Article

International migration research, progress update: May 2023

An update on international migration methods and research.

Contact:
Kerry Miller and Dominic Webber
pop.info@ons.gov.uk
+44 1329 444661

Release date:
25 May 2023

Next release:
To be announced

Table of contents

1 . Overview of migration statistics transformation

This article provides an update on the research we have been undertaking through our migration statistics transformation programme to develop admin-based migration estimates (ABMEs), deliver incremental quality improvements, and expand the range and granularity of our statistics.

We have introduced a number of improvements to our bulletin Long-term international migration, provisional: year ending December 2022 published in May 2023. Estimates are now available up to the year ending (YE) December 2022.

This article includes further information on:

- improvements to methods for producing non-EU migration estimates (Section 2: Improvements to methods for non-EU migration using Home Office data)

- implementation of measuring asylum applicants, resettlement scheme arrivals and irregular migration (Section 3: Asylum applicants, resettlement scheme arrivals and irregular migration)

- improvements to methods for producing EU migration estimates (Section 4: Improvements to EU-based migration estimates)

- investigation into emigration by reason for non-EU nationals (Section 5: Emigration by reason for non-EU nationals)

- research into EU visa holders and EUSS (Section 6: EU Visa holders and EUSS in Exit Checks)

- progress updates on the potential of machine learning techniques to improve the accuracy of provisional estimates (Section 7: Using machine learning (ML) to produce more accurate provisional immigration estimates)

- proposals for uncertainty measures (Section 8: Uncertainty)

We continue to work on a longer-term programme of research set out in our previous International migration statistical design report, with the aim of delivering further improvements alongside the next release of migration statistics, which is scheduled to be published in November 2023. We are also supporting the Dynamic Population Model (DPM) research to develop timely and coherent migration and population estimates at a national and local authority level.

2 . Improvements to methods for non-EU migration using Home Office data

Home Office Borders and Immigration data (including visa information and travel data) are currently our best source for estimating non-EU immigration and emigration. The latest methodology is based on research published in our previous reports (Exploring international migration concepts and definitions with Home Office administrative data from April 2021 and February 2020). Our understanding of this complex source of information continues to expand which is enabling us to refine and develop our methods.

Immigration estimates

We have improved the estimates of long-term immigration for our most recent Long-term international migration, provisional: year ending December 2022 bulletin. This method classifies individuals as long-term immigrants if the difference between first arrival and last departure in a visa period is more than 12 months. We have refined the method to exclude any journeys on long-term visit visas to ensure these visits are no longer included in our estimates. Those on long-term visit visas are only eligible to stay in the UK for up to 6 months per visit, but as the visas can be valid up to 2, 5 or 10 years, the method would incorrectly identify them as long-term migrants.

For those individuals whose first arrival occurred within the 12 months before the end of the reference period (currently December 2022), we do not yet have enough information to see a long-term stay of 12 months or more. We have continued to develop our adjustment for the overestimation of long-term immigration among our most recent arrivals. To improve this we have broken down each visa type to account for different migrants having different behaviours. This adjustment has reduced the long-term immigration estimate by around 143,000 for the year ending (YE) December 2022.

We have also implemented an adjustment for those arriving on the Ukraine schemes and British National (Overseas) arrivals where we do not have enough information to suggest how many will go on to have a long-term stay of 12 months or more. Analysis using Home Office Borders and Immigration data has shown that there are a number of individuals from these groups who have been out of the UK with no subsequent rearrival. We have used these data to implement an adjustment for these groups and removed these arrivals from our long-term immigration estimates where they have been outside the UK for 8 weeks or more without a subsequent arrival. This adjustment has removed 39,000 from Ukraine Schemes arrivals and 3,000 for British Nationals (Overseas) (BN(O)) for the YE December 2022.

Emigration estimates

We have continued to adjust for the overestimation of emigrants within the most recent years. This is because we do not currently have enough data to see when an individual, who is assumed to have made a last departure from the UK, may in fact return within the next 12 months. This assumption may incorrectly classify approximately 2% of identified emigrants. For the YE December 2022, this has removed 3,000 individuals from the emigration estimate.

