Quality & Quantity
https://doi.org/10.1007/s11135-023-01757-x

Multiple systems estimation for studying over-coverage and its heterogeneity in population registers

Eleonora Mussino[1] · **Bruno Santos**[2] · **Andrea Monti**[1] · **Eleni Matechou**[2] · **Sven Drefahl**[1]

Accepted: 15 September 2023

Abstract
The growing necessity for evidence-based policy built on rigorous research has never been greater. However, the ability of researchers to provide such evidence is invariably tied to the availability of high-quality data. Bias stemming from over-coverage in official population registers, i.e. resident individuals whose death or emigration is not registered, can lead to serious implications for policymaking and research. Using Swedish Population registers and the statistical framework of multiple systems estimation, we estimate the extent of over-coverage among foreign-born individuals' resident in Sweden for the period 2003–2016. Our study reveals that, although over-coverage is low during this period in Sweden, we observed a distinct heterogeneity in over-coverage across various sub-populations, suggesting significant variations among them. We also evaluated the implications of omitting each of the considered registers on real data and simulated data, and highlight the potential bias introduced when the omitted register interacts with the included registers. Our paper underscores the broad applicability of multiple systems estimation in addressing and mitigating bias from over-coverage in scenarios involving incomplete but overlapping population registers.

Keywords Over-coverage · Sweden · Multiple-systems estimation · Population registers · Foreign born

1 Introduction

The demand for policies grounded in rigorous research and compelling evidence has never been greater, but the ability of researchers to provide such evidence is highly dependent upon the availability of high-quality data. In particular, the COVID-19 pandemic has highlighted the need for accurate estimates of population sizes, not only for the overall population but also for distinct subgroups within it. Sweden has long been known for its excellent population and vital registration and was one of the first countries to establish a system

✉ Eleonora Mussino
eleonora.mussino@sociology.su.se

[1] Stockholm University, Stockholm, Sweden

[2] University of Kent, Canterbury, United Kingdom

of interconnected registers for administrative and research purposes. For a long time, data quality was taken for granted, but with increased international mobility, recent findings by Statistics Sweden and the Swedish National Audit Office (Statistics Sweden 2015; 2018; Swedish National Audit 2017; Swedish Tax Authorities, 2018) have raised some data quality concerns and indicated a need to re-evaluate the quality of the population data that underpins all research and policy decisions.

Over-coverage represents the most salient potential error source for register data systems that form the basis of official national statistics, population forecasts, academic research (e.g. Statistics Sweden 2015), and survey sampling (Salentin 2014). Over-coverage occurs when failing to administrate the emigration or death of individuals, leading to population overestimation of the resident population (e.g. Monti et al. 2020)[1] With more countries having moved to register-based systems in their collection of population data, the relevance of over-coverage bias is increasing.

Every member of a resident population is subjectable to measurement error. Nevertheless, the propensity for error related to emigration is disproportionally higher among populations that are more mobile and consequently more difficult to capture in data sources, namely international migrants. In Sweden, international migrants constitute now one fifth of Sweden's resident population (Statistics Sweden 2023). Increases in return, onward and circular migration (De Haas et al. 2019; Jeffery & Murison 2011; Monti et al. 2020) have surged the potential for over-coverage.

The errors introduced by population over-coverage—and its consequences—should not be under-estimated. In particular, over-coverage can impact upon three main aspects: the size, composition and outcomes of national—and especially migrant—populations. Firstly, we likely over-estimate the number of resident migrants (and consequently the total resident population) as current estimates include individuals who have already left. Secondly, if the propensity for over-coverage is selective on key demographic factors, such as age and sex, then we likely provide an inaccurate picture of the composition and, therefore, the needs of the migrant population. Thirdly, over-coverage likely introduces bias into the numerators (if over-coverage is selective on the outcome e.g. if the outcome is education and the less educated are more susceptible to over-coverage) and the denominators (population bases) of national and migrant-specific estimates. This distortion is particularly misleading when comparing migrants directly with native-born populations, who typically exhibit lower mobility and are therefore less likely to be over-covered. In a recent study by Monti et al. (2020), mortality rates among migrants living in Sweden aged 20–30 years were up to 2.5 times greater when correcting over-coverage; fertility levels were up to 1.5 times greater among immigrants in their 20s. With these magnitudes of error in mind, over-coverage could undermine most of what we think we know about the outcomes of migrant populations in their host countries. In the best-case scenario, a specific finding remains valid, and correcting over-coverage only modifies the size of pre-existing differences between migrants and native-born populations for a given socio-demographic outcome. In the worst-case scenario, errors introduced through over-coverage accumulate to

[1] The primary source of over-coverage is individuals who emigrate without delisting, although we anticipate rare occasions of over-coverage due to unrecorded mortality. However, in the context of Sweden, where this study is conducted, death registration within the country is virtually complete. There are incentives for listing deaths that occur abroad. Nevertheless, unregistered deaths of individuals abroad remain a part of unregistered emigrations.

generate differences between migrants and native-born that misdirect research, misinform policy, and may fuel harmful public narratives.

As official authorities are just recently starting to acknowledge the need for continuous approaches to identify over-coverage (cf. Swedish Tax Authority 2018), no common approach exists for identifying, let alone correcting, over-coverage across countries, nor for assessing the potential consequences for funded social science research. Sweden offers an ideal research context given its world-renowned population registers, extensively used by the international research community, in conjunction with its diverse migrant population and relatively high rates of re-emigration (Monti 2020).

Here, we consider the use of generalized linear models for contingency tables, as employed within the framework of multiple systems estimation (MSE, Bohning et al. (2017), Bird and King (2018), Cruyff et al. (2021), van der Heijden et al. (2019)). The MSE modelling approach uses overlapping but incomplete population registers, also referred to as lists in the literature and in this paper, as variables (covariates), whilst taking into account individual socio-demographic characteristics, and hence estimates the number of individuals from each combination of characteristics that are part of the population. This allows us, for the first time, to identify heterogeneity in over-coverage across sub-populations. To do so we employ a Bayesian approach, similar to King et al. (2014), relying on the R package *conting* (Overstall and King 2014) for inference. Additionally, we repeat the analysis after omitting each list in turn and quantify the effect on the corresponding estimate of population size and hence over-overage. We present an extensive simulation study that highlights the potential bias introduced in the estimation of population size when omitting a list that interacts with the included lists. Finally, we compare our new findings to those obtained using existing approaches designed to correct for over-coverage.

2 State-of-the-art

Over-coverage is relevant for all countries that have a sizeable immigrant resident population, as accounting for it is necessary to obtain unbiased population estimates. Core parts of the over-coverage problem are the unknown scale and variation across countries, which also relates to different definitions of international long-term migrants across countries.

2.1 Previous over-coverage studies

Over-coverage in population registers has created concerns in many different countries, in particular when population registers are used to estimate the *de Jure* population; see for instance Pettersen et al. (2018) for Norway and Statistics Denmark (e.g. 2014) for Denmark, but no general or unified method has been established on how to deal with this challenge. Most of the studies did not really focus on creating a replicable measure of over-coverage. Their methods were specific to their unique data sources used, limiting their transferability to other contexts and aiming only to explain the lower mortality among migrants vs their host populations (e.g., Syse et al. 2016 for Norway; Wallace and Kulu 2014 for England and Wales; and Turra and Elo 2008 for US). Following the same approach, Wallace and Wilson (2022) confirm that around 20–25% of the lower mortality among migrants could be explained by over-coverage, with substantial variation by birth country. Recently, complementary approaches have emerged. Rampazzo et al. (2021)

augmented the UK Labour Force Survey with Facebook's advertising platform to attempt to estimate the actual migrant stock even in a context "where there are no ground truth data".

