

Distr.: General
11 October 2022

English

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Migration Statistics

Geneva, Switzerland, 26–28 October 2022

Item A of the provisional agenda

Improvements in use of administrative data for migration statistics

Using administrative data to produce timely estimates of migration for the UK

Note by Office for National Statistics |

Abstract

The Office for National Statistics (ONS) has launched a transformation programme that aims to measure international migration with administrative data. Historically international migration statistics for the UK were measured using the International Passenger Survey (IPS), a face-to-face survey which asks migrants their intention to migrate. It has long been acknowledged that measuring migration using the IPS has stretched it beyond its original purpose.

Estimates published in May 2022 were the first official, albeit experimental, international migration statistics that are predominantly based on administrative data. These estimates marked an important milestone in our transformation journey, and a substantial shift away from previous data and methods. These new methods estimate separately the migration of EU, non-EU, and British Nationals, using the best available data, sourced from Home Office and the Department for Work and Pensions.

Estimating international migration with administrative data has brought significant challenges such as producing a consistent back series, measuring uncertainty, and producing timely estimates. We have an ambitious programme of work to improve the quality of UK international migration statistics and our next iteration of improved estimates are due to be published in November 2022.

*Prepared by Laura Cheatham

NOTE: The designations employed in this document do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

I. Introduction

1. International migration statistics for the UK have historically been measured using the International Passenger Survey (IPS), a face-to-face survey at most air and sea-ports in the UK. The IPS measures a migrant's intention to stay in or depart from the UK long-term (for 12 months or more). Previous analysis has demonstrated that certain groups of migrants have a higher degree of uncertainty in their intentions to move to and from the UK (ONS, 2019). Therefore, it has long-been acknowledged that the reliance on the IPS for migration statistics stretched it beyond its original purpose.
2. These findings led the ONS to launch a transformation programme in 2019, focussing on measuring migration using administrative data first and foremost.
3. In 2020 when the coronavirus (COVID-19) pandemic stopped international travel, the IPS was also suspended. In response, we accelerated our approach for transforming migration statistics using administrative data supported by statistical modelling.
4. The IPS resumed operation in January 2021 but, the migrant focussed shifts, aimed at specifically boosting the number of migrant contacts, are no longer being run and the data is even more limited than it was previously. Therefore, we continue to focus on using administrative data to estimate international migration.
5. Estimates published in May 2022 for the Year ending June 2021 were the first official, albeit experimental, estimates of international migration that were predominantly based on administrative data. This represents a major milestone in the transformation of migration statistics.

II. Estimating International Migration during the coronavirus pandemic

April 2021

6. With the suspension of the IPS we needed an alternative method for measuring international migration during the COVID-19 pandemic more quickly than we anticipated when we began the transformation programme back in 2019.
7. We explored time series modelling because of the strong seasonal trends that are evident in international migration over time.
8. We developed a state space model (SSM) where we projected the trends and seasonality of the previous IPS data forward and adjusted it by the structural shift seen in the Home Office visa data for non-EU citizens. This version included assumptions on EU nationals having different travel options during lockdowns. When airports were closed, there was an increase in travel via ferries and Eurotunnel and we hypothesised that motivated EU nationals would have used these transport routes to travel to and from the UK (ONS, 2021a).
9. We used a Delphi Panel approach to gather expert opinion on our model assumptions and modelled estimates. We invited experts to give their views on our assumptions and provide any other evidence that we should consider in our models.
10. This first version of the model provided migration estimates for March 2020 and Quarter 2 2020 (April to June).

November 2021

11. In November 2021 we updated the modelled estimates and published two additional quarters up to the year ending December 2020.
12. We continued to use a time series approach to model international migration during Quarter 3 (July to September) and Quarter 4 (October to December) 2020.
13. For Quarter 2 2020 we applied an adjustment to reflect EU nationals having different travel options during lockdowns when airports were closed. After expert advice we turned off this adjustment from July 2020, when the proportion of cross Channel travel (rail and ferry) reduced as air travel resumed near normal proportions.
14. For non-EU migration, expert consensus supported the modelled approach. For EU and British migration, in the absence of alternative timely data, we continued to model immigration and emigration using non-EU migration trends based on Home Office Exit Checks data. For EU we incorporated an additional adjustment to the model using the ratio of EU and non-EU IPS data. For British we modelled immigration (repatriation) using non-EU departure data, and vice-versa.

