Chapter 4

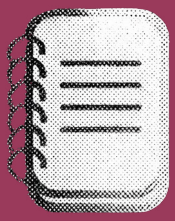
What are good quality data on a phenomenon that is hard to measure?

Denis Kierans and Lalaine Siruno

To cite: Kierans, D., and Siruno, L. (2025). What are good quality data on a phenomenon that is hard to measure? 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-vq79-04

Keywords: Data quality, Irregular migration, measurement uncertainty, indicators, estimates, critical data assessment, FAIR data principles

What are good quality data on a phenomenon that is hard to measure?



Key points

- Irregular migration is difficult to measure, and the data that exist are often limited, inconsistent, or outdated. This chapter introduces a practical framework to help users assess the quality and credibility of such data, rather than taking estimates at face value.
- It distinguishes between key data types—stocks vs flows, estimates vs indicators—and highlights how conceptual ambiguity, observational gaps, and poor documentation can undermine how irregular migration data are interpreted and used.
- When applied to over 250 estimates across 14 countries, the framework reveals significant variation in quality. While some countries produce relatively robust and transparent figures, many rely on outdated, methodically weak or poorly documented estimates. Still, pockets of good practice exist across North America and Europe, which can be built on.
- The chapter argues that responsible use of irregular migration data depends not only on improving data systems, but also on the ability of users to critically assess what data mean, how they were produced, and whether they are fit for purpose.

Introduction: Measuring the unmeasurable?

Policymakers often point to data to justify their decisions, particularly in contested policy spaces, such as immigration (Boswell, 2009; Kraler & Reichel, 2022). Irregular migration, while a point in case, poses distinct challenges to this practice. Although irregular migrants are the subject of intense political and media scrutiny in many countries, information about them is notoriously scant and unreliable (Vollmer, 2011).

This chapter examines what constitutes "good quality" data in this complex landscape. It introduces the MIrreM framework for assessing irregular migration estimates and indicators and discusses how uncertainty (Box 4.1) and other criteria shape data quality. Rather than assuming these figures are fit for purpose, we suggest users interrogate them, asking whether the data are credible, transparent, and suited to question at hand. We conclude with a checklist based on the five MIrreM criteria to support this process.

Box 4.1: Uncertainty in irregular migration data

Denis Kierans

There are three main sources of uncertainty in irregular migration data:

- **Conceptual uncertainty** arises from disagreements over who should count as an irregular migrant. Legal definitions vary across jurisdictions, media usage is inconsistent, and there is no universal agreement on terminology in academic circles, either. The MIrreM project distinguishes three categories: (1) migrants in an irregular situation, (2) migrants with a provisional status or a claim to one, and (3) EU citizens without residence rights. This taxonomy helps make explicit the assumptions behind different datasets and more detail on this can be found in Chapter 2.
- **Ethical uncertainty** emerges from the tension between producing data, such as estimates and indicators, on irregular migration and protecting individual rights. Certain types of data collection run the risk of compromising privacy and creating risks for vulnerable people. Furthermore, due to the political sensitivity of this area, these numbers – irrespective of their quality – can contribute to negative public sentiment and actions towards migrants in general and be used to justify reactionary and punitive policymaking. More information about ethics and data protection can be found in Chapter 3.
- **Practical uncertainty** is related to observability. Many irregular migrants avoid contact with authorities and may not appear in traditional sources of data on migration, such as administrative systems or surveys. Indicators based on events (e.g., border apprehensions, deaths occurring during the migration process in the Mediterranean, the English Channel or along the US – Mexico border) only capture the visible portion of the phenomenon. Estimating what is not directly observed requires the use of methods underpinned by assumptions, some more plausible than others.

This chapter focuses primarily on the conceptual and practical aspects of uncertainty. However, it is important to consider all elements when assessing and using irregular migration data.

What types of irregular migration data are there?

Data on irregular migration can be categorised broadly in two important ways. First, as either stock or flow data:

- **Stocks:** Quantitative estimates of the number of irregular migrants present in a given country or area at a particular point in time. Stock data help to answer questions such as "how many irregular migrants are living in Country X today?"
- **Flows:** Data that describe movements into, within, or out of irregular status over a specified period of time. Flow data help to answer questions like "how many irregular migrants were apprehended at the border in the past month?"

Second, irregular migration data can be distinguished as an estimate or an indicator:

- **Estimates:** quantifying a population group that cannot be directly observed or counted. Estimates help to answer questions such as "how many irregular migrants we living in the United States in 2020?"
- **Indicators:** Counts of observed events, such as border apprehensions or asylum decisions. Indicators help to answer questions like "how many migrants were intercepted by border control last quarter?" Indicators can track trends and inform planning but do not capture full populations. They often feed into estimates.

