Chapter 5
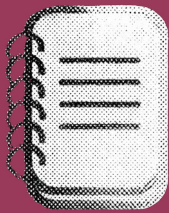
Innovations in methodological approaches to estimate irregular migrant stocks and flows

Alejandra Rodríguez-Sánchez and Jasper Tjaden

Innovations in methodological approaches to estimate irregular migrant stocks and flows

Key points

- This chapter builds on Rodríguez-Sánchez and Tjaden (2023), who reviewed the main methods for estimating irregular migrant stocks and flows, spanning both established and more experimental approaches.

- Measuring irregular migration remains a fundamental challenge: the population is difficult to observe, and even widely used methods such as residual estimation or capture–recapture provide only partial pictures.

- Traditional techniques continue to form the backbone of the field, but improvements have often come from incremental innovations, such as using mortality data to refine life-course approaches, or expanding residual methods with large government databases and machine learning.

- More novel directions, such as exploiting consular registers, driver's licence data, or online search behaviour, show promise in filling gaps, though these remain context-specific and experimental.

Introduction

The scientific study of irregular migration, its description, and its estimation are closely connected with the search for appropriate measurements and empirical observations related to this form of migration. Aside matters related to the definition of irregular migration, treated in other chapters in the book, and especially in Kraler and Ahrens (2023), there are multiple challenges associated with measuring irregular migration stocks and flows and attaining estimates of the size of these quantities.

Some of these challenges were highlighted by the seminal CLANDESTINO project – an EU–funded project (2007-2009) which reviewed data and methodologies on irregular migration over a decade ago (Jandl, 2011). Key obstacles include irregular migrants' reluctance to disclose their status in surveys or censuses, the absence of adequate sampling frameworks, and their elevated mobility patterns—all necessitating alternative research approaches. In MIrreM, we seek to update this review with advances in literature in terms of data

sources and methodologies. For a comprehensive and detailed overview of each method, see the review paper by Rodríguez-Sánchez and Tjaden (2023). In that review, a detailed explanation of how the different methods work, typical databases used, and their strengths/weaknesses are documented. Depending on which definition of irregular migration we employ, some methods might be better suited than others to capture the different operationalizations, especially as these will be based on different data sources
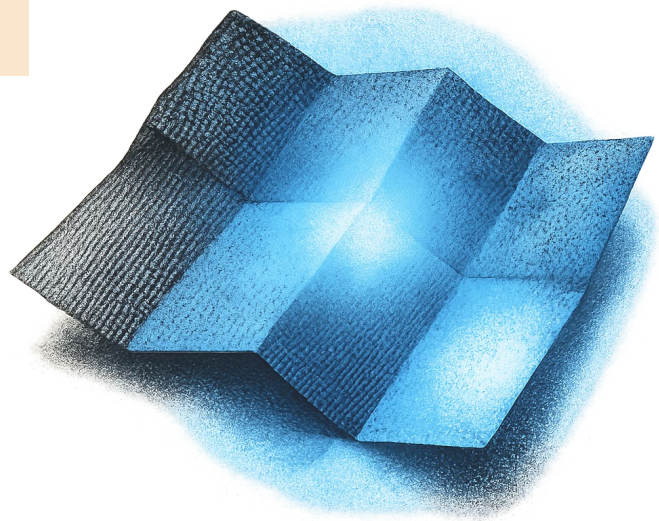
> **Box 5.1: Traditional and innovative approaches**
> *Alejandra Rodríguez-Sánchez and Jasper Tjaden*
>
> By "Traditional" we refer to approaches covered by previous methodological overviews (Jandl, 2011; Pinkerton et al., 2004) on which our overview builds. These are well established methods that are used to estimate irregular migration across the world. We included these traditional methods in our review out of a desire to be comprehensive, but also because some of the innovations build from well-established methods, like the residual approach. We defined "Innovative" approaches as those methods that either use novel data sources (e.g., digital behavioural data) or apply a new estimation method to standard data sources. These approaches improve upon some of the limitations of established methods. The innovative approaches were identified through literature review and discussions with experts.