There has also been increased attention to this problem in Sweden, a country with a long standing culture on using register based research. Kirwan and Harrigan (1986) had access to both Swedish and Finnish data and they found, for Finnish migrants in Sweden, an over-coverage rate of about 2.5 percent and concluded that an error of that magnitude is unlikely to bias conclusions for their studied outcomes. Already in the late 1990s, Statistics Sweden estimated that among Nordic migrants residing in Sweden over-coverage was about 1 percent, while for other migrants it was about 2.8 percent (Qvist 1999). More recently, Ludvigsson et al. (2016) concluded that over-coverage is a minor bias for the entire Swedish population (about to 0.25–0.5%) but substantial among migrants born outside the Nordic countries (4–8%). The difference between Nordic and non-Nordic immigrants is closely related to different registration procedures and cooperation between the authorities responsible for the population list in the Nordic countries, as elaborated later. Monti et al (2020) compared different indicators of over-coverage present in the literature (see next section) and found that over-coverage in Sweden among foreign-born residents is increasing over time. Moreover, over-coverage varies by country of origin, reflecting registered levels of emigration. This variation across subgroups explains how over-coverage levels appear low when looking at a national scale, but worryingly high when concentrating at some origin groups (i.e. those with high registered emigration levels). For example, possible over-coverage rates reached over 14% among highly mobile migrants from US, Canada, Australia and New Zealand as well as migrants from Denmark within the early and mid-2000's (Monti 2020).

2.2 Existing measures of over-coverage in sweden

In numerous OECD countries, coverage surveys are implemented to test for over- and under-coverage of the census (Brown et al. 2011). This practice is particularly prevalent in countries that are transitioning from traditional to register based censuses (Bijak et al. 2021; Righi et al. 2021). However, this is not the case of Sweden. Since 2011, following the development of the dwelling register, the population census in Sweden, akin to other Nordic countries, is entirely register based.

Throughout the centuries-long evolution of population registers in Sweden, the approach has been oriented towards estimating the *de jure* as opposed to the *de facto* population (Andersson et al. 2023). Nonetheless, for the 2011 census, the European Commission (Commission Regulation no. 1151/2010) requested that Member States provide an estimate of under-coverage and of over-coverage of the population census. On this occasion, Sweden carried out a Post Enumeration Survey (PES) to assess the accuracy of the population registration based on the newly formed dwelling register. This aimed to verify the data quality of household type and household size as recorded by the census register. The results indicate that the number of smaller households is under-estimated and the number of larger households is over-estimated in the population registers (Statistics Sweden 2013).

Previously, over-coverage in register-based research has occasionally been addressed by implementing an income-based exclusion method (Aradhya et al. 2017; Weitoft et al. 1999). Using this exclusion method, it is presumed that every individual without any economic activity any given year can be assumed to not live in the country, and thus should be excluded from the study population. Whilst this relatively straightforward approach might

seem appealing, the precision is low and might even introduce bias by non-random exclusion of individuals with higher risk of not showing economic activity, as for example newly arrived immigrants (Monti et al 2020).

More holistic approaches to estimate over-coverage, commonly referred to as "register-trace" approaches, have been developed by Statistics Sweden. The first of these methods, that we call the "cross-sectional" register trace approach, tracks a larger number of activities in different Swedish administrative registers for a given year. Activities added involve household income, active unemployment, international migration, internal moves, studies, change of civil statuses, deaths and change of citizenship (Statistics Sweden 2015). *Inactive* individuals are considered over-covered. A second method, which we refer to as the "longitudinal" approach as it uses several observation years, has been developed in an attempt to refine the model. Using the second method, all individuals not found active in the first step are further analyzed in relation to 24 "indicators", scenarios[2] that would either strengthen or weaken the suspicion of over-coverage. For example, *inactivity* followed by *activity* at the same address suggests the individual to have lived in the country the whole time and thus not be over-covered, whereas *studies* followed by *inactivity* suggests the individual has emigrated and thus has correctly been classified as over-covered. Each indicator is weighted on a scale of one to three by a subjective notion of their relevance. Individuals are thereby classified as either belonging to the actual population or being over-covered based on the weighted sum of indicators.

After further revising their second method, Statistics Sweden (2018) currently distributes weights using coefficients from a logistic regression, where the 24 indicators are included as independent variables. The outcome variable in this model is built on the future status of inactive individuals (Swedish Tax Authority 2018). That is, among the inactive individuals in a given year, some are eventually found active in the registers in following years, whilst others are manually de-registered by the Swedish Tax Agency. The future status of the either active or de-registered individuals then serves as the outcome variable of the model, assuming that the status was the same throughout the whole period.

Whilst improving the longitudinal register trace approach by adding the use of parametric weights, some caveats still apply. First of all, the parametric distribution of weights in a given year depends on conditions only observed in the future, which necessarily introduces some degree of error. The method also assumes that those who do not register their emigration have the same characteristics as those who do register their move. As a consequence of the necessary waiting time to obtain the required future information, estimates are not available for the most recent years. After six years (which is the time period used), about a fifth of individuals are neither found active nor de-registered, and are excluded from the model (Swedish Tax Authority 2018). Given the non-random processes behind appearing as active in the registers, become de-registered or not found at all, there is great potential for biased inference.

In 2020, Monti et al. compared the two register trace approaches in terms of their inference on estimated prevalence with the zero personal income approach. They found that using zero personal income alone will likely overestimate over-coverage to a large extent as compared with the register-trace approaches and that the differences between the approaches have increased over time. However, although more promising, the register-trace

[2] After refining the list of scenarios there are currently 24 scenarios or "indicators" as referred to by official agencies (Swedish Tax Authority 2018).

approaches are not immune to similar misclassifications. Moreover, they are difficult to replicate in countries with lower-quality registration systems.

3 Data and methods

We use administrative register data of the total Swedish population, including the annual registers of Register of Total Population (RTB), the Longitudinal Integrated Database for Health Insurance and Labor Market Studies (LISA), the Intergenerational register as well as event registers of birth, death, international migration and internal moves (Statistic Sweden 2017a, b, 2019, 2022). The registers comprise detailed annual information on the registered population, including registered immigration and emigration, collected by different agencies and provided by the Swedish national institute of statistics (Statistic-Sweden-SCB). In Sweden, all individuals with the actual or planned primary residence within the country for at least one year (and with the legal right to do so) are required to register their presence within the Swedish Tax Authority. Incentives to register are very high, since practically all formal contact with public authorities and institutions as well as the capacity to participate in society, for example by opening a bank account or obtaining a mobile phone, will require individuals to have a personal identification number, given upon registration. Hence, under-coverage of the *de jure* population is minimal, and restricted to individuals waiting for the administrative filing of their immigration. Upon emigration from Sweden, individuals are equally required to de-register from the population registers if planning on living outside the country for at least one year. However, the knowledge and incentives to de-register are much lower than for immigration.

Our observation period covers the years 2003–2016[3] For each year, the study population consists of those foreign-born individuals aged 18 and older with registered presence in the country on the 31st of December of the previous year, excluding those who emigrated or died in the current year.

To detect over-coverage, we initially follow the cross register-trace approach proposed by Statistics Sweden (2018), placing an emphasis on compiling evidence of individual presence in the country by looking at officially recognized activities across the different registers during one year. This approach reasoned that if someone resides in Sweden, some form of activity should be visible in at least one of the registers (see Statistic Sweden 2015; Monti et al 2020). We include in our models the following activities/registers: internal moves, citizenship acquisition, marriage, divorce, the birth of a child, employment, active unemployment, enrollment in higher education and household income[4] (see Table 1 in the Appendix). Secondly, we consider a model based on MSE (see next section) that estimates the number of undetected individuals, that is individuals who do not appear in any register (referred to as lists in the MSE framework) even though they were resident that year, given their characteristics (age, sex, region of origin and time in Sweden). The total population size is then estimated as the number of individuals detected in at least one list, plus the number of estimated undetected individuals. The register trace approach would

[3] We start from 2003 because from this year all the definitions of the lists used in the paper are constant and the quality of the registers considered good (Statistic Sweden 2019).

[4] Differently from previous approaches, we did not use death and emigration as lists but instead we use them to define the population. This was a modelling issue. These variables do not overlap with any other lists by design, so they cannot be considered in this analysis.

classify these undetected individuals as over-covered because they do not appear in any of the considered lists. In our paper, we refer to these (estimated) undetected individuals as *false positives*, because with previous approaches, that do not consider the probability of not detecting individuals even though they are present, they would have been falsely classified as over-covered and hence excluded from any subsequent analysis.

To summarize, we have the official total population present in RTB at the end of each year (Fig. 6 in the appendix). We expect that there will be individuals present in RTB who emigrated previously and have not deregistered. Thus, the number of individuals in RTB provide an upper bound for the number of individuals present (UBP). This quantity forms our denominator in our calculation of over-coverage.

Then for each year we obtain the estimate of the population size from the register trace approach, which corresponds to the number of individuals that are seen in at least one register that year (lower bound of the population size: LBP). The over-coverage for the register trace approach is then calculated as (UBP—LBP)/UBP.

