



Using Swedish administrative registers to study immigrants and their descendants

Identifying immigrants, measuring their migration background, and linking data across generations

Ben Wilson



Stockholm
University

Using Swedish administrative registers to study immigrants and their descendants:

Identifying immigrants, measuring their migration background, and linking data across generations

Ben Wilson ^{1,2}

1: Department of Sociology, Stockholm University

2: Department of Methodology, London School of Economics

Abstract

This article extends the work of prior publications and publicly available documents to examine in detail how Swedish register data can be used to study immigrants and their descendants. Although there is a wealth of prior research on immigrants and their descendants in Sweden, there is a lack of research that has demonstrated how Swedish administrative registers can be used to identify these populations and measure different aspects of their migration background. There are a large number of registers that can be used to study this topic in Sweden but here we focus on the most common registers that are used to define populations and measure their key demographic characteristics. We examine different aspects of migration background—including country of birth, immigration and emigration events, citizenship and residence permits—as well as parental migration background based on intergenerational linkages. Throughout, we provide examples and analysis using collections of data that are available to researchers as part of the REFU-GEN project. We include new analysis to illustrate key issues with respect to data quality and the strengths and weaknesses of the Swedish register data for research on this topic. However, our aim is that the guide is useful for anyone seeking to understand administrative data, not only in Sweden but also in other countries.

Keywords: *Immigrants, descendants of immigrants, immigration, emigration, country of birth, citizenship, residence permits, refugees, register data, Sweden*

Stockholm Research Reports in Demography 2024:16

ISSN 2002-617X

© Ben Wilson



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Contents

1.	Introduction.....	4
2.	Coverage of the population.....	8
3.	Using register data to measure migration background.....	11
4.	Country of birth.....	13
5.	Immigration and emigration events	15
6.	Citizenship	19
7.	Residence permits and categories of admission.....	20
8.	Intergenerational linkages	25
9.	Descendants of immigrants and their parental migration background.....	29
10.	Conclusion	31
11.	References.....	32
12.	Appendix.....	34

Acknowledgements:

There are many people who should be acknowledged in contributing to the collective knowledge that this article is based upon, including (in alphabetical surname order): Gunnar Andersson, Linus Andersson, Siddhartha Aradhya, Erik Carlsson, Sven Drefahl, Daniela Foresta, Julie Fournier, Raffaele Grotti, Lisa Harber-Aschan, Wooseong Kim, Anastasia Lam, Andrea Monti, Eleonora Mussino, Elena Pupaza, Caroline Uggla, Matthew Wallace, Rosa Weber, and Frankseco Yorke. With thanks to these colleagues and the many other researchers who have been affiliated, or working, with Stockholm University Demography Unit (SUDA) and have dedicated time and effort to sharing their knowledge about Swedish register data. This research is partly funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, grant agreement No 948727 (REFU-GEN), as well as the Swedish Research Council for Health, Working life and Welfare (FORTE), grant numbers: 2016-07105, 2018-00310 and 2019-00603; and the Swedish Foundation for Humanities and Social Sciences (Riksbankens Jubileumsfond, RJ), grant registration number M18-0214:1.

1. Introduction

As with similar data in other contexts, Swedish register data holds a number of advantages when used to study immigrants and their descendants. A key advantage is that it covers almost the entire population. This enables high quality research to be carried out without the need for population inference, (although there are some sources of bias, as outlined below). It also facilitates that study of population subgroups that are too small to study when using other data sources (e.g. surveys), such as immigrants from individual countries of birth or with specific characteristics. Whole-population data also enables researchers to create aggregate measures, for example summary measures of area-level characteristics (that might otherwise be estimated using traditional census data), and then to link these aggregate measures to all individuals in the register.

Another advantage of Swedish register data is the fact that everyone in the population registers has a unique identification number, which is provided in pseudo-anonymised format when the data are made accessible for research. This identification number is consistent across registers (i.e. it refers to the same person in different registers), which enables data from different sources to be linked for the same individual so that researchers can carry out longitudinal analysis. In addition, the way that identification numbers are administered makes it possible to link individuals with each other, for example if they share the same household or are parents of the same child. The benefits of this are discussed further below, but they not only include the ability to link immigrants with their descendants, but also to link longitudinal measurements of their trajectories in different domains of life.

The use of Swedish register data for research purposes requires ethical approval for a specific research project, and is based on collections of administrative data that are provided by government agencies, in particular Statistics Sweden.¹ There are a range of registers that researchers can gain access to, including data on health, education, income, and many other aspects of life. This includes the core demographic data that most researchers use to study immigrants and their descendants, such as age, sex and migration background. However, although there is a wealth of prior research that has examined the migration background of the Swedish population, there is a lack of research that has shown how Swedish administrative registers can be used to identify immigrants and their descendants, in particular when considering the complex task of processing the data prior to statistical analysis. At first glance, the measurement of migration background may appear to be an easy task, especially for those unfamiliar with register data. Yet, as we demonstrate here, this is far from the case.

The aim of this article is therefore to show how Swedish administrative registers can be used to study immigrants and their descendants, with an emphasis on identifying immigrants, measuring their migration background, and linking data across generations. We build on publicly available documents and prior publications, such as those that describe the use of Swedish register data for medical research (e.g. Ludvigsson et al. 2016), and a recent study that assesses the use of Swedish and Danish registers for migration research (Careja and Bevelander 2018).² In contrast to what is available at the time of

¹ In order to obtain Swedish register data, or access to these data via Statistics Sweden's secure remote-access system (MONA), researchers currently need to apply for ethical approval at least twice: once from the Swedish Ethical Review Authority (Etikprövningsmyndigheten) and once from Statistics Sweden, in addition to additional approvals that may be required if data are requested from other agencies, such as data from the National Board of Health and Welfare (*Socialstyrelsen*) (Etikprövningsmyndigheten 2023; Ludvigsson et al. 2015, 2016; National Board of Health and Welfare 2023; Statistics Sweden 2023a). Statistics Sweden obtain and process data from various other agencies, much of which can be provided without additional ethical approval from those agencies. Most researchers who use Swedish register data do so via the collections of data that are organized by Statistics Sweden (SCB) and/or the National Board of Health and Welfare (Socialstyrelsen, who are responsible for the majority of registers relating to health and have a separate ethical review process). As it stated on SCB's website in October 2023: "Statistics Sweden's microdata on individuals and enterprises is covered by secrecy, but you can access anonymised microdata following a confidentiality assessment, provided that Statistics Sweden considers that you have the grounds to process the data" (for more info, see: <https://www.scb.se/en/services/ordering-data-and-statistics/ordering-microdata/>).

² Most researchers who use Swedish register data do so via the collections of data that are organized by Statistics Sweden (SCB) and/or the National Board of Health and Welfare (Socialstyrelsen). As it stated on SCB's website

writing, we focus on the Swedish register data and examine different aspects of migration background—including country of birth, immigration and emigration events, citizenship and residence permits—as well as parental migration background based on intergenerational linkages. Throughout, we provide examples and analysis (using collections of data that are available to researchers as part of the REFUGEN project), including new analysis to illustrate key issues with respect to data quality and the strengths and weaknesses of the Swedish register data for research on this topic. Even when familiar with register data, it is useful to understand the quality of register data in relation to the topic under examination. At the same time, our aim is that this article is useful for anyone seeking to understand administrative data, in Sweden and in other countries.

1.1. Different sources of register data

A common question among new users of register data is ‘*when does the data begin?*’. However, this is not an easy question to answer. It is hard to say exactly when the register data begin because they have their origin in disparate administrative sources that have been held, historically, by different government agencies. Some key events in the history of the Swedish registers are shown in *Table 1.1*, although more detailed information is available elsewhere (G. Andersson et al. 2023). The collection of population data and the administration of censuses in Sweden has a long history, stretching back to the middle of the 18th Century (G. Andersson et al. 2023). Yet it was not until the middle of the 20th Century that the first foundations for a population register were put in place.

The ability to link data across different registers (i.e. different administrative sources) is dependent upon the existence and maintenance of unique identification numbers. Since 1947, the Swedish authorities have recorded a ten-digit personal identity number for all individuals who have resided in Sweden (Statistics Sweden 2017a). This population registration system was computerised in 1967, and then used by Statistics Sweden in 1968 to establish an electronic register of the Swedish population—the Total Population Register (RTB)—which contains information about the population and how it changes over time. Since this time, RTB (*‘Registret over TotalBefolkningen’* in Swedish) has formed the basis for the production of Swedish population statistics, and is one of the key sources of data for most researchers using the registers. RTB is based on notifications provided by the Swedish Tax Agency (from their population register), for example notifications about births, deaths, immigration and emigration. The Swedish Tax Agency retains overall responsibility for population registration, with their registration system updated on a daily basis, and information passed to Statistics Sweden as required. For each year, Statistics Sweden prepares an annual version of RTB that includes all members of the population usually resident in Sweden on 31 December. It is annual RTB data that are made available for research, including information on citizenship and country of birth if requested.

Since many research studies rely on RTB—for defining their study population or providing variables for analysis—the first annual RTB file (1968) often represents the earliest period that can be studied. More specifically, 1968 is the start of the annual Swedish stocks data with complete coverage of the usually resident population.³ Yet, it is sometimes possible (and/or desirable) to incorporate data prior to 1968. This is not just for more historical studies, but also when studies include people who were born before 1968, or when trying to make use of intergenerational linkages, for example to study the role of parental background (e.g. using data that exist to identify people born or dying before 1968 and to link them to other members of the population after 1968). As such, each researcher will decide the appropriate starting point of their study, bearing in mind various issues regarding data availability and data quality.

in October 2023: “Statistics Sweden’s microdata on individuals and enterprises is covered by secrecy, but you can access anonymised microdata following a confidentiality assessment, provided that Statistics Sweden considers that you have the grounds to process the data.” <https://www.scb.se/en/services/ordering-data-and-statistics/ordering-microdata/>

³ Although we note that there are some people living in Sweden who are not part of this usually resident population, as defined by Statistics Sweden. There are also stocks data available before 1968, in particular the census data for 1960 and 1965, which are digitised with personal identifiers and available for research.

Table 1.1: Key dates for register data relating to immigrants and their descendants

Period	Event	Description
1947-onwards	Introduction of a centrally maintained unique identifier	The Swedish Tax Agency (Skatteverket) has maintained a ten-digit-identification number ('person number') for all individuals who have resided in Sweden since 1947.
1960s	Digitalisation of population data	Starting in the 1960s, there was a coordinated programme of the digitalisation of population data. Population registration was computerised in 1967, and then used by Statistics Sweden in 1968 to establish a register of the entire population.
1961-onwards	Inclusion of Swedish residents in the Multi-Generation Register	The Multi-Generation Register provides information linking the identification numbers of individuals to those of their parents. To be included, index persons must have been registered by Statistics Sweden at some time since 1961 and born in 1932 or later.
1961-onwards	Existence of comprehensive registers for vital events	Registers on events that took place during the period 1961–1967 exist for births, deaths and emigrations. However, the quality of these events is considered by Statistics Sweden to be lower than for events from 1968 onwards, in part because there is no stock register (until 1968) via which the quality of these events data can be assessed.
1968-onwards	Availability of an electronic population register for the whole population	The Total Population Register (RTB) was created by Statistics Sweden in 1968 and contains information about the population and how it changes over time. For each year, Statistics Sweden prepares an annual version of RTB that includes all members of the population usually resident in Sweden on 31 December. It is annual RTB data that are made available for research, including information on citizenship and country of birth if requested.
1987-onwards	Beginning of the registers on residence permits	Information on admission categories that is held by Statistics Sweden is based on residence permits that are recoded by the Swedish Migration Agency. Based on these records, Statistics Sweden holds a register of annual data on 'reason for residence (<i>grund för bosättning</i> , GFB) from 1987 onwards.
1997-onwards	Beginning of the longitudinal database for integration studies, STATIV	The STATIV database was developed to study different aspects of the population from an integration policy perspective. When the Swedish Integration Board closed on 1 July 2007, the responsibility for the database was transferred to Statistics Sweden. Annual data is available from 1997. (For more information on the variables that are included in STATIV please see <i>Section 3</i>)

Note: See the text for the references that were used to create this table.

In connection with the establishment of RTB, the 1960s began an extensive set of digitalisation activities at Statistics Sweden, some of which concerned data that were added to RTB, and some of which concerns data that are held in separate registers. For example, there are registers of data on births, deaths and emigrations events that took place during the period 1961–1967 (Statistics Sweden 2006), including all deaths in Sweden since 1952 (Brooke et al. 2017). However, the quality of these events is considered by Statistics Sweden to be lower than for events from 1968 onwards, in part because there is no stock register (until 1968) via which the quality of these events data can be assessed. After 1968, the quality of vital events data is considered to be excellent. With respect to immigration, everyone entering Sweden is obliged to register with the authorities within one week of arrival if they intend to stay for more than one year and/or if they are required to do so under the terms of their admission category. From 1967–1997, date of immigration is based on the date of registration with the county administrative board, while from 1998 onwards, date of immigration is either based on date of residence permit or date of registered arrival, with late notifications recorded as such (Statistics Sweden 2006).

The Multi-Generation Register provides information linking the identification numbers of individuals to those of their parents (Ekbom 2011; Statistics Sweden 2017a). To be included, index persons must have been registered by Statistics Sweden at some time since 1961 and born in 1932 or later. Before 1961 there was no population register in automatic data processing (ADP) format (which also explains why vital events are not systematically available prior to 1961). If a person was 15 or younger in 1947 (when the Tax Agency began to maintain the identification numbers), their parents were recorded in their personal record. This means that parental information is only rarely available for those persons born before 1932.

In addition to the Multi-Generation Register, the Total Population Register, and other registers that contain vital events, Statistics Sweden manages a wide variety of registers that contain microdata from administrative registers held by different official agencies. Often these registers are focussed on specific topics, such as those that focus on educational or geographic information. Related to immigration, information on admission categories is held by Statistics Sweden in an annual register of data on ‘reason for residence’ (grund för bosättning, GFB) from 1987 onwards. The source of these data is detailed information on residence permits (categories, decisions and dates) that are recoded by the Swedish Migration Agency before being provided to Statistics Sweden.

At the same time, there are also registers that Statistics Sweden create through a synthesis of different sources, including other (more specific) registers. An example of this is STATIV, a longitudinal collection of register data, available from 1997 onwards, that was developed to study migration and integration. Some information on the variables that are included in STATIV is given in *Section 3*. Another example of a ‘synthesised’ register created by Statistics Sweden is LISA, which is the longitudinal database for health insurance and labour market studies that begins in 1990 (Ludvigsson et al. 2019; Statistics Sweden 2023b).⁴ Both STATIV and LISA essentially consist of annual data that are collected from various registers (including RTB), except with a slightly different focus. Indeed, there is even some overlap between the two.

