

Chapter 7

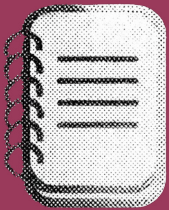
Register data sources on migrant stocks

Laura Peitz

To cite: Peitz, L. (2025). Register data sources on migrant stocks. In D. Kierans and A. Kraler (eds), *Handbook on Irregular Migration Data. Concepts, Methods and Practices*. Krems: University of Krems Press.
<https://doi.org/10.48341/g31s-vq79-07>

Keywords: Register data, administrative sources, migrants under an obligation to leave, toleration, Duldung, Irregular migrant stocks, Data integration, AZR, Germany

Register data sources on migrant stocks



Key points

- Analysing irregular migrant stocks using register or other administrative data can prove challenging given the usually undocumented nature of the phenomenon; yet, some recent efforts highlight the potential of register data.
- The German Central Register of Foreigners (AZR) provides longitudinal data on non-Germans staying or having previously stayed in the country, including subsets of irregularly staying migrants, allowing for in-depth analyses of various research questions.
- Throughout, the chapter provides real-life examples of how register data has been used in irregular migration research across different countries.

Introduction

Counting the number of irregular migrants or more closely investigating this target group using administrative data presents profound challenges. In most national contexts, registration in population registers or similar systems is tied to a legal residence status. Consequently, migrants without the legal right to stay are typically excluded from such databases altogether. Moreover, even if they could be registered, irregularly staying migrants might deliberately avoid contact with public authorities to minimize the risk of detection and possible deportation. This further limits their visibility in administrative data systems. Single administrative data sources may contain data on irregular migrants interacting with particular

public institutions – for example for schooling or urgent healthcare, or upon regularization or police force encounters – but this data is oftentimes incomplete, fragmented, inconsistent, and usually not linked to broader administrative registers.

While these obstacles pose significant barriers to the statistical inclusion of irregular migrant populations, some efforts have recently emerged to provide details on irregular migrant stocks based on administrative data sources (see also UNECE, 2025). In **Italy**, for instance, irregularly staying migrants can be identified by comparing data from various administrative sources and applying the Signs of Life method (see Box 7.1).

A similar approach has recently been applied in **Chile**, where the number of irregular migrants is estimated by integrating data from post-census administrative records on education, tourist stays, and police reports, and comparing this data against the baseline of residence permit applicants (see Box 7.2). In **Spain**, everybody is encouraged to register in the municipal population registers of their municipality (*Padrón Municipal*). The registration is a prerequisite for accessing basic rights and public services, such as health care and schooling. It is independent of legal status and – crucially – is not used for immigration control. As a consequence, the *padrones* even include irregularly staying migrants. When comparing or linking the *padrón* data to other administrative datasets, it is possible to assess questions around migrant irregularity, such as deriving the number of irregularly staying migrants

from a comparison of the *padrones* with the database of legal stay permits (González-Enríquez, 2016). In the **UK** and **Poland**, recent efforts have been undertaken to produce a time series of the number of irregularly staying migrants based on the ethnic economies approach and non-linear count regression models. The assumption is that regularly settled ethnic groups provide support for individuals of similar ethnicity from their countries of origin to circumvent national restrictions on migration rules regarding work. Based on this, the numbers of detentions extracted from official police and border enforcement data are scaled up to the regularly residing foreign population using non-linear count regression models to estimate the number of irregularly staying migrants per country of origin (Beręsewicz, 2024).¹

Box 7.1: Applying the ‘Signs of Life’ method: The case of Italy

Marco Marsili

In Italy, the National Institute of Statistics (Istat) identifies the number of irregularly staying migrants by applying the Signs of Life (SoL) approach. To this end, in a first step, data on migration (changes of residence) are drawn from the centralized population register (ANPR, managed by the Ministry of the Interior). These data are subjected to standard control and correction procedures. In general, the quality of the data is quite high; in case of partial non-response, the information is filled using donor hot-deck methods of imputation or by retrievals from the previous year’s census, where available.