Estimating irregular migrant stocks and, especially, irregular migration flows, despite important advances, remains a challenging endeavour. Methods estimating stocks measure the total number of irregular migrants residing in a country at a specific point in time (e.g., the year 2022), whereas methods estimating flows capture changes in that population over a defined period (e.g., 2015–2020), attaining measures of inflows or outflows. MIrreM's innovative pilot studies which are summarized in the next chapter, and which were based on this overview, are aimed at tackling some of these challenges.

This chapter of the Handbook is intended to highlight some of the most innovative aspects of approaches we found through our scoping review for measuring both stocks and flows. A quick overview of the methodologies we found in this search can be found in Table 1. We grouped the methods based on their core data (e.g., government data, non-government data, survey, mixed data, and digital data) and estimation strategies, plus a brief description of their main idea, rather than focusing on minor differences across methods. Our review encompasses both traditional and innovative methods (see Box 1).

Type	Government administrative & census	Description
Traditional	Capture-Recapture / Multiple Systems Estimation	The many varieties of this approach are based on a statistical method that estimates population size by comparing how many individuals appear in multiple data sources. For irregular migration, it matches people between enforcement records (like police apprehensions) and general population databases. The detection rate from this overlap is used to estimate the total irregular migrant population. Variations include Multiple Systems Estimation and repeated capture-recapture, all using the same core principle of linking individuals across datasets.
	Regularisation & formal status adjustments	Amnesties or regularizations grant legal status to irregular migrants, revealing population size as a byproduct. Case-by-case adjustments, called "silent regularization," happen continuously without formal announcements, also providing population estimates.
	Life course events	This method compares observed demographic rates (e.g., births, deaths, hospitalizations) among migrants to expected rates in non-migrant populations. When migrant rates exceed projections, the excess suggests the presence of unregistered irregular migrants who are not captured in official statistics.
	Labour demand models	This approach estimates labour demand from economic output, then subtracts reported foreign workers to find irregular migrant workers. A similar method estimates the shadow economy size and assumes migrants' share of irregular work.
	Flow-Stock models	Here, the main idea is that irregular migrant population can be obtained by understanding the initial size of this population at some point, and adjust by inflows (adding subsequent overstayers, border crossings) and outflows (subtracting deportations, emigration, deaths, status changes).
	Irregular border crossings & apprehensions	Irregular migration flows are measured by counting people apprehended at borders or after entry, through checks within a country.
	Database systems enabling the identification of visa overstayers	This method calculates irregular migrants by subtracting legal departures from legal arrivals plus status changes in tracking systems. For instance, the US ADIS system matches entry records with departure and status change data for air, land, and sea travelers. Visa overstayers are identified as those who entered legally but neither left nor gained new legal status. The approach may use biometric and biographical verification to detect overstayers.
Innovative	Using consular registers of migrant communities in combination with other data-sources	This method combines consulate card registrations with survey data. For example, Mexican consulates in the US register nationals and issue cards by region. These counts estimate total resident Mexican nationals, then survey data on undocumented percentages from the American Community Survey is applied to calculate irregular Mexican migrants.
	Drivers' licenses	In some states in the US, comparing the number of drivers licenses before and after law changes that grant irregular migrants access to driver's licenses (e.g., the AB60 reform in California, US) could be indicative of the size of undocumented present, similarly to regularization given the widespread use of cars.
	Consumption data	In this method, estimates of expected food consumption (like rice) based on development levels and daily consumption is compared to actual consumption levels. The excess consumption, unexplained by the official population, is attributed to irregular migrants and converted to population estimates using consumption pattern assumptions.
	Biometrics & surveillance	This method uses AI facial recognition and biometric data to identify irregular migrants by comparing faces against government databases. For example, migrants who entered the country irregularly or overstayed the terms of their visa will not appear in matching official biometric records. Large-scale application across complete databases enables population estimation.