One additional register worth mentioning is *Historiska befolkningsregistret*—the Historical Population Register (HPR)—which was created by Statistics Sweden as a way of producing population data more efficiently, rather than repeatedly combining data from annual registers like RTB with data from various registers of vital events (Statistics Sweden 2006). The register is essentially a combination of data that measure *stocks* (i.e. individual circumstances at a specific moment in time) and data that measure *flows* (i.e. events that occur), with a focus on measuring the population and producing population statistics (Statistics Sweden 2006). The HPR is often used by Statistics Sweden to define and extract the population when researchers make a request for microdata (as is the case for REFUGEN). It is a flexible register by virtue of its considerable scope, but is essentially an aggregation of data that are also included in more specific registers, such as the annual versions of RTB (Statistics Sweden 2006). There are many other registers that are available for research, and new registers are in the process of being created; either with administrative data that has not previously been used for statistical or research purposes, or through the digitalisation of older data sources such as census or parish registers (some of which are already available, e.g. in Umeå and Skåne). Nevertheless, the aforementioned registers are those that are most likely to be central to the study of immigrants and their descendants, in particular when trying to define study populations and their key characteristics.

⁴ Longitudinell integrationsdatabas för sjukförsäkrings- och arbetsmarknadsstudier (LISA)

2. Coverage of the population

When trying to design, carry out, or interpret research on immigrants and their descendants, a key question is who is (and who is not) part of the population. This is particularly important where researchers seek to make claims about the ‘whole’ population or to harness the strengths of register data. People enter the population register when they are born, if they are born in Sweden, or when they receive a resident permit or register their immigration. All members of the population have a unique identification number, which is available in the data that Statistics Sweden (or other agencies) provide to researchers as a pseudo-anonymized identification number.

To understand who is in the population, it is useful to consider *coverage*, which is the extent to which the observed population matches the ideal (target) population. Depending on its aims, research can be biased when coverage is not achieved sufficiently. Under-coverage refers to the omission of units belonging to the target population (e.g. because their data have not yet been processed), while over-coverage refers to the inclusion of elements that do not belong to the target population (e.g. because they have emigrated but their emigration has not been recorded).

Two aspects of coverage are useful to consider: (1) defining the target population/subpopulation, and (2) defining the target time period (and/or cohort). While the latter may be trivial when using some data sources, it is important when using register data because some time periods do not cover some (or all) groups of the population very well. For example, a lot of information about immigrants who arrived before the 1960s is missing in the Swedish registers, essentially because the Total Population Register (RTB) did not begin until 1968 (and census data before 1960 has yet to be digitised and linked to other registers).

2.1. Population coverage and individual identification numbers

When analysing the data on immigrants and their descendants in Sweden, researchers will make various choices about how to define their study population, where the population can be limited by age, period, birth cohort, arrival year, (parental) country of birth, and numerous other characteristics. However, in making these choices, it is important to recognise that the population may be restricted by the history of the Swedish registers and the coverage of the data. After receiving ethical approval, the collections of register data that researchers are able to access from Statistics Sweden (and/or other agencies) will all vary (in terms of content and coverage), in particular because of differences between research projects (and it is the research project that justifies the ethical approval and the scope of the research that is allowed). Here we describe data that can be made available to researchers who request data on the whole population, although not all projects will have access to such data.

The practical (and conceptual) starting point for many projects—like REFU-GEN—is an exhaustive table of personal identifiers that includes all people who have ever been part of the Swedish population over the study period (i.e. those who are in the collections of register data that are made available to that specific project).⁵ The REFU-GEN project uses data that cover the entire Swedish population from 1968, but also uses data prior to 1968 where these data are available, notably in order to measure the migration background of an individual’s ancestors.⁶ One aim of the project is to study the inequalities that are experienced by immigrants who arrived in Sweden as refugees, as well as their children, including as compared with other members of the population.

⁵ Note that this list can include some individuals who are not part of the usually resident population in any year.

⁶ The initial extract of register data for REFU-GEN was delivered to us (researchers at Stockholm University) by Statistics Sweden in summer 2022. This extract included data up to 17 March 2022, which is the date for which intergenerational relationships were calculated. However, different registers were delivered covering different periods, for example because annual files are not ready until sometime after the end of the year (e.g. although RTB was provided for 2021, LISA was only provided up to and including 2020).

The exhaustive table of personal identifies in REFU-GEN includes almost 17 million observations (16,918,983), the majority of which correspond to a unique pseudo-anonymised identification number (based on the actual Swedish person numbers administered by the Swedish Tax Agency). Although the Swedish population has never been more than 11 million at the time of writing, the number of identifiers is considerably larger because the exhaustive table includes all individuals who have ever lived in Sweden over the study period.⁷ This is illustrated by the summary statistics in *Table 2.1*. In the data made available to researchers under REFU-GEN, there are 16,393,700 individuals (or more specifically, individual identification numbers) who were registered in the population in at least one year from 1968-2021, and 407,283 who were not. This can be contrasted with the total number of identification numbers in the population in a specific year, for example 10,379,263 in 2020.

Table 2.1: Population by year and duplicate status

Coverage	Unique IDs (frequency)	Duplicate IDs (frequency)	Duplicate IDs (% of total)	Total IDs (frequency)
Population 1970	8,080,755	323	<0.01	8,081,078
Population 1980	8,317,423	463	0.01	8,317,886
Population 1990	8,589,493	1,023	0.01	8,590,516
Population 2000	8,880,514	2,167	0.02	8,882,681
Population 2010	9,386,254	29,246	0.31	9,415,500
Population 2020	10,288,785	90,478	0.87	10,379,263
Population 1970-2020	15,904,731	105,742	0.66	16,010,473
Not in population 1970-2020	780,836	9,674	1.22	790,510
Population 1968-2021	16,279,783	113,917	0.69	16,393,700
Not in population 1968-2021	405,784	1,499	0.37	407,283
All register data in REFU-GEN	16,685,567	115,416	0.69	16,800,983

Source: Authors' analysis of Swedish register data.

Table 2.1 also introduces one aspect of register data quality that is rarely discussed in academic articles that use register data. Not all of the 16.9m identification numbers in REFU-GEN's exhaustive table are unique, and this is true of the register data in general. Some identification numbers are duplicates, i.e. they include multiple versions of the same number. To further illustrate the issue, the number of these in the initial delivery of REFU-GEN data are given in *Appendix Table A1.1*. Duplicates have arisen mainly due to changes of identity numbers, and these changes are made because some part of the personal identity number is incorrect. It is the Swedish Tax Agency that decides that a personal identity number should be changed and then Statistics Sweden are notified about this change.

The most common reasons are:

- (a) registration of incorrect birth dates in connection with immigration or birth, and
- (b) recording of incorrect gender in connection with immigration or birth.

Less common reasons include:

- (c) gender reassignment,
- (d) two persons have the same identity number, and
- (e) one person has two personal identity numbers.

⁷ This is not only all those who have been part of the usually resident population at the end of every year (from 1968) but also some individuals who have immigrated and emigrated in the same year, as well as infants who have emigrated or died in the year that they were born.

As shown in *Table 2.1*, the number of duplicates has increased over time. This is largely due to the fact that most duplicates are related to the registration of foreign-born individuals (see also *Table 4.2*). The registration of incorrect birth dates in connection with immigration is one of the most common reasons why an individual's personal identity number needs to be changed.⁸ Thus, persons born outside Sweden are the group with the highest percentage of changes to personal identity numbers. There are many reasons why birth dates may be recorded incorrectly, including administrative errors and reporting errors (e.g. preferences for certain dates, such as the 1st day of the month, when dates are uncertain or unknown).

In the data that are available to researchers, there is a variable to indicate all identity numbers that have changed (*LopNrByte*). In addition, there is a variable to indicate numbers have been reused (*AterAnv*). Reused personal identity numbers are those that appear to have belonged to more than one individual, based upon other information available to Statistics Sweden. There are 26,163 identification numbers in the REFU-GEN data extract that have been reused, almost one third of which are duplicates.⁹ The most common reason for reuse is that a foreign-born individual is assigned the identification number of a previously deceased person or the number that a person had previously but which has been changed. The proportion of reused numbers is larger among those that are not in the population in any year from 1968-2021 (0.33% as compared with 0.12% for those in the population 1968-2021). More information on person numbers is available in published articles (Ludvigsson et al. 2009) and from Statistics Sweden (for some information in English see Statistics Sweden 2006, in particular the appendix, and for information in Swedish see: 2016a).

Unfortunately, it is very difficult to distinguish between duplicates in any specific analysis. This is because it is hard to tell which version of a duplicate is which in different registers. As such, researchers tend to drop individuals with duplicate identification numbers (in a similar way to complete case analysis for missing data). It is also possible to retain duplicates, which in some cases is made easier by an identifier that is provided by Statistics Sweden to show which version of a duplicate is the last person number (*SenPnr*). However, this identifier does not exist in most registers. In any event, we expect that it is rare that the approach to duplicates has a material impact on the research design and its findings, essentially due to the relatively small number of identification numbers with duplicates.

2.2. Who is not a member of the population?

Another aspect of the register data that is worth highlighting is that it does not include certain individuals who are currently living in Sweden. As noted above, people enter the population register when they are born, if they are born in Sweden, or when they receive a resident permit or register their immigration. This inevitably results in certain groups of individuals being excluded from the population, although to a large extent this is by design (and aligns with the definitions that are the basis on which official statistics are created). According to Statistics Sweden (2006, p. 101):

“Persons who, on immigration to Sweden, intend to be resident in Sweden for less than one year are not entered in the population register. Citizens of countries outside the Nordic region must have a residence permit to be entered in the population register. Foreign citizens at foreign embassies or consulates and their staff are not registered in Sweden. Persons in the above-mentioned categories are not therefore included in the TPR register.”

⁸ The registers also include a variable to identify incorrect personal identity numbers (*FelPersonNr*). Typically, these are personal numbers in SCB's material that are incorrect, i.e. incomplete or otherwise do not correspond to a real personal identity number. They are not unique to the individual and may be judged to be incorrect based on other data derived from the personal identity number, such as date of birth. In REFU-GEN, there are 159,459 incorrect personal identity numbers in the exhaustive table. One third of these are also duplicates, and are not the last version of that identification number, so they therefore likely to be dropped in most analyses. Of the rest (52,558), 99.99% are missing any information on country of birth, and 99.47% are not in the population in any year 1968-2021, which means that they are also likely to be dropped, in particular when studying immigrants and their descendants.

⁹ See the Appendix for additional tables of statistics on reused and incorrect identification numbers.

It is therefore the case that anyone awaiting a residence permit are not entered in the population register until they receive a residence permit. For example, asylum-seekers are not entered into the register until their claim has been approved, and those whose claim is not approved (including those who are appealing previous decisions) will remain outside the register. Apart from those who do not intend to reside in Sweden, such as tourists and short-term migrants, the other group of individuals who are excluded from the register are those who have not registered with the authorities. Given that registration is required in order to access almost all institutions and public services in Sweden, the number of unregistered people living in Sweden is thought to be very low.

2.3. Emigration and over-coverage

Another issue that concerns data quality, and the definition of study populations, relates to emigration. Data on recorded emigrations cover the period from 1961 onwards¹⁰, and these data can be used to include or exclude non-resident individuals or to incorporate emigration into any analysis as befits the research design. However, it is well-known that some individuals remain in the register, despite no longer being (alive or) resident in Sweden. This is often referred to as over-coverage (as discussed above). Estimates of over-coverage and its consequences vary, but its consequences can be material, depending upon the type of analysis that is carried out (Monti et al. 2020; Mussino et al. 2023; Wallace and Wilson 2021). For example, although over-coverage among foreign-born individuals for the period 2003–2016 is low, there is distinct heterogeneity in over-coverage across various sub-populations (Mussino et al. 2023).

3. Using register data to measure migration background

There are various different ways of measuring ‘migration background’, which is a term often used to refer to individuals who have immigrated to Sweden, and usually those born abroad. However, there are other dimensions of migration background, which might be used to differentiate individuals, as summarised in *Tables 3.1a* and *3.1b*.

Table 3.1a: Different ways of measuring migration background

Time constant	Time varying
Country of birth	Citizenship
First admission category	Current residence permit status
Year of first arrival	Year of last international migration
Parent’s country of birth	Parent’s citizenship

Table 3.1b: Key variables and data sources

Variables	Data source ¹
Country of birth	Population register
Immigration (and emigration) dates	International migration register
Country of previous (or next) residence	International migration register
Citizenship	Population register
Residence permits	Permit register (SIV) / STATIV
Anonymised identifier of parents	Multigenerational register

1: Although these are listed as the data sources, all of these variables may appear (in the same or a different format) in other registers. For example, information on country of birth is available in STATIV and in LISA (another longitudinal collection of register data). When the data are provided to researchers by Statistics Sweden they may try to minimise the provision of duplicate variables, not least to save storage space on each project’s database.

¹⁰ A small number of emigrations are recorded prior to 1961 but there is an extremely high level of missingness.

Other variables are available in various registers that enable researchers to measure different aspects of immigrant's lives. With respect to migration background, some of these exist in separate registers, including those held by agencies who do not typically provide data to researchers (notably the Swedish Migration Agency: *Migrationsverket*), and some of these are held by Statistics Sweden. The most important of these is STATIV, which is a longitudinal collection of register data that was essentially created to study migration and integration (Statistics Sweden 2022, 2023c).¹¹ *Table 3.2* indicates the types of data that are in STATIV.

Table 3.2: Types of data in STATIV

Migration background	As listed in <i>Tables 3.1a</i> and <i>3.1b</i>
Migration background of parents	As listed in <i>Tables 3.1a</i> and <i>3.1b</i> , but for parents
Residence permits	Dates of registration and permit decisions, types of permit including refugee categories
Demographic characteristics	Age, sex and birth cohort
Family background	Partnership status, children, whether adopted or unaccompanied
Residence	Dwelling type, information on reception by municipalities after being granted a residence permit (data and location)
Education	Level of education and attendance of education in Sweden, including language training
Employment	Employment status and occupation
Income	From all major sources, including social allowances and social-insurance transfers
Information on support programmes	Introduction programmes, labour market training, other support
Other information about parents	Including education, employment and income

Note: Much of the information in STATIV is also available in other registers that can themselves be obtained by researchers. For example, REFU-GEN has more detailed information on education than is available in STATIV, which is available in educational registers that are also held by Statistics Sweden.

Considering this wealth of data on migration background, it is useful to make distinctions between different types of data that researchers may want to use. One of these is the difference between data that measure *stocks* (i.e. individual 'states' or circumstances at a given moment in time) and data that measure *flows* (i.e. events that occur, usually implying a change between states).