In a second step, the information of the ANPR is integrated into a demographic data system (Midea-Anvis, Micro-DEmographic Account - Virtual Statistical register of the population) which, in addition to migrations, also incorporates data of other population changes (births, deaths, acquisitions of citizenship). Midea-Anvis is a counting system based on micro-data, in which all data are integrated with each other and with respect to the population of the last census, in order to verify the stock-flow coherence of the information acquired.

The last step is comparing Midea-Anvis with a large set of administrative archives (AIDA, Integrated Archive of Administrative Data), including, among others, the tax, social security, energy consumption, and education registers as well as the Cadastre of buildings and constructions. Each administrative archive in AIDA provides life signals on habitually resident persons who have spent a significant amount of time in Italy over the last three years. The comparison between AIDA and Midea-Anvis produces three distinct datasets:

1. individuals present in Midea-Anvis and confirmed as residents through the life signals system in AIDA (the so-called “usual resident population”);
2. individuals not present in Midea-Anvis but with strong life signals in AIDA (under-coverage);
3. individuals present in Midea-Anvis but without life signals in AIDA (over-coverage).

¹ This research was led by Brendan Georgeson (Office for National Statistics, UK) and Maciej Beręsewicz (Poznań University of Economics and Business, Poland).

The second of these datasets is relevant information about irregular migrants. It comprises all those individuals who, despite not having specific authorization to reside in the national territory (for example, because they have an expired residence permit) show signs of administrative life in Italy (e.g. because of working, studying or awaiting residence permit renewal). Hence, as regards irregular or undocumented migrants, the current structure of the data production system allows to correctly focus on a specific group of irregularly staying migrants. On the other side, the evaluation of fully undocumented migrants is most challenging as, by definition, they do not show any sign of life. Nonetheless, Istat also produces national estimates of fully undocumented migrants on a yearly basis. The sources used to produce these estimates have varied over the years, depending on data availability, including sample surveys and data from administrative sources. In recent years, the methodology has been improved by also integrating data of the Ministry of the Interior relating to police stops on the territory or at the border, as well as data relating to actual repatriations to countries of origin.

In contrast, **Germany** presents a unique case in this regard, as it has been using a Central Register of Foreigners (AZR) for over 70 years, which includes comprehensive data on the majority of non-nationals staying in the country, even parts of those without legal residency status. The reason is that in Germany, many migrants who are formally obliged to leave the country are issued a *Duldung* (tolerated status) while their removal is temporarily suspended due to either actual obstacles (e.g.

missing travel documents or illness) or legal reasons (e.g. family unity) preventing deportation. These migrants are well captured in the Central Register of Foreigners, meaning that the register is suited for detailed analyses aimed at specific subgroups of irregular migrants. The rest of this chapter will present the Central Register of Foreigners and the potentials and pitfalls of using its administrative data in irregular migration research.

Box 7.2: Chile's experiences in integrating data for estimating the foreign population with irregular migration status²

Julibeth Rodríguez and Felipe Mallea

Since 2014, Chile has witnessed an increasing migrant flow, which has meant that the country must assume the challenge of officially measuring the phenomenon. To this end, the National Statistics Institute (INE) and the National Migration Service (Sermig) have developed a methodology for estimating the number of foreign nationals residing in Chile between censuses by linking border-control data with residency applications after the 2017 Census.

This study employs a methodology that integrates baseline data from the census with data from post-census administrative records by linking microdata of various government institutions. A critical component is using administrative records to identify populations with a potentially irregular migration status.

2 This work was conducted by the Studies Department at the National Migration Service of Chile and the Demography Subdepartment at the National Institute of Statistics of Chile (team members: Gabriel Santander, Consuelo Salas, Marisol Opazo, Pablo Roessler, Felipe Hugo, Luis Rodríguez, Miguel Ojeda, Francisco González). More details are available at: <https://serviciomigraciones.cl/estudios-migratorios/estimaciones-de-extranjeros/> and <https://www.ine.gob.cl/estadisticas/sociales/demografia-y-vitales/demografia-y-migracion>.

The methodology acknowledges the complexities in precisely quantifying irregular migration, which are due to the multifaceted character and the dynamic fluidity of migration status. Consequently, the scope of the estimation is deliberately confined to a specific subset of the population with irregular status, a delimitation necessitated by the availability of relevant data sources and by the objective of clearly distinguishing the populations with regular and irregular migration status.

