Article

International migration: developing our approach for producing admin-based migration estimates

A summary of our current research into estimating international migration using administrative data focusing on the Registration and Population Interaction Database (RAPID), which includes data from the Department for Work and Pensions (DWP) and HM Revenue and Customs (HMRC); and Home Office Border data.

Table of contents

1 . Overview

This article presents research into a new methodology to measure international migration. Estimates presented are not official statistics on migration, nor are they used in the underlying methods or assumptions in the production of any other official statistics. This article and the accompanying report on the methods for using the administrative data source Registration and Population Interaction Database (RAPID) to measure international migration into and out of the UK are to help inform users on the progress of our approach for transforming migration statistics using administrative data and our plans to further develop these in the future.

Using administrative data to measure international migration presents a substantial change in methodology, and it is important that this is recognised when comparing this new approach against previous estimates. Until now, estimates based on the International Passenger Survey (IPS) record migrants' intentions to remain in or out of the UK in the next 12 months. Our new approach, based upon administrative data, reports on the observed length of activities within these datasets to estimate migration into and out of the UK. This has the benefit of removing uncertainty around migrant intentions where administrative data are based on actual observed patterns of behaviour. Both approaches align with the UN definition of a migrant, relating to a change in the usual place of residence for 12 months or more.

This report does not cover the period that saw the main impact of the coronavirus (COVID-19) pandemic. A separate methodological article on the use of statistical models to estimate international migration for the period January to June 2020 has been published alongside this report. We have published an update on what has changed with migration and mobility since the COVID-19 pandemic which summarises available data sources on international migration and mobility patterns.

RAPID provides a single coherent view of interactions across the breadth of benefits and earnings datasets for anyone with a National Insurance number (NINo). The Migrant Worker Scan (MWS) shows all non-UK nationals who have registered for a NINo from 1975 onwards along with their NINo registration date and self-reported date of arrival. This information alongside the interactions within RAPID can be used to infer long-term migration of non-UK nationals into the UK where this activity extends over 12 months or more, and long-term migration out of the UK where this activity stops for over 12 months (as per the UN definition).

Border data from the Home Office provide further insight into the cross-border movement of people. Initial estimates from this source are based on actual travel patterns of non-European Economic Area (EEA) nationals, as those are the people who were covered by immigration rules at the time the data was collected. Future work will aim to build further understanding of the movement of EEA nationals in this data, as this information is now collected following the UK's exit from the EU.

While the coverage of RAPID data is extensive for most migrants because of the wide range of data sources included, there are some populations less well covered in the data. For example, students who do not hold a NINo will not be included and students who do hold a NINo may not be identified as a long-term migrant if they do not undertake any activity, such as working alongside their studies. To address the coverage gaps identified, we have applied a series of adjustments to estimates from RAPID using other available administrative data. In addition, migrants under the age of 16 years and UK nationals have been removed from the dataset used in this analysis, understanding the migration patterns of these groups make up an important part of our future development of admin-based migration estimates (ABMEs).

Our analysis of RAPID data showed that both EU and non-EU migration trends broadly mirror those estimated by the IPS. However, estimates from RAPID showed higher net migration for EU nationals and lower net migration for non-EU nationals compared with Long Term International Migration (LTIM) estimates based on the IPS.

One of the main drivers for the higher estimates of EU net migration based on RAPID is likely to be the uncertainty of these migrants in their intentions to move to and from the UK. This makes it difficult to measure their migration patterns using the IPS. This suggests that differences between IPS figures and RAPID may be related to RAPID being based on actual observed patterns of behaviour.

The lower non-EU net migration estimates based on RAPID are mainly driven by the lower inflow estimates from RAPID across the timeseries. One of the main drivers for the lower estimates of non-EU net migration based on RAPID is likely to be the measurement of students migrating to and from the UK. While we have adjusted estimates from RAPID to better capture the volume of student migration to and from the UK, it is likely that this adjustment does not fully address this coverage gap and that some students are still missing from the estimates.

Our ABME development so far has focused on understanding the potential of RAPID and Home Office border data to estimate migration into and out of the UK. We are still at an early stage in our journey towards a transformed migration statistics system based on administrative data.