Table 3.3: Examples of data on stocks and flows relating to migration background

Stocks	Flows
Population by country of birth	Immigrations in a given period
Population by citizenship	Emigrations in a given period
Population by admission category	Births to immigrants
Population by year of arrival	Deaths of immigrants

Table 3.3 provides some specific examples of stocks and flows with respect to data that relates to migration background. For example, the annual RTB data that are made available to researchers enable them to identify members of the population usually resident in Sweden, as recorded at the end of the year (31 December). By analysing variables that identify an individual's characteristics (e.g. their country of birth), researchers can identify the *stocks* of individuals who have been resident in Sweden each year from 1968 onwards. At the same time, researchers might analyse registers of recorded immigrations, i.e. *flows* into Sweden. By combining these data on stocks and flows, in different ways, researchers can give a rich picture of each individual's life, including their migration background.

¹¹ As SCB state in their English report: "The STATIV database has been developed by Statistics Sweden together with the Swedish Integration Board to provide a basis for illustrating the situation and development within different areas of society from an integration policy perspective. When the Swedish Integration Board closed on 1 July 2007, the responsibility for the database was transferred to Statistics Sweden. The first year of data is from 1997. The database is updated each year during the spring and a new year of data is added."

4. Country of birth

Data on country of birth will be provided to different research projects with varying levels of detail. In part this is because country of birth is highly disclosive. Some projects will therefore receive data on individual countries of birth, for some or all countries, whereas others will receive data that groups countries of birth, for example by continent or region of the world.

Although the quality of the data on country of birth is generally excellent, in particular after 1968 (when RTB begins), there is some missingness. However, as shown in *Table 4.1*, this missingness is less than 0.02% of the registered population every year. There is a higher level of missingness (around 1%) for those in the population registers overall (i.e. if researchers examine an exhaustive extract of register data as described in *Section 1*)¹². Yet it is unlikely that this would be the level of missingness in most studies of contemporary Sweden, essentially because most studies will condition on membership of the population at some point over the last fifty years.¹³

Table 4.1: Population by year and nativity

Coverage	Foreign-born	Swedish-born	Missing ¹	Missing (% of total)	Total
Population 1970	533,122	7,546,813	1,143	0.01	8,081,078
Population 1980	626,317	7,690,688	881	0.01	8,317,886
Population 1990	789,194	7,800,317	1,005	0.01	8,590,516
Population 2000	1,002,657	7,878,828	1,196	0.01	8,882,681
Population 2010	1,383,875	8,030,598	1,027	0.01	9,415,500
Population 2020	2,045,214	8,332,530	1,519	0.01	10,379,263
Population 1968-2021	3,245,869	13,129,046	18,785	0.11	16,393,700
Not in population 1968-2021	57,352	187,944	161,987	39.77	407,283
All register data in REFU-GEN	3,303,221	13,316,990	180,772	1.08	16,800,983

1: Missing includes where there is no information on country of birth or where individuals are recorded as unknown or born in international territory.

Source: Authors' analysis of Swedish register data.

As stated in *Section 1*, It may be useful to note that of those who are missing country of birth, 29% (52,555) are recorded by Statistics Sweden as having an incorrect person identification number (*FelPersonNr*). When this error has been discovered a correct identifier is given to the individual. As such, it may be reasonable to assume that individuals with incorrect identifiers are covered in the data even after incorrect identifiers are dropped from any analysis (which itself seems like a reasonable course of action, both on account of this reason and the fact that they are not likely to be linked to any useful information in other registers).¹⁴ *Section 1* also notes that incorrect identifiers are one of the main reason for duplicate identifiers, which are far more common for foreign-born individuals (*Table 4.2*).

¹² We note that the very few (less than 10) identifiers missing information on sex are also missing country of birth.

¹³ Where this is not the case, it may be useful to note that of the 180,772 identifiers that were missing country of birth, 121,202 were recorded as born before 1940, 41,789 between 1941 and 1970, and 17,755 after 1970.

¹⁴ Incorrect identifiers are nevertheless retained by Statistics Sweden for quality assurance purposes, and given to researchers for the same reason.

Table 4.2: Quality of data on country of birth

Country of birth	Unique identifier	Duplicate	Total
Foreign-born	3,190,178	113,043	3,303,221
Swedish-born	13,315,373	1,617	13,316,990
Missing ¹	180,016	756	* 180,772
Total	16,685,567	115,416	16,800,983

1: Missing includes where there is no information on country of birth or where individuals are recorded as unknown or born in international territory.

Source: Authors' analysis of Swedish register data.

Table 4.3: Top 20 largest countries of birth by population size, 31 December 2020

Rank	Country	Frequency	%	Cum. %	SCB estimate	Difference
1	Syria	193,560	9.5	9.5	193,594	34
2	Iraq	146,442	7.2	16.6	146,440	-2
3	Finland	140,339	6.9	23.5	140,337	-2
4	Poland	93,763	4.6	28.1	93,762	-1
5	Iran	81,300	4.0	32.1	81,301	1
6	Somalia	70,176	3.4	35.5	70,184	8
7	Yugoslavia	63,390	3.1	38.6	63,419	29
8	Afghanistan	60,854	3.0	41.6	60,858	4
9	Bosnia and Herzegovina	60,158	2.9	44.5	60,161	3
10	Turkey	52,632	2.6	47.1	52,628	-4
11	Germany	51,432	2.5	49.6	51,434	2
12	Eritrea	47,143	2.3	51.9	47,156	13
13	Thailand	44,341	2.2	54.1	44,339	-2
14	India	42,788	2.1	56.2	42,790	2
15	Norway	41,058	2.0	58.2	41,062	4
16	Denmark	38,927	1.9	60.1	38,929	2
17	China	35,995	1.8	61.8	36,023	28
18	Romania	32,744	1.6	63.4	32,741	-3
19	UK	31,035	1.5	64.9	31,035	0
20	Lebanon	28,890	1.4	66.4	28,885	-5

Note: The total foreign-born resident population in 2020 is estimated to be 2,045,214 in the REFU-GEN data. The comparative estimate for 2020 published by SCB in 2023 is 2,046,731 (The source for this and the SCB estimates in the table is Statistics Sweden 2023d). These estimates are not expected to be identical, but it is reassuring that they are almost the same. Source: Authors' analysis of Swedish register data.

Further information on the Swedish population by country of birth is available from Statistics Sweden, and this can be used by researchers to check that they have similar frequencies when analysing their study population (Statistics Sweden 2023d). An example of this is shown in *Table 4.3*, which shows the top twenty largest countries of birth in the 2020 resident population by population size. Estimates are not expected to be identical given constant updates and corrections that are made to the registers. However, it is reassuring that they are almost the same when comparing Statistics Sweden's figures for 2020, published in 2023, with those from the REFU-GEN data, which is an extract of register data that was created in 2022. The top twenty countries accounted for around two thirds of the foreign-born population in 2020, with around one third being accounted for by the top five: Syria, Iraq, Finland, Poland and Iran. This can be contrasted with the situation fifty years earlier, in 1970, when the top twenty countries accounted for 95% of the population and the top five countries were: Finland, Norway, Germany, Denmark and Yugoslavia (see *Appendix Table A4.1*). Since that time there has been a diversification of migration, such that immigrants who have ever resided in Sweden were born in all regions of the world (see also *Appendix Table A4.2*).

5. Immigration and emigration events

A register of immigration and emigration is maintained by Statistics Sweden with the aim of including all international migration events in a given year. It is based on population registration information provided by the Swedish Tax Authority (as for RTB), and the data are available for research, including exact dates of migration (if requested), as well as information on country of previous residence (for immigrations) and country of next residence (for emigrations). The quality of the country data is lower than the quality for the migration events, which is generally excellent, bearing in mind the issues of under-coverage and over-coverage mentioned in *Section 1*. As also explained in *Section 1*, it may be useful to note that people are only registered as immigrating to Sweden when they receive a resident permit or register their immigration. This means, for example, that Statistics Sweden's data on immigration events do not include those who have not (yet) registered with the authorities, or asylum seekers who are awaiting a decision on their claim.

Using the data on international migration events, it is possible to calculate the first recorded arrival date for the foreign-born population. In general, the quality of this variable is very good. However, although it can be calculated for almost everyone in the registers, it is missing for some identification numbers, albeit for different reasons (*Table 5.1*). One reason, for 180,772 identification numbers, is that they are missing information on country of birth, as well as information on year of first arrival in the immigration events register. Almost 90% (161,987) of these identification numbers are not in the population in any year from 1968 onwards. The rest (18,785) are recorded in RTB (more than half before 1970), making it possible to impute their arrival year (or at least know the maximum value that it might be) based upon the first year in which they appear in the total population. Yet it remains impossible to be sure whether they are Swedish-born or foreign-born due to missing data on country of birth. There are also 246,171 identification numbers that are known to be foreign-born (i.e. not missing country of birth) and missing year of first arrival in the events register. Of these, 90% (221,810) are recorded as being resident in Sweden in 1968 (in RTB), and 6% (14,121) are recorded as being resident in at least one year from 1969-2021. Of the remaining 4% (10,240), more than half have a first recorded emigration, almost all (5,465) prior to 1969 (leaving a residual 4,775).

Table 5.1: Missing data on first year of arrival for the foreign-born

Explanation for missingness	Frequency
Missing country of birth and year of first arrival	180,772
Missing year of first arrival but <u>not</u> country of birth and recorded as resident in the population in 1968	* 221,810
Missing year of first arrival but <u>not</u> country of birth and recorded as resident in the population after 1968	** 14,121
Missing year of first arrival but <u>not</u> country of birth and <u>not</u> recorded as resident in the population in 1968, with a first recorded emigration before 1969	5,465
<i>Residual identification numbers with missing first year of arrival in the register of immigration events</i>	4,775
Total missing	426,943

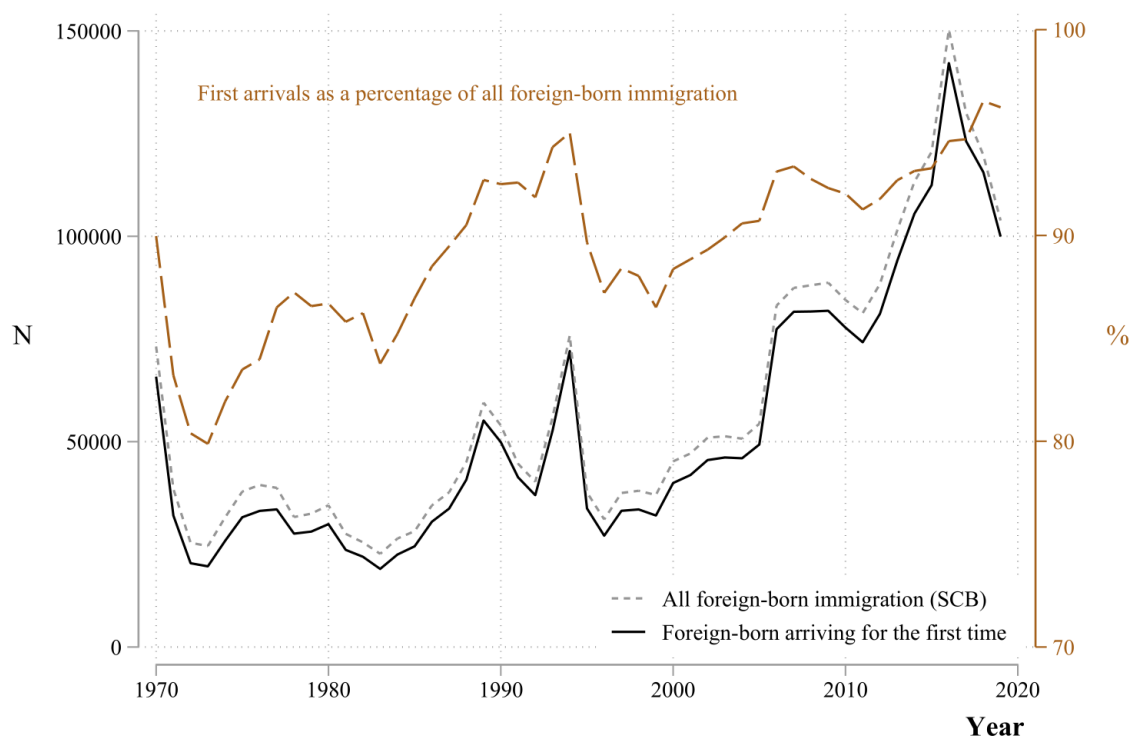
* Given that they are in the population in 1968, these individuals can at least be assumed to have first arrived in Sweden before 1969. ** 98% (15,898) of these are recorded as emigrating between 1961 and 1968 (i.e. between the 1960 Census and the start of the annual RTB files. Source: Authors' analysis of Swedish register data.

All of this information can be used by researchers to impute year of first arrival (or at least a maximum possible value). For example, those who are resident in the population in 1968 can be assumed to have first arrived in Sweden before 1969. As such, they can be combined with others who arrived before 1969 and analysed as a group, or they can be dropped legitimately from any analysis when the study population excludes those arriving prior to 1969. Indeed, the same approach can be used to impute year of first arrival in Sweden (using RTB) when it is missing for any individual. A similar approach to imputation can be carried out using information on emigration (or information on residence permit dates, although these are missing for almost all of the residual 4,775 in *Table 5.1*). In summary,

these imputation approaches will result in far fewer foreign-born individuals with missing first immigration year, albeit with some assumptions about missingness and the quality of the imputation data.

In the rest of this report, we do not impute any values (for any variable). Instead, we examine the quality of (and describe) the observed values. One way to do this is to compare with official statistics published by Statistics Sweden. However, it can be difficult to compare trends in first arrival with official statistics because they focus on all immigrations (and not first immigrations), including repeat/circular migration (i.e. emigrating from Sweden and then coming back). More generally, such definitional differences often explain discrepancies between different analyses with register data. Nevertheless, a comparison between all immigrations of the foreign-born (from Statistics Sweden) and first immigrations of the foreign-born (from REFU-GEN) suggests that they are not too dissimilar, and that they tend to become more similar over time (Figure 5.1). In other words, the proportion of immigrations that are repeat migrations has become smaller over time. This reflects changes in the composition of the foreign-born population, such as the decreasing proportion of immigrants from other Nordic countries (who are more easily able to emigrate and then come back to Sweden).

Figure 5.1: All immigrations and first immigrations of the foreign-born population 1970-2019



Note: First immigrations (foreign-born immigrants arriving in Sweden for the first time) are estimated from REFU-GEN. All immigrations are taken from SCB's published estimates (Statistics Sweden 2023e). The long-dashed brown line shows first immigrations as the percentage of all immigrations (with the axis on the right-hand side). Source: Authors' analysis of Swedish register data.

