

Article

# International migration research, progress update: November 2022

An update on international migration methods and research.

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

Release date:  
24 November 2022

Next release:  
To be announced

## Table of contents

1. [Overview of migration statistics transformation](#)
2. [Improvements to methods for non-EU migration using Home Office data](#)
3. [Improvements to EU RAPID-based migration estimates](#)
4. [Irregular migration](#)
5. [Non-UK born population](#)
6. [Developing alternative definitions of international migration](#)
7. [Using machine learning to produce more accurate provisional estimates](#)
8. [Related links](#)
9. [Cite this article](#)

# 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 with our [provisional long-term international migration estimates](#) published today (24 November 2022). Estimates are now available up to the year ending June 2022, a six-month improvement in the timeliness of these data. A new method for measuring non-EU emigration has been implemented and further improvements have been made to the methods used to estimate EU migration. We have also made revisions to previous estimates for earlier time periods.

This article includes further information on:

- the improvements that have been implemented to methods for producing both non-EU and EU migration estimates ([Section 2](#) and [Section 3](#) )
- ongoing research to understand the available data sources for measuring irregular migration and the non-UK born population ([Section 4](#) and [Section 5](#))
- progress updates on longer-term research into alternative migration definitions and the potential of machine learning techniques to improve the accuracy of provisional estimates ([Section 6](#) and [Section 7](#))

We continue to work on the longer-term programme of research set out in our [previous report](#), with the aim of delivering further improvements alongside the next release of migration statistics scheduled to be published in May 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 Border Systems 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 February 2020](#) and [April 2021](#)). Our understanding of this complex source of information continues to expand enabling us to refine and develop our methods.

### Visa periods

Within the Home Office Border Systems Data we look at first arrival and last departure to and from the UK within a visa period to determine whether a person is immigrating long-term (12 months or more). Visa periods are constructed by linking together any consecutive or concurrent visas held. If either the first arrival or last departure information is missing, then visa start or end dates are used as a proxy.

The previous iteration of this method stated that if there was a gap of any duration between visas, then a new visa period was started. Our research has shown that many people have a short gap between a visa ending and a new one starting. Our improved method takes account of this, by allowing a gap of up to and including seven days between visas to be within the same visa period.

For example, previously, an individual with two nine-month visas with just one day between them, would not have been classified as a long-term migrant despite being in the UK for 18 months. Using our improved method, both of these visas are now within the same visa period, with the combined length of stay classifying the individual as a long-term migrant.

## Immigration estimates

We have improved the estimates of long-term immigration for the most recent time points in the data series. The method classifies individuals as long-term immigrants if they have stayed in the UK for at least 12 months following their first arrival in a visa period.

For those individuals whose first arrival occurred within the 12 months before the end of the dataset (currently June 2022), we do not yet have enough data to see a long-term stay of 12 months or more. Instead we use their visa end date as a proxy for a future departure date. All individuals in this group with a visa lasting at least 12 months are therefore counted as long-term immigrants.

However, we recognise that this assumption inherently leads to an overestimation of long-term immigration among our most recent arrivals. On average, over the three-year period from year ending June 2018 to year ending June 2020, around 17% of individuals with visas lasting 12 months or more actually left within 365 days of their arrival date and would have been initially misclassified as a long-term migrant. We have therefore reduced estimates for the latest time points using this proportion (broken down by visa type: work, study and other) to account for the likely level of overestimation.

This adjustment based on historical patterns was not applied to new Ukraine Sponsorship Scheme and Ukraine Family Scheme visa-holders, or those who immigrated under the Afghan Resettlement Programme.

## Emigration estimates

Previously, our long-term emigration estimates have been calculated using aggregated Home Office Border Systems Data. From this we calculated the ratio between emigration (numerator) and immigration (denominator) on a monthly basis and applied it to the calculated non-EU immigration estimates to estimate emigration.