The estimation of the population with irregular status specifically includes individuals who do not possess any type of residency permit application, who have been in the country for at least six months, and who have no recorded departure for the period ending on December 31, 2023. The sources for the estimation are as follows:

1. The biometric control system (between June and December 2023)
2. Expired tourist visa extensions or police reports (including both formal denunciations for unauthorized entry and self-reported clandestine entries)
3. Primary and secondary student enrollment in Chilean educational institutions of those who are assigned a provisional identifier because they lack a national identification number (RUN)

By including a wide range of administrative records, we can account for the two main areas that form the basis of irregular migration of foreign nationals in Chile: (1) those who enter the country clandestinely and who cannot apply for a residency permit, and (2) those who enter the country legally and who cannot apply for a residency permit. With these two areas and their combination with records from border control, it can be determined whether the person was in the country for the period ending on December 31, 2023. In processing the data, 33,251 people who left the country were excluded. In contrast, we included those whose presence and residence in Chile were shown by the records of their administrative acts to be subsequent to their exit from the country.

The final dataset for the population with irregular migration status comprised 336,984 individuals, whose information was categorized by primary source: 261,449 from police reports, 10,217 from expired tourist visas, and 65,318 from official enrollments without a national identification number (RUN).

In conclusion, this study presents a methodology that contributes to international migration statistics by integrating census data with diverse administrative data. While recognizing the inherent uncertainties in estimating the population with irregular migration status, the results offer valuable insights for targeted public policy design and demonstrate potential for adaptation in other countries, which would thereby improve the quality and comparability of regional migration data.

Structure and contents of the AZR

The *Ausländerzentralregister* (AZR, Central Register of Foreigners) is Germany's primary administrative register for non-German nationals living in Germany. Established in 1953 and governed by law (*Gesetz über das Ausländerzentralregister*), the AZR plays a central role in federal and local migration governance, strategic planning as well as in daily

migration-related administrative activities. The data is entered into the register primarily by local immigration offices (*Ausländerbehörden*) as well as other public institutions such as the Federal Office for Migration and Refugees (*Bundesamt für Migration und Flüchtlinge*, BAMF) or federal and state police forces. Various public authorities use

the data stored in the AZR to support their case-by-case decision-making, operational planning, and political decision-making. Over the past years, the register has become an important, though complex, information source on various migration-related questions for public administration, the government, the media, the broader public, and for researchers interested in understanding patterns and dynamics of migration and residence (Brückner, 2019; Peitz, 2025; Tanis, 2022; Weber, 2022).

The AZR covers all non-German nationals who reside in or have resided in Germany for more than three months. In addition, it includes data of individuals who have filed an asylum claim and of those who have been issued residence law decisions, such as expulsion or deportation orders. The data recorded is stored in the AZR for the duration of an individual's stay in Germany, and usually for ten years after their departure (five years after death). All data entries are deleted from the AZR upon naturalization, without the possibility of further tracking these individuals given the lack of a central population register in Germany.

The AZR contains various data attributes per individual. Which types of attributes are stored depends on the specific group of migrants. Only rudimentary information is stored in the case of EU citizens, while the most comprehensive data is collected on individuals entering the asylum system.

The variables contained include:

- Personal data: unique AZR identifier, full name, date of birth, gender, nationality, marital status
- Border crossings: entries, voluntary departure, forced return
- Residence status: temporary and permanent residence titles, *Duldung*, obligations to leave the country
- Asylum procedure: application filed, asylum status, rejection

In addition to this “core” data, the AZR has in recent years been expanded by multiple additional variables, including language skills, education and profession, postal address, and integration course information, but the quality of these variables varies (see below).

With exception of time-invariant personal data, data entries in the AZR are usually location- and time-stamped: They contain the date of the respective data entry as well as the municipal level of the executive authority (which usually corresponds to individuals' place of residence), along with the respective federal state. Importantly, whenever new information is entered for many of the ‘core’ variables, the previous data entry is not overwritten. Instead, all previous information on these variables is kept as long as an individual's data is stored in the AZR (see Gleiser & Hinz, 2024, p. 8). This way, the AZR data allows for longitudinal and flow analyses.