The "first arrival, last departure" emigration method is dependent on having a completed visa period within which to identify a last departure. Anyone who has an open visa period at the end of the reference period is not counted as an emigrant, even if they left the UK over a year ago. This assumption means that emigration estimates will be an underestimate.

For years other than the most recent 12 months (currently 2018 to 2021), we have added an adjustment to include individuals in the emigration estimate who have been absent for 52 weeks or more, and whose exits have not been captured because of the visa period still being valid. For the final year (currently 2022) not enough time has passed to see if an individual has left for 52 weeks or more with no re-arrival. Therefore, to adjust for the final year, we have identified individuals who have left the UK and not subsequently returned and who have up to three months left on their visa after the data extract (end of December 2022).

This identifies individuals who have left the UK and are near the end of their visa period and therefore are unlikely to return. For example, someone who left in September 2022 and has not subsequently returned with a visa period that expires in January 2023. For the year ending December 2022, this adds on 73,000 to our emigration estimate. This adjustment will reduce the future scale of revisions to our estimates for the year ending December 2022.

Next steps

There are a number of individuals within the Home Office Borders and Immigration data that have travel information that is not matched to a visa. Those with unmatched travel have been removed from the estimates while we develop our understanding of this group further, and because of the assumption that most of this group should not be counted as long-term migrants.

We do not currently have a suitable methodology to include those who hold Indefinite Leave to Remain (ILR) in our estimates. Our next steps are to investigate how we can integrate this group into our estimates.

Table 1: Impact adjustments have on YE December 2022 estimates

Adjustment	Impact on estimate YE December 2022 (+/-)
Immigration early leavers	- 143,000
Immigration early leavers British Nationals (Overseas) (BN(O)) and Ukraine Schemes	- 42,000
Emigration rearrivals	- 3,000
Emigration early exits	+73,000
Immigration arriving on an Long-term international migration (LTIM) visa and applying for asylum	-13,000

Source: Office for National Statistics, Home Office

3 . Asylum applicants, resettlement scheme arrivals and irregular migration

We have used data published by the Home Office to add the total number of asylum applicants and resettlement scheme arrivals to the international migration estimate. This does not include arrivals from the Ukraine Schemes and British Nationals (Overseas) (BN(O)) as these groups are already included in the non-EU estimates. For assumptions related to this work, refer to our Methods to produce provisional long-term international migration estimates methodology.

When including asylum applicants in the international migration estimate, we made an adjustment to address potential double counting. As we used published data on the number of asylum applications, we were not able to assess how many of these applicants also appear in our non-EU estimates produced using the Home Office Borders and Immigration data. As a provisional method we have used analysis provided to us from the Home Office Migrant Journey data, which show the proportion of asylum applicants who held another form of leave within seven days of lodging their application (the proportion for 2022 is around 14%). We have removed this proportion from the number of asylum applications.

We have also investigated records for those with a visa expiring in 2022, who applied for asylum in 2022, to understand whether there was double counting of these cases. An adjustment of less than 1,000 was made for this in the data for YE December 2022. However, we did not have data available for previous years, so this will be explored in the future.

The resettlement scheme arrivals include individuals resettled under the Afghanistan Citizens Resettlement Scheme and the Afghan Relocations and Assistance Policy, as well as other pre-existing resettlement schemes.

We added the total number of asylum returns to the international emigration estimate using nationality data to split the numbers into EU and non-EU.

In the future, rather than using the published Home Office counts, we are exploring the possibility of receiving the data on asylum applications together with the Home Office Borders and Immigration data. This will enable us to more accurately remove individuals that are double counted in our estimates; for example, people with more than one application. We are exploring, with the Home Office, the feasibility of using Home Office Borders and Immigration data, in addition to the published data, to improve these estimates.