Access to more timely record-level Home Office Border Systems Data has now allowed us to extend the “first arrival, last departure” method to produce estimates of international emigration.

The latest method identifies previous long-term immigrants with a last departure from the UK and records them as a long-term emigrant if they do not return to the UK within 12 months, or if they only return for a short-term stay. We have more confidence in this approach as it is consistent with that used to measure immigration. The more detailed record-level data also enables us to produce direct estimates of non-EU emigration from administrative data for the first time.

This method will overestimate emigrants within the most recent leavers, because we do not currently have enough data to see when an individual assumed to have made a last departure from the UK may in fact return within the next 12 months. Evidence from the period year ending June 2018 to year ending June 2020 suggests that, on average, this assumption may misclassify approximately 2% of identified emigrants. Therefore we have included an adjustment to reduce estimates for the latest data time points using this proportion.

## 3 . Improvements to EU RAPID-based migration estimates

We consider the Registration and Population Interactions Database (RAPID) to be the best data source, at this moment in time, with the most complete coverage for estimating EU migration. We also use RAPID to produce an estimate of non-EU migration trends, which allows us to validate the trends seen in the Home Office data. We [previously adjusted for some of the coverage gaps in RAPID](#) and have been continuing to build our understanding of our sources and limitations, making improvements to how we adjust the data.

## Student adjustment

RAPID uses [interactions with the breadth of benefits and earnings systems](#) in the Department for Work and Pensions (DWP) and HM Revenue and Customs (HMRC) to estimate migration into and out of the UK. Any [students who do not work alongside their studies will not be identified as a long-term migrant](#) using RAPID. For our [latest estimates released in November 2022](#), we have [improved our previous method to adjust for this undercoverage of students](#).

To identify EU students immigrating into the UK long-term, we continue to use Higher Education Statistics Authority (HESA) data as the best available data source. Our latest method links this to newly acquired HMRC Pay as You Earn Real Time Information (PAYE RTI) data to better understand how many international students are in employment alongside their studies.

Our latest analysis shows that between the academic year ending 2016 and the academic year ending 2018, around 39% of inflowing EU students are thought to have had some employment activity during their studies. This is compared with around [61% of EU students in employment activity](#) based on our previous analysis. This decrease in the proportion of students working has resulted in an increase in the number of students being added to the RAPID estimate.

In our new method, the international student inflows for the academic year ending 2016 to the academic year ending 2018 are linked to their corresponding records in PAYE RTI from April 2014 to April 2019 via the Demographic Index. Anyone without any payments from employment in the tax year of arrival and the following tax year are considered to not be included in our RAPID estimates. We calculated the proportion of “students not working” for the three academic years and used an average of these years for years of HESA data where PAYE RTI is not yet available.

HESA data for the academic year ending 2022 are not available until early 2023. Trends from the International Passenger Survey (IPS) suggest that the proportion of EU immigration arriving to study remained broadly flat. Where we were unable to generate student inflow for the most recent year, we have rolled forward the inflow from the previous year. As more data become available, we can review this assumption, including investigating the potential to use the Universities and Colleges Admissions Service (UCAS) data as a leading indicator.

This analysis also provides the following improvements to the student adjustment by:

- including PAYE RTI data containing more detailed employment activity than the annual summary data, which allows us to create bespoke measures of activity after arrival
- including postgraduates, students from the whole of the UK and re-arrivals in our analysis for the first time; therefore “students not working” proportions are more representative of the population they are applied to
- better linkage of the HESA and PAYE RTI data using the Demographic Index rather than the [Statistical Population Dataset V2.0 used previously](#)

For the outflow adjustment, we have continued to use the [method first developed in April 2021](#).