Irregular migration stocks and flows based on AZR data

The AZR can provide indicators on irregular migration stocks and flows. However, one needs to carefully delineate the groups of irregular migrants who are, and who are not, included in AZR data. Being an administrative register utilized and filled by public authorities, the AZR, virtually by definition, contains only data on migrants with contact to the authorities. Based on the MİRreM taxonomy

(Kraler, 2023), the following groups of irregularly staying migrants (migrants with an obligation to leave the country) can be identified using the AZR: individuals who are issued a return decision, whose status is expired or revoked, and whose removal is formally suspended.³ The following flows into and out of irregularity can be traced based on AZR data: immigration, being born into an obligation to

³ Due to the specific filter functions in the AZR, identifying these groups is possible in the most current cross-sectional dataset. It is, however, not necessarily possible for all these groups retrospectively in the longitudinal dataset.

leave, loss of status, death, voluntary and enforced departure, and regularisation (including the permit issued when entering new asylum procedures).

Undocumented migrants who have never come into contact with authorities remain outside the scope of the AZR, making the register incomplete for fully irregular populations. Similarly, data for irregular migrants who are identifiable using the AZR is only up-to-date as long as individuals show administrative signs of life in the AZR or have undoubtedly terminated their irregular status (by a registered departure or by regularization). In the case of individuals who no longer show signs of life in the AZR and who do not have a documented termination of irregularity, it is not always clear from the data whether they have left the country without official knowledge or gone into hiding.

However, given Germany's *Duldung* system and the wide spread of this provisional 'status', which regularly needs to be renewed (Schütze, 2023), many migrants staying in Germany irregularly are in contact with the authorities and have recurrent positive data entries in the AZR. In addition, a large group of irregularly staying migrants in Germany are rejected asylum seekers, who always have been in contact with authorities at some point in time, and, when possessing a *Duldung*, subsequently

are. As a consequence, the AZR contains the entire residence history of virtually the full sample of individuals obliged to leave the country and showing signs of life in the AZR (e.g. by being in possession of a *Duldung*). And, given AZR storage policies, it contains not only data of those present in Germany at the time of data extraction, but also of those who have been present within the previous ten years.

This highlights the unique advantages of using AZR data in analyses on irregularly staying migrants: Being able to analyse close to a full sample of the specified sub-group, usually relatively promptly and without temporal delay, and in greater detail than with other administrative data sources. This includes the availability of time- and location-stamped data allowing for longitudinal analyses with a dynamic perspective, such as analyses related to the duration of stay or timing of key transitions in great temporal granularity (for an example, see Box 7.3). The AZR can also be used to investigate period effects (such as new laws, political or societal events) with a quasi-experimental framework (e.g. Peitz & Carwehl, 2025). Also, the AZR permits interregional comparison or the assessment of interregional mobility patterns, and serves as a solid sampling frame to draw representative samples for survey research.

Box 7.3: Using AZR data to analyse pathways out of irregularity: An application example

Laura Peitz

A recent analysis of pathways out of an obligation to leave presents a compelling case for using longitudinal administrative data to understand temporal dynamics in irregular migration (Peitz, 2025). Drawing on data from the German Central Register of Foreigners (AZR) between 2013 and 2022, the study examines the trajectories of over 400,000 individuals whose asylum claims were finally rejected. This rich dataset allows for a nuanced temporal analysis of how individuals exit irregularity – through voluntary departure, deportation, or regularization.

By applying event history analysis within a competing risk framework, the study shows that temporal dynamics significantly shape outcomes. Voluntary departures are most likely in the first two years following a final rejection, while regularizations increase in likelihood with the length of stay. In contrast, deportations remain relatively rare and largely concentrated in the early years following a final rejection.