Finally, our MSE model estimates the probability that individuals, based on their demographic characteristics, appear in any list (register) combination, and hence also the probability that they do not appear in any list, and therefore estimates the number of unseen individuals for every combination of categorical covariate values (estimated population of undetected individuals: EP). Our new probabilistic model yields estimates of the number of individuals falsely classified as over-covered: false positive. Our estimate of total population in the country year is given by LBP + EP, which is in between the lower bound and the upper bound. Our over-coverage estimate is calculated as 1—(LBP + EP)/UBP.

For each individual we also have information about their sex, age, region of birth and time since migration. Previous studies have found that these are important factors to estimate in the probability of leaving the country (Monti 2020) as well as to do not de-register upon emigration (Monti et al 2020). Because we are considering contingency tables (see below), all variables need to be treated as categorical factors. Given their age, individuals were grouped in three intervals: between 18 and 35, between 36 and 60 and more than 60 years old. Regarding their time in Sweden, we considered three categories: between 0 and 5, between 6 and 10 and more than 10 years in the country. Countries of birth are also grouped as (1) Denmark and Norway, (2) Iceland and Finland, (3) Eastern Europe, (4) Western, Europe, (5) Middle East and North Africa (MENA), (6) United States of America (USA), Canada and Oceania and (7) rest of the World.

3.1 MSE approach

The modelling approach considered in this paper corresponds to the so called multiple systems estimation (MSE) method, where individuals that are uniquely identifiable appear in one or more lists (registers), which are incomplete but overlapping, and interest lies in estimating the number of individuals that do not appear in any of the available lists i.e. the undetected individuals. MSE has gained a lot of attention recently for estimating the size of cryptic populations, such as drug users (King et al. 2014), victims of human trafficking (Cruyff et al. 2017) and victims of modern slavery (Silverman 2020).

We consider probabilistic models using the theory of MSE, based on log-linear models for contingency tables (Bird and King 2018). All lists are binary, with value 1 corresponding to the event of being observed in that list. Categorical variables correspond to individual characteristics and can have more than two levels. Each cell in the contingency table in this case corresponds to the number of people with a certain combination of the

categorical variables and the lists considered, jointly referred to here as covariates. The log-linear model framework allows us to estimate the main effects of the covariates, as well as of all the two-way interactions, on the expected number of individuals in each cell. This gives rise to an estimate of the number of people in the unobservable cells, which corresponds to the individuals who have not appeared in any list in a given year (undetected). We note here that the number of unobservable cells is equal to the number of combinations of the categorical variables in the data set, i.e. if there are C categorical variables, with variable X_i having k_i levels, there are $U = \prod_{\{i=1\}}^{C} k_i$ unobservable cells and hence corresponding estimated numbers of undetected individuals.

In the case of L lists and C categorical variables, there are $D = 2^L U$ cells, with cell d having count n_d, $d = 1, \ldots, D$. We model n_d as a realization of a Poisson distribution with mean μ_d, with μ being the vector of means. In our model we assume $\log \mu = X\beta$, where X corresponds to the design matrix, with columns including the dummy variables representing the levels of the different covariates, including the interaction terms, and β the parameter vector of the log-linear model.

We obtain our parameter estimates considering a Bayesian approach, using the R package *conting* (Overstall and King 2014, version 1.7), where our estimates are based on samples from the posterior distribution of the parameter vector. This parameter vector includes the model coefficients, β, as well as the vector with the number of individuals in each of the U unobservable cells. For each one of these U we have then an estimate of undetected individuals based on the posterior mean of our draws of the posterior distribution. As for the total number of individuals present each year, this is estimated as the number of individuals observed in at least one list (LBP), plus the sum of the posterior means of the inferred counts in the U cells. We summarize the uncertainty around our estimates using 95% posterior credible internals.

We study the importance of each used list in the estimation of over-coverage by removing one list at a time for the year of 2015 and estimating the over-coverage without the information for that particular list. This is a similar study to the one considered by Sharifi Far et al. (2020), who analysed data on modern slavery. In turn, we remove one list at a time from the set of the covariates, and hence all the interactions in which that list was involved, in our probabilistic model, as if this information was not available in the data. As a consequence, in this study, all the quantities discussed previously such as LBP and EP change accordingly. This analysis provides the information on how the removal or omission of the different lists available can have an impact on the estimates of over-coverage.

Additionally, to validate our model and to test its performance when the model considered is correctly specified, as well as when it is mis-specified by omitting one of the lists, we performed an extensive simulation study, presented in the supplementary material.

4 Results

Figure 1a shows the trend of over-coverage in Sweden between 2003 and 2016 according to the cross register-trace approach (black line) and our MSE- approach (blue line). An initial comparison between the deterministic and probabilistic approaches shows that over-coverage has been overall low during this period in Sweden, with both methodologies depicting very similar temporal trends. The MSE approach provides, as expected, a consistently lower estimate than the register-trace approach. This reflects the robustness of the MSE approach, with the disparity between the two estimates fluctuating within a

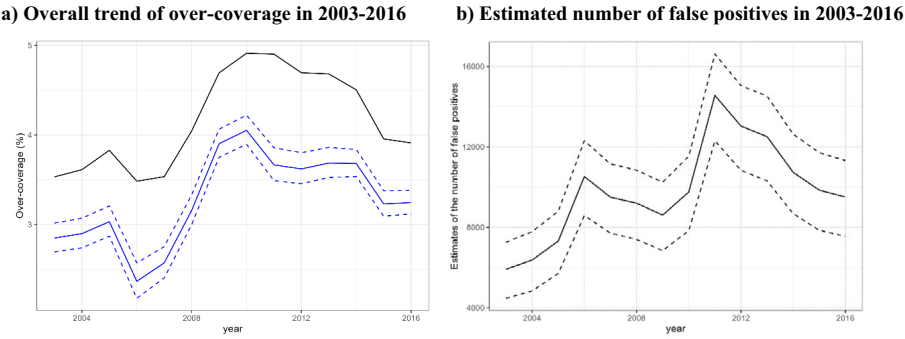

Fig. 1 a Overall trend of over-coverage in 2003–2016 b Estimated number of false positives in 2003–2016 *Note* Own elaborations using Swedish population registers **a** The register-trace approach is represented by the black line and the MSE approach by the blue line. **b** The solid line is the sum of the posterior mean for all U unobservable cells in each year, while the dashed lines represent the corresponding 95% posterior credible interval

consistent range over the years, despite the models being estimated independently. A peak in over-coverage is observed for both approaches in 2010, which is likely due to the continued consequences of the economic crisis, for example increased re-emigration Alderotti et al 2022). In the final years of our observation period, over-coverage has been decreasing, with the MSE providing an estimation of approximately 3% "overestimation" for the most recent years. Figure 1b shows the estimated number of false positives for the same period, which is calculated as the difference in number of individuals over-covered between the two approaches.

Figure 2 illustrates the same trend for each of the main socio-demographic characteristics included in our model. In general, it seems that over-coverage is proportionally higher among men, young adults, recent migrants, and migrants from neighboring countries such as Norway and Denmark, as well as those from regions such as the United States of America, Canada and Oceania. However, we observe a contrasting pattern of lower over-coverage for migrants from Finland.

While over-coverage is more prevalent among men, the overall trends are similar in both sexes. We also observe distinct historical patterns by age and time since migration. Trends in over-coverage vary greatly by time since migration. Among newly arrived, over-coverage is highest, peaking in 2010, and decreasing sharply afterwards. Among migrants who lived in the country for six to 10 years we observe a considerably lower level of over-coverage, which is slowly increasing over time. Among migrants who have been in Sweden for more than 10 years we observe a very low and stable proportion of over-coverage. The stark contrasts between the three groups might reflect changes in the composition of the immigration pattern throughout the observation period in Sweden[5] as well as the effect of the *Swedish introduction program* (Qi et al. 2021). The 2010 reform of the program introduced several changes aimed to make refugees' integration quicker and more effective, which potentially affected the propensity of those migrants to be active in society and consequently be part of population registers.

[5] Different inflows by country of origin also reflect different propensities of emigration.