III. Using administrative data to estimate migration into and out of the UK

15. Our latest estimates published in May 2022 cover the time period up to the year ending June 2021.
16. Early insights from the 2021 Census results indicated that the modelled estimates were not as robust as previously thought. We are committed to continually improving our methods to estimate international migration and therefore we moved to a new method that relied less on IPS data and statistical modelling and made greater use of administrative data. This represents a major milestone in the ambition to estimate international migration using mainly administrative data.
17. This uses different data sources and methods for each nationality grouping, using the strengths of each data source. We currently publish estimates on immigration, emigration and net migration for non-EU nationals, EU nationals and British Nationals.
18. Methods developed for all nationality groups align with the UN definition of a long-term migrant which is a person who moves to a country other than that of their usual residence for at least a year. Therefore, across our data sources we are looking for interactions within the administrative data that indicate someone has been in the UK or out of the UK for 12 months or more.

A. Estimating the migration of non-EU nationals

19. Estimates of non-EU nationals are derived from Home Office Exit Checks data. This data combines visa information with travel movements into and out of the country, via the Initial Status Analysis (ISA) system. Long-term migrants are identified using arrival and last departure dates within a visa period as an approximation for length of stay in the UK.

20. We consider this to be the best source of data, with the most complete coverage, for measuring non-EU migration. However, our methods for estimating immigration using HO data are more developed than for emigration (ONS, 2021b).
21. To align to the UN definition of a long-term migrant the first step is to identify those on long-term visas. We use arrival and last departure dates within a visa period as an approximation for length of stay in the UK. Short trips abroad over the course of an extended period of residence in the UK are excluded. If either the first arrival or last departure information is missing, then visa start or end dates are used as a proxy.
22. Visa periods are constructed by linking together any consecutive or concurrent visas held. If there is a gap between visas, then a new visa period is started. We look at previous visa periods to determine if this is a new long-term immigrant or one who has previously been in the country. If no presence is identified in the country during the 12 months preceding first arrival on a given visa, or if the previous visa period had a length of stay of less than 12 months, then this pattern of travel will be considered as identifying a new long-term immigrant.
23. Levels of immigration are comparable between Home Office data and estimates derived from data sources from the Department for Work and Pensions. However, there are greater disparities in estimates of non-EU emigration across the two sources (ONS, 2022a).
24. Currently we do not have an equivalent method for measuring non-EU emigration from the Home Office data. Therefore, we calculate a ratio between emigration (numerator) and immigration (denominator) on a monthly basis from the aggregated travel data from the ISA system. We then apply this ratio to the calculated non-EU immigration estimates to estimate emigration. This assumes the trends in the aggregate data for the immigration and emigration series are similar to the trends in the Exit Checks data used to estimate long-term immigration. This is a reasonable assumption as both datasets are derived from the same source.
25. We are developing our understanding of information in the Exit Checks data which can be used to estimate emigration. This includes investigating imputation methods to apportion estimates for the numbers of “true emigrants” depending on departure information, leave type, and other characteristics.
26. In May 2022 record level data was not yet available for the final 8 months of the time series (November 2020 to June 2021). This missing time period was estimated using an aggregated version of the travel data from the ISA system which contained data up to November 2021. We applied the pattern of change observed in the aggregate data set (low frequency series), using the Denton-Cholette method to predict the record level dataset (high frequency) for the missing series. This provided us with a rate of change for arrivals on a month-to-month basis.
27. For estimates that we are publishing in November 2022, we are due to have record level data for the whole timeseries up to June 2022. However, we intend to use this method if the same situation arises in the future.

C. Estimating the migration of EU nationals

28. Currently, we cannot estimate the migration of EU nationals using Home Office data. This is because of free movement between the EU and UK until January 2021 and continued free

movement for EU nationals who have been granted residency through the EU Settlement Scheme (EUSS). With the introduction of new immigration routes for EU nationals through European Economic Area (EEA) visas, these individuals will be present in the Home Office data from January 2021 onwards. We are working with the Home Office to understand if this may allow us to estimate migration of EEA nationals, especially new migrants who are not part of the settlement scheme. This is similar to current methods used for non-EEA nationals.