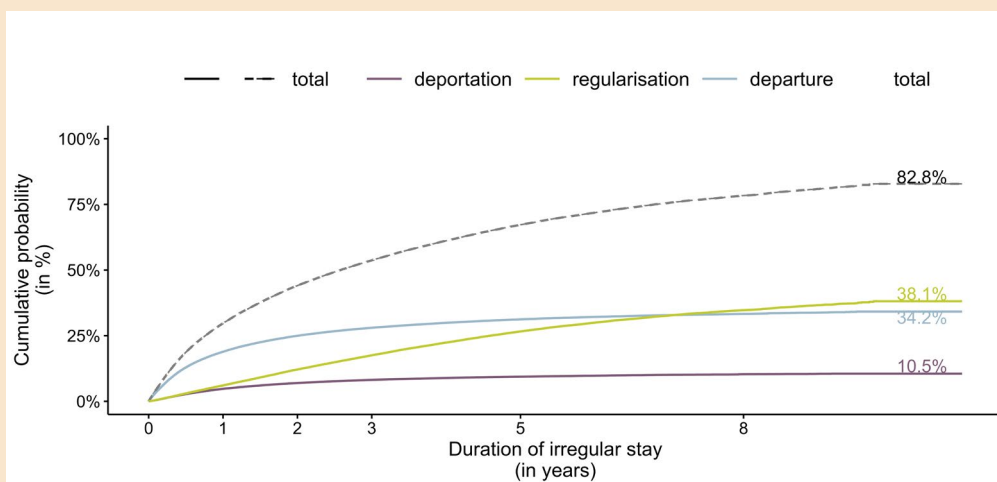


Figure 7.1: Cumulative incidence of competing pathways out of irregular status (Source: Peitz 2025, p. 19)

The analysis also reveals how policy instruments interact with time. For instance, the impact of different types of Duldung is distinctly time-sensitive: the restrictive ‘Duldung light’ for individuals with unclear identity tends to prolong irregularity, while employment and training-related types of Duldung facilitate regularization – but only after years of legal limbo. The study also shows that while designating countries of origin as “safe” aims to accelerate return, the actual timing of exits varies more by structural factors than by policy labels alone.

In sum, administrative longitudinal data unlocks critical insights into when and how irregular migrants transition out of legal limbo. Such data enables the evaluation of migration policies over time, offers evidence for reforming regularization schemes, and underscores the importance of integrating temporal dimensions into migration research and governance.

References:

Peitz, L. (2025). Return or regularization? A temporal analysis of rejected asylum seekers in Germany (RSC Working Paper 2025/06). <https://hdl.handle.net/1814/78047>

At the same time, working with AZR data for scientific purposes comes with some pitfalls that need to be carefully addressed. The AZR is not primarily a scientific database. Rather, its structure reproduces the legal complexity of Germany's migration governance system, and data quality can reflect the local administrative capacity of individual authorities and the effectiveness of data exchange between agencies. For instance, some variables, such as residence permits, can carry a multitude of values that need to be combined into meaningful categories during the data management phase.

Despite the recent introduction of automated data entries, implausibility checks during data entry, and automated interfaces between the applications used by different agencies, some issues potentially impairing data quality remain. These can be due to delayed reporting, incomplete status transitions, differences in interpretation or practices across federal and local agencies, or incorrect data entry practices. As a consequence, the raw data can show implausible entries that need to be carefully addressed in data cleaning and processing. Therefore, both for data management and for the interpretation of data and results, detailed knowledge on the legal and administrative procedures is needed.

Conclusion

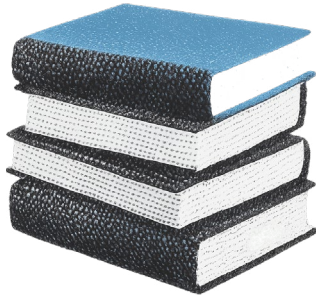
The AZR systematically and comprehensively captures several subgroups of irregular migrants – a category of persons who are usually not well captured by administrative data. Despite the challenges that arise when using such administrative data for scientific analysis, the AZR remains one of the most wide-ranging administrative databases on non-nationals in Europe, and a valuable source for informative irregular migration analyses. In the future, record linkage with other administrative data sources could even enhance the potential of AZR data for research around irregular migration. Prerequisite for such record linkage is a legal basis

Data quality is usually fairly high on 'core' variables, as they are vital for administrative processes and their systematic collection is legally required. In contrast, (some of the) additional variables are not essential for daily administrative tasks, their data collection is legally required only for specific groups (e.g. asylum seekers), or they are handled using other applications than the AZR. Data on these variables is therefore not necessarily entered and kept in the quality that would be necessary for scientific purposes. When wanting to exploit information of such variables, researchers should critically balance the data's informational value against potential bias.