Part of this work involves looking at the coverage of irregular migration in our counts. There are two groups of irregular migrants: those who arrive legally and become irregular by overstaying their visa, and those who arrive irregularly. Those in the first group are counted in our estimates according to their original reason for migration, for example, if they have arrived on a study or work visa. If they overstayed their visa, they will be assumed to leave at the time of their visa expiry and recorded as a new entry when they apply for asylum. Of the second group (for example, small boat arrivals), most claim asylum. We believe that the statistics now capture the majority of irregular migrants, given that a considerable number of asylum applicants come from these two groups of irregular migrants. However, we will miss those who enter irregularly and never claim asylum, and we may undercount those who enter on non long-term international migration (LTIM) visas (e.g., visit visas) and overstay. We will continue to work with the Home Office to determine the feasibility of including these irregular migrants in our estimates.

4 . Improvements to EU-based migration estimates

We consider the Registration and Population Interactions Database (RAPID) to be the best and most complete data source, at this moment in time, for estimating EU migration. We have continued our research into improving the coverage adjustments that are required to complement the EU estimates obtained from RAPID.

Student adjustment

An improvement to the adjustment for estimating student immigration was developed and discussed in our International migration research, progress update: November 2022. This methodology has now been extended to cover the adjustment to the emigration estimates.

Removing C3/C4 arrivals from RAPID estimates

In RAPID, we previously included 4 arrival categories. C1/C2 arrivals most closely align with the UN definition of a long-term migrant. However, when we initially developed admin-based migration estimates we expanded on this definition of long-term activity in order to reflect the complexity of people's lives. This created two further categories, which are C3/C4 arrivals; these groups only make up a small proportion of arrivals. However, as C3 /C4 arrivals do not align with the UN definition of a long-term migrant, we have since removed these from our estimates. This is supported by analysis from Census 2021, which suggested that including these arrivals expanded too far on the UN definition of long-term migration.

Forecasting methods

RAPID data are made available to the Office for National Statistics (ONS) on an annual basis in Quarter 3 (Jul to Sep) for the previous tax year. ONS estimates extend beyond the end date of this dataset. Currently we publish bi-annual international migration estimates, which requires forecasting RAPID for 3 or 9 months. Our forecasting approach generates figures beyond the timeframe of RAPID data, using signals and trends in a higher frequency time series. We use the International Passenger Survey (IPS) for this as it helps to incorporate seasonality and trends of migration flows.

The estimates produced from this forecasting are point estimates. We are currently undertaking research to understand what other timeseries methods we could use that would provide uncertainty intervals; an essential component for all point estimates. We are exploring two methods, the first being exponential smoothing. For this a value is forecasted by calculating a weighted average of past observations, with the weights exponentially decreasing as the observation gets older. The second method is Autoregressive integrated moving average (ARIMA) modelling. This approach aims to describe the autocorrelations found within the data. The term "autoregressive" means that the value is forecasted using a linear combination of its past values. A moving average model uses past forecast errors in a regression-like model. This work will be completed in the coming months and implemented in the YE Jun 2023. We expect estimates to be published in November 2023.

5 . Emigration by reason for non-EU nationals

Through the use of Home Office Borders and Immigration data we have made good progress in identifying, within our migration estimates, the reasons why non-EU nationals come to the UK. The "first arrival, last departure" (FALD) method for estimating immigration and emigration (which we have used to form our estimates since May 2022) assigns a person's reason for migration based on the first visa they were granted to enter the UK.

The "reason for migration" that is provided for non-EU migrants is based on their visa type, as opposed to the self-declared reason, as reported by the IPS. When administrative and survey data records are linked at the individual level, the self-declared reason for migration does not always match the reason they have been granted permission to enter, as shown by [research by the Swiss Federal Statistical Office (PDF, 370KB)](). Users should therefore exercise caution when comparing breakdowns between non-EU, EU and British nationals.