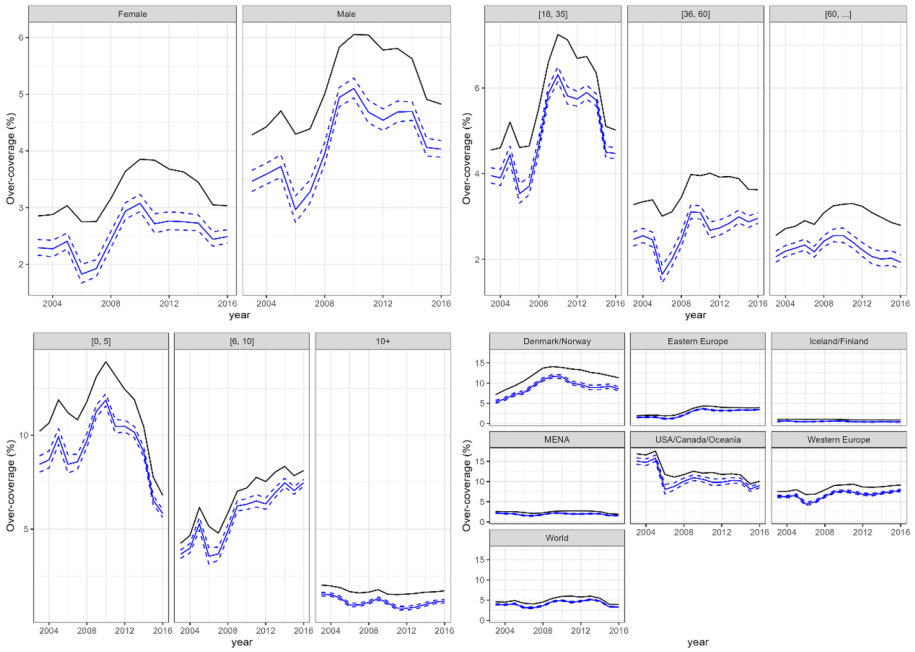

Fig. 2 Trend of over-coverage in 2003–2016 by sociodemographic characteristics *Note* Own elaborations using Swedish population registers the register-trace approach is represented by a black line) and our MSE by a blue line

The results by country of birth are in line with previous research (Monti et al 2020): over-coverage is lowest among Finnish migrants as well as among migrants from MENA countries or "other" countries of origin. It is highest among migrants from Norway and Denmark, followed by migrants from Western Europe and the US / Canada / Oceania. Among Eastern-European migrants, over-coverage has been increasing slowly since the enlargement of the European Union, probably because circular migration and return migration are generally facilitated in a context of free mobility.

Figure 3 shows the comparison in estimated over-coverage between the register trace approach (dashed line) and our MSE approach (solid line) considering all the possible combinations of the sociodemographic characteristics (sex, age, region of birth, and time since migration) for each region of origin, where the upper-bound of the population size (UBP) in each of these combinations is shown in appendix Figure 7. Over-coverage tends to be highest among recently arrived migrants for most regions of origin, particularly within the oldest age bracket. However, this group of migrants also tends to be the smallest. Migrants from neighboring Norway and Denmark demonstrate consistently higher over-coverage, even for those migrants who stayed in Sweden for less than 10 years across all age categories. Similar trends are found for Western European migrants. For migrants from Iceland and Finland, we observed higher over-coverage only for older migrants and soon after their arrival. Similar observations can be made for migrants born in Eastern Europe. However, since the expansion of the European Union, the percentage of over-coverage increased also among younger individuals that stayed in the country for less than 10 years. Intriguingly, for migrants from the US and

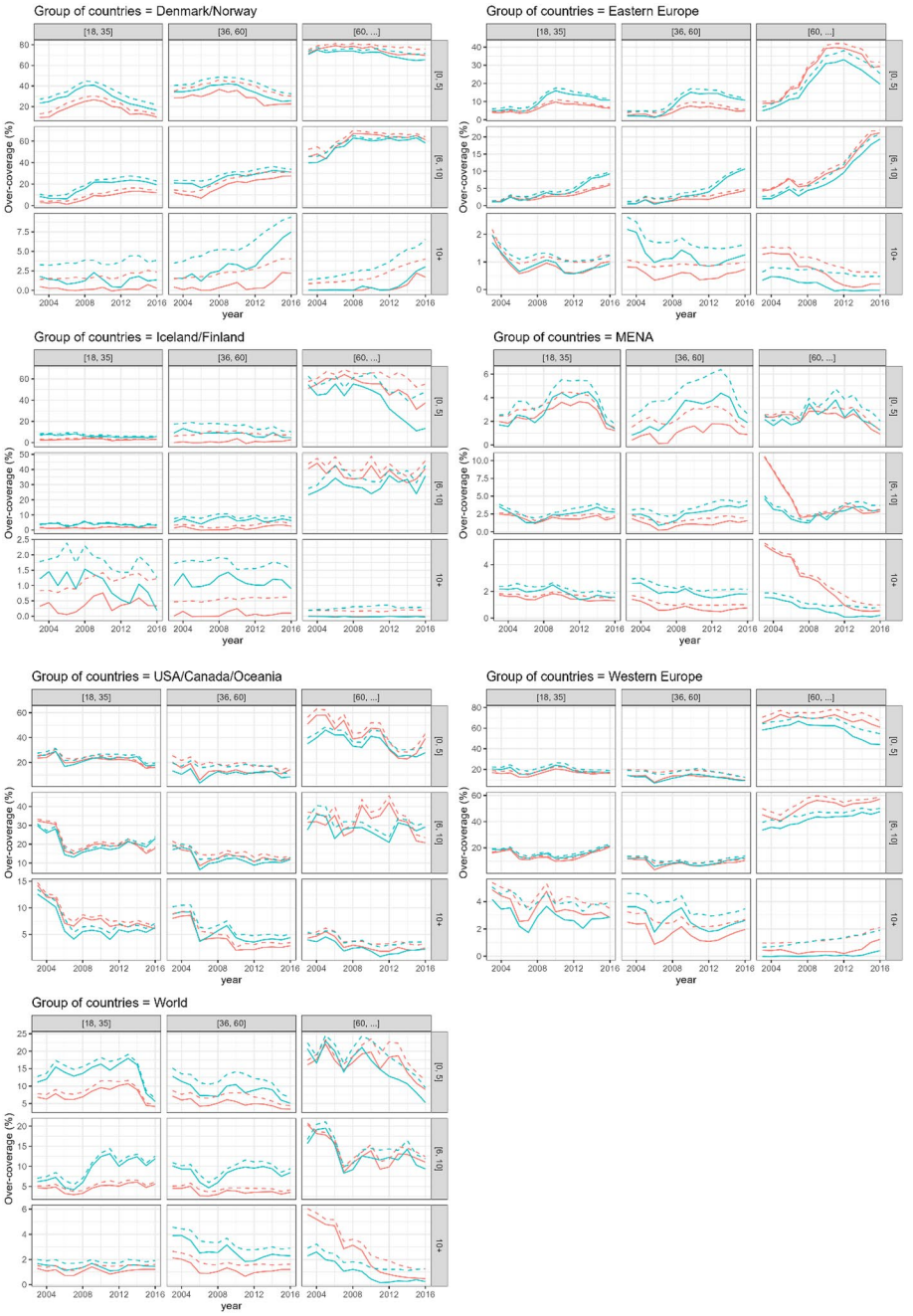

Fig. 3 Trend of over-coverage in 2003–2016 by combination of sociodemographic characteristics (sex, age, region of birth, and time since migration) *Note* Own elaborations using Swedish population registers women are represented by a red line and men by a blue line

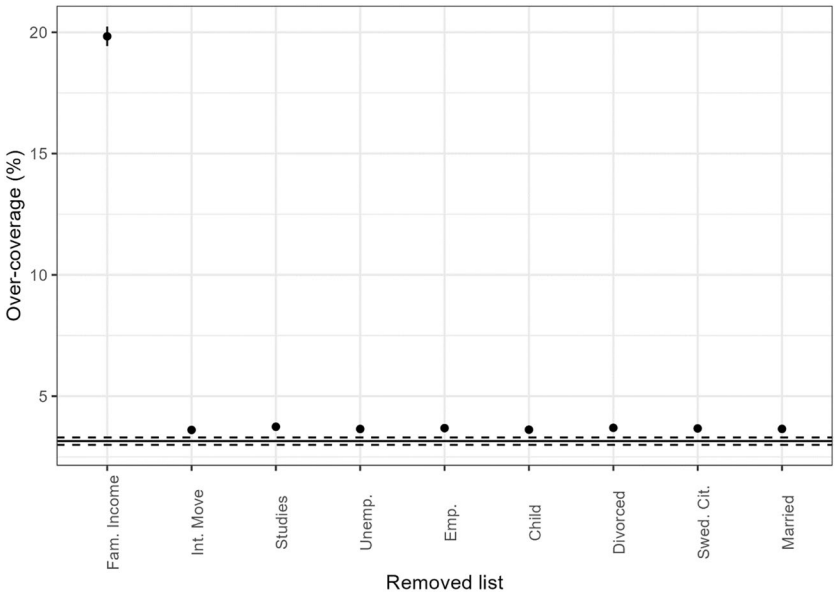

Fig. 4 Overall effect of removing a list in 2015 *Note* our elaborations using Swedish population registers