Type	Non-government administrative data	Description
Innovative	Institutional registers: College enrolment, prison population, and human trafficking	Institutional registers (such as universities, prisons, trafficking victim services) capture legal status data. Universities, for example, can identify irregular students through admissions when documentation is missing. Prison systems, in turn, record migrants' legal status after they have committed crimes. NGOs can provide information on trafficking victims. Hence, these sources can provide direct counts of irregular migrants within specific populations potentially useful in combination with other approaches.
	Missing, disappeared or deceased migrants	This method tracks individuals who are missing, disappeared, or died while attempting irregular border crossings or migration journeys. Although this is not a count of irregular migrant flows, it might provide further evidence to understand dynamics of irregular migration flow.
	Survey data	
Traditional	Self-identification in surveys	Surveys can help determine respondents' legal status either by asking directly about residence permits or indirectly by inquiring about rights or other characteristics linked to having/or not having a legal status.
	Expert surveys / Delphi surveys	Researchers survey experts who estimate the population size in specific areas. The Delphi method, for example, involves multiple survey rounds where experts see colleagues' responses and discuss until reaching consensus through iterative feedback.
Innovative	Statistical imputation	This approach uses survey data with direct or indirect legal status questions to train predictive models, then applies these models to larger national surveys lacking such legal status-related questions to identify irregular migrants. Various techniques are used including logical edits, logistic regression, or machine learning.
	Retrospective surveys	In this method surveys on regular migrants can be used to retrospectively infer their previous irregular status periods, allowing for an estimation of previous legal statuses, especially if samples are sufficiently large.
	Mixed data sources	
Traditional	Multiplier/ simple extrapolation	Find the ratio of irregular migrants compared to all foreign-born individuals, then project this proportion onto different geographic scales or contexts.
	Residual estimation method	This method calculates irregular migrant stocks by taking the total migrant count or foreign-born population from census data and subtracting the estimated number of legal migrants. The difference represents the irregular migrant population size.
Innovative	Alternative residual: labour force survey and social security register	This method is based on a comparison of employment data from surveys and official registers. When survey data shows more immigrant workers in certain sectors than official registrations, it suggests irregular migrant workers in those sectors.
	Digital data	
Innovative	Online search behaviour	Search engine data (like Google Trends) tracking searching behaviour of people, could employ information on trends for queries such as "will I be deported" to gauge policy impacts. Analysing search volume changes after events like the "Muslim ban" could help in specific groups of irregular migrants and their responses to policy shifts.

Table 5.1: Overview of methodological approaches covered in Rodríguez-Sánchez & Tjaden (2023) review (Note: Author's own elaboration)

Review criteria

Our review of each methodology was based on a series of criteria we deemed fundamental to understanding the scope of each method, meaning which population the method is able to produce estimates for, and the quality of its estimates. The selection of which features of the methodologies to highlight was based on existing common standards for the evaluation of scientific evidence, such as the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), the Grading of Recommendations, Assessment, Development and Evaluation (GRADE), among others (Guyatt et al., 2008; Page et al., 2021).

Each of the methods was evaluated in Rodríguez-Sánchez and Tjaden (2023) according to the following criteria:

- **Main Idea:** Explains the core concept of the method in simple terms for non-experts.
- **Data Source:** Identifies main data types (e.g., administrative, surveys, census) and gives examples.
- **Coverage / Definition:** Describes which sub-groups of irregular migrants are included or excluded, highlighting potential biases.
- **Estimation Assumptions:** Outlines key assumptions needed for the method to estimate the total irregular migrant population accurately.
- **Reliability:** Assesses whether the method gives consistent results over time.
- **Scalability:** Evaluates whether the method can be applied in different countries.
- **Ethical Issues:** Flags ethical concerns in data use, collection, and potential risks to migrants.
- **Examples:** Provides references to studies that apply the approach.