29. The current method for estimating the migration of EU nationals uses the Registration and Population Interactions Database (RAPID) from the Department for Work and Pensions (DWP) (ONS, 2022b).
30. This database is created by DWP to provide a single coherent view of citizens' interactions across the breadth of earnings and benefits systems in DWP, HM Revenue and Customs (HMRC) and local authorities via housing benefit. It covers everyone with a National Insurance Number (NINo), and for each person, the number of weeks of "activity" within these systems is summarised within a tax year.
31. To estimate long-term international migration to and from the UK using this data firstly we use information from the Migrant Worker Scan (MWS), which identifies all non-UK nationals registering for a NINo from 1975 onwards. This gives us further information including self-reported date of first arrival, date of registration for a NINo, nationality at registration and previous country of residence.
32. Both long-term and short-term migrants can be issued with a NINo. To determine the long-term immigration of non-UK nationals we use a combination of data from the MWS showing when a NINo was issued, alongside the "activity" in the underlying earnings and benefits datasets. To align to the UN definition, we are looking for "activity" to occur for 12 months or more.
33. To identify sustained "activity" over 12 months or more we amalgamate all the tax year datasets into one single longitudinal dataset. This shows all activity since 2010 or since first arrival for those who arrived after 2010. Activity is defined as the number of weeks of interactions with the earnings and benefits systems, where interactions show that person is "active" within the source systems, therefore we use this to show "activity" within the administrative data.
34. Our research has shown that people's lives are complex, therefore we created four categories defining patterns of activity of long-term arrivals. The first two categories most closely align with the UN definition of a long-term migrant whereby we are looking for sustained long-term interactions after arriving in the UK and these make up the largest proportion of long-term arrivals (over 90%). We have included two further categories that expand on this definition of long-term activity, to reflect the complexity of people's lives, although these groups only make up a small proportion of arrivals (less than 10%).
35. Category 1 arrivals: the number of weeks of activity in the registration year and registration year plus 1 are a total of at least 52 weeks, therefore suggesting they are resident for 52 weeks or more over that two-year period.
36. Category 2 arrivals: looking at the period between arrival and registration, plus the duration of activities in registration year and registration year plus 1, the total is over 52 weeks, therefore suggesting they are resident for 52 weeks or more over that time period.
37. Category 3 arrivals: activity occurred in three consecutive years from registration (where registration is counted as an activity), and where the 52-week activity criteria is not met.

However, the presence of activity across multiple tax years suggests they are a resident long-term.

38. Category 4 arrivals: where the number of weeks between the registration date and the end of the tax year, plus the activity in the registration year plus 1 is over 52 weeks in total, where there must be at least one week of activity in the registration year plus 1.
39. It is assumed that to continue to be resident in the UK someone would be present in at least one of the source earnings and benefits system that feed into RAPID. Therefore, to measure long-term emigration we need to determine individuals who no longer have activity in the data and are therefore no longer resident in the UK. Anyone who has a whole tax year of inactivity against all source systems in RAPID are counted as a long-term emigrant.
40. RAPID also estimates re-arrivals using the same methodology although only Category 1 and Category 3 rules apply. This is because Category 2 considers the time between arrival and registration for a NINo which only applies to first time arrivals. Category 4 considers the time between registration for a NINo and any activity, which only applies to first time arrivals. Anyone who has a period of inactivity, and a subsequent period of activity will be counted as a re-arrival.
41. The coverage of RAPID is extensive for most migrants due to the wide range of data sources included, however, there are some populations where activities within the source dataset are less well covered. For example, migrants who come to the UK with the sole purpose of studying. Students who do not hold a NINo will not be included in the benefits and earnings data and those who only work for some of their time in the UK may not have enough “activity” to be captured by one of the four arrival categories.
42. Therefore, we apply a student adjustment using data from the Higher Education Statistics Agency, linked to HM Revenue and Customs (HMRC) earnings data to inform us about employment and economic activity of international students in Higher Education. We identify the proportion of first year students who are not in any employment during their studies, and by applying this proportion to the HESA first year inflow we can estimate the number of first year students who are likely to not be captured by RAPID.
43. Like the Home Office data, data from DWP was only available up to April 2021. Therefore, again using the Denton-Cholette method we applied the predicted IPS series to the financial year RAPID estimates to both disaggregate it to a monthly series and then predict this RAPID-based measure for April to June 2021. We intend to repeat this for the YE June 2022 international migration estimation due to be published in November 2022 where the latest record level data will cover the period up to April 2022.
44. The current method identifies the total number of weeks of activity in a tax year but does not distinguish when this activity occurred. Therefore, we want to improve the categorisation of long-term migrants to look at continued activity during each month or at the end of each month, accounting for short periods of inactivity. This would also help produce more timely estimates of emigration, we are currently reliant on a whole tax year of inactivity, but using these monthly indicators of activity we could instead identify rolling 12 month periods of inactivity.
45. We are improving the student adjustment by linking Higher Education Statistics Agency (HESA) data to HM Revenue and Customs (HMRC) real time information (RTI). By linking these data, we can identify the extent to which students have economic activity within a tax year, thereby allowing us to estimate the number of international students not present on RAPID.