Our immediate next steps will focus on refining these methods, extending our use of RAPID data to include UK nationals and working closely with the Home Office as data on EU nationals becomes available from their new migration system. Looking beyond this, we plan to publish a further ABME publication in early 2022. This will cover the period up to April 2021, reporting data beyond the point of the last IPS estimates and including some data covering the onset of the coronavirus (COVID-19) pandemic. It will also include our latest progress in the development of ABME methods.

We are also exploring how multiple data sources (including RAPID and Home Office border data) can be integrated, considering their relative strengths and limitations, to provide the best possible estimates of international migration. As well as exploring how we can make use of statistical modelling techniques to understand the period affected by the coronavirus pandemic and how the timeliness of ABMEs can be improved.

2 . Transformation of migration statistics

We have long acknowledged that the International Passenger Survey (IPS), which underpins our existing international migration estimates, has been stretched beyond its original purpose and we need to consider all available sources and methods to fully understand international migration. In March 2020, the IPS was suspended because of the impact of the coronavirus (COVID-19) pandemic on survey operations and no further IPS data on migration have been collected. Therefore, we have accelerated our approach for transforming migration statistics using new methods and administrative data.

Our previous research identified a range of data sources held across government that can help us to measure migration including immigration, income, benefits and education data. However, some of these data sources can only tell us about migration into the UK. Therefore, we have focused our current research on those that can tell us about both immigration and emigration. So far, the two sources of administrative data which have shown greatest potential for the measurement of long-term immigration and emigration are:

- Department for Work and Pensions (DWP) Registration and Population Interaction Database (RAPID)

- border data from the Home Office exit checks dataset

In this first iteration of our development of admin-based migration estimates (ABMEs), we have explored each data source individually, developing methods to estimate long-term migration within each source. However, it is important to recognise that we are still at an early stage in transforming migration statistics and there remains further development before we will be able to produce official estimates of international migration using administrative data sources. At this stage, estimates are based on aggregate RAPID data, and we have applied broad initial adjustments to take account of coverage gaps. The estimates presented in this article should be interpreted with this in mind. There will be a level of uncertainty around these adjustments and as other new data sources become available, we will continue to refine these adjustments and will reflect these in future research outputs. The methods used in this first iteration of ABMEs are likely to change and evolve as we explore these data sources in more detail.

As part of the delivery of a transformed admin-based approach to migration statistics, by 2023 we will aim to bring together more detailed data across a range of data sources, linking sources together where possible. This will allow us to draw on the strengths of each source in understanding migration and minimise the impact of coverage issues in any one individual data source.

3 . Measuring international migration with administrative data

Delivering new measures of international migration using administrative data sources presents a substantial change in the measurement of migration. Until now estimates of international migration have been based on the International Passenger Survey (IPS), which interviewed migrants to record how long they were intending to remain in or out of the UK in the next 12 months. Administrative data on the other hand are retrospective and tell us about actual activity that has already happened. It is important to understand this shift in the underlying data and the methodology when comparing estimates based on administrative data to the previously published IPS estimates.

As administrative data are based on actual observed patterns of behaviour there will be time lags before we can use these sources to determine an arrival to the UK. Migrants may not register for public services or come into contact with government systems immediately and consequently, will not be present in the administrative data until they do. We also need to wait for arrivals or departures to be active or inactive in the data sources for at least 12 months to be considered a long-term migrant (as stated in the UN definition).

Our work so far has focused on what administrative data tell us about historical trends of international migration, this includes an adjustment to the latest two tax years to take account of the time needed to determine if someone is a long-term migrant. However, as part of our development of admin-based migration estimates (ABMEs) we are considering alternative modelling approaches to provide more timely estimates of migration. See the transformation overview for more information on the progress of this work.

Registration and Population Interaction Database (RAPID)

The Registration and Population Interaction Database (RAPID) has been developed by the Department for Work and Pensions (DWP) to provide a single coherent view of interactions across the breadth of systems in DWP, HM Revenue and Customs (HMRC) and local authorities via Housing Benefit. These interactions include benefits, employment, self-employment, pensions and in-work benefit. RAPID contains a record of everyone who has a National Insurance number (NINo). For each person, the number of weeks of "activity" (interactions) within these systems is summarised for each year from tax year ending 2011 to tax year ending 2020.

While RAPID has not been developed with the sole purpose of measuring international migration, non-UK nationals can be identified in RAPID using the Migrant Worker Scan (MWS) dataset. This records information on overseas nationals registering for a NINo from 1975 onwards. The MWS also collects information on migrants' self-reported date of first arrival, their NINo registration date, their nationality at time of application and their previous country of residence.