What we found

Among the traditional approaches, covering both indirect and direct approaches as classified by Jandl (2011), we review multiplier or simple extrapolation, the capture-recapture or multiple system estimation approach, the residual estimation method, self-identification in surveys, and expert or Delphi surveys. Although these methods possess important drawbacks we highlight in the literature, these methods are well-known and considered standard. We found important innovations regarding the multiplier, in Drbohlav and Lachmanová (2023), which document the results of implementing the multiplier in practice; and also, innovative work in the residual method, with the use of machine learning, larger government databases on social programs, and the evaluation of robustness of the method to core methodological assumptions (van Hook et al., 2021).

Moreover, among traditional approaches, we included the use of specific events, such as large-scale regularization and formal status adjustment programs (Sabater & Domingo, 2012), and life course events. Changes in legislation have offered the opportunity to understand the number of individuals lacking legal status in the past (Kraler, 2019). In turn, the life course events, in which administrative or register data sources on births, deaths, or hospitalizations, can offer important clues as to the sizes of populations as long as these can be extrapolated to the larger population. In particular, we highlight the potential of data on mortality (Surkyn et al., 2023), an approach which holds promise to be implemented in various countries relying on similar data.

Another class of traditional approaches we cover follows statistical modelling practices. For example, labour demand and the flow-stock modelling. In the labour demand modelling (Hess, 2006), only irregular migrant workers are estimated on the basis of reported economic output based on administrative data. In the flow-stock model (Fazel-Zarandi, Feinstein & Kaplan,

2018; Rodilitz & Kaplan, 2021), in turn, used in the US, and which have been criticized for providing implausible estimates that go orders of magnitude beyond existing estimations (Capps et al., 2018), information on cumulative inflows (visa overstayers, irregular border crossings/apprehensions) minus cumulative outflows (deportations, voluntary emigration, mortality, status changes) is used to derive an estimate.

Finally, among traditional approaches we also included the use of official, administrative, and commercial databases that allow for the estimation of irregular flows or stocks. Although this could be considered partly innovative, as new data sources have become available, irregular border crossings (Savatic et al., 2021; FRONTEX, 2022), data bases on asylum claims and refugee status (Ghui & Blangiardo, 2019), migrant deaths and apprehensions, and database systems enabling identification of visa overstayers. In the US, for example, overstayer events are estimated by considering all arrivals through air, sea, and land, matched to records of exits (Department of Homeland Security, 2022; Warren, 2017).

Advances and challenges

Among the innovative approaches, we document the development of important approaches. For example, in terms of databases, institutional registers on college enrolment (Hsin & Reed, 2020), as well as driver license register data (Lueders & Mumper, 2022), and consular registers of migrant communities (Bhandari et al., 2021). On their own, these databases cannot, on their own, be used to estimate the total number of irregular migrants, but when put in combination with other data sources they have the potential to provide important clues about irregular migrant population size.

On the more methodological side, the use of statistical imputation in large databases, often in connection to the residual method (Gálvez-Iniesta, 2020), constitutes an important innovation worth mentioning (Borjas & Cassidy, 2019; Ro & van Hook, 2022). Also, the use of innovative data sources on consumption and online search behaviour can be highlighted among the most important innovations (Nixon, 2022; Böhme, Gröger, & Stöhr, 2020).

New databases that also inform irregular migrant flows, for example, have been created as a result of important citizen-driven projects (UNITED for Intercultural Action. (n.d.), or the creation of more encompassing surveillance programs. An example of the first is the Missing Migrants Project collecting data on the population of migrants dead and presumed missing while en route (García-Borja & Black, 2022). Knutson (2021) and Molnar (2019) discuss the uses of socio-technical systems based on artificial intelligence that enable facial recognition in the enforcement of migration.