Canada, even if there is a decreasing trend over time, it appears that those who migrated as children or young adults and who have resided in Sweden for a relatively extended period can still exhibit over-coverage. For migrants coming from MENA, over-coverage is very low, except at the beginning of the observation period and for older migrants who stayed in Sweden for more than 5 years. However, the trend decreases over time. Comparing males to females, over-coverage is usually higher in the latter group in the youngest age groups and generally residing in the country for less than 10 years. This difference is less pronounced in Western European migrants. For most years, the estimate of over-coverage is usually larger for the female population for migrants who are over 60 years, with the exception of Nordic countries.

Figure 4 and Table 2 show the estimates of over-coverage using MSE when each of the lists is removed in turn. In Fig. 4, the solid line shows the estimate with all the lists being considered, while each point denotes the estimate when we remove that particular list. The idea is that both approaches presented in this paper, the existing register-trace and the new MSE approach, require information on a number of lists that might not be accessible to all researchers. Figure 4 shows the contribution of each single list and helps other researchers understand the magnitude of disparity between estimates that are expected when that list is not considered in the analysis. In the x-axis, the lists are ordered according to the increasing lower bound of the population size (LBP) values when each list is removed, i.e., when we remove household income we have the lowest number of individuals appearing in at least one list, while when we remove information about marriages we have the largest number of individuals seen at least once. The largest effect in terms of the estimate of over-coverage is observed when the largest

list, which is household income, is removed from the analysis. Not including this list would lead to an estimate of over-coverage of more than 15% in Sweden in 2015. Similar results have been found for all the years studied in this paper. Overall, removing any list would give higher estimates of over-coverage (difference between the line and the different dots), though this result can be different when we look at all the possible combinations of the socio-demographical variables, available in the appendix (Fig. 8). The ordering of LBP when we remove each list does not explain entirely the biases observed in Fig. 4 (values of this figure available in the appendix, Table 2). For instance, although the analysis without the register of internal moves gives the second lowest LBP, the bias without this list is not the second largest, as this seems to be the case with the removal of the register related to studies. As demonstrated in our simulation study, shown in the supplementary material, the effect of list omission is influenced not only by the size of the list, or the level of overlap between lists, but also whether the omitted list interacts with the lists included in the analysis.

However, the effect of omitting different lists also varies according to socio-demographic characteristics. Figure 5 shows that the exclusion of the household income, in particular, affects the estimation of over-coverage by age, duration of stay, and sex. Older migrants are less economically active and the omission of this list would over-estimate the over-coverage, while the impact for younger individuals is more similar to the other lists. Despite over-coverage being higher among new and medium-term migrants, the effect of omitting the family household list is much stronger for longstanding migrants. Not surprisingly, inference on the number of women present is more affected by the removal of this specific list. If we examine country of birth, this list appears less relevant for migrants that from MENA and other parts of the world, while its effect is greater for those coming from Iceland and Finland.

To validate our approach, demonstrate its generalizability and highlight the effects of list omission under different scenarios, we present an extensive simulation study in the supplementary material where we simulated data for three different sizes of lists, one small, one medium and one large in terms of the proportion of individuals that appear in each list. Our findings demonstrate that, if the lists are independent of each other, that is if the probability of appearing in a list does not depend on whether an individual has appeared in any other list, then removing any of the lists does not introduce bias in the estimation. However, it increases uncertainty, as expected, since removing a list leads to a lower number of detected individuals. On the other hand, if there is dependence between the list that is removed and at least one of the other two lists that are considered in the model, then there can be substantial bias in the estimation of the population size, and hence of over-coverage as we define it in this paper. Our simulation shows that the size of the bias depends on both the size of the list that is removed as well as on the size of the dependence between the lists, while the direction of the bias is not consistent across the different scenarios that we considered. Therefore, our results demonstrate that, as is the case with any statistical model, the validity of the results depends on whether the model is correctly specified, which in the case of log-linear models for contingency tables is linked to the dependence between lists and whether that is being modelled correctly.

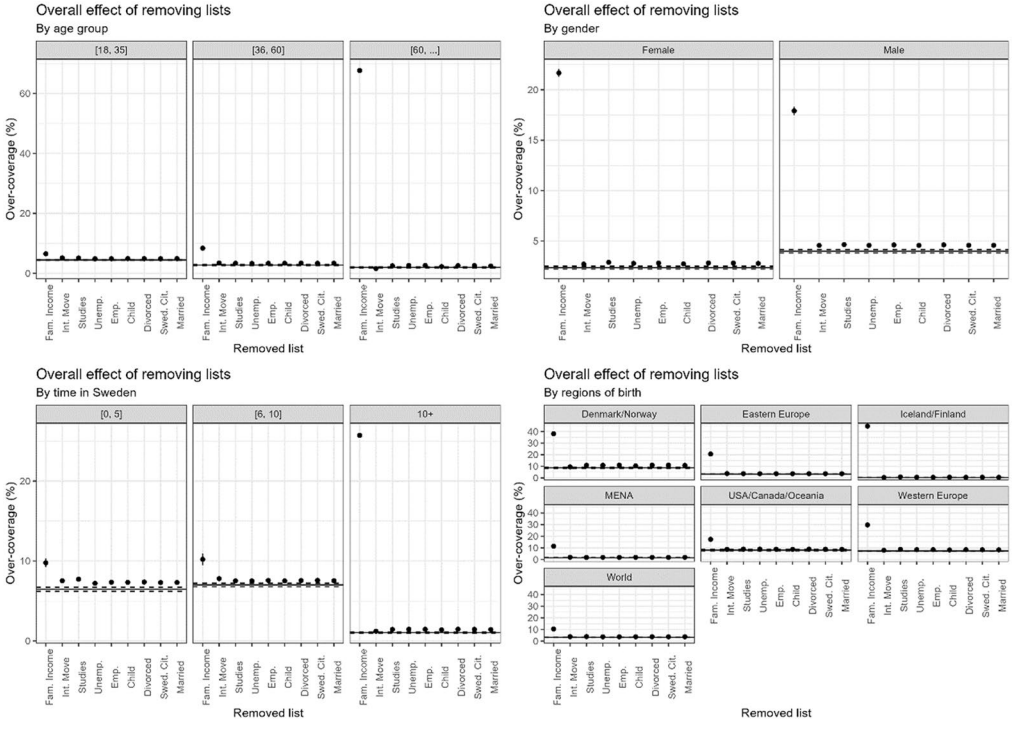

Fig. 5 Overall effect of removing a list in 2015 by socio-demographic characteristics *Note* our elaborations using Swedish population registers

5 Discussion

This study contributes to the well-known challenge of measuring international migration stocks and flows (Bilsborrow et al 1997; Bijak and Wiśniowski 2010; Willekens 1994, 1999). In particular, we aim to estimate the over-coverage of the migrant population in the Swedish Population Registers and to identify heterogeneity in over-coverage in sub-populations. This is a timely problem because it has been argued that the change and the proliferation of the migration process with increases in return-, onward-, and circular migration (Castles et al. 2009; Jeffery and Murison 2011) might affect the count of the migrant resident population and the potential for errors. This is due to the fact that in many countries, including Sweden, incentives to register emigration are comparably low. Additionally, many countries are moving towards register-based data collection systems including register-based censuses. Thus, the Swedish case, with high quality and long-term experience with population registers and a comparably high share and heterogeneity of immigrants offers a great case study to estimate over-coverage.

Using a multiple systems estimation (MSE) approach, we estimate the number of undetected individuals each year, that is a direct measure of over-coverage, and compare it with the cross register-trace approach used in the literature (Monti et al 2020; Statistic Sweden 2018), and study the variation across socio-demographic characteristics.