Figure 5.2a: Resident population 1970-2020 by year of first arrival 1940-2020

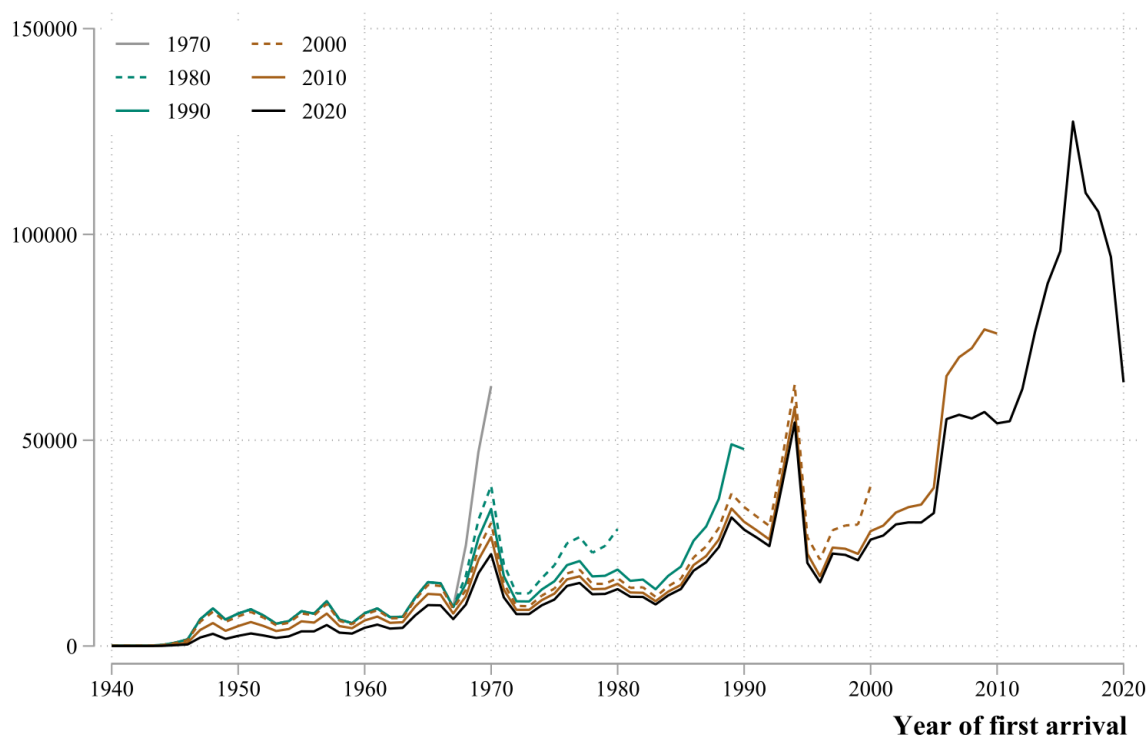
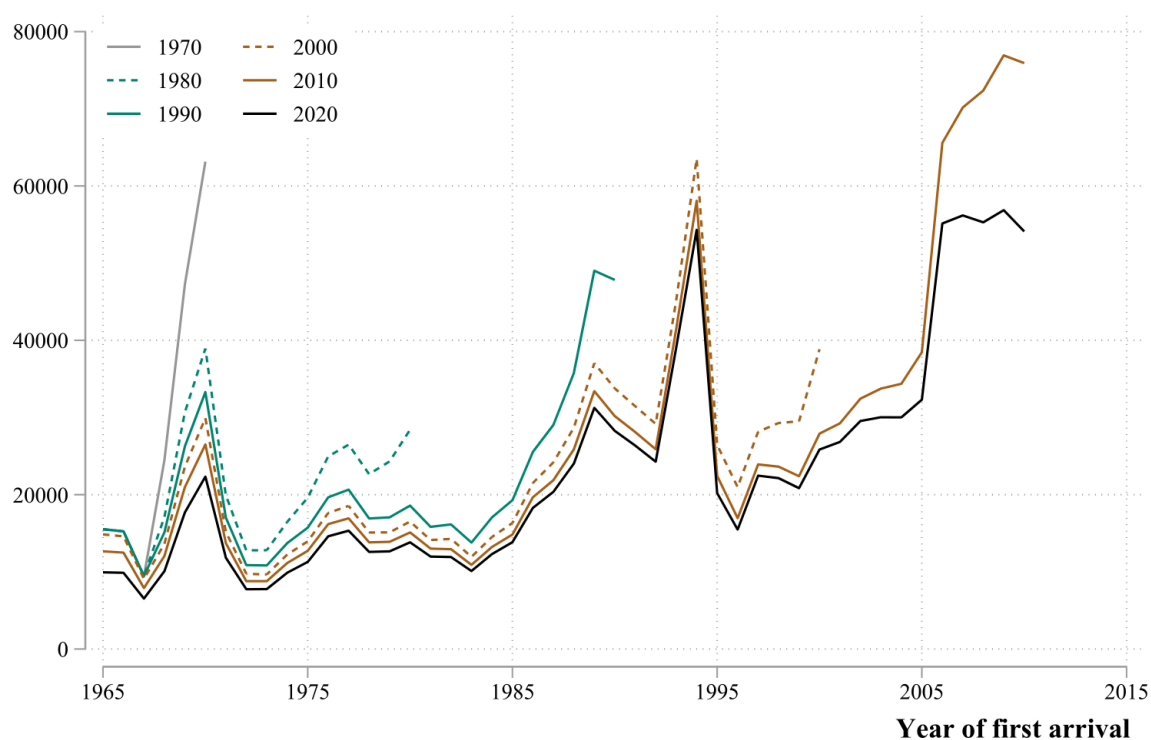


Figure 5.2b: Resident population 1970-2020 by year of first arrival 1965-2015

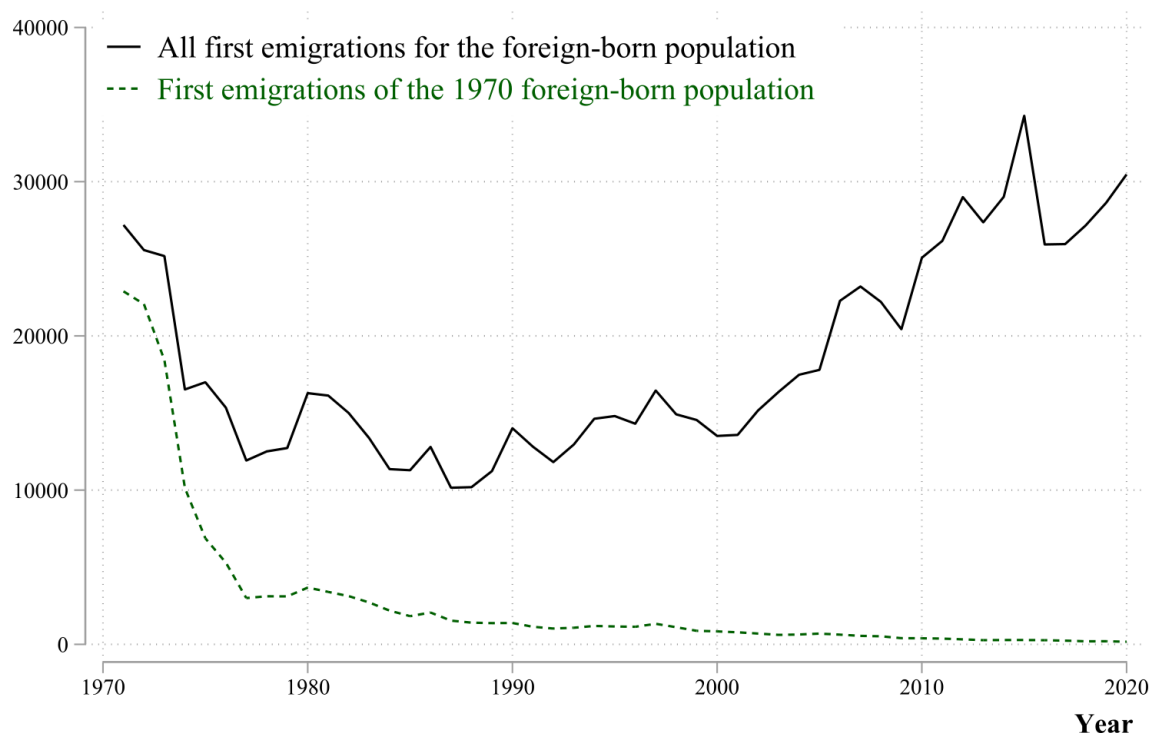


Note: Each line is plotted for the population estimates of that year. For example, the solid black line is the profile of years of arrival for the foreign-born population who were resident in Sweden in 2020. The figures not only show the broad trends in immigration (of new residents who have never lived in Sweden) but also trends in emigration after first arrival. For example, the difference between the lines for 1970 and 2020 shows that of the more than 60,000 people who arrived in 1970 and were resident in Sweden that year, only slightly more than 20,000 were still resident in Sweden in 2020. Although some of them died in the intervening period, most emigrated from Sweden between 1970 and 2020. Source: Authors' analysis of Swedish register data.

One useful way to understand the longitudinal structure of the register data—and its flexibility for studying immigrants—is to combine data on stocks and flows. *Figures 5.2a and 5.2b* show different estimates of the foreign-born resident population for each decade, from 1970 to 2020, by year of first arrival. Each estimate (i.e. each line) indicates how the foreign-born population is composed of a diverse range of immigrants who arrived in different years over a long period of time. In addition, by comparing the estimates for different years, it is possible to see the evolution of the population—and its composition by year of first arrival—over time. For example, the difference between the lines for 1970 and 2020 shows that of the more than 60,000 people who arrived in 1970 and were resident in Sweden that year, only slightly more than 20,000 were still resident in Sweden in 2020. Although some of them (around a quarter) died in the intervening period, most of them emigrated from Sweden between 1970 and 2020 (more than half in the first five years).

In addition to immigration events (for all residents), the registers also include registered emigrations. Like immigrations, the same individual can have multiple events, and it can be useful to distinguish between first recorded emigrations and repeat emigrations. Figure 5.3 shows the trend in first emigration events for the foreign-born population from 1971-2020. To put this in context, and illustrate how stocks and flows can be combined in different ways, the figure also shows first emigrations for the foreign-born population who were resident in 1970. Comparing the two trends, it is clear that emigrations tend to relate to more recently arrived individuals (as we know from previous research, e.g. Monti 2020). For example, we can infer from the figure that the majority of first emigrations among the foreign-born in 1980 were events for immigrants who arrived after 1970 and had resided in Sweden for less than ten years. Similarly, of all foreign-born residents in 1970 who emigrated (at least once) from 1970-2020, more than half (58%) did so between 1971 and 1975.

Figure 5.3: Emigration of the foreign-born population 1971-2020



Note: All emigrations of the foreign-born are shown using the solid black line. The short-dashed green line shows the first recorded emigrations of the resident population in 1970. Source: Authors' analysis of Swedish register data.

6. Citizenship

Swedish citizenship policy has a long and complex history, and compared to many other countries confers fewer civil rights, although citizenship is linked to voting in Sweden and rights in other countries, notably the other members of the EU (see Bernitz 2012 for a detailed overview). With respect to the registers, annual data on citizenship have been collected by Statistics Sweden since the beginning of the annual population registers (RTB) in 1968. There is one major interruption in the way that the data are recorded. Prior to 1998, a two-digit alphabetical code is given for foreign citizens or (if Swedish citizenship has been acquired then) acquisition of Swedish citizenship is indicated by the year in which the acquisition took place. This single variable is replaced by two variables from 1998 onwards, one for country of citizenship and one for the date that this citizenship was acquired. People who acquired Swedish citizenship before 1967 are usually recorded as missing the date. A few people have a date before 1967, however this information is only partial in terms of coverage.

In general, the information on citizenship has lower quality and coverage when trying to identify country of citizenship before people became Swedish citizens. This is essentially because the data is truncated for many individuals. For example, there is no information about citizenship on arrival for all immigrants who arrived before 1968 but were Swedish citizens in 1968 (when the annual citizenship data begins). It is also the case that citizenship data provided to researchers includes only one country per person-year (although the country can change from year to year, enabling longitudinal analysis). When an individual has more than one citizenship, the recorded country is chosen based on a set of prioritising rules, which are broadly as follows (Statistics Sweden 2006, p. 74):

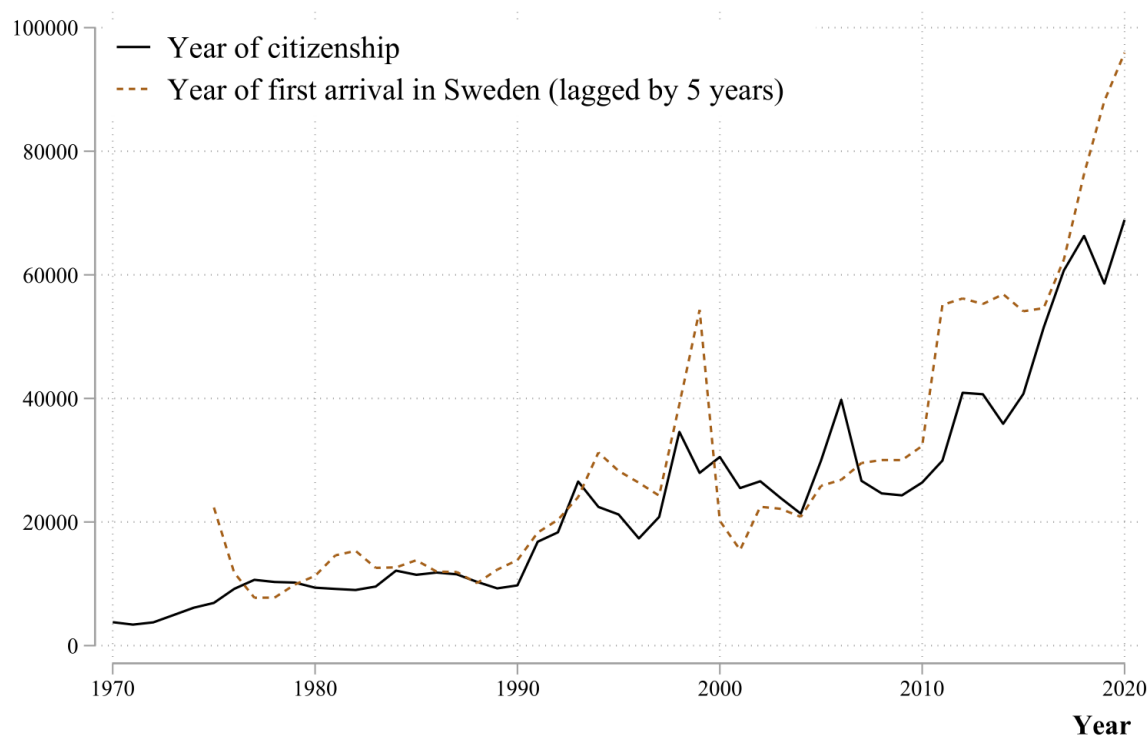
1. Swedish citizenship
2. Citizenship of an EEA country
3. Citizenship of a non-EEA country with which Sweden has a mutual social insurance agreement
4. Other citizenship

If an individual has multiple citizenships from countries with the same priority, then the earliest obtained citizenship is chosen. It follows from these rules that for those who have both foreign citizenship and Swedish citizenship, only the Swedish citizenship is registered. For a Swedish citizen, the registers therefore do not show if they are also a citizen of another country. It is for this reason that the data are often used at a level of aggregation that only distinguishes between whether individuals do (or do not) have Swedish citizenship, and the year at which this Swedish citizenship was obtained.