Estimating the migration of British Nationals

46. While historically the migration of British Nationals has been a smaller component of total international migration, this group remains the most challenging to measure using administrative data. This is because there is no requirement for these individuals to interact with administrative data sources to inform them of an intention to emigrate or subsequently return.
47. The IPS data are still our main source of information on the migration of British Nationals. The IPS was reinstated in January 2021, and we use these data as our estimates for January 2021 onwards. However, to cover the period when the IPS was suspended (March to December 2020), we use the SSM time series analysis. This takes the available IPS and administrative data and uses the relationship between them to estimate the missing IPS data. We assume that the pattern of British nationals' immigration to the UK is equivalent to non-EU nationals' emigration from the UK (measured using visa data) and vice versa.
48. To move away from IPS data it is likely that combination of sources will be the most appropriate. We are assessing whether the earnings and benefits data from DWP can be used to measure migration of British Nationals. This will involve identifying whether extended periods of inactivity could indicate emigration and whether subsequent activity could indicate immigration.
49. This is more difficult because there are many reasons someone can stop interacting with earnings and benefits system and still be living in the UK. For example, this can include students who stop working while in higher education, young people Not in Education Employment or Training (NEET) or someone taking a career break. Therefore, as part of our work we will be creating rules in the dataset to keep individuals resident in these scenarios.

IV. The challenges of administrative data

A. Why we cannot count people in and out at the border

50. A common misconception is that it is easy to measure international migration by simply counting people in and out as they cross the border. There are many reasons why this is difficult.
51. Some people hold two passports and use different passports for incoming and outgoing journeys.
52. The UK and Ireland belong to a free travel zone called the Common Travel Area, where people can travel freely between the two countries and movements across the land border between Northern Ireland and the Republic of Ireland are not tracked.
53. Our users want to know not only how many people migrate but where they have moved to or from within the UK, counting people as they cross the border doesn't provide any information on where they go on to settle or where they have come from. Therefore, we are dependent on alternative data sources to provide this information.

C. Producing timely estimates of migration into and out of the UK

54. Due to the timeliness of the different data supplies, and time needed to process the data, migration estimates based on administrative data will be provided to users with a lag of around five to six months.
55. Data from the Home Office are supplied around 3 months after the end of the reference period and are supplied quarterly. Data from the Department for Work and Pensions are supplied around 3 months after the end of the tax year and is an annual dataset.
56. Users have requested much more timely insights of migration patterns and change. This includes an almost real-time estimate of international migration.
57. To produce more timely estimates, we are exploring a continuation of the State Space Model (SSM) where we have an absence of data. This would use signal data to nowcast international migration. We will be assessing the strengths of multiple data sources of signal data. This will include published visa data and data from Advanced Passenger Information from the Home Office. We will also assess the use of employment counts by nationality, or GP registrations.
58. As the State Space Model can use multiple data sources together, we can assess different combinations to evaluate which data sources provide the most accurate results.
59. In addition to the lag from the timeliness of the data supplies we also have a lag due to moving from intentions-based estimates to estimates based on actual behaviour. The IPS recorded a persons intention to migrate which allowed for more timely estimates, whereas administrative based methods require at least 12 months to have passed in order to observe long-term interactions to determine if that individual is a long-term migrant. Although, in many cases significantly more than 12 months of data are required.
60. Provisional estimates using Home Office data can be produced for the latest reference period using proxy information from visa end dates instead of actual length of stay where necessary. Provisional estimates allow more timely and up-to-date estimates to be made available. However, these should be interpreted with some caution as we do not yet have a long enough time series of data to fully assess the quality of provisional estimates compared with those produced using the full method.
61. Provisional estimates using data from DWP can be produced by applying an adjustment to the latest two years of RAPID inflow and the final year of RAPID outflow to account for this. These adjustments estimate the proportion of recent arrivals who become long-term migrants based on previously seen patterns in the estimates from RAPID. The same methodology is used to estimate the number of long-term migrants who are expected to have left in the latest year. However, caution should be taken as we know these proportions based on historic patterns may not hold true during the COVID-19 pandemic and with ongoing changes in migration policy following Britain's exit from the EU.
62. To improve the provisional estimates of long-term migration where not enough time has passed to observe actual long-term interactions, we are exploring how a predictive machine learning model may provide timely prediction at the aggregate level before the rule-based classification in the record level data. A separate paper has been submitted to the UNECE Migration Statistics Expert group on this topic "A machine learning approach to classifying UK long-term international migrants using administrative data".