While this new student adjustment is an improvement over our previous method there are still limitations. These include:

- HESA data may contain some students who do not arrive in the UK as it provides data on student enrolments but not on whether they arrive in the UK; arrival in the UK is inferred from a number of indicators including term-time postcode and location of study
- Masters students on a one-year course are included in our estimates but some may not be resident in the UK for 12 months
- there is the possibility of linkage error between the two data sources that we are unable to measure the impact of at this time

We are committed to improving our understanding of the interactions between international students and the labour market. Our next steps are to update our analysis with the latest HESA and PAYE RTI data when they become available and to focus on improving the student outflow adjustment.

## Under-16s adjustment

In April 2021, we acknowledged that [there is a coverage gap in RAPID for those under 16 years old](#). Since then, we have only published migration estimates for EU nationals for those over the age of 16 years. In November 2022, we have introduced a new adjustment to RAPID to help fill this coverage gap.

The adjustment uses an adult-to-child ratio derived from the IPS. Where IPS data are not available (namely in 2020), a five-year average ratio (2016 to 2019, 2021) is applied. This ratio is then applied to RAPID to estimate the number of under-16s to add to the RAPID estimate. This ratio is calculated separately for both inflow and outflow.

This is our first step into looking at this cohort and we will continue to investigate and develop our understanding of migration patterns of EU nationals under 16 years of age.

## 4 . Irregular migration

Irregular migration has become an increasingly topical area of interest. We have carried out research to understand the irregular migrant data journey and assess the viability of adding irregular migration estimates to our migration publications.

Data on irregular migrants are collected upon UK disembarkation, by the Ministry of Defence and the Home Office, as well as throughout humanitarian and immigration processing by the Home Office.

The primary route for irregular migration into the UK is by Channel crossings by small boat. In the year ending June 2022, data published by the Home Office show that approximately [35,000 people arrived by small boat](#) and that from 2018 to 2022, 94% of irregular migrants arriving via this route have applied for asylum. Just over three-quarters (76%) of the initial decisions in the year ending June 2022 were grants of asylum, humanitarian protection or alternative forms of leave. Of those who applied for asylum in 2021, 46,814 cases (94%) were still awaiting an outcome at the end of June 2022.

In future we plan to account for irregular migration in our estimates of long-term international migration by including a component measuring those claiming asylum. Research is ongoing to ensure there is no double-counting with our existing methods and that we include the correct cohort of those claiming asylum (given the time lag between their arrival and a decision on their application being made). We aim to be able to update further on our planned approach for our next release in May 2023.

### Data coverage and quality

Data on irregular migrants making Channel crossings are collected from all persons intercepted and transported to reception and processing sites. These journeys are constrained to an area of the Channel that is covered by a large-scale surveillance operation. This provides confidence that authorities have accurate awareness of the counts of escorted and non-escorted landings. Other routes of irregular migration exist but make up a small proportion of the estimated total.

The data are considered to be of high quality, going through several layers of validation and processing, with regular communication between organisations to reconcile underlying datasets. Reconciliation between Home Office systems has an error rate of 0.1% to 0.2%, with work ongoing to lower this to 0.01% of individuals processed.

Statistics covering these alternative travel routes and small boat crossings can be found on the Home Office website: [Irregular migration to the UK statistics](#).

## 5 . Non-UK born population

Historically we measured non-UK born population estimates using the Annual Population Survey (APS) but on 27 October 2022 we announced that we have [discontinued the series using the APS](#). This is because of an underlying [data issue with the Migrant Worker Scan](#), which means changes in non-UK population will not represent real changes beyond June 2021. As part of the transformations of [population and migration statistics](#) and the [Labour Force Survey \(LFS\)](#), we are reviewing the best methods to produce estimates of the UK population.

Our first [migration statistics from the Census 2021 for England and Wales](#) were published on 2 November 2022, which are our best estimate of the population of England and Wales by country of birth and passports held.