Similar to country of birth, a number of identification numbers are missing information on citizenship, either country of citizenship, year of citizenship, or both. Of the 3.3m foreign-born individuals who have ever been in the registers (up to the time of the REFU-GEN data extract), only 59,052 are missing information that prevents researchers from identifying whether they have ever been Swedish citizens or not, and of those who can be identified as Swedish citizens, only 7,385 are missing data on the year that this was obtained. Although it should be noted that 187,268 have an imprecise date that can only be determined approximately as ‘before 1969’ because they are already Swedish citizens when the data begin in 1968.

A time series of citizenship acquisition is shown in *Figure 6.1*. An additional means of checking the quality of the data is to compare with the time series of new arrivals. Although some foreign-born immigrants who arrive in Sweden for the first time have Swedish citizenship (around 2.5% of those who arrived for the first time from 1970-2020), some may neither want nor need citizenship, and many may emigrate before being eligible to apply, this is not true for the vast majority. For example, of the foreign-born population in 2020, almost 80% who first arrived in the decade 2000-2010 had claimed citizenship by 2020. *Figure 6.1* compares the overall trends, but we note that the trend for new arrivals is shown 5 years later than the actual period it concerns. It is clearly evident that there tends to be a delay of approximately 4-6 years between first arrival and obtaining Swedish citizenship, which is unsurprising considering that eligibility for Swedish citizenship has typically ranged between 3-5 years over the last fifty years, and that it can take between 1-2 years for applications to be processed and approved.

Figure 6.1: Year of obtaining Swedish citizenship for the foreign-born population in 2020



Note: The solid black line shows the year of obtaining Swedish citizenship for the foreign-born population who were resident in Sweden in 2020. For comparison, the dashed line shows the year of first arrival in Sweden for the same population in 2020, but lagged by five years to highlight the correlation between the trends allowing for the time it takes to be eligible and apply for citizenship. Source: Authors' analysis of Swedish register data.

7. Residence permits and categories of admission

Not everyone who migrates to Sweden requires a residence permit, but a permit is often requisite for living in the country. The type of residence permit is also a useful piece of information for migration scholars, not least because it indicates reason for migration and provides some information about selection into migration and opportunities on arrival.

7.1. Trends in immigration, policies and data relating to residence permits

The rules governing who does and who does not require a permit have changed over time, are very detailed. There have also been various changes in the way that data have been collected and made available to researchers. Some of the key events—in terms of trends, policies, and data availability—are summarised below (for more details, see R. Andersson et al. 2010; Bevelander 1999, 2011; Statistics Sweden 2022).

Prior to 1945 and the Second World War

With respect to foreign-born arrivals, the majority of immigration to Sweden prior to 1945 was intra-Nordic migration. This changed during the Second World War when a sizeable number of non-Nordic refugees came to Sweden, in particular from nearby countries such as Estonia, alongside an increase in Nordic immigration (from Denmark, Norway and Finland). Toward the end of the war there were more refugees who arrived from other European countries (including from concentration camps), however many war refugees emigrated in 1945, either returning home or moving to other countries. Some refugees remained in Sweden, but it is not possible to identify them directly in the data (e.g. using data on refugee status, which did not become available until the mid-1980s), except by focussing on specific countries of birth and years of arrival.

From 1945 until the end of the 1960s

Excluding the period during and immediately after the Second World War, the vast majority of immigrants to Sweden from 1945 until the early 1970s were labour migrants. Partly, this reflects the fact that Sweden's main immigration policy was to actively seek labour migrants, who were needed due to excess demand for jobs (Bevelander 2011). In the 1950s and 1960s, the majority of foreign-born immigrants still came from other Nordic countries, in particular Finland, but there was an increase in non-Nordic immigration. Up until the 1960s, this was predominantly skilled labour from other European countries, such as the Netherlands, Belgium, West Germany, Austria, Italy and Greece (R. Andersson et al. 2010). In the 1960s, there was an increase in lower-skilled immigration, from the same origins and other countries like Yugoslavia and Turkey. During the 1960s, the Swedish Government began to change the rules governing entry into Sweden, partly in response to criticisms about the impact of immigration on the labour market (Bevelander 1999). The new rules began to apply from 1968 and made it much more difficult for non-Nordic immigrants to arrive in Sweden as labour migrants. Nevertheless, considering the whole period from 1945 until the end of the 1960s, it is reasonable to assume that most immigrants arrived as labour migrants, even though it is not possible to differentiate immigrants by admission category in the data (until the mid-1980s). However, this assumption is more likely to hold for particular groups of immigrants—e.g. by country of origin and year of arrival—depending upon specific details with respect to their migration history. For example, the majority of refugees (approximately 60%) who migrated to Sweden from 1950-67 were from Hungary or Yugoslavia (see Table 4.2 in R. Andersson and Solid 2003, which is itself sourced from SOU 1982), yet not all immigrants from these countries were refugees.

The 1970s and mid-1980s

Although there were immigrants who arrived as refugees before the 1970s, it was not until the early 1970s that Sweden began to receive a sizeable number of refugees who were born outside Europe (R. Andersson and Solid 2003; SOU 1982). From 1973, immigrants who were fleeing persecution began to arrive from South America—notably from Chile but also from other countries (Statistics Sweden 2016b). Partly as a reaction to this, the Swedish multicultural policy was introduced in 1974/75, with a focus on immigration and integration (R. Andersson et al. 2010). Refugee and conflict-induced migration continued through the early 1980s, when there was an increase in the number of immigrants arriving from the Middle East, notably Iranian and Kurdish refugees (R. Andersson et al. 2010). With respect to register data, information on admission categories is not available prior to the mid-1980s (with the exception of some lower-quality retrospective information that covers only part of the population). As such, it is only possible to study refugees or other categories (e.g. labour migrants or students) by using other information in the registers and making some assumptions, for example by focussing on specific countries of birth and years of arrival.

From the mid-1980s until the mid-1990s

In 1985, Sweden introduced a refugee dispersal policy, partly as a means of managing the increasing number of refugees who were arriving (R. Andersson et al. 2010). Immigrants from South America and the Middle East continued to arrive in relatively large numbers, in particular from Iran and Iraq (Statistics Sweden 2016b). There were also a growing number of arrivals from Africa, but the largest group of arrivals were immigrants from (former) Yugoslavia and Bosnia-Herzegovina after the start of the Yugoslav Wars in mid-1991 (Statistics Sweden 2016b). From 1987, data are available from Statistics Sweden on 'reason for residence' (*grund för bosättning*), which is based on the administrative registration of residence permit categories. The variable is described in considerable detail in the Swedish-language manual for STATIV (Statistics Sweden 2022), although a translation of part of this material is also available (Statistics Sweden 2017b). As it notes: "*This variable indicates the most recent recorded reason for residence in Sweden. The variable is collected from the Swedish Migration Board.*" (Statistics Sweden 2017b, p. 18). In the REFU-GEN data extract, there are three sources of data on residence permits: (i) annual flows data from 1987 (*SIV flode*), (ii) annual stocks from 1990 (*SIV stocks*), and (iii) annual data from 1997 onwards in STATIV. These data can be analysed and harmonised to derive many variables, including first recorded residence permit (admission category) and whether individuals have ever received a specific permit type.

From the mid-1990s until the present day

In 1995, Sweden became a member of the European Union, and in 1996 a member of the Schengen agreement, which marked the beginning of a new wave of immigration from other European countries (R. Andersson et al. 2010). As the EU has expanded since then, there has been a growing group of European immigrants. Since 1995, there have been various changes in migration policy. Key events include reforms to the dispersal policy in 1994 and the setting up of a new government department—the Board of Integration—in 1998 to support and monitor integration in municipalities (although this was subsequently closed in 2007, see: R. Andersson et al. 2010). More recently, significant changes to Sweden’s immigration policy have been made, partly in response to the large number of immigrants who arrived during and after 2015. At present, you must have a residence permit or right of residence in order to stay in Sweden for more than three months, unless you are a Swedish or other Nordic citizen (Statistics Sweden 2023c). Citizens of EU member states, or EEA countries, and members of their families have a right of residence, so do not need a residence permit, while citizens from other countries (outside the EU and EEA countries) must apply for a residence permit (Statistics Sweden 2023c).

7.2. Changes in residence permit data over time

Data on residence permits is available in annual format from 1987. Although these annual files include some information on permits obtained prior to 1987, this information is only partial, such that high-quality data on residence permits is only really available from 1987 onward. As mentioned in *Section 7.1*, the REFU-GEN data extract has three sources of data on residence permits. For this report, these data have been harmonised to create a list of permits for each individual, and then to derive a variable indicating each individual’s first recorded residence permit.

As shown in *Figures 7.1 and 7.2*, some information on permits is missing for some individuals. As expected, this missingness is particularly prevalent prior to 1987 (i.e. before data were collected and made available routinely). It is also related to migration background, in particular country of birth and citizenship. For example, there is a much lower level of missingness among non-Nordic immigrants (*Figure 7.1b*) and those who were not born in Europe or to parents born in Europe (*Figure 7.2b*).

Figure 7.1a: Residence permits for all foreign-born immigrants by (broad) admission category and year of first arrival 1940-2020

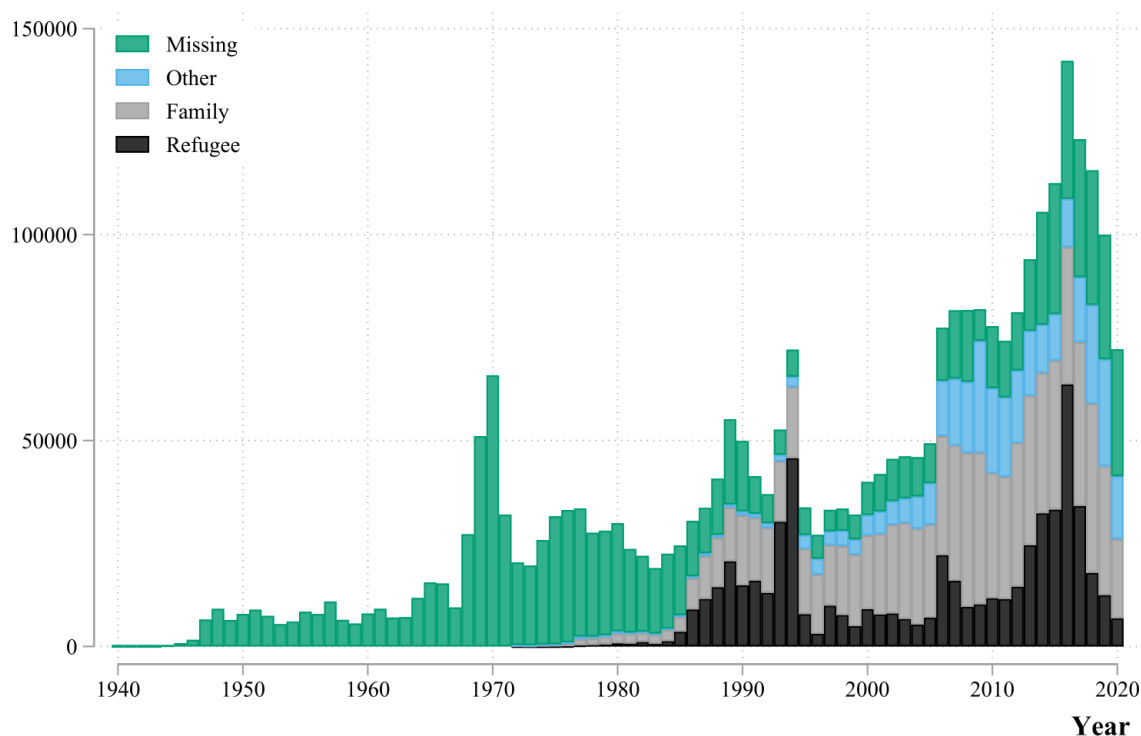
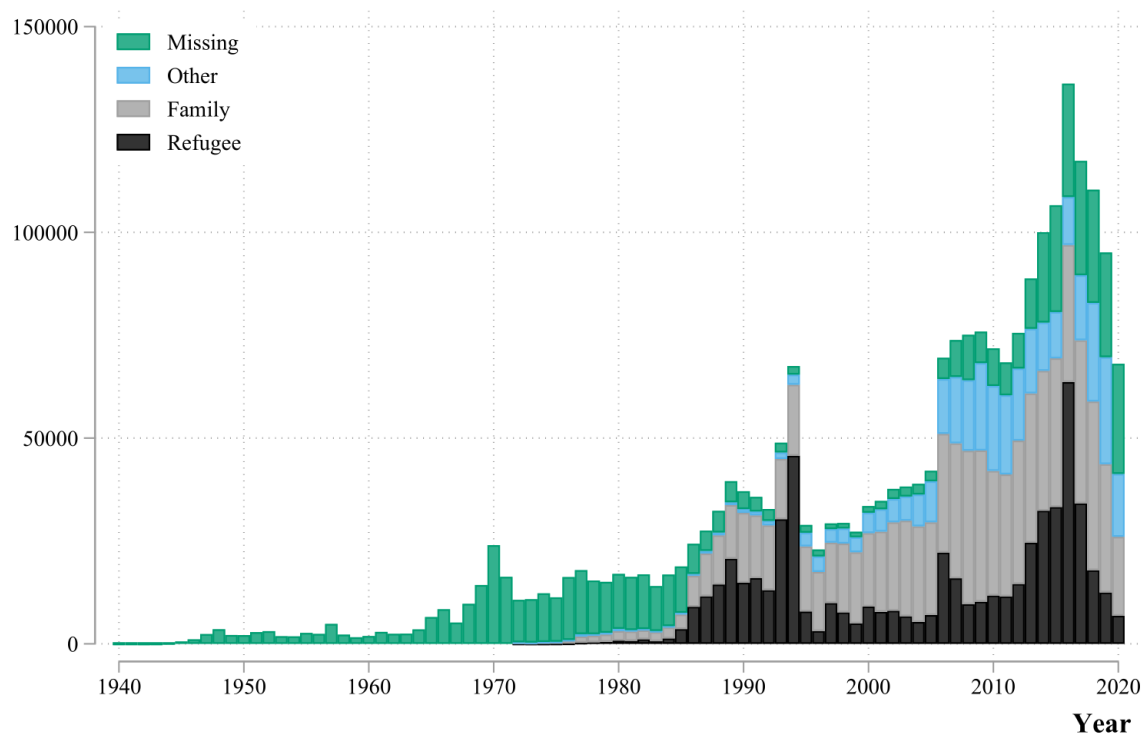


Figure 7.1b: Residence permits for all foreign-born immigrants by (broad) admission category and year of first arrival 1940-2020, **excluding immigrants born in other Nordic countries**



Source: Authors' analysis of Swedish register data.