Both long-term and short-term migrants can be issued with a NINo, therefore the process of being issued with a NINo is not enough to indicate long-term migration. To determine long-term immigration of non-UK nationals in RAPID we use a combination of data from the MWS showing when a NINo was issued alongside the "activity" within DWP and HMRC datasets. Using the self-reported date of arrival, we can measure the number of weeks of subsequent activity after arrival to determine those who stay in the UK long-term. All our research using administrative data has shown that people's lives are complex, therefore we have created multiple categories[1] of long-term interactions to account for this complexity.

We have created four categories defining patterns of activity of long-term arrivals (Figure 1). 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. It is important to note that RAPID does not specify that this activity is continuous, however, we have assumed the total activity measured to be enough to indicate long-term presence in the UK. These two categories make up the largest proportion of long-term arrivals in RAPID (over 90%).

We have also included two further categories that expand on this definition of long-term activity, although it is important to note that each these groups only make up a small proportion of arrivals.

- Category 1: activities in the registration year and registration year plus one suggest they are resident for 52 weeks or more over that two-year period

- Category 2: the period between arrival and registration, plus the duration of activities in registration year and registration year plus one, suggest they are resident for 52 weeks or more

- Category 3: 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 but where the activity profile suggests they are resident for 52 weeks or more

- Category 4: where the number of weeks between the registration date and the end of the tax year, plus the activity in the registration year plus one suggest they have been resident for 52 weeks or more; there must be at least one week of activity in the registration year plus one

Figure 1: Illustrative examples of identifying long-term international arrivals using RAPID data
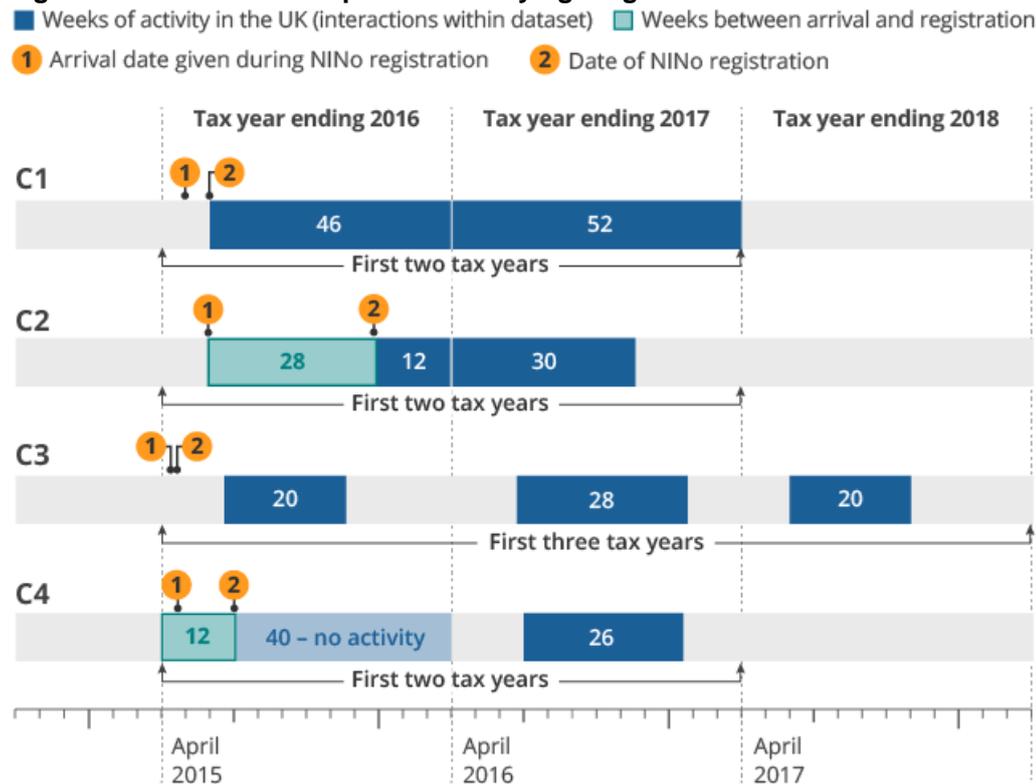
Source: Department for Work and Pensions - Registration and Population Interaction Database

Notes:

1. National Insurance number (NINo).

2. Where activity extends over multiple tax years RAPID does not specify that this activity is continuous, however we have assumed the total activity measured to be enough to indicate long-term presence in the UK.