Finally, when publishing and communicating results that are based on register data such as the AZR, it is crucial to clearly state the specific subgroup of irregularly staying migrants that are covered by the data and to be transparent about the informed but inevitably arbitrary decisions made during the data cleaning process. Also, in light of potential implausibilities and case incompleteness, it may be more appropriate to highlight patterns, interrelationships, and dynamics rather than presenting absolute figures that risk conveying a false sense of certainty.

as well as a careful design in accordance with ethical considerations and data protection regulations.

The example of the AZR also shows the added value of making administrative migration databases exploitable for more advanced statistical research, such as longitudinal analyses. For this reason, a random sample of the AZR data is now available to researchers for scientific purposes, via the Research Data Centre of the German Federal Office for Migration and Refugees (Gleiser & Hinz, 2024; Gleiser et al., 2024).



References

Beręsewicz, M. (2024). Estimation of the number of irregular foreigners in Poland using non-linear count regression models. Economic Commission for Europe.

Brückner, G. (2019). Das Ausländerzentralregister als Längsschnittdatenquelle. WISTA, 2019(1), 35–67. https://www.destatis.de/DE/Methoden/WISTA-Wirtschaft-und-Statistik/2019/01/auslaenderzentralregister-laengsschnittdatenquelle-012019.pdf?__blob=publicationFile&v=1

Gleiser, P., & Hinz, T. (2024). Datenreport zum AZR-Forschungsdatensatz 2023 (Veröffentlichungen des BAMF-FDZ 01/2024). <https://doi.org/10.48570/bamf.fdz.voe.01/2024.d.2024.azrdatenreport.1.0>

Gleiser, P., Salas Escobar, C., & Hinz, T. (2024). Codebuch zum AZR-Forschungsdatensatz 2023 (Veröffentlichungen des BAMF-FDZ 02/2024). <https://doi.org/10.48570/bamf.fdz.voe.01/2024.d.2024.azrcodebuch.1.0>

González-Enríquez, C. (2016). Spain: Irregularity as a Rule. In A. Triandafyllidou (Ed.), Irregular Migration in Europe: Myths and Realities (pp. 247–266). Routledge. <https://doi.org/10.4324/9781315589848>

Kraler, A. (2023). Taxonomy of Migrant Irregularity (MirreM Deliverable D1.2). University for Continuing Education Krems. <https://doi.org/10.5281/zenodo.7875543>

Peitz, L. (2025). Return or regularization? A temporal analysis of rejected asylum seekers in Germany (RSC Working Paper 2025/06). <https://hdl.handle.net/1814/78047>

Peitz, L., & Carwehl, A.-K. (2025). Zwischen Duldung und Bleiberecht: Wie und für wen wirkt das Chancen-Aufenthaltsrecht? (BAMF-Kurzanalyse 03/2025). <https://doi.org/10.48570/bamf.fz.ka.03/2025.d.2025.chaufr.1.0>

Schütze, T. (2023). The (Non-)Status of ‘Duldung’: Non-Deportability in Germany and the Politics of Limitless Temporariness. Journal of Refugee Studies, 36(3), 409–429. <https://doi.org/10.1093/jrs/feac056>

Tanis, K. (2022). Die Wohnhistorie Geflüchteter in Deutschland (BAMF Kurzanalyse 01/2022). <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-79874-5>

UNECE. (2025, March 12). Measuring Hard-to-Reach Groups in Administrative Sources. United Nations Economic Commission for Europe. <https://unece.org/sites/default/files/2025-04/HardToReachGroups%20Consult.pdf>

Weber, J. (2022). Binnenmobilität von Geflüchteten mit Schutzstatus in Deutschland: Eine explorative Analyse auf Basis des Ausländerzentralregisters (Forschungsbericht No. 39). https://www.bamf.de/SharedDocs/Anlagen/DE/Forschung/Forschungsberichte/fb39-binnenmobilitaet.pdf?__blob=publicationFile&v=10