Figure 7.2a: Residence permits for all foreign-born immigrants by (broad) admission category and year of first arrival 1970-2020

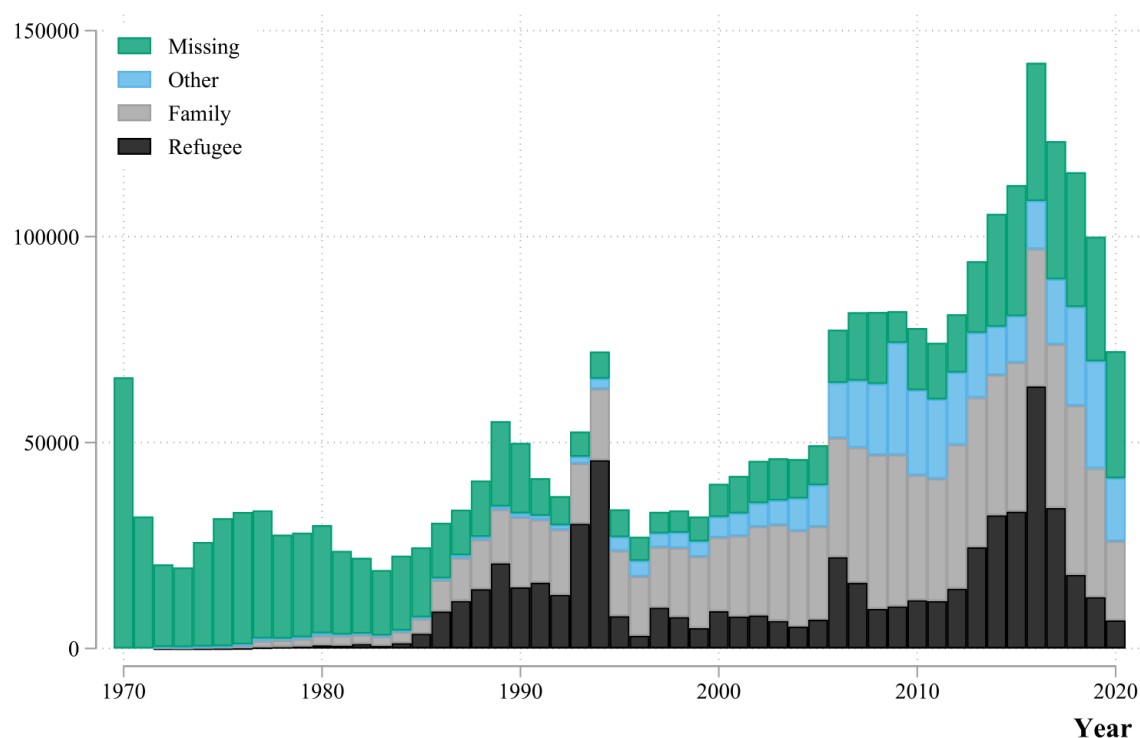
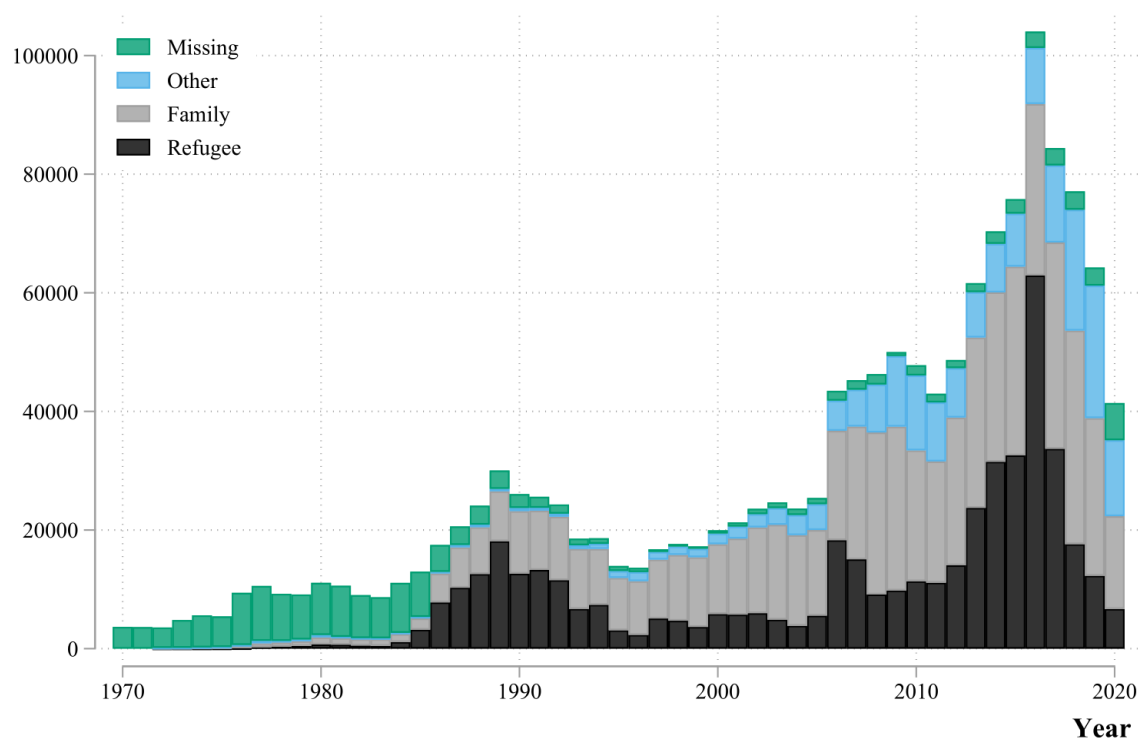


Figure 7.2b: Residence permits for all foreign-born immigrants by (broad) admission category and year of first arrival 1970-2020, excluding immigrants born in Europe and immigrants with parents born in Europe



Source: Authors' analysis of Swedish register data.

8. Intergenerational linkages

One of the significant benefits of using Swedish register data is the ability to link information about parents with information about their children. For example, this enables researchers to identify children of immigrants and categorise them according to their parents' migration background. The linkage is made possible thanks to data tables that connect the identification numbers of individuals to the identification numbers of their biological parents (mothers and fathers).¹⁵ Information from Statistics Sweden describes how this linkage is generated (Statistics Sweden 2017a), and its quality is also discussed in peer-reviewed publications (Ekbom 2011; Hamm et al. 2021). The linkage is only possible if the parents have ever appeared in the Swedish register, which means that their parents will need to have been resident in Sweden at some point. Missingness therefore relates to the history and development of the registers, and is particularly high for older cohorts (*Figure 8.1*).

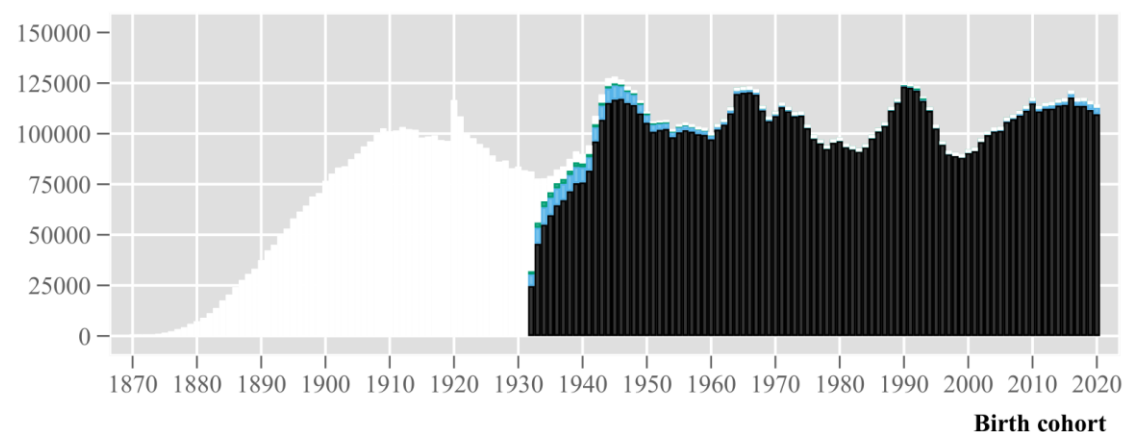
More specifically, there are several reasons why parental identification numbers are missing. Personal identification numbers were introduced in 1947, and if a person was 15 or younger in 1947, then their parents were recorded in their individual record (Statistics Sweden 2017a). This means that parental information mainly exists for Swedish-born individuals if they were born from 1932 onwards. This is shown in *Figure 8.1*, which also shows that for those born in Sweden after 1932, linkage is generally of a very high quality, but for those born abroad missingness is high, unless they arrived as children. This is because the multigenerational register does not include information on parents unless they have ever lived in Sweden. Most foreign-born immigrants who arrive as adults do not arrive with either of their parents or have any parents who live in Sweden (or migrate to Sweden at a later date), which helps to explain why so many foreign-born individuals have missing information on both parents. The exception is foreign-born immigrants who arrive as children. For example, most foreign-born immigrants who first arrived in Sweden aged under-18 can be linked to both their parents, or at least to either their mother or father (see *Figure 8.1*).

Given that the registers enable parents and children to be linked, the same information can be used to identify grandparents and/or grandchildren (as well as siblings, cousins and other family members who can be linked using a series of connections between parents and children). Similar to parental identification numbers, the existence of grandparental identification numbers is primarily related to birth cohort, and the data are generally of high quality for those born in Sweden after 1970 (*Figure 8.2*). *Figure 8.3* shows that more than three quarters of individuals born after 1970 have information observed for specific grandparental relationships: their mother's mother, mother's father, father's mother and father's father. As with parental identification numbers, missingness is also related to nativity (being born abroad), in this case parental nativity.

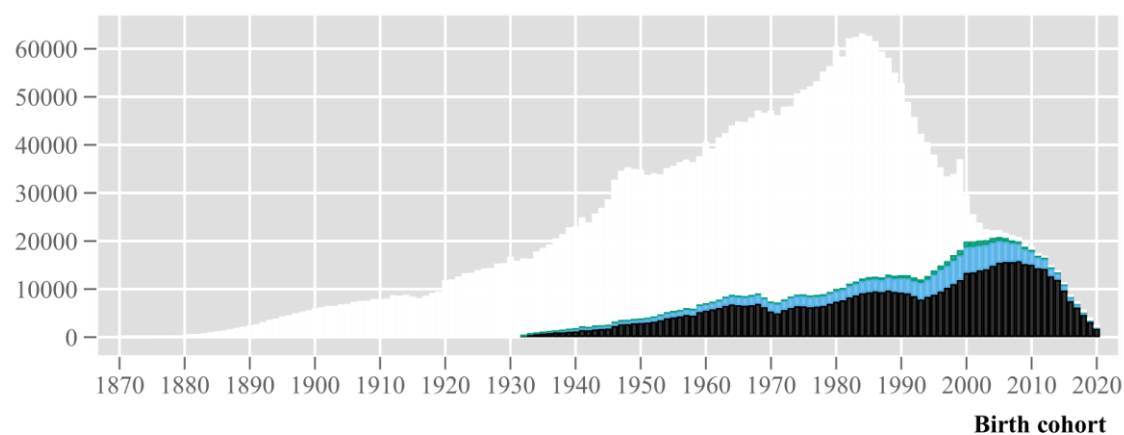
¹⁵ There are also links between children and their adoptive or foster parents, but they are not discussed or analysed here.

Figure 8.1: Parental identification number linkage status, birth cohorts 1870-2020

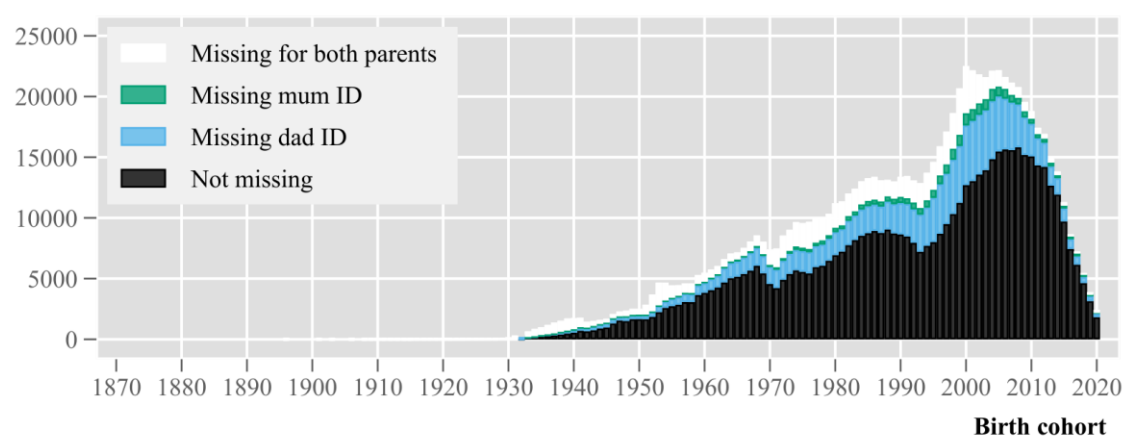
All Swedish-born



All foreign-born

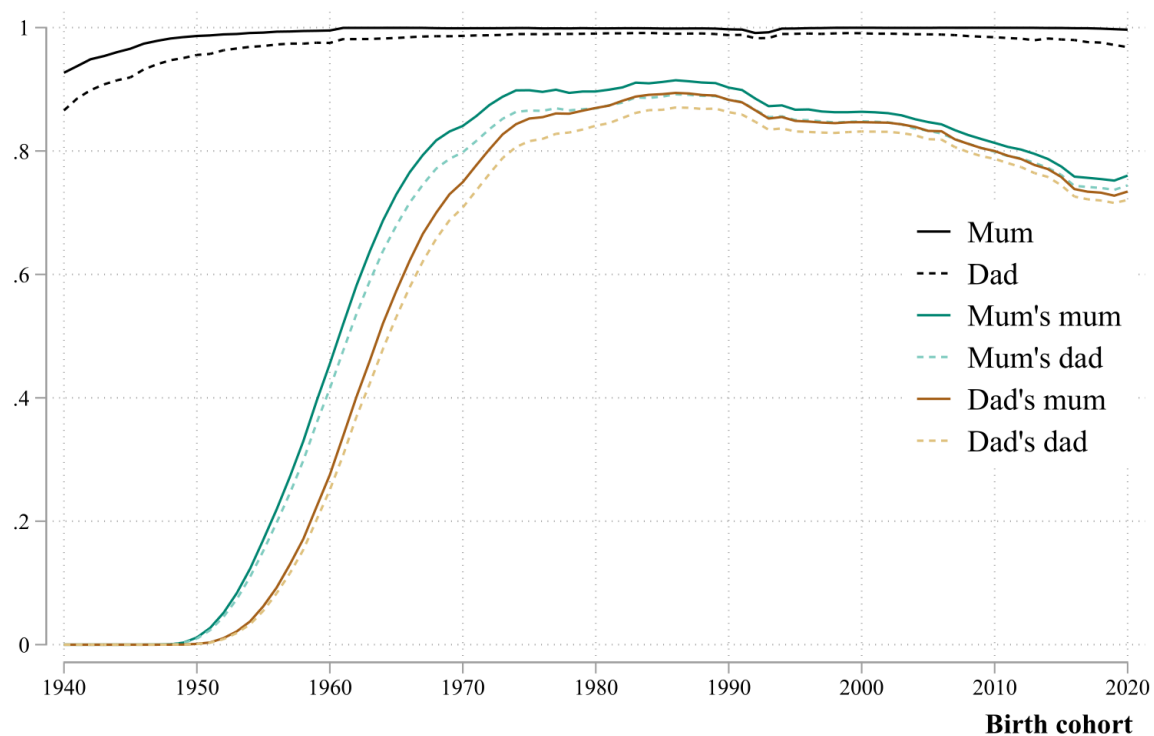


Foreign-born who arrived aged under-18



Source: Authors' analysis of Swedish register data.

