The Registration Error – A Method Report on Estimation of Over-Coverage, Under- Coverage and Registered at the Wrong Address]. Örebro: SCB BV/REG and PMU/MIÖ.

Statistics Sweden. 2018b. Inrikes- och utrikes födda per region, ålder och kön. År 2000–2016 [Swedish and foreign- born population by region, age and sex. Year 2000–2016]. Available: http://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__BE__BE0101__BE0101E/InrUtrFoddaRegAlK on/ (accessed: 29 January 2018).

Strozza, S. 2004. Estimates of the Illegal Foreigners in Italy: A Review of the literature, *International Migration Review* 38(1): 309–331.

Swedish Tax Authorities. 2006. *Kartläggning av folkbokföringsfelet* [Mapping of the population registration error]. Solna, Swedish Tax Authorities 2006: 7.

Syse, A., B. H. Strand, Ø. Næss, Ó. A. Steingrímsdóttir, and B. Kumar. 2016. Differences in all-cause mortality: a comparison between immigrants and the host population in Norway 1990–2012. *Demographic Research* 34(1): 615–656.

Tiit E-M. &E. Maasing. 2016. Residency index and its application in censuses and population statistics. Eesti Statistika Kvartalikiri, *Quarterly Bulletin of Statistics Estonia* 16(3): 53–60.

Turra, C. M. and I. T. Elo. 2008. The impact of salmon bias on the Hispanic mortality advantage: new evidence from social security data. *Population Research and Policy Review* 27(5): 515–530.

United States Census Bureau. 2018. Selected Characteristics of the Native and Foreign-Born Populations 2012–2016. American Community Survey 5-Year Estimates. Available: https://www.census.gov/topics/population/foreign-born.html (accessed: 29 January 2018).

van der Heijden, P. G. M., G. van Gils, M. Cruijff, and D. Hessen. 2006. *Een schatting van het aantal in Nederland verblijvende illegale vreemdelingen in* 2005 [An Estimate of the Number of Illegal Aliens Staying in the Netherlands in 2005]. Utrecht: IOPS Universiteit Utrecht.

Wallace, M. and H. Kulu. 2014. Low immigrant mortality in England and Wales: a data artefact? *Social Science & Medicine* 120(20): 100–109.

Weitoft, G. R., A. Gullberg, A. Hjern, and M. Rosen. 1999. Mortality statistics in immigrant research: method for adjusting underestimation of mortality, *International Journal of Epidemiology* 28(4): 756–763.

Woodrow, K. A., and J. S. Passel. 1990. Post-IRCA undocumented immigration to the United States: Assessment based on the June 1988 CPS, in F. D., Bean, B. Edmonston, and J. S. Passel (eds), *Undocumented Migration to the United States: IRCA and the Experience of the 1980s* (Vol. 7). The Urban Institute, pp. 33–76.

Appendix

Figure A1 Proportions emigrating among the foreign-born population aged 18–75, by country (group) of birth, Sweden 1990–2012
Source: Authors' calculations based on Swedish register data.

Table A1 Indicators used in the longitudinal register-trace approach.

			Weight by Statistics Sweden (2015 p. 45)
	All indicators refer to year t, and only to those individuals not regarded active in the cross-sectional register-trace approach.		
1	Death the year after (t + 1)	Indicates correct registration (individual resides in the country)	2
2	Internal move, change in civil status or citizenship the year before and after the inactive year (t−1 and t + 1)		2
3	Active again the year after and with no internal move within Sweden in between (t + 1)		2
4	Foreign citizen who immigrated after age 60		2
5	Reason of residence permit: enough financial capital to support themselves		2
6	Household (household) income over yearly national base amount (calculated in relation to consumer price index) (time t)		1
7	Emigration the year after (delayed registration) (t + 1)	Indicates over-coverage (individual does not reside in the country)	3
8	Enrolled in tertiary education the year before (t−1), and a foreign citizen (time t)		2
9	Reason of residence permit: Studies		2
10	Immigration two years before (t−2), followed by a positive personal income the first year (t−1)		3
11	Positive personal income the year before (t–1), positive income the year after (t + 1) and a new address (t + 1)		2
12	No known address (time t)		2
13	Not registered in the Swedish Total Population Register the year after (t + 1), without any notification of death or emigration		3
14	A positive personal income the year before (t−1)		1
15	Reason of residence permit: Work		2
16	A registered death the year before (t−1)		3
17	A registered emigration the year before (t−1)		3
18	A registered immigration the year after (t + 1)		3

Notes: All indicators correspond to, but are not equal to, indicators listed by Statistics Sweden (2015, pp. 41–5). Weights are the same as the SCB weights.

Table A2 Distribution of socio-demographic characteristics of the foreign-born population marked as over-covered, by measurement approach, Sweden 2010

	Measurement approach					
	Zero personal income	Personal plus household income	Personal and household income plus citizenship change	Cross-sectional register-trace	Longitudinal register-trace	Total foreign-born population
Sex						
Male	54	59	60	61	62	49
Female	46	41	40	39	38	51
Total percentage	*100*	*100*	*100*	*100*	*100*	*100*
Age						
18–25	18	10	10	21	15	6
26–30	21	21	23	23	22	9
31–35	15	17	18	15	16	10
36–40	11	12	13	10	11	12
41–50	17	18	18	15	16	25
51–60	11	12	13	9	11	19
61–75	7	11	5	7	9	19
Total percentage	*100*	*100*	*100*	*100*	*100*	*100*
Citizenship						
Swedish	13	9	11	11	12	55
European	49	52	54	59	56	25
Non-European	37	39	34	29	32	19
Unknown	1	1	1	1	1	1
Total percentage	*100*	*100*	*100*	*100*	*100*	*100*
Resident permit						
Work	8	9	9	10	11	6
Family	29	21	21	21	21	30
Asylum	5	5	5	5	6	20
Student	18	23	21	15	17	4
Other	5	5	5	6	3	2
No need/missing	34	36	39	43	42	39
Total percentage	*100*	*100*	*100*	*100*	*100*	*100*
Latest year of immigration						
Earlier than 1990	7	5	6	7	7	33
1990–99	8	6	8	9	9	20
2000–04	11	10	14	16	16	12
2005–06	8	8	11	12	11	7
2007–08	19	20	27	27	27	10
2009	16	17	22	15	13	6
Missing	31	34	12	15	18	13
Total percentage	*100*	*100*	*100*	*100*	*100*	*100*
Total N	124,382	96,082	70,940	5,042	40,559	989,997

Source: Authors' calculations based on Swedish register data.