Box 4.2: Irregular migration to the UK: A Home Office statistical overview

Jon Simmons and Lucy Swinnerton

The Home Office publishes a wide range of statistics that reflect different aspects of irregular migration to the UK, drawn from the department's operational systems.

Daily counts of small boat arrivals in the English Channel provide a near real-time operational snapshot of people detected arriving in the UK this way. These figures, published on GOV.UK,[1] with a public dataset of daily arrivals, are manually collated by Border Force officers and offer a timely but provisional view of one of the most visible forms of irregular entry.

The Home Office's comprehensive publication of statistics on the operation of the UK immigration system are the quarterly 'Immigration System Statistics',[2] which bring together administrative data from operational and case working systems to provide a wide range of information on irregular migration. This includes demographic details of those arriving to the UK by small boat (age, sex and nationality), as well as information on asylum claims made by small boat arrivals, decisions on those claims, and referrals to the National Referral Mechanism (NRM) for potential victims of modern slavery.

Publicly available data on irregular arrivals are not limited to those entering the UK on small boats. Data on inadequately documented air arrivals (individuals arriving without valid or with fraudulent documents) and detections in-country and at ports are also recorded on Home Office systems and included in published statistics. The department also publishes statistics on immigration enforcement activity, including return of small boat arrivals with no right to remain in the UK. This information helps to contextualise the experiences and outcomes of individuals arriving to the UK via irregular routes and compliments other forms of research.

Data on irregular arrivals are derived from live operational systems and so numbers may change as more information is added. The statistics sometimes present a partial picture of most recent events due to time required for data validation or additional safeguarding assessments, or the time necessary to come to a decision on often complex individual cases. Although the statistics are drawn from live systems and may be subject to revision, they do nonetheless provide a consistent and broad basis for understanding patterns over time.

The Home Office statistics are accredited Official Statistics published according to the Code of Practice for Statistics overseen by the UK Statistics Authority. The published 'User Guide to Immigration System Statistics'[3] provides documentation to support understanding of the data, including definitions, data sources, and known limitations.

While the statistics published by the Home Office provide a wide range of valuable insights, they cannot capture the full extent of irregular migration. Some individuals evade detection entirely while others may see their status change. Others of course regularise their status, for example through a successful application for refugee status. The Home Office has published some initial information in relation to those who arrive on a visa but subsequently claim asylum (including individuals who overstay their visa and claim after their visa status is no longer valid),[4] but this is novel and complex data and so unlike the regular quarterly outputs at present it is regarded as 'statistics in development'.

It is also possible that the same person may be detected at multiple times in different operational systems, which cannot always or simply be linked. These statistics therefore offer a partial view of irregular migration and users should interpret data with care, carefully considering the limitations and assumptions that underpin them.

1 See https://www.gov.uk/government/publications/migrants-detected-crossing-the-english-channel-in-small-boats.

2 See https://www.gov.uk/government/collections/immigration-statistics-quarterly-release.

3 See https://www.gov.uk/government/publications/user-guide-to-home-office-immigration-statistics--9/user-guide-to-immigration-statistics.

4 See https://www.gov.uk/government/publications/source-of-asylum-claims-in-2024/source-of-asylum-claims-in-2024.

How do you assess the quality of irregular migration data?

To assess the quality of these data, MIrreM used a structured set of criteria which are informed by the FAIR data principles: Findability, Accessibility, Interoperability, and Reusability (Wilkinson et al., 2016). The following framework outlines the five criteria against which we assessed the quality of the irregular migration estimates.

MIrreM criteria	High (3 points)	Medium (2 points)	Low (1 point)
Accessibility	All raw data used to construct the estimate is publicly available and electronically accessible with no permissions required.	At least some of the raw data used to construct the estimate is only available on request from relevant authorities. If some of the data is not available at all, then give 1 point.	At least some of the raw data used to construct the estimate is not available for most potential users.
Documentation	Full documentation about data and methods are available and accessible. The level of information allows for replication of the estimates.	Limited information on data, estimation methods, and quality are available and accessible. Insufficient details to replicate the estimates.	Information on data and estimation methods is neither available nor accessible.
Reliability	Analysis includes demonstrated reliability indicators, with limitations clearly specified (e.g. ranges, alternative calculations, characterisation as minimum or maximum estimate).	Some discussion of reliability, but no indicators in quantitative terms.	Missing a discussion of reliability.
Methodology	Methodology is adequate and comprehensive including, but not limited to, rigorously implemented multiplier or residual studies.	Methodology is adequate, even if not comprehensive, including but not limited to: (1) Simple multiplier calculations; (2) Simple residual estimates; (3) Adjustment of older estimates with partly insufficient data; (4) Aggregate estimates for different groups, partly relying on plausibility calculations.	Inadequate method and application of the method; resulting estimate lacks foundation
Data	The analysis relies on an adequate dataset not likely to have a considerable bias, including no bias for any group estimates. There are no strong assumptions regarding the data.	The analysis relies on a biased dataset. There are plausible adjustments and assumptions. This includes cases in which the dataset does not provide the information necessary or it is necessary to make strong assumptions.	The analysis relies on a biased dataset, without proper adjustments. The assumptions regarding data are not plausible.