The census provides the best picture of society at a moment in time every 10 years. However, there is a need for more timely and frequent statistics that make the best use of all available data and enable us to understand our population and how it changes on an ongoing basis. We have explored provisional measures that can roll forward the Census 2021 data to produce an even more up-to-date picture. For example, if we wanted to look at the levels of non-UK born living in England and Wales in June 2022:

- to start, we use the [census population estimates](#) for England and Wales — there were 3,643,000 EU-born and 6,375,000 non-EU born in March 2021
- remove [deaths](#) occurred since then (22 March 2021 to 30 June 2022) – 26,000 EU born, and 50,000 non-EU born
- add net non-UK migration since then (as given in our [statistical bulletin](#) released today) – negative 72,000 EU nationals and 539,000 non-EU nationals; these are based on nationality and flows in and out of the whole of the UK
- from this we can estimate that the population level for EU-born in June 2022 was 3,545,000 and for non-EU born was 6,864,000
- these estimates are [experimental](#) and provisional so there is a degree of [uncertainty](#) around them (estimates have been rounded to the nearest 1,000)

These provisional estimates indicate a rise to nearly 7 million for the non-EU population. In contrast, the EU population has slightly fallen since the census.

This method is for England and Wales only. We need to wait for the Scottish census estimates to be published before we can produce a UK estimate. In addition, the method does not yet factor in internal migration, for example, someone born abroad moving from Scotland into England. It uses UK flows when in future we will disaggregate to reflect England and Wales.

In the longer-term and working with the devolved administrations to produce a UK estimate, we are considering the following approaches.

### Use of the Dynamic Population Model

This uses new methods to produce near real-time estimates of the population size. [The Dynamic Population Model \(DPM\)](#) will use a statistical modelling approach to draw strength from a range of data sources such as administrative and survey data. We anticipate using the DPM to support the production of these statistics from 2023.

### Use of the transformed LFS instead of the APS

This could look very similar to the previous series. Based on the planned timelines, this would mean our first set of statistics using the [transformed LFS](#) would provide an estimate year to June 2024, released in autumn 2024 – allowing one full year of data collection.

## 6 . Developing alternative definitions of international migration

We are aware that our long-term migration estimates do not cover all types of migration or fully reflect the reality of dynamic movement and mobility seen in today's population.

Previous [feedback from our users](#) highlighted a need for a broader range of estimates that encompasses the diversity of migration patterns. Statistics on flows of short-term and temporary migrants are needed to better understand and plan for interim populations at a national and local level. There is also strong interest from our users in receiving real-time estimates on the numbers of visitors and short-term migrants within the UK and the likelihood of these migrants staying long-term.

In response, we are continuing to develop methods to meet the increasing need for flexibility in defining a migrant. Our work builds on our previous research on supplementary estimates of migrants based on actual time spent here in the UK, [Exploring international migration concepts and definitions with Home Office administrative data](#) (February 2020).

We are currently at the proof-of-concept stage, focusing on immigrants and using our existing administrative sources to identify evidence of:

- during and end of the month activity within the Department for Work and Pensions' (DWP's) Registration and Population Interaction Database (RAPID)
- travel dates recorded within Home Office Border Systems Data and how we can use them to build up a picture of alternative lengths of stay and types of migration to the UK

We are continuing to gain insights from the DWP's RAPID by looking at the breakdown of activity by month, rather than by activity within a tax year. This offers us the potential to refine our classifications based on patterns of behaviour and length of activity, contributing to our evidence base for alternative definitions of migration.

For our supplementary counts of non-EU immigration we are focusing on Home Office travel information, (rather than visa information) calculating the length of stay in days for each separate journey. By linking together the journeys undertaken by the same individual, we can total up the number of days each individual was in the country over a set period of time. This calculation can then be applied to any alternative definitions of migration by setting three time parameters, the:

- target period — the time within which a journey must either start and/or end
- migrant period — the cumulative number of days that an individual needs to be in the UK during the target period, in order to be categorised as a migrant
- new migrant period — the length of time an individual needs to have been out of the UK before their first arrival in the target period to qualify as a new migrant.