In our previous [Long-term international migration, provisional: year ending June 2022 bulletin](), we provided estimates that included the "reason for immigration" by broad visa classification group (Study, Work and Other). This enabled users to see a more detailed picture of reasons for immigration of non-EU nationals entering the UK.

In our [Long-term international migration, provisional: year ending December 2022 bulletin](), we have provided more in-depth categories for the reason for immigration. In addition, we have provided estimates of emigration by visa type for non-EU nationals for the first time. While this does not provide information on their reason for emigration, it provides emigration estimates by reason for immigration. This means that we are able to track an individual migrant's journey from their first visa through to when they leave the UK. While this does not directly estimate the reason for emigration, it means that emigration estimates by "reason" are broadly comparable to the immigration estimates by reason.

Providing immigration and emigration estimates based on a person's initial visa to enter the UK allows more reliable comparisons to be made. However, this does not tell the complete story regarding how individuals behave during their time in the UK; particularly students. For more information see our [Population and migration estimates – exploring alternative definitions: May 2023 article]().

6 . EU Visa holders and EUSS in Exit Checks

The Registration and Population Interactions Database (RAPID) remains the best data source for estimating the migration of EU nationals at this time. However, since January 2021, EU nationals have been required to obtain a visa in order to enter the UK, or if they were already residing in the UK then they were allowed to apply for the EU settlement scheme (EUSS). The Office for National Statistics (ONS) has been provided with data for EU nationals requiring a visa as well as those on the EUSS. These data are linked to their travel data to begin research into the viability of using this data source to form estimates for EU nationals. From this we have made good progress in developing a methodology similar to what is used when we produce migration estimates for non-EU nationals.

While research is still in its early stages, we have made progress in identifying EU nationals who hold visas and have started to apply a similar logic to determine whether they should be considered long-term immigrants in the UK; this is similar to the methodology used to produce migration estimates of non-EU nationals. We will compare these estimates with other data sources such as RAPID. Further development will also include addressing the gaps in this new data source when compared with others. For example, addressing the common travel area and to what extent data are missing from this method of travel and how it affects migration estimates.

In addition, through these new experimental Home Office Borders and Immigration data, we can identify EU nationals on the EU settlement scheme who have immigrated long term. We will continue to work with the Home Office to understand this group further and develop this strand of work.

7 . Using machine learning (ML) to produce more accurate provisional immigration estimates

In our last [International migration research progress update article (November 2022)](#), we described the work that we are doing to develop supervised machine learning models. This is so that we can produce more accurate provisional predictions of whether, at record level, an international immigrant is likely to be classified under long-term international migration (LTIM) status for recently arrived immigrants. We have used three classification models: logistic regression, random forest, and [XGBoost](#). These models produce probabilities to facilitate binary true or false predictions for an individual's LTIM status. Our analysis uses Home Office Borders and Immigration data from 2015 to 2019 to remove the impact of changing behaviour during the coronavirus (COVID-19) pandemic. Also, it only includes potential long-term immigrants whose last departure had not occurred by the end of each annual reference period.

We are comparing against the current [first arrival, last departure](#) method with an adjustment for early leavers currently used to estimate provisional immigration where the last departure has not yet occurred. Our initial models were less likely to make false positive errors, where the model predicts LTIM but the truth is non-LTIM, than the current method. However, our initial models were more likely to make false negative errors, where the model predicts non-LTIM but the truth is LTIM, and overall resulted in less accurate estimates of total long-term immigration. Therefore, we have worked towards model improvements and optimisation so our models now outperform the current method at both record and aggregate level in our historical test data (Table 2).

Table 2: Average performance of the experimental ML methods and the current method for predicting LTIM status across 3 annual cohorts (2015-2019)

Method	Precision	Recall	F1 score	Mean absolute error (%)
Current method [Note 3]	0.8	0.99	0.88	15.16
Logistic regression	0.90	0.90	0.90	3.19
Random forest	0.90	0.82	0.86	7.79
XGBoost	0.90	0.83	0.87	6.13

Source: Office for National Statistics, Home Office

Notes

Precision [Note 1] and recall [Note 2] are calculated at the record level for each cohort and then averaged across cohorts (F1 score is an average of precision and recall that gives a single point of comparison). Mean absolute error is calculated as the absolute percentage difference between the sum of the actual and predicted LTIM status in each cohort, averaged across cohorts.