Table 4.1: Criteria for the MIrreM quality evaluation of estimates (Source: Kierans and Vargas-Silva, 2024)

Alongside scoring each estimate on five individual criteria, we also developed an aggregate quality score. It combines the different dimensions into a single number to help users quickly understand the overall robustness of an estimate.

Although we hope it is helpful, we recognise it is not objective. It reflects deliberate decisions about what matters most. Some criteria – such as the soundness of methods or how uncertainty is handled – are more central to quality than, for example, whether data are fully open-access. We therefore weighted the criteria: methodology and reliability counted more than data access or documentation. We also introduced thresholds to stop poor methods or unreliable outputs from boosting the total score, reflecting feedback that no amount of documentation or access can compensate for fundamental flaws.

A slightly different quality framework was used for the assessment of indicators. In particular, the methodology and data criteria were excluded. This is because, unlike estimates, indicator data are observed or registered event collected and reported by national statistical institutes or enforcement authorities like FRONTEX. Instead, we grouped validity and reliability to assess whether an indicator is in fact measuring the type of irregular flow it is supposed to measure.

MIrreM Criteria	High (3 points)	Medium (2 points)	Low (1 point)
Accessibility	Data is publicly available and electronically accessible with no permissions required	Data is available on request from relevant authorities	Data is available, but access and use are exclusive to authorities
Documentation	Sufficient and transparent information on data and methods are available and accessible; a comprehensive quality report is also available	Limited information on data, methods, and quality are available and accessible	Information on data, methods, and quality are neither available nor accessible
Validity and reliability	Data is representative of the phenomenon it is supposed to measure and adequately reflects the type of irregular migration being measured; data is relatively complete (not highly selective) and does not indicate internal contradictions	Data is selective and points to some internal contradictions	Data is neither valid nor reliable

Table 4.2: Criteria for the MIrreM quality evaluation of indicators (Source: Siruno et al., 2024)

We understand others may have different views on how to assess these data, which we welcome. Our aim is to encourage critical reflection and debate around irregular migration data. More information about the MIrreM data collection process and tools are available in Vargas-Silva et al. (2025).

What do high and low quality data look like?

We found wide variation in the quantity and quality of national estimates (n=260). For instance, the United States produces multiple annual estimates of high reliability. The UK and Germany, which have the largest estimated irregular migration populations in Europe, rely on relatively outdated figures. Austria excels in access to the underlying data but does worse against other criteria. Canada and Portugal have few estimates, all scored poorly. No estimates were located for Bosnia and Herzegovina, Morocco, Serbia or Türkiye.

Country	Aggregate	Access	Documentation	Reliability	Method	Data
Germany	**11.1**	2.0	2.9	3.0	3.0	2.5
France	**10.7**	2.8	2.8	2.8	2.8	1.9
United States	**10.3**	1.8	3.0	2.6	2.7	3.0
United Kingdom	**10.1**	1.9	3.0	2.7	2.7	2.0
Italy	**10.1**	1.7	2.4	2.4	3.0	3.0
Spain	**10.0**	2.6	2.0	2.9	2.6	2.5
Belgium	**9.5**	2.2	2.7	2.0	2.7	2.7
All countries	**9.1**	**2.0**	**2.4**	**2.4**	**2.5**	**2.2**
Greece	**9.1**	2.1	2.1	2.7	2.0	2.9
Austria	**8.3**	3.0	2.0	2.3	2.1	1.3
Netherlands	**7.9**	1.2	2.2	2.2	2.2	1.3
Ireland	**7.8**	1.6	2.2	2.4	1.9	1.6
Poland	**7.6**	2.4	1.9	2.1	2.4	1.0
Portugal	**7.2**	1.8	1.8	1.8	1.8	1.9
Finland	**5.9**	1.9	1.6	1.9	1.6	1.4
Canada	**4.0**	1.0	1.0	1.0	1.0	1.0

Table 4.3: Average MIrreM quality point scores for irregular migration stock estimates, criteria and aggregate, by country. Note: Aggregate scores range from 4 (low) points to 12 (high) points; the other criteria range from 1 to 3 points. (Source: Kierans et al., 2024)

The MIrreM project analysed six of the most frequently discussed indicators related to irregular migration flows, all of which are published by Eurostat.