This paper contributes to the literature in several ways. First, we demonstrated how the general class of log-linear models within the MSE framework can be used to estimate the number of individuals that were not observed in a given year using population registers. These models are typically employed in the context of populations that are difficult to be monitored, such as drug users or victims of modern slavery (King et al. 2014; Silverman 2020). However, their use with population registers, which as previously noted are now widely used in social science, is uncommon.

The use of these models with population registers brings new challenges to this modelling effort, as the choice of lists and covariates has an impact on the estimation process. Different than in other applications of MSE, the number of lists available is unusually large, which increases the number of parameters to be estimated, taking into account all the possible interactions between the lists and covariates. In our application, the choice of lists and covariates was based on prior knowledge of the system, as well as on practical considerations. The removal of deaths and emigration as one of the lists, for instance, was due to the fact that these lists did not have any intersection with others by construction and this hampered the estimation of the interaction terms. We considered all two-way interactions between covariates (including lists), but no higher-order interactions, and did not consider variable-selection methods. The models fitted for each year are considerably high-dimensional ($300+$ parameters/latent variables), so any higher-order interactions would substantially increase computational cost. The R package we employed includes the option of Bayesian variable selection through reversible jump MCMC (Green 1995), which can be used to explore the model space. However, in our case, the dimension of the contingency table and hence of the model space resulted in very slow mixing of the algorithm and posterior model probabilities that were low ($<20\%$) for all models, hence not providing support for any particular model in this case.

Second, this paper not only estimates the prevalence of over coverage and socio-demographic variations, it also includes an intersectional perspective considering all these aspects in the same model. Our results are in line with previous studies (Monti et al 2020; Statistic Sweden 2018). We found large variations in over-coverage by region of origin and

we confirm that a higher prevalence of over-coverage can be observed in the context of free mobility and in particular among neighboring countries such as Norway and Denmark, where it is easy to regularly, or even daily, cross the border. Despite Nordic agreements in place to automatically report all intra-Nordic migration, this is only partially reflected by our results. On one hand, we have low over-coverage levels for Finnish and Icelandic-born migrants. Conversely, in the cases of Norway and Denmark —countries not only in close geographic proximity to Sweden but also with population centers favorably situated to facilitate daily migration and transnational life —the system appear to be less effective in detecting a significant number of erroneous registrations.

Previous research discussed that errors in the registration system might accumulate at older ages (Monti et al 2020), however, this current study shows that over-coverage is lowest among individuals 60 + and among migrants who resided in Sweden for more than 10 years.

Third, this study investigated the effect of omitting each list in turn in terms of estimation of over-coverage. In this case, removing household income, the largest list considered, also had the largest effect. The generalizability of this third finding may be subject to scrutiny. As of now, a data infrastructure set up that also includes a household or family list, is common only in the Nordic countries and few other select countries. Nonetheless, we anticipate this will not remain the case in the near future, with many countries (e.g. Italy) progressing towards a similar population register collection infrastructure. Additionally, our simulation study confirmed that estimation using MSE is reliable when the model is correctly specified, regardless of what included lists represent. Therefore, this approach holds potential applicability in different contexts, as it does not solely rely on this specific list of lists but just on a set of incomplete yet overlapping lists.

Fourth, comparing the deterministic approach, such as the register-trace, with our MSE approach, we are able to estimate the number of false positives, which is individuals that despite not being active in Sweden they are still living there and would be classified as over-covered by existing approaches. This is an advantage of the MSE model, which allows us to estimate probabilities of being observed in the country, based on the information of age, duration of stay and region of origin.

Lastly, this paper also contributes to the political debate: first it shows how over-coverage impacts our understanding of society, underscoring the necessity for knowing the population under study and the administrative processes of data collection to mitigate potential bias. Additionally, it contributes to the debate on the suitability of the definition of a 'usual resident'. Showing how definitions and administrative processes become greatly important if we consider residents in regions that are close to the national borders, or transnationally mobile people. As such, the findings are relevant in the context of compiling data and making statistical comparisons between countries within the European Union, as it illustrates the ambiguities behind definitions related to *de jure* and de facto populations. Presently, register-based systems of population data build largely on those legally registered in a country (e.g. Sweden), whereas the EU definition relies on who can de facto be considered living in a country. Thirdly, this paper demonstrates how existing registers can work to overcome issues related to *de jure*/de facto definitions, by showing how certain individuals

are de facto active though not registered as such in a country. Hence, the paper showcases the potential of population registers for self-validation.

Appendix

See Figs. 6, 7, 8 and Tables 1 and 2.

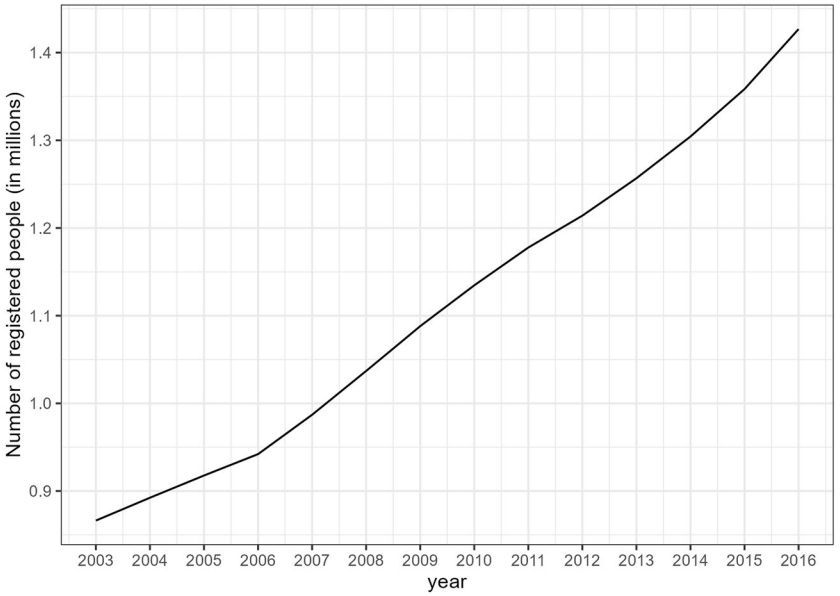

Fig. 6 Total foreign-born population aged 18 and older present in RTB at the end of each year *Note* Our elaborations using Swedish population registers

Fig. 7 Individuals by the combination of the sociodemographic characteristics *Note* our elaborations using Swedish population registers women are represented with a red line and men with a blue line