1. Precision measures the proportion of positive model predictions that are true positives (individuals are actual LTIMs, and model predicts that they are), e.g., 0.90 precision score means 90% of the immigrants the model predicts to stay long term turn out to do so.

2. Recall measures the proportion apart from the false negatives (individuals are actual LTIMs, but the model predicts that they are not), e.g., 0.90 recall score means the model correctly identifies 90% of all LTIMs (false negative is 0.10 in this case).

3. Note that this performance is not directly comparable with published estimates as we have conducted our analysis on a subset of potential long-term immigrants only, and excluded immigrants whose LTIM status can be confirmed in each annual cohort.

We have achieved this improvement through three main developments. Firstly, we engineered new input features and variables that describe LTIM and early-leaver rates in the previous cohort disaggregated by age and reason for migration. Secondly, we restricted the reference period of the training data to only include the immediate previous cohort. For example, training on the 2017 to 2018 cohort, to test the 2018 to 2019 cohort, to remove the influence of older and less relevant data. Thirdly, we oversampled LTIM observations in the training data so that the models are more heavily penalised for making incorrect predictions for LTIM observations during training.

We also implemented recursive feature elimination using SHAP (Shapley Additive explanation) values (PDF, 865KB), hyperparameter optimisation using cross-validated grid search, and a second-stage logistic regression model for calibration, but none of these resulted in meaningful improvements. We have also tried reframing LTIM prediction as a regression problem where the models predict a provisional length of stay instead of classifying LTIM status directly. Our initial findings suggest that regression models may be more suited to this task than classification models, we will report further on this in our next update.

Our next steps are to finalise model selection, assessing our chosen model's performance consistency over time, and quantifying its uncertainty and sensitivity. This will involve assessing survival analysis as a competing method. We will investigate how each model performs over the coronavirus (COVID-19) and post-EU exit period to see how they are affected by important changes to the migration policy context and migrant behaviour. We will also investigate a weighting approach or alternative cost functions as alternatives to oversampling, with the overall goal still being to penalise false positive predictions more heavily during model training. Our next major milestone is to produce a recommendation on the use of ML for record-level provisional long-term immigration prediction in November 2023.

8 . Uncertainty

The Office for Statistics Regulation (OSR) has recommended that users need a clear understanding of uncertainty associated with international migration estimates and guidance on how the estimates can be used appropriately. We at the Office for National Statistics (ONS) are committed to developing uncertainty measures for our international migration estimates.

This is a complex area of methodology, which requires careful and iterative development of methods. We will be publishing a working paper on 1 June, 2023 to share our research progress. This will focus on quantifying uncertainty specifically associated with adjustments, modelling, and survey-based estimates. In the working paper, we present a simulation-based method to determine the impact of some of the individual sources of uncertainty in the statistical system. We will describe initial results of this method for these individual sources of uncertainty, and the impact on migration estimates. We will also present initial results for our first composite measure of uncertainty for the migration of EU nationals. Certain assumptions and subjective decisions are necessary for estimating international migration using administrative data. We aim to test the impact of these assumptions and subjective decisions through sensitivity analysis.

Our future work will extend our quantification of uncertainty in international migration estimates as we analyse more sources of uncertainty in the statistical system – those sources that are, for simplicity, assumed to be zero in our upcoming working paper. We will develop methods to quantify the uncertainty in the administrative data sources themselves, and uncertainty in the methods applied to the administrative data to establish if a person is a long-term migrant. This will allow us to develop a more complete composite measure of uncertainty for international migration. We will also explore alternative methods alongside our simulation approach for quantifying uncertainty.