Type of flow	Indicator – inflow	Indicator – outflow
Geographic	(1) TCNs refused entry at the external borders	(3) TCNs ordered to leave
	(2) TCNs found to be illegally present inland because of illegal entry	(4) TCNs returned to a third country following an order to leave
Status-related	(5) Negative first instance asylum decisions	
	(6) Decisions withdrawing status granted at first instance decision	

Table 4.4: Selected irregular flow indicators

We found these indicators to be of relatively good quality. However, it remains difficult to establish whether they are fully accurate and reliable. For example, double counting and missing information can affect measurement precision, especially when disaggregated by age and sex. Furthermore, because the data are typically produced by bureaucracies with limited oversight, there are few opportunities to validate the numbers by cross-checking them with other information.

Criteria	Quality	Explanation
Accessibility	High	Data are publicly available and electronically accessible from Eurostat with no permissions required. General alignment with the FAIR Data Principles.
Documentation	High	Sufficient and transparent information on data and methods available and accessible; comprehensive quality report for almost all countries covered.
Validity and reliability	Medium	The data is generally representative of the specific flows being measured, but there is some variation re: external validity and reliability (percentage of missing values and inter-item correlation tests).

Table 4.5: Summary of quality assessment of selected Eurostat indicators (Source: Siruno et al., 2024)

We hope that these examples highlight the value of systems of critical appraisal by helping users understand and distinguish between different sources of irregular migration data.

Conclusion

Measuring and estimating irregular migration will always be difficult. But better data and their use is possible. A step in the right direction is to incorporate critical appraisals of the data as a matter of course, especially for those shaping policy and the public debate. To this end, the MIrreM quality criteria can be used or adapted as a preliminary, rapid-fire assessment tool for irregular migration data.

Not all estimates or indicators will meet all criteria fully. However, if you are unable to answer the relevant questions on the checklist, we recommend learning more about the data before you use it in your work or draw conclusions from the data.

- **Accessibility**: Are the raw data publicly available?
- **Documentation**: Are the methods and sources clearly explained?
- **Reliability**: Are uncertainty and limitations explicitly addressed?
- **Methodology**: Is the *estimate's* approach comprehensive, appropriate and replicable?
- **Data quality**: Are the data underpinning the *estimate* relatively unbiased and fit for purpose?
- **Validity and reliability**: Are you clear on what the *indicator* measures and does it do so accurately and consistently?

Figure 4.1: Checklist for rapid assessment irregular migration estimates and indicators

To conclude, we emphasise that assessing the quality of irregular migration data is not merely an academic or technical matter. Given irregular migration data's uneven quality and limited availability – combined with the political and public sensitivity of the issue – it is easy to misinterpret, with potentially serious consequences. Avoiding this requires investment not only in data systems, but also in the capacities of individuals and institutions to interpret and use irregular migration data responsibly.

References

Boswell, C. (2009). The Political Uses of Expert Knowledge: Immigration Policy and Social Research. Cambridge University Press.

Kierans, D. & Vargas-Silva, C. (2024). The Irregular Migrant Population of Europe. MIrreM Working Paper No. 11. Krems: University for Continuing Education Krems (Danube University Krems). https://doi.org/10.5281/zenodo.13857073

Kraler, A. & Reichel, D. (2022). Migration Statistics. In: Scholten, P. (eds) Introduction to Migration Studies. IMISCOE Research Series. Springer, Cham. https://doi.org/10.1007/978-3-030-92377-8_27

Siruno, L., Leerkes, A., Hendow, M. & Brunovksá, E. (2024). Working Paper on Irregular Migration Flows. MIrreM Working Paper No. 9. Krems: University for Continuing Education Krems (Danube University Krems). https://doi.org/10.5281/zenodo.10702228

Vargas-Silva, C., Leerkes A., Kierans, D., Siruno, L. & Kraler, A. (2025). Tools for collecting information on irregular migration estimates and indicators [version 1; peer review: 2 approved]. Open Res Europe 2025, 5:176. https://doi.org/10.12688/openreseurope.20695.1

Vollmer, B. A. (2011). Policy Discourses on Irregular Migration in the EU-'Number Games' and 'Political Games'. European Journal of Migration and Law, 13(3), 317-339. https://doi.org/10.1163/157181611X587874

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3. https://doi.org/10.1038/sdata.2016.18