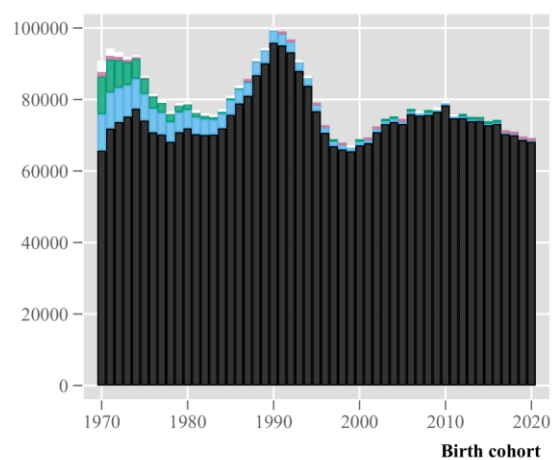
Figure 8.2: Percentage of observed identification numbers for parents and grandparents by ancestral relationship for all **Swedish-born individuals, birth cohorts 1940-2020**



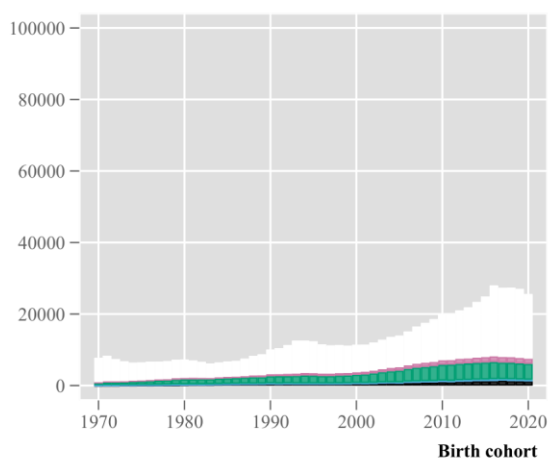
Source: Authors' analysis of Swedish register data.

Figure 8.3: Number of grandparental identification numbers that are missing for the whole population by parental country of birth, birth cohorts 1970-2020

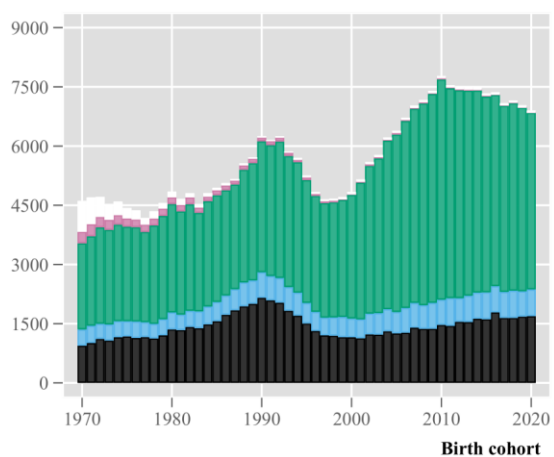
Both parents Swedish-born



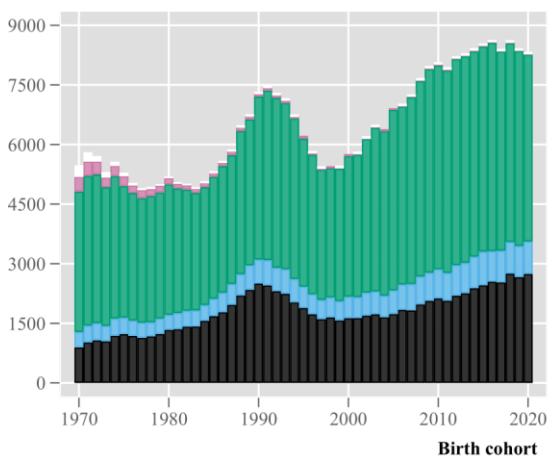
Both parents foreign-born



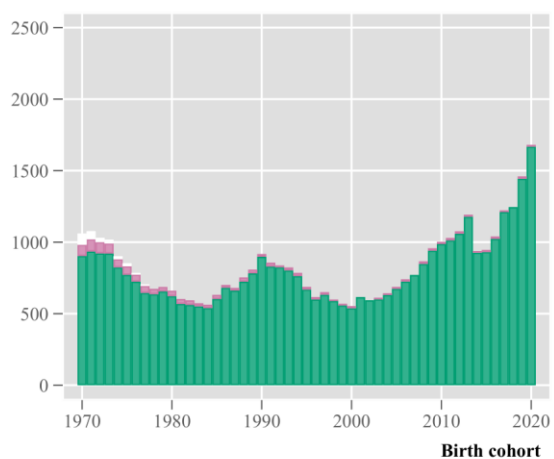
Mum foreign-born, dad Swedish-born



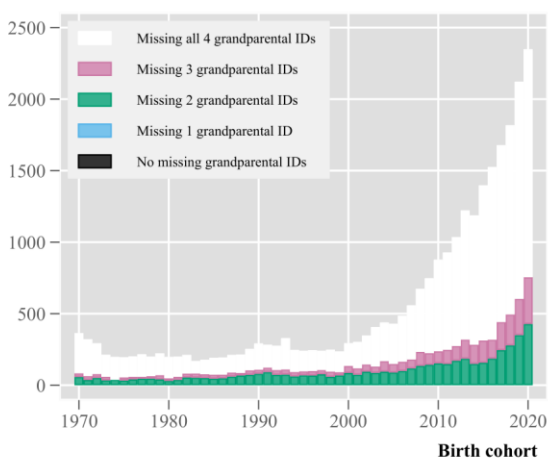
Dad foreign-born, mum Swedish-born



One Swedish-born, one missing nativity



One foreign-born, one missing nativity



Source: Authors' analysis of Swedish register data.

9. Descendants of immigrants and their parental migration background

In this final section of the report, we provide a brief overview of the ways that linkages between parents and children (described in *Section 8*) can be combined with the population-level data on migration background (described in *Sections 4-7*) in order to generate analysis of parental migration background. The quality of these data is linked to many of the same issues discussed in the previous sections. For example, since data on residence permits are only of good quality for immigrants arriving after 1985, data on parental residence permits are only of good quality for immigrants whose parents arrived after 1985.

The Swedish register data only include information about parental migration background for individuals whose parents are also in the population registers. This means that it is not possible to examine the parental migration background of most foreign-born individuals, unless they arrived as children (see *Figure 8.1* for information on parental linkage by nativity). For example, 15.1% of people who were resident in Sweden in 2020 were unable to be linked to any of their parents. The equivalent percentage for the Swedish-born population was 1.9%, while it was 69% for the foreign-born.

Table 9.1: Migration background of the Swedish population in 2020

	Frequency	% of total
All foreign-born	2,045,214	19.7
who first arrived as adults (18 or older)	1,357,881	13.1
who first arrived as children (under-18)	659,162	6.4
who are missing age at arrival	28,171	0.3
All Swedish-born	8,332,530	80.3
who have two foreign-born parents	637,718	6.1
who have one foreign-born and one Swedish-born parent	766,154	7.4
who have two Swedish-born parents	6,527,376	62.9
who are missing parental country of birth *	401,282	3.9
Missing country of birth	1,519	<0.1
total	10,379,263	100

* Of the 401,282 Swedish-born people who are missing parental country of birth, 39% cannot be linked to either parent (i.e. are missing the identifier for both parents), 33% have a Swedish-born parent but cannot be linked to their other parent, 20% have one Swedish-born parent and can be linked to their other parent, but that parent is missing information on country of birth, and the remaining 8% are a mixed category where one parent is missing information on country of birth, often because they have no Swedish identification number. Source: Authors' analysis of Swedish register data.

Many analyses have used Swedish register data to study the country of birth of immigrants and the parental country of birth of their children. This is sometimes referred to as 'migration background' (although the term is also used more broadly) because all foreign-born individuals have migrated internationally at some point in their life, (so they and their children might be referred to as having a migration background). As shown in *Table 9.1*, 19.7% of the Swedish population in 2020 were born abroad, around one third of whom first arrived as children. Of the more than 8m people who were born in Sweden, the vast majority had two Swedish-born parents, while 6.1% of the total population had two foreign-born parents and 7.4% had one. Although a low percentage of the Swedish-born population have missing country of birth (3.9%), this level of missingness is higher than for many other variables in the register (and this is related to the linkage issues discussed above, including in *Section 8*).

Table 9.2: Parental country of birth and year of arrival for the Swedish-born population in 2020

Mother	Father			Status unknown ²	Total
	Swedish-born	Foreign-born, arrived < 1987 ¹	Foreign-born, arrived 1987 +		
Swedish-born	6,527,376	243,834	144,533	170,043	7,085,786
Foreign-born, arrived < 1987 ¹	254,352	172,611	18,911	10,652	456,526
Foreign-born, arrived 1987 +	123,089	41,110	404,712	15,500	584,411
Status unknown ²	39,876	983	1,194	163,754	205,807
Total	6,944,693	458,538	569,350	359,949	8,332,530

1: Mothers/fathers are assumed to have first arrived before 1987 if, prior to 1987, they are recorded as having resided in Sweden, emigrated, died, received Swedish citizenship, or received a first residence permit, even if their actual year of arrival is not recorded; 2: Status unknown includes those cases where mothers/fathers are missing either country of birth or year of arrival. Source: Authors' analysis of Swedish register data.

Researchers may also wish to analyse other aspects of parental background that are linked to migration. As mentioned, this is more difficult for the foreign-born population, in particular those who arrive as adults. *Table 9.2* instead focuses on the children of immigrants, specifically those who were born in Sweden and resident in the Swedish-born population in 2020. It highlights several issues that may arise when analysing both parental country of birth and parental year of arrival (or more generally other combinations of parental variables). First, (as in *Table 9.1*), there are a number of individuals who are missing information on one or both parents. In some cases, there are substantive reasons for this, for example because they are adopted (note that 'parents' here refers to biological parents). Second, the fact that people typically have two parents mean that there are a range of combinations, (each of which may have different levels of relevance for the researcher or research-user). For example, the highlighted cells in *Table 9.2* indicate the 749,049 Swedish-born people who lived in Sweden in 2020 and had a foreign-born parent who arrived in Sweden for the first time after 1986. However, only 54% of these had two foreign parents who arrived in Sweden for the first time after 1986. The rest had parents with a more dissimilar migration background from each other (or one parent whose migration background was at least partially unknown).

There are many more ways that researchers can analyse parental background other than examining country of birth and year of arrival. With respect to factors relating to migration, the list of other potential aspects includes: age at migration, country of previous residence, citizenship, and emigration, as well as residence permits and categories of admission. As is well-known among migration scholars, groups that appear to be homogeneous are often heterogeneous when disaggregated by other factors, and this is also true of *Table 9.2*. For example, the REFU-GEN project includes a focus on the children of immigrants who arrived in Sweden as refugees. We include further analysis in the Appendix (*Table A9.1*), which shows how the analysis in *Table 9.2* might be expanded to include mother's and father's first residence permits, focussing on those who have at least one foreign-born parent who arrived after 1986 (which is when the high-quality data on residence permits begins to be available, see *Section 7*). This analysis shows that many individuals have mothers and fathers who arrived in Sweden with different types of admission category (first residence permit). For example, many people had mothers who arrived as refugees but fathers who were admitted under family ties or with another residence status, in addition to those with fathers who were Swedish-born. This heterogeneity of parental background would only be further enriched analytically by considering additional factors or more detailed categories such as specific parental countries of birth.

10. Conclusion

This article has provided an overview of the various ways that Swedish register data may be used to study immigrants and their descendants. Using new analysis, it has illustrated key issues with respect to data quality and the strengths and weaknesses of Swedish register data for researching this topic. The first two sections introduced the data and explained how it can be used to measure migration background. One of the key issues that it raised was coverage. This is a clear strength of the register data because it includes almost all members of the (registered) population. However, the data are not without quality issues, including with respect to duplicates and over-coverage.

One of the key aspects of working with register data is to understand how the availability and quality of the data has changed over time. Sections 4-8 describe changes over time in four key aspects of the data for studies of immigrants: country of birth, immigration and emigration events, citizenship, and residence permits (categories of admission). For the first three of these categories, data are of high quality for the population who were resident in Sweden after the population registers began in 1968. For residence permits, the data are of high quality from 1987 onwards, when the data are available for researchers in annual format. In projects like REFU-GEN that have a focus on refugees, this makes it hard to study those who arrive as refugees, or under other admission categories, prior to 1987. However, it is possible to focus on specific types of immigrant using knowledge of Sweden's migration history, combined with other information in the registers such as year of arrival and individual country of birth.