It is assumed that to continue to be resident in the UK someone would be present in at least one of the source systems that feed into RAPID and therefore have activity in the RAPID dataset, either through claiming benefits, or through their earnings or pension[2]. Therefore, to measure long-term emigration we need to determine individuals who no longer have any "activity" in the RAPID dataset and are therefore no longer resident in the UK. Anyone who has a whole tax year of inactivity against all source systems that feed into RAPID are counted as a long-term emigrant.

Re-arrivals are defined as anyone who has had a period of inactivity (therefore is a long-term emigrant) and subsequently has a period of activity again, even after many years. RAPID measures re-arrivals using the same methodology as first time arrivals although only Category 1 and Category 3 rules apply.

For further details of the rules used to define arrivals, departures and re-arrivals see the accompanying methodological note.

The challenges and limitations of using RAPID to measure migration

As RAPID covers everyone with a National Insurance number (NINo), it includes migrants from EU and non-EU countries as well as UK nationals. The methodology for estimating migration currently removes UK nationals, understanding the migration patterns of this group makes up an important part of our future development. Any non-UK nationals arriving in the UK will need to apply for a NINo in order to work, claim benefits or apply for a student loan. The coverage is extensive for most migrants because of the wide range of data sources included, however, there are some populations less well covered in the data (for example, students). In addition, there are populations with differences in how groups are measured (in particular, those migrants who have taken on UK nationality since arriving in the UK). There are further differences in the coverage of RAPID compared with the coverage of the IPS, including for migrants who are accompanying friends or family and migrants of pensions age. For further information see the accompanying methodological note.

Migrant children aged under 16 years are not covered in RAPID. Children who arrive into the UK do not need to register for a NINo in the same way as adults. As such, they are not recorded or captured by the MWS. While Child Benefit data are contained within RAPID, they do not provide any evidence of the nationality of the child and are not suitable for the analysis of migration into or out of the UK. Therefore, those under the age of 16 years have been removed from the dataset and the estimates from RAPID presented in this report concentrate on those aged 16 years and over. See Section 9: Future developments for more information on our future work plans to measure the migration patterns of those under the age of 16 years.

Visiting students who do not hold a NINo will not be included in RAPID. Any students who do hold a NINo will be included in RAPID, however, they may not be identified as a long-term migrant if they do not undertake any activity that verifies their long-term presence in the UK, for example, some form of work alongside their studies.

In addition, there are also some differences in how RAPID records migrants compared with other data sources such as the IPS and Home Office border data. RAPID identifies migrants using the MWS and captures their nationality at the point of registration. This means that anyone who has moved to the UK and applied for a NINo since 1975 will be recorded in RAPID as a non-UK national. This is because RAPID does not currently include any data on applications for UK citizenship or indefinite leave to remain and therefore does not update the nationality for these people. Consequently, these people would be counted as outflow of a non-UK national even though they have subsequently gained UK citizenship.

In addition to the population coverage challenges, there is also the time gap caused by the time required to assess whether activity is long-term. Where administrative data are based on patterns of activity with the underlying datasets, it takes at least 12 months of activity to take place before a measure of long-term duration can be calculated. In most cases more than 12 months of data are needed to determine long-term residency.

To address the coverage gaps identified and account for the time needed to assess whether records in RAPID are long-term migrants, we have applied a series of adjustments to estimates from RAPID using other available administrative data.

Adjustments currently applied to the estimates from RAPID use existing administrative data and research to demonstrate how alternative data can be used to fill the gaps in RAPID. There will be a level of uncertainty around these estimates and as other new data sources become available, we will continue to refine these adjustments and will reflect these in future research outputs. For a detailed explanation see the accompanying methodological report.

Student inflow adjustment

RAPID is reliant on migrants interacting with the employment and benefits systems to capture their presence in the UK. Our previous analysis has shown that not all students work alongside their studies and that the proportion that do work varies by nationality group. Using this previous analysis, alongside data from the Higher Education Statistics Agency (HESA) showing the number of students enrolling in first year courses, we can estimate a number of students arriving in the UK to study who are unlikely to already be captured on RAPID. This adjustment is then added onto the RAPID estimates to account for these arrivals.