Significant progress has been achieved in refining estimation methods to address the shortcomings of traditional techniques like the residual method. These improvements have been driven by the adoption of novel methodologies and the growing availability of diverse data sources (Vespe et al., 2017). However, estimation approaches remain largely fragmented, often shaped by the specific type of data source employed.
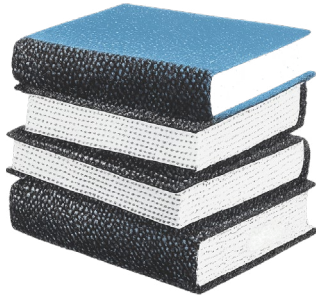
References

handari, R., Feigenberg, B., Lubotsky, D., & Medina-Cortina, E. (2021). Projecting Trends in Undocumented and Legal Immigrant Populations in the United States (No. NB21-16). NBER. https://www.nber.org/programs-projects/projects-and-centers/retirement-and-disability-research-center/center-papers/nb21-16

Borjas, G. J., & Cassidy, H. (2019). The wage penalty to undocumented immigration. Labour Economics, 61, 101757. https://doi.org/10.1016/j.labeco.2019.101757

Böhme, M. H., Gröger, A., & Stöhr, T. (2020). Searching for a better life: Predicting international migration with online search keywords. Journal of Development Economics, 142, 102347. https://doi.org/10.1016/j.jdeveco.2019.04.002

Capps, R., Gelatt, J., Van Hook, J., & Fix, M. (2018). Commentary on "The number of undocumented immigrants in the United States: Estimates based on demographic modeling with data from 1990-2016". PloS one, 13(9), e0204199. https://doi.org/10.1371/journal.pone.0204199

Department of Homeland Security. (2022). FY 2020 Entry/Exit Overstay Report (Entry/Exit Overstay Report). Department of Homeland Security. https://www.dhs.gov/sites/default/files/2021-12/CBP%20-%20FY%202020%20Entry%20Exit%20Overstay%20Report_0.pdf

FRONTEX. (2022, October 7). Risk Analysis for 2022/2023. Publications. https://frontex.europa.eu/publications/risk-analysis-for-2022-2023-RfJIVQ

Gálvez Iniesta, I. (2020). The size, socio-economic composition and fiscal implications of the irregular immigration in Spain [WorkingPaper]. https://e-archivo.uc3m.es/handle/10016/30643

Ghio, D., & Blangiardo, G. C. (2019). Exploring the link between irregular migration and asylum: the case of Italy. Genus, 75(1), 14. https://doi.org/10.1186/s41118-019-0060-3

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. J. (2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ, 336(7650), 924–926. https://doi.org/10.1136/bmj.39489.470347.AD

Hsin, A., & Reed, H. E. (2020). The Academic Performance of Undocumented Students in Higher Education in the United States. International Migration Review, 54(1), 289–315. https://doi.org/10.1177/0197918318825478

Jandl, M. (2011). Methods, Approaches and Data Sources for Estimating Stocks of Irregular Migrants: Estimating stocks of irregular immigrants. International Migration, 49(5), 53–77. https://doi.org/10.1111/j.1468-2435.2011.00701.x

Kraler, A. (2019). Regularization of Irregular Migrants and Social Policies: Comparative Perspectives. Journal of Immigrant & Refugee Studies, 17(1), 94–113. https://doi.org/10.1080/15562948.2018.1522561

Kraler, A., & Ahrens, J. (2023). Conceptualising migrant irregularity for measurement purposes. In MIrreM Working Paper No. 2 (Version 2). Krems: University for Continuing Education Krems (Danube University Krems). https://doi.org/10.5281/zenodo.7788961

Lueders, H., & Mumper, M. (2022). Driving while unauthorized: Auto insurance remains unchanged when providing driver licenses to unauthorized immigrants in California. Journal of Risk and Insurance, 89(3), 669–696. https://doi.org/10.1111/jori.12382