C. Using Advanced Passenger Information for international migration statistics

63. Advanced Passenger Information (API) contains information that passengers are required to provide to their airline or travel company before travelling. This includes personal data such as name, gender, date of birth and passport number. We conducted a study to assess whether API microdata can provide an additional data source to improve the quality of international migration estimates.
64. API data could be used to track individuals arriving and departing from the UK. API has potential to add value for assessing arrivals and departures of British nationals; a group that are not well covered in existing administrative data sources. Further value could be realised through access to real-time data to monitor changing trends in travel patterns as they happen, for example those resulting from COVID-19 or global conflicts. While not necessarily conforming to our standard definitions of migration, this may provide a timely signal to nowcast our headline estimates.
65. This data is highly complex, and our initial feasibility assessment has highlighted a range of caveats and challenges to its use for improving international migration estimates that will need to be considered going forward. We intend to work closely with ONS data scientists as well Home Office data experts to develop understanding of the data; improved metadata; and data linkage.

How we are improving the granularity of migration estimates

66. Users have highlighted the need to not only know how many people migrate to and from the UK, but why. The IPS asked migrants their reason for migration, which allowed us to disaggregate our previous IPS based estimates. However, this is more difficult to estimate with administrative data.
67. We plan to make further use of visa data from the Home Office to give an indication of the reason for migration as the new immigration system matures. However, we must use this data in conjunction with other sources, to provide a fuller picture. EU nationals who have registered for pre-settled or settled status as part of the EUSS are not subject to immigration control. Therefore, we will not have the same information about their reason for migration as visa-requiring nationals.
68. We are assessing the use of data from the Higher Education Statistics Agency (HESA) linked to HMRC earnings data to understand the migration patterns of students at the end of their studies. This can indicate how many students stay in the UK to work at the end of their studies. And of those who leave the UK how many subsequently return to the UK to work.
69. As part of our transformation programme, users have highlighted a need for granular estimates of international migration by age and sex at national and local authority level. We are exploring the use of administrative data to provide these granular estimates. One of the benefits of using administrative data is that it typically has large population coverage, allowing for reasonably robust disaggregation. For instance, both the data from DWP and the Home Office contain information on the age and sex of those in the data. Estimates at Local Authority level are more complex. While we have some limited address information from RAPID, data from the Home Office is based on visa information and travel data that does not provide an indication of where migrants settle. One option we are exploring is to

estimate geographical proportions in RAPID for EU, non-EU, and British national migrants, and apply these to alternative data sources.

V. Conclusion and Future Developments

70. The next long-term international migration provisional estimates up to the Year Ending June 2022 are due to be published in November 2022.
71. This will provide estimates up to the year ending June 2022, where our latest published estimates cover the period up to June 2021. We will be including data on irregular migration into our administrative based estimates for the first time. In addition to this we will be delivering improvements to our methods.
72. We are improving the emigration method for non-EU nationals using Home Office data. This includes investigating imputation methods to apportion estimates for the numbers of “true emigrants” depending on departure information, leave type and other characteristics.
73. For EU nationals using the earnings and benefits data we are working on improving the student adjustment, this will use data from the Higher Education Statistics Agency (HESA) linked to HMRC earnings data. By linking these data, we can identify the extent to which students have economic activity within a tax year, thereby allowing us to estimate the number of international students not present on the earnings and benefits data.
74. In addition, for non-EU nationals we are improving the methods used to identify long-term activity within the earnings and benefits data. The current method identifies the total number of weeks of activity but does not distinguish when this activity occurred. Therefore, we want to look at indicators of activity during each month, identifying continued activity, accounting for short periods of inactivity.

VI. References

75. Office for National Statistics (2019) ‘Understanding different migration data sources: August 2019 progress report’, available from:
<https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/articles/understandingdifferentmigrationdatasources/augustprogressreport>
76. Office for National Statistics (2021,a) ‘Using statistical modelling to estimate UK international migration’, available from:
<https://www.ons.gov.uk/methodology/methodologicalpublications/generalmethodology/onsworkingpaperseries/usingstatisticalmodellingtoestimateukinternationalmigration#exploratory-data-analysis>
77. Office for National Statistics (2021,b) ‘ Developing our approach for producing admin based migration estimates’, available from:
<https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/articles/internationalmigrationdevelopingourapproachforproducingadminbasedmigrationestimates/2021-04-16#home-office-border-data-and-international-migration>
78. Office for National Statistics (2022a), ‘Long-term international migration, provisional: Year ending June 2021’, available from:
<https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/bulletins/longterminternationalmigrationprovisional/june2021#comparisons-with-other-data-sources>

79. Office for National Statistics (2022b), 'Methods to produce provisional long-term international migration estimates', available from:
<https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/methodologies/methodstoproduceprovisionallongterminternationalmigrationestimates#the-method-for-our-latest-estimates> |