9 . Impact of methodological changes

Table 3 and Table 4 outline the improvements made to our estimates since we published in November 2022. Please note that these numbers cannot be added to or subtracted from the previously published estimates to create the new estimates. They are included to give an indication of the size of the changes.

Table 3: Improvements and impacts of methodological changes to immigration estimates

Immigration	Previously published figure	Improvement	Impact	Numerical Impact	Newly published figure
Non-EU Estimates	YE June 22 704,000	Improved methods and understanding of new data [Note 2]	Increase	+29,000	YE June 22 808,000
		Asylum applicants and resettlement schemes	Increase	+87,000	
		Improvement to immigration early leavers	Increase	+<1000	
		Immigration early leavers immigration adjustment for British Nationals (Overseas) (BN(O)) and Ukraine scheme	Decrease	-13,000	
EU Estimates [Note 1]	YE March 22 229,000	Removing C3/C4 arrivals from RAPID estimates	Decrease	-17,000	YE March 22 200,000
		Asylum applicants and resettlement schemes	Increase	+<100	
		Updated student adjustment	Decrease	-13,000	
British Nationals Estimates	YE June 22 135,000	Removing BN(O) from IPS based estimates of British Nationals	Decrease	-26,000	YE June 22 109,000

Source: Office for National Statistics, Department for Work and Pensions, Home Office

Notes

1. For EU estimates we are only able to outline the impact of individual method changes on year ending March 2022 estimates. As the Registration and Population Interactions Database (RAPID) is an annual dataset at year ending March, we make method changes and adjustments to this time point then use statistical modelling to disaggregate to other year ending periods.

2. We have made a number of improvements to our first arrival, last departure method in the Home Office Borders and Immigration data as well as a slight change in the record-level data received.

Table 4: Improvements and impacts of methodological changes to emigration estimates

Emigration	Previously published figure	Improvement	Impact	Numerical Impact	Newly published figure
Non-EU Estimates	YE June 22 195,000	Improved methods and understanding of new data [Note 2]	Decrease	-46,000	YE June 22 170,000
		Asylum applicants and resettlement schemes	Increase	+2,000	
		Emigration re-arrivals adjustment	Decrease	-1,000	
		Emigration early exits adjustment	Increase	+20,000	
EU Estimates [Note 1]	YE March 22 266,000	Removing C3/C4 arrivals from RAPID estimates	Decrease	-26,000	YE March 22 236,000
		Student emigration adjustment	Decrease	-3,000	
		Asylum applicants and resettlement schemes	Increase	+<20	
British Nationals Estimates	YE June 22 90,000	No change			YE June 22 90,000

Source: Office for National Statistics, Department for Work and Pensions, Home Office

Notes

1. For EU estimates we are only able to outline the impact of individual method changes on year ending March 2022 estimates. As the Registration and Population Interactions Database (RAPID) is an annual dataset at year ending March, we make method changes and adjustments to this time point then use statistical modelling to disaggregate to other year ending periods.

2. We have made a number of improvements to our first arrival, last departure method in the Home Office Borders and Immigration data as well as a slight change in the record-level data received.

10 . Related links

[International migration hits new high but there are signs of change](#)
Blog | Released 25 May 2023
Jay Lindop, Deputy Director of the Centre for International Migration, provides context to the migration numbers released today, including key drivers and improvements to methods.

[Long-term international migration, provisional: year ending December 2022](#)
Bulletin | Released 25 May 2023
Experimental and provisional estimates of UK international migration, 2018 to 2022, covers the period since COVID-19 travel restrictions eased.

[Methods to produce provisional long-term international migration estimates](#)
Methodology | 25 May 2023
An explanation of the methods used to produce the latest provisional experimental statistics on migration flows into and out of the UK.

[Population and international migration statistics revisions policy](#)
Methodology | 26 May 2022
This policy explains how we will deal with revisions specific to population and international migration statistics.

11 . Cite this article

Office for National Statistics (ONS), released 25 May 2023, ONS website, article, [International migration research: progress update, May 2023](#)