For example, if we are interested in the number of travellers who are present in the UK for the majority of a year, our target period would be for 365 days (such as July 2018 to June 2019) and our eligibility criteria for a migrant period would be half a year plus one day, or 184 days.

If we considered these migrants to be new migrants only if they had not been present in the UK 12 months before their first arrival in the target period, then the time period for the last parameter would also be 365 days. This method has great potential to be adapted to target periods of different lengths, such as 16 months or 24 months, and vary the length of stay needed to be classified as a migrant and a new migrant.

The granularity and flexibility of calculating journey lengths from arrival and departure dates also presents the greatest limitation of this method. As discussed previously, not all journeys in the Home Office Border Systems Data have an arrival date or departure date ([Exploring international migration concepts and definitions with Home Office administrative data February 2020](#)). We will be investigating the extent of incomplete journeys and various options to impute missing travel dates as we continue to develop this proof of concept.



## 7 . Using machine learning to produce more accurate provisional estimates

One of the challenges to producing timely long-term international migration (LTIM) estimates from administrative data is predicting whether a new migrant will become a long-term migrant when their activity is first recorded. In previous statistics using the International Passenger Survey (IPS) we used migrant intentions as the predictor of length of stay, but administrative data do not record intentions.

In our Admin-Based Migration Estimates (ABMEs), we apply operational definitions of LTIM using deterministic classification rules to retroactively classify migrants based on their recorded activity in admin data (in line with the UN definition for LTIM). However, we cannot apply these rules with certainty to migrants who have not been recorded in data long enough to determine a known outcome. We are now exploring whether supervised machine learning (ML) methods could produce accurate provisional predictions of LTIM status before classification rules can be applied.

We are currently developing proof-of-concept ML models to predict LTIM status for non-EU immigrants in Home Office Border Systems Data. In current ABMEs, potential long-term non-EU immigrants are assigned a provisional LTIM status using our “first arrival, last departure” rule (see [Section 2](#)) with their future visa end date used as a provisional last departure date. Appropriately trained ML models may be able to make more accurate predictions where the visa end date is not a good predictor of length of stay.

Supervised ML methods may be able to more accurately account for the likelihood of under- or over-staying a visa across migrant groups by taking into consideration a wide range of migrant characteristics. For example, immigrants on student visas may have a different likelihood of overstaying their visa than those on family visas, and these likelihoods might change over time.

It may also be possible to produce less-biased provisional LTIM flows by using classification probabilities from the models as fractional weights, that is, enumerating provisional long-term migrants in proportion to their model-predicted probability of being genuine long-term migrants.

We plan to assess the accuracy of model-based LTIM predictions against known outcomes at the individual record level, as well as their impact on the accuracy of provisional admin-based immigrant flows against the confirmed data back-series. We will also consider the degree to which these models need to be retrained to maintain accuracy over time. If these proof-of-concept models are successful, we hope to extend this work to non-EU emigrants and EU migrants. We aim to provide an update on this research next year.



## 8 . Related links

[Long-term international migration, provisional: year ending June 2022](#)

Bulletin | Released 24 November 2022

Experimental statistics on UK international migration, including the impact of the coronavirus pandemic.

[International migration statistical design progress report: July 2022](#)

Article | Released 14 July 2022

An update on the future of international migration statistics in 2022 and beyond.

[Dynamic population model for England and Wales: July 2022](#)

Article | Released 14 July 2022

The development of new methods to produce near-real time estimates on the size of the population, using a dynamic population model.

[Dynamic population model case studies: November 2022](#)

Article | Released 23 November 2022

A selection of local authority case studies including estimates of population for up to June 2022, using modelling techniques. This article shares initial research outputs.

[Transforming the way we produce population statistics.](#)

Blog post | Released 14 July 2022

How we are transforming the way we produce population statistics, including an update on the transformation of international migration and the role of information from Census 2021.

## 9 . Cite this article

Office for National Statistics (ONS), released 24 November 2022, ONS website, article, [International migration research: progress update, November 2022](#)