Nixon, S. (2022). Of Rice and Men: Rice Consumption-Based Estimates of Undocumented Persons in Malaysia. International Migration Review, 019791832211264. https://doi.org/10.1177/01979183221126466

Pinkerton, C., McLaughlan, G., & Salt, J. (2004). Sizing the illegally resident population in the UK (No. 58/04). Migration Research Unit University College London. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=949bc2f103e29ec74a63070eaf24fd4f9fccdb4b

Sabater, A., & Domingo, A. (2012). A New Immigration Regularization Policy: The Settlement Program in Spain. International Migration Review, 46(1), 191–220. https://doi.org/10.1111/j.1747-7379.2012.00884.x

Savatic, F., Thiollet, H., Jualin, T., Mesnard, A., & Senne, J.-N. (2021). Borders Start with Numbers. Measuring Migration in Times of Crisis (MigrAtion Governance and AsYlum Crise-MAGYC D 8.2). https://www.magyc.uliege.be/upload/docs/application/pdf/2021-12/d8.2v1october2021.pdf

Surkyn, J., Ahmad Yar, A. W., Ceylan, D., Vanthomme, K., & Bircan, T. (2022). New publication 'An analysis of mortality rates for estimating undocumented migrants in Belgium' — Hummingbird (HumMingBird Project 870661 – H2020). Retrieved 28 September 2023, from https://hummingbird-h2020.eu/news/news-items/D7.2

Van Hook, J., Morse, A., Capps, R., & Gelatt, J. (2021). Uncertainty About the Size of the Unauthorized Foreign-Born Population in the United States. Demography, 58(6), 2315–2336. https://doi.org/10.1215/00703370-9491801

Vespe, M., Natale, F., & Pappalardo, L. (2017, September 7). Datasets on Irregular Migration and Irregular Migrants in the EU. JRC Publications Repository. https://publications.jrc.ec.europa.eu/repository/handle/JRC107749

Warren, R. (2017). DHS Overestimates Visa Overstays for 2016; Overstay Population Growth near Zero during the Year. Journal on Migration and Human Security, 5(4), 768–779. https://doi.org/10.1177/233150241700500403

Fazel-Zarandi, M. M., Feinstein, J. S., & Kaplan, E. H. (2018). The number of undocumented immigrants in the United States: Estimates based on demographic modeling with data from 1990 to 2016. PLOS ONE, 13(9), e0201193. https://doi.org/10.1371/journal.pone.0201193

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. International Journal of Surgery, 88, 105906. https://doi.org/10.1016/j.ijsu.2021.105906

Ro, A., & Van Hook, J. (2022). Comparing the Effectiveness of Assignment Strategies for Estimating Likely Undocumented Status in Secondary Data Sources for Latino and Asian Immigrants. Population Research and Policy Review, 41(2), 449–464. https://doi.org/10.1007/s11113-021-09658-3

Rodilitz, S., & Kaplan, E. H. (2021). Snapshot Models of Undocumented Immigration. Risk Analysis, 41(9), 1643–1661. https://doi.org/10.1111/risa.13658

Rodriguez Sanchez, A., & Tjaden, J. (2023). Estimating Irregular Migration – A Review of Traditional and Innovative Methods. In MIrreM Working Paper No.4 (version 2). Krems: University for Continuing Education Krems (Danube University Krems). https://doi.org/10.5281/zenodo.8380854

Hess, S. (2006). The demand for seasonal farm labor from Central-and Eastern European countries in German agriculture (CIGR Journal). Institute of Agricultural Economics. https://cigrjournal.org/index.php/Ejounral/article/download/647/641/0

UNITED for Intercultural Action. (n.d.). About the "UNITED List of Refugee Deaths". UNITED Against Refugee Deaths. Retrieved June 9, 2025, from UNITED for Intercultural Action's website