The final analytical sections (8 and 9), explain the intergenerational linkages that enable researchers to study the descendants of immigrants and their parental migration background. On the one hand, these data have a specific set of issues that must be understood in order to understand why data are 'missing', or at least unavailable, for certain individuals. In some cases, this may impact the validity of research. On the other hand, the intergenerational linkages in the Swedish register data enable a host of possibilities for studying the characteristics of immigrants' descendants. This includes many aspects of parental background that too numerous to consider here, as well as the possibilities of using linkages between parents and children to study different types of family relations, such as siblings or cousins. Considering this, and the many other benefits of the Swedish register data that we have highlighted here, it seems reasonable to conclude that they are a world-leading resource for studying immigrants and their descendants. With this in mind, we hope that this article helps to facilitate future research and improve its ability to study important questions on the topic as rigorously as possible.

11. References

- Andersson, G., Monti, A., & Kolk, M. (2023). *Vem bor här?: en ESO-rapport om gamla och nya folkräkningar* (Rapport till Expertgruppen för Studier i Offentlig ekonomi.). Regeringskansliet: Statens offentliga utredningar. https://eso.expertgrupp.se/wp-content/uploads/2022/03/ESO-rapport-2023_2_webb.pdf. Accessed 1 February 2024
- Andersson, R., & Solid, D. (2003). Dispersal policies in Sweden. In *Spreading the "Burden"?: A Review of Policies to Disperse Asylum Seekers and Refugees* (1st ed., pp. 65–102). Bristol University Press. <https://doi.org/10.46692/9781847425782>
- Andersson, R., Turner, L. M., & Holmqvist, E. (2010). Contextualising ethnic residential segregation in Sweden: welfare, housing and migration-related policies.
- Bernitz, H. (2012). *Country Report: Sweden*. EUDO Citizenship Observatory: European University Institute, Florence Robert Schuman Centre for Advanced Studies.
- Bevelander, P. (1999). The employment integration of immigrants in Sweden. *Journal of Ethnic and Migration Studies*, 25(3), 445–468. <https://doi.org/10.1080/1369183X.1999.9976695>
- Bevelander, P. (2011). The Employment Integration of Resettled Refugees, Asylum Claimants, and Family Reunion Migrants in Sweden. *Refugee Survey Quarterly*, 30(1), 22–43. <https://doi.org/10.1093/rsq/hdq041>
- Brooke, H. L., Talbäck, M., Hörnblad, J., Johansson, L. A., Ludvigsson, J. F., Druid, H., et al. (2017). The Swedish cause of death register. *European Journal of Epidemiology*, 32(9), 765–773. <https://doi.org/10.1007/s10654-017-0316-1>
- Careja, R., & Bevelander, P. (2018). Using population registers for migration and integration research: examples from Denmark and Sweden. *Comparative Migration Studies*, 6(1), 19. <https://doi.org/10.1186/s40878-018-0076-4>
- Ekbom, A. (2011). The Swedish Multi-generation Register. In J. Dillner (Ed.), *Methods in Biobanking* (pp. 215–220). Totowa, NJ: Humana Press. https://doi.org/10.1007/978-1-59745-423-0_10
- Etikprövningsmyndigheten. (2023). Protecting the individual in research. *Etikprövningsmyndigheten*. <https://etikprovningsmyndigheten.se/en/>. Accessed 11 October 2023
- Hamm, N. C., Hamad, A. F., Wall-Wieler, E., Roos, L. L., Plana-Ripoll, O., & Lix, L. M. (2021). Multigenerational health research using population-based linked databases: an international review. *International Journal of Population Data Science*, 6(1), 1686. <https://doi.org/10.23889/ijpds.v6i1.1686>
- Ludvigsson, J. F., Almqvist, C., Bonamy, A.-K. E., Ljung, R., Michaëlsson, K., Neovius, M., et al. (2016). Registers of the Swedish total population and their use in medical research. *European Journal of Epidemiology*, 31(2), 125–136. <https://doi.org/10.1007/s10654-016-0117-y>
- Ludvigsson, J. F., Håberg, S. E., Knudsen, G. P., Lafolie, P., Zoega, H., Sarkkola, C., et al. (2015). Ethical aspects of registry-based research in the Nordic countries. *Clinical Epidemiology*, 7, 491–508. <https://doi.org/10.2147/CLEP.S90589>
- Ludvigsson, J. F., Otterblad-Olausson, P., Pettersson, B. U., & Ekbom, A. (2009). The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. *European Journal of Epidemiology*, 24(11), 659–667. <https://doi.org/10.1007/s10654-009-9350-y>
- Ludvigsson, J. F., Svedberg, P., Olén, O., Bruze, G., & Neovius, M. (2019). The longitudinal integrated database for health insurance and labour market studies (LISA) and its use in medical research. *European Journal of Epidemiology*, 34(4), 423–437. <https://doi.org/10.1007/s10654-019-00511-8>
- Monti, A. (2020). Re-emigration of foreign-born residents from Sweden: 1990–2015. *Population, Space and Place*, 26(2), e2285. <https://doi.org/10.1002/psp.2285>
- Monti, A., Drefahl, S., Mussino, E., & Härkönen, J. (2020). Over-coverage in population registers leads to bias in demographic estimates. *Population Studies*, 74(3), 451–469. <https://doi.org/10.1080/00324728.2019.1683219>
- Mussino, E., Santos, B., Monti, A., Matechou, E., & Drefahl, S. (2023). Multiple systems estimation for studying over-coverage and its heterogeneity in population registers. *Quality & Quantity*. <https://doi.org/10.1007/s11135-023-01757-x>

- National Board of Health and Welfare. (2023, February 8). Registers. *Socialstyrelsen*.
<https://www.socialstyrelsen.se/en/statistics-and-data/registers/>. Accessed 11 October 2023
- SOU. (1982). *Invandringspolitiken, Bakgrund: Delbetänkande av invandrarpolitiska ommittén (Immigration Policy, Background: Interim report of the Committee on Immigration Policy)*. Stockholm: Statens offentliga utredningar.
- Statistics Sweden. (2006). *Historic population register*.
https://share.scb.se/ov9993/data/publikationer/statistik/_publikationer/be9999_2006a01_br_b_e96st0603.pdf. Accessed 19 October 2023
- Statistics Sweden. (2016a). *Bakgrundsfakta – Personnummer (Background Facts – Personal identity number)*.
- Statistics Sweden. (2016b). *From Finland to Afghanistan – immigration and emigration since 1970 for persons born in different countries*.
http://www.scb.se/Statistik/_Publikationer/BE0701_2016A01_BR_BE51BR1601.pdf.
 Accessed 17 October 2016
- Statistics Sweden. (2017a). *Multi-Generation register 2016: A description of contents and quality*.
- Statistics Sweden. (2017b). *Documentation of STATIV: 1997-2015*.
<https://www.scb.se/contentassets/659b9a5233dc4dd49b22630b2745ca57/dokumentation-en-stativ.pdf>. Accessed 25 October 2023
- Statistics Sweden. (2022). *STATIV: Longitudinellt register för integrationsstudier [A longitudinal database for integration studies]*.
- Statistics Sweden. (2023a). Ordering microdata. *Statistiska Centralbyrån*.
<https://www.scb.se/en/services/ordering-data-and-statistics/ordering-microdata/>. Accessed 11 October 2023
- Statistics Sweden. (2023b). Longitudinell integrationsdatabas för sjukförsäkrings- och arbetsmarknadsstudier (LISA). *Statistiska Centralbyrån*. <https://www.scb.se/vara-tjanster/bestall-data-och-statistik/bestalla-mikrodata/vilka-mikrodata-finns/longitudinella-register/longitudinell-integrationsdatabas-for-sjukforsakrings--och-arbetsmarknadsstudier-lisa/>. Accessed 18 October 2023
- Statistics Sweden. (2023c). Registret för integrationsstudier (STATIV). *Statistiska Centralbyrån*.
<https://www.scb.se/vara-tjanster/bestall-data-och-statistik/bestalla-mikrodata/vilka-mikrodata-finns/individregister/registret-for-integrationsstudier-stativ/>. Accessed 18 October 2023
- Statistics Sweden. (2023d). Population by country of birth, age and sex. Year 2000 - 2022. *Statistikdatabasen*.
http://www.statistikdatabasen.scb.se/pxweb/en/ssd/START__BE__BE0101__BE0101E/FodelselandArK/. Accessed 19 October 2023
- Statistics Sweden. (2023e). Immigration and emigration 1970–2022 and projection 2023–2070. *Statistiska Centralbyrån*. <https://www.scb.se/en/finding-statistics/statistics-by-subject-area/population/population-projections/population-projections/pong/tables-and-graphs/immigration-and-emigration-by-sex-and-country-of-birth-and-projection/>. Accessed 23 October 2023
- Wallace, M., & Wilson, B. (2021). Age variations and population over-coverage: Is low mortality among migrants merely a data artefact? *Population Studies*, 1–18.
<https://doi.org/10.1080/00324728.2021.1877331>

12. Appendix

Table A1.1: Duplicates in the Swedish registers

	Frequency
Unique identification numbers	16,685,567
Identification numbers with more than one version (including surplus duplicates)	233,416
Identification numbers with more than one version (excluding surplus duplicates)	115,416
Surplus identification numbers (to be removed leaving one version for each different number)	118,000
Final list of identification numbers after keeping one version of each duplicate number	16,800,983

Source: Authors' analysis of Swedish register data.

Table A1.2: Population by year and whether IDs have been reused

Coverage	Reused IDs (frequency)	Reused IDs (% of total)	Total IDs (frequency)
Population 1970	6,985	0.09	8,081,078
Population 1980	5,788	0.07	8,317,886
Population 1990	5,700	0.07	8,590,516
Population 2000	7,076	0.08	8,882,681
Population 2010	9,479	0.10	9,415,500
Population 2020	8,688	0.08	10,379,263
Population 1970-2020	19,048	0.12	16,010,473
Not in population 1970-2020	2,106	0.27	790,510
Population 1968-2021	19,792	0.12	16,393,700
Not in population 1968-2021	1,362	0.33	407,283
All register data in REFU-GEN	21,154	0.13	16,800,983

Note: Populations are based on individual identification numbers (IDs) excluding surplus identification numbers (leaving one version for each different number). Source: Authors' analysis of Swedish register data.

Table A1.3: Population by year and whether IDs are incorrect

Coverage	Incorrect IDs (frequency)	Incorrect IDs (% of total)	Total IDs (frequency)
Population 1970	77	0.00	8,081,078
Population 1980	33	0.00	8,317,886
Population 1990	13	0.00	8,590,516
Population 2000	0	0.00	8,882,681
Population 2010	10	0.00	9,415,500
Population 2020	15	0.00	10,379,263
Population 1970-2020	265	0.00	16,010,473
Not in population 1970-2020	52,293	6.62	790,510
Population 1968-2021	277	0.00	16,393,700
Not in population 1968-2021	52,281	12.84	407,283
All register data in REFU-GEN	52,558	0.31	16,800,983

Note: Populations are based on individual identification numbers (IDs) excluding surplus identification numbers (leaving one version for each different number). Source: Authors' analysis of Swedish register data.

Table A4.1: Top 20 largest countries of birth by population size, 1970

Rank	Country	Frequency	%	Cum. %
1	Finland	235,845	44.2	44.2
2	Norway	44,371	8.3	52.6
3	Germany	41,636	7.8	60.4
4	Denmark	39,154	7.3	67.7
5	Yugoslavia	30,785	5.8	73.5
6	Estonia	18,266	3.4	76.9
7	USA	12,724	2.4	79.3
8	Greece	12,023	2.3	81.6
9	Poland	10,898	2.0	83.6
10	Hungary	10,646	2.0	85.6
11	Austria	7,728	1.5	87.1
12	Czechoslovakia	7,447	1.4	88.5
13	Italy	7,288	1.4	89.8
14	USSR	7,256	1.4	91.2
15	UK	5,394	1.0	92.2
16	Spain	3,758	0.7	92.9
17	Turkey	3,754	0.7	93.6
18	Latvia	3,265	0.6	94.2
19	Netherlands	2,896	0.5	94.8
20	France	2,452	0.5	95.2

Source: Authors' analysis of Swedish register data.

**Table A4.2: Population by country of birth,
whole extract of data in REFU-GEN**

Country of birth	Frequency	Per cent	Cumulative
Nordic	738,975	22.37	22.37
Northern Europe	135,399	4.10	26.47
Western Europe	200,977	6.08	32.55
Southern Europe	351,549	10.64	43.20
Eastern Europe	312,532	9.46	52.66
Northern Africa	53,861	1.63	54.29
Eastern Africa	185,773	5.62	59.91
Western Africa	30,237	0.92	60.83
Central Africa	12,935	0.39	61.22
Southern Africa	6,871	0.21	61.43
Western Asia	534,882	16.19	77.62
Central Asia	11,559	0.35	77.97
Southern Asia	299,662	9.07	87.04
Eastern Asia	96,915	2.93	89.98
Southeastern Asia	109,760	3.32	93.30
Oceania	15,928	0.48	93.78
Northern America	77,210	2.34	96.12
Caribbean	7,216	0.22	96.34
Central America	14,465	0.44	96.78
South America	106,515	3.22	100.00
All foreign-born	3,303,221		

Source: Authors' analysis of Swedish register data.

Table A9.1: Parental country of birth, year of arrival and first residence permit for the **Swedish-born population in 2020 with at least one parent who arrived after 1986**

	Father		Foreign-born, arrived 1987+			Unknown ³	Total
	Swedish-born	Foreign-born, arrived < 1987 ¹	...first residence permit as refugee	...first residence permit as family ties	...other first residence permit status ²		
Mother							
Swedish-born			26,407	79,374	38,752		144,533
Foreign-born, arrived < 1987 ¹			3,511	13,026	2,374		18,911
Foreign-born (FB), arrived 1987+							
...first residence permit as refugee	12,729	4,102	94,047	45,351	3,963	8,045	168,236
...first residence permit as family ties	75,645	33,738	113,234	71,664	34,638	5,923	334,842
...other first residence permit status ²	34,715	3,270	3,581	12,300	25,935	1,532	81,333
Unknown ³			611	355	228		
Total	123,089	41,110	241,391	222,070	105,889	15,500	749,049

1: Mothers/fathers are assumed to have first arrived before 1987 if, prior to 1987, they are recorded as having resided in Sweden, emigrated, died, received Swedish citizenship, or received a first residence permit, even if their actual year of arrival is not recorded; 2: Those with 'other' first residence permit status includes those who do not require a residence permit (e.g. based on their citizenship) as well as those who had a different type of first residence permit (e.g. for employment or studies); 3: Status unknown includes those cases where mothers/fathers are missing either country of birth or year of arrival. Source: Authors' analysis of Swedish register data.

Stockholm Research Reports in Demography

Stockholm University,
106 91 Stockholm,
Sweden

www.su.se | info@su.se | ISSN 2002-617X



Stockholm
University

Demography Unit