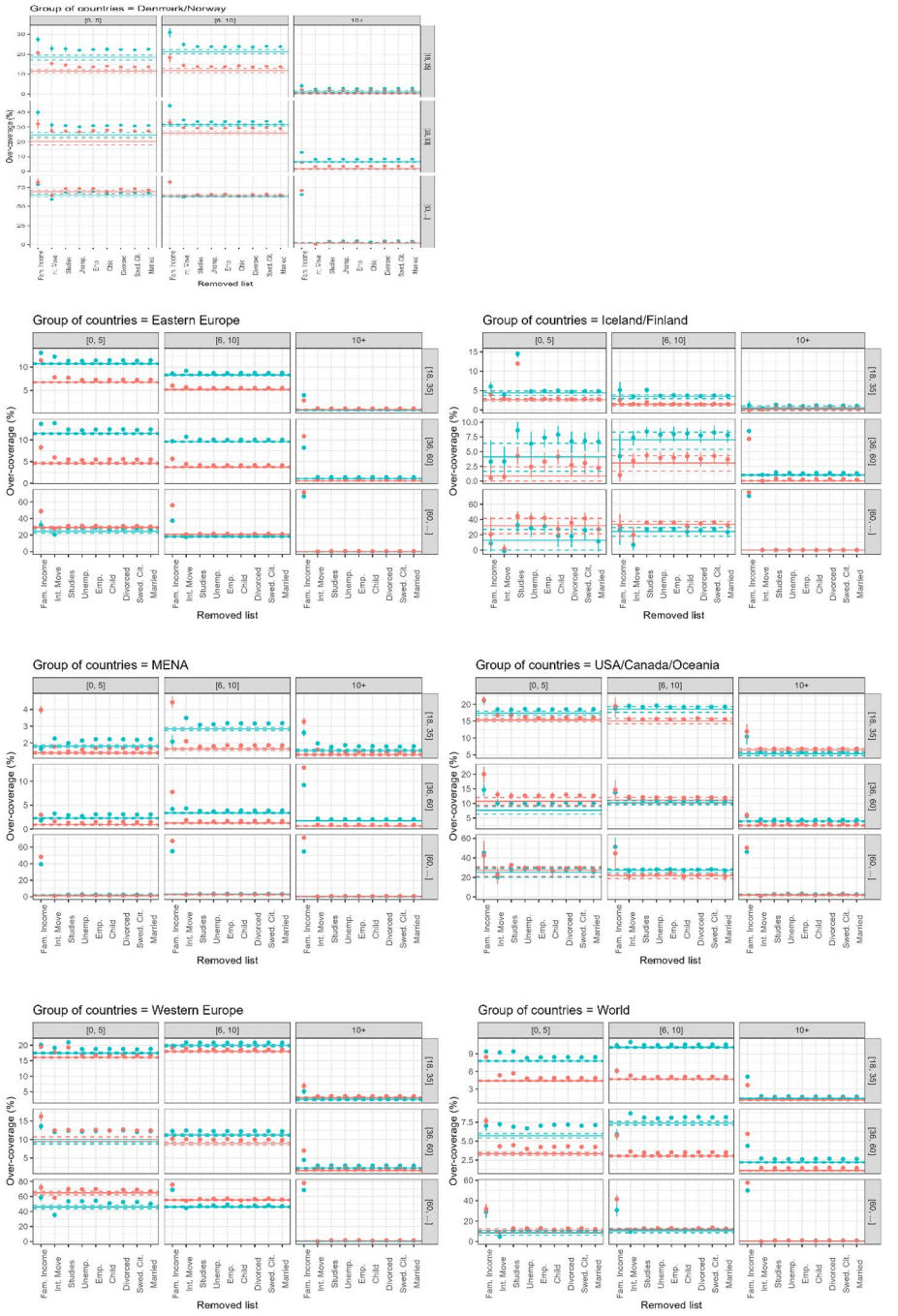

Fig. 8 Overall effect of removing a list in 2015 by combination of socio-demographic characteristics *Note* our elaborations with Swedish population registers women are represented with a red line and men with a blue line

Table 1 Lists and original source of data

Indicator (i.e. list, activity)	Administrative register	Variables	
Child born	Intergenerational register	Time constant	Intergenerational identifier
	Birth register [Födelseuppgifter]	Time constant	Birth year of child
Obtaining Swedish citizenship	Total Population Register [RTB]	Annual	Citizenship if different from year-1 [medlandgrp]
Marriage			Civil status if different from year-1 [civil]
Divorce			
Internal move	Internal moves register [Inrikes flyttar]	Event	Year of internal move
Family income	Longitudinal Integrated Database for Health Insurance and Labour Market Studies [LISA]	Annual	Household identifier [lopnr_famid]
			Income variables:
			Old age pensions [aldpens]
			Unemployment benefits [arblos; ampol]
			Parental leave benefit [forled]
			Employment related earnings [forvers]
			Social benefits/allowances [socink; bostbidrpersf; bostbidrpersf04; socbidrfam]
			Student allowances [stud]
			Sick leave [fortid; sjukre; sjukpp]
Employment			Employment [syssstatg; syssstat; syssstatj; syssstat11]
Unemployment			Swedish Public Employment Service [alkod; iakod; ak14dag; adeldag; astudag; asysdag; alosdag]
Higher education			Enrollment in higher education [studdelt]

Table 2 Overall effects of removing a list in 2015

List removed	Register trace%	MSE approach%	95% Credible interval	
None	3.9565	3.1559	3.3021	3.0014
Fam. Income	24.3844	19.8357	20.2345	19.4359
Int. Move	4.3886	3.6129	3.7719	3.4598
Studies	4.1838	3.7449	3.8560	3.6312
Unemp	4.1156	3.6523	3.7629	3.5435
Emp	4.0505	3.6879	3.7848	3.5922
Child	3.9900	3.6239	3.7221	3.5207
Divorced	3.9874	3.6997	3.7857	3.6098
Swed. Cit	3.9793	3.6749	3.7619	3.5818
Married	3.9790	3.6551	3.7472	3.5567

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11135-023-01757-x.

Author contributions EMu initiated the paper and drafted the manuscript. Together with EMa they developed the analytical strategy. BS developed the Statistical model with the support of EMa and they wrote the methods section in the manuscript. AM created the dataset with the support of SD. All authors contributed to the study conception and design, as well all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open access funding provided by Stockholm University. This research was also supported by the Swedish Research Council for Health, Working Life and Welfare (FORTE), grant numbers 2016-07105; and the Swedish Research Council (VR), grant 2021-00875.

Data availability This study is produced under the Swedish Statistics Act, where privacy concerns restrict the availability of register data for research. Aggregated data can be made available by the authors, conditional on ethical vetting. Code to fully reproduce all the results of the simulation study is provided.

Declarations

Conflict of interests No competing interest.

Ethical considerations The article is based on human demographic data, derived from administrative Swedish register data of the total population of Sweden. Research on this data falls under the act on ethical review and the personal data act and associated rules and guidelines. This research has been approved by a Swedish national ethical review board, and data have been made available by Statistics Sweden. All the information in the Statistics Sweden data source has been anonymized and contains no direct identifying information.

References

Alderotti, G., Mussino, E., Comolli, C.L.: Natives' and migrants' employment uncertainty and childbearing during the great recession: a comparison between Italy and Sweden. Eur. Soc. **25**(4), 539–573 (2022). https://doi.org/10.1080/14616696.2022.2153302

Andersson, G., Monti, A., Kolk, M.: Vem bor här? En ESO-rapport om gamla och nya folkräkningar. Rapp. till Expertgr. Stud. i Offent. Ekon. **2023**, 2 (2023)

Aradhya, S., Scott, K., Smith, C.: Repeat immigration: a previously unobserved source of heterogeneity? Scand. J. Public Health **45**(Suppl 17), 25–29 (2017). https://doi.org/10.1177/1403494817702334

Bijak, J., Wiśniowski, A.: Bayesian forecasting of immigration to selected European countries by using expert knowledge. J. r. Stat. Soc. Ser. A Stat. Soc. **173**(4), 775–796 (2010). https://doi.org/10.1111/j.1467-985X.2009.00635.x

Bijak, J., Bryant, J., Golata, E., Smallwood, S.: Preface Introduction to the special issue on population statistics for the 21st century. J. off. Stat. **37**(3), 533–541 (2021)

Bilsborrow, R.E., Hugo, G., Oberai, A.S.: International migration statistics guidelines for improving data collection systems. International Labour Organization, Switzerland (1997)

Bird, S.M., King, R.: Multiple systems estimation (or capture-recapture estimation) to inform public policy. Annu. Rev. Stat. Appl. **2**(5), 95–118 (2018). https://doi.org/10.1146/annurev-statistics-031017-100641

Bohning, D., Van der Heijden, P.G.M., Bunge, J. (eds.): Capture-recapture methods for the social and medical sciences. CRC Press, Florida (2017)

Brown, J., Abbott, O., Smith, P.A.: Design of the 2001 and 2011 census coverage surveys for england and wales. J. r. Stat. Soc. Ser. A Stat. Soc. **174**, 881–906 (2011). https://doi.org/10.1111/j.1467-985X.2011.00697.x

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

Cruyff, M., van Dijk, J., van der Heijden, P.G.M.: The challenge of counting victims of human trafficking: Not on the record: a multiple systems estimation of the numbers of human trafficking victims in the Netherlands in 2010–2015 by year, age, gender, and type of exploitation. Chance **30**, 41–49 (2017)

Cruyff, M., Overstall, A., Papathomas, M., McCrea, R.: Multiple system estimation of victims of human trafficking: model assessment and selection. Crime Delinq. **67**(13–14), 2237–2253 (2021)

De Haas, H., Castles, S., Miller, M.J.: The age of migration: international population movements in the modern world. Guilford Press, New York (2019)

Green, P.J.: Reversible jump markov chain monte carlo computation and bayesian model determination. Biometrika **82**(4), 711–732 (1995)

van der Heijden, P.G.M., Smith, P.A., Whittaker, J., Cruyff, M., Bakker, B.F.M.: Dual- and multiple-system estimation with fully and partially observed covariates. In: Zhang, L., Chambers, R.L. (eds.) Analysis of integrated data, pp. 137–168. CRC Press, Florida (2019)

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

King, R., Bird, S.M., Overstall, A.M., Hay, G., Hutchinson, S.J.: Estimating prevalence of injecting drug users and associated heroin-related death rates in England using regional data and incorporating prior information. J. r. Stat. Soc. Ser. A **77**, 209–236 (2014)

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

Ludvigsson, J.F., Almqvist, C., Bonamy, A.K., Ljung, R., Michaelsson, K., Neovius, M., Stephansson, M., Ye, W.: Registers of the Swedish total population and their use in medical research. Eur. J. Epidemiol. **31**(2), 125–136 (2016). https://doi.org/10.1007/s10654-016-0117-y

Monti, A.: Re-emigration of foreign-born residents from Sweden: 1990–2015. Popul. Space Place **26**(2), e2285 (2020). https://doi.org/10.1002/psp.2285

Monti, A., Drefahl, S., Mussino, E., Härkönen, J.: Over-coverage in population registers leads to bias in demographic estimates. Popul. Stud. **74**(3), 451–469 (2020). https://doi.org/10.1080/00324728.2019.1683219

Overstall, A.M., King, R.: Conting: an R package for bayesian analysis of complete and incomplete contingency tables. J. Stat. Softw. **58**(7), 1–27 (2014). https://doi.org/10.18637/jss.v058.i07

Pettersen, S.V., Åmberg, J., Tønnessen, M., Vassenden, K.: Underestimation of emigration—an alternative approach. Presentation at the International Forum on Migration Statistics, Paris, January 2018 (2018)

Qi, H., Irastorza, N., Emilsson, H., Bevelander, P.: Integration policy and refugees' economic performance: evidence from Sweden's 2010 reform of the introduction programme. Int. Migr. **59**, 42–58 (2021). https://doi.org/10.1111/imig.12813

Qvist, J.: Täckningsproblem I Registret over totalbefolkningenRTB---Skattning av övertäckning med en indirectt 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)

Rampazzo, F., Bijak, J., Vitali, A., Weber, I., Zagheni, E.: A framework for estimating migrant stocks using digital traces and survey data: an application in the United Kingdom. Demography **58**(6), 2193–2218 (2021). https://doi.org/10.1215/00703370-9578562

Righi, P., Falorsi, P.D., Daddi, S., Fiorello, E., Massoli, P., Terribili, M.D.: Optimal sampling for the population coverage survey of the new italian register based census. J. off. Stat. **37**(3), 655–671 (2021). https://doi.org/10.2478/jos-2021-0029

Salentin, K.: Sampling the ethnic minority population in Germany. the background to "migration background." Methods Data Anal. **8**(1), 25–52 (2014). https://doi.org/10.12758/mda.2014.002

Sharifi Far, S., King, R., Bird, S., Overstall, A., Worthington, H., Jewell, N.: Multiple systems estimation for modern slavery: robustness of list omission and combination. Crime Delinq. (2020). https://doi.org/10.1177/0011128720951429

Silverman, B.W.: Multiple-systems analysis for the quantification of modern slavery: classical and Bayesian approaches. J. r. Stat. Soc. A **183**, 691–736 (2020). https://doi.org/10.1111/rssa.12505

Statistics Denmark: Undervurdering av udvandringer. Note by Thomas Nielsen (TMN) 7. February 2014 (2014)

Statistics Sweden: Background Facts, Population and Welfare Statistics. Multi-generation register 2016. A description of contents and quality. https://www.scb.se/contentassets/95935956ea2b4fa9bcaab51afa259981/ov9999_2016a01_br_be96br1702eng.pdf (2017b).

Statistics Sweden: LISA, Longitudinal integrated database for health insurance and labour market studies. https://www.scb.se/contentassets/f0bc88c852364b6ea5c1654a0cc90234/lisa-bakgrundsfakta-1990-2017.pdf (2019). Accessed on

Statistics Sweden: Flyttningar inom kommun, inom län och mellan län 2002—2021. Demografiska rapporter 2022:10. https://www.scb.se/contentassets/1a0f506a930d4df992f9689cecb01c3a/be0701_2022a01_br_be51br2210.pdf (2022)

Statistics Sweden: Overcoverage in the total population register—a register study. Backgr. Facts. Popul. Welf. **2015**, 1 (2015)

Statistics Sweden: Background Facts, Population and Welfare Statistics. Multi-generation register 2016. A description of contents and quality. https://www.scb.se/contentassets/95935956ea2b4fa9bcaab51afa259981/ov9999_2016a01_br_be96br1702eng.pdf (2017b)

Statistics Sweden: Census 2011 round (cens_11r) National Reference Metadata in Euro SDMX Metadata Structure (ESMS). https://ec.europa.eu/eurostat/cache/metadata/EN/cens_11r_esmscs_se.htm (2013)

Statistics Sweden: Description of the Register: The Total Population Register. https://www.scb.se/contentassets/8f66bcf5abc34d0b98afa4fcbfc0e060/rtb-bar-2016-eng.pdf (2017a)

Statistics Sweden: The Registration Bias—A Methodological Report on the Estimation of Over-coverage, Under-coverage and Registration at the Wrong Address. (Swedish: "Folkbokföringsfelet. En metodrapport om skattning av övertäckning, undertäckning och folkbokförda på fel adress.") Örebro: SVB BV/REG and PMU/MIÖ (2018)

Statistics Sweden: Flyttningar inom kommun, inom län och mellan län 2002—2021. Demografiska rapporter 2022:10. https://www.scb.se/contentassets/1a0f506a930d4df992f9689cecb01c3a/be0701_2022a01_br_be51br2210.pdf (2022)

Statistics Sweden. https://www.scb.se/en/finding-statistics/statistics-by-subject-area/population/population composition/population-statistics/ (2023)

Statistics Sweden: LISA, Longitudinal integrated database for health insurance and labour market studies. https://www.scb.se/contentassets/f0bc88c852364b6ea5c1654a0cc90234/lisa-bakgrundsfakta-1990-2017.pdf (2019)

Swedish National Audit Office: Population registration—a quality work uphill. NAO 2017:23. (Swedish: Riksrevisionen: Folkbokföringen, - ett kvalitetsarbete i uppförsbacke. RIR 2017:23) (2017)

Syse, A., Strand, B.H., Næss, Ø., Steingrímsdóttir, Ó.A., Kumar, B.: Differences in all-cause mortality: a comparison between immigrants and the host population in Norway 1990–2012. Demogr. Res. **34**(1), 615–656 (2016). https://doi.org/10.4054/DemRes.2016.34.22

Swedish National Audit Office: Population registration—a quality work uphill. NAO 2017:23. (Swedish: Riksrevisionen: Folkbokföringen, ett kvalitetsarbete i uppförsbacke. RIR 2017:23) (2017)

Swedish Tax Authorities: Quality Control in the Total Population Register. Nr. 200 398792–18/113. (Swedish: Skatteverket. Kvalitetsuppföljning i folkbokföringsregistret. Diarienummer 200 398792–18/113) (2018)

Turra, C.M., Elo, I.T.: The impact of salmon bias on the Hispanic mortality advantage: new evidence from social security data. Popul. Res. Policy Rev. **27**(5), 515–530 (2008). https://doi.org/10.1007/s11113-008-9087-4

Wallace, M., Kulu, H.: Low immigrant mortality in England and Wales: a data artefact? Soc Sci Med **120**, 100–109 (2014). https://doi.org/10.1016/j.socscimed.2014.08.032

Wallace, M., Wilson, B.: Age variations and population over-coverage: is low mortality among migrants merely a data artefact? Popul. Stud. **76**(1), 81–98 (2022). https://doi.org/10.1080/00324728.2021.1877331

Weitoft, G.R., Gullberg, A., Hjern, A., Rosen, M.: Mortality statistics in immigrant research: method for adjusting underestimation of mortality. Int. J. Epidemiol. **28**(4), 756–763 (1999). https://doi.org/10.1093/ije/28.4.756

Willekens, F.: Monitoring international migration flows in Europe. Eur. J. Popul. **10**, 1–42 (1994). https://doi.org/10.1007/BF01268210

Willekens, F.: Modeling approaches to the indirect estimation of migration flows: from entropy to EM. Math. Popul. Stud. **7**(3), 239–278 (1999). https://doi.org/10.1080/08898489909525459

Willekens, F.: Evidence-Based Monitoring of International Migration Flows in Europe J. Off. Stat. **35**(1), 231–277 (2019). https://doi.org/10.2478/jos-2019-0011