Student outflow adjustment

As we have applied a student inflow adjustment, it is important for net migration that an outflow adjustment is also made. The student outflow adjustment is based on the method used as part of the IPS adjustment introduced in August 2019. For non-EU students we have adjusted the method to take into account the proportion of students working alongside their studies. The adjustment for non-EU students uses travel data collected at the UK border (Home Office exit checks) to identify the proportion of students leaving at the end of their studies. As Home Office exit checks only cover non-EEA migrants, the EU student adjustment adopts a similar approach utilising analysis from the Graduate Outcomes data published by the Department for Education.

UK citizenship adjustment

RAPID is currently unable to identify migrants who go on to become UK citizens, and continues to classify them as non-UK nationals. This means the estimates of non-UK outflow will include some people who have since become UK nationals. By using Home Office Migrant Journey data to look at non-EU migrants who were issued a visa by the Home Office, we can estimate the proportion who go on to gain UK citizenship within a set period of time (10 years). Using outflow estimates from RAPID of non-EU nationals who have lived in the UK for at least 10 years prior to outflow we can apply an adjustment to account for the fact that an estimated proportion of these will now have become UK citizens. Therefore the outflow of these UK citizens has been removed from the non-UK outflow total. As we develop our approach for estimating international migration of UK nationals, we will also include those removed from non-UK outflow through this adjustment.

Provisional inflow and outflow

As administrative data are based on actual observed patterns of behaviour there will be time lags before we can use these sources to determine arrivals to the UK. We need at least 12 months of observed activity within the data to determine if a migrant is arriving or departing long-term and for some arrival categories we need up to three years of data to assess if that person is in the UK long-term (Figure 1). In order to produce more timely analysis, we have applied 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.

Notes for Measuring international migration with administrative data

1. These are applied in order, and once categorised, a person is not re-categorised. That is, if a person fulfils the criteria for Category 1 and Category 3, they remain classed as Category 1.

2. There are some instances where someone may be inactive in the datasets but still resident in the UK. For example, an adult in receipt of a Child Benefit claim which ends because of the child reaching the age where the benefit is no longer payable. In these instances RAPID applies rules to keep these people resident and they are not counted as emigrating, see the accompanying methodological note for more information.

4 . Results and analysis of RAPID data

This section presents our first analysis of RAPID (Registration and Population Interaction Database) data to estimate international migration flows of non-UK nationals from the year ending March 2012 onwards[1]. It looks at trends over time and presents these alongside previous Long-Term International Migration (LTIM) estimates based on the International Passenger Survey (IPS). For further breakdowns by EU and non-EU country groups see the accompanying datasets.

Estimates of international migration using administrative data sources require a change in the methodology for estimating migration. Estimates from RAPID data are based on actual long-term patterns of behaviour of migrants after arriving in the UK whereas the IPS is based on migrants' intentions to stay in or out of the UK in the following 12 months. Be aware of the main data source and methodological differences when comparing estimates from the two sources.

Long-term international migration (LTIM) estimates presented here do not match those previously published in the Migration Statistics Quarterly Report. Analysis presented in this report has removed those aged under 16 years of age from the IPS component of LTIM. We have also presented the IPS confidence intervals to illustrate the uncertainty in these estimates. It is important to note that because of the adjustments applied there will be a level of uncertainty around the estimates from RAPID. While it is not possible to measure this uncertainty at this time, as part of our ongoing development we will be exploring approaches to measuring uncertainty of admin-based migration estimates (ABMEs).

There are small differences between the annual time periods covered by each data source. Estimates from RAPID are tax years (ending 6 April) and LTIM estimates are year ending 31 March. Since these are broadly comparable time periods and for the purpose of clarity, for each data source we refer to a common annual period of "year ending March".

Understanding EU migration data in RAPID

For EU nationals, both estimates from RAPID and LTIM data show a similar trend in the number of arrivals into the UK (Figure 2). However, for the year ending March 2012 onwards, estimates from RAPID show consistently higher numbers of long-term arrivals, with RAPID estimates being nearly double those from LTIM.

While the two sources are not directly comparable because of the methodological differences in measuring long-term migration, estimates from RAPID data suggest they are between 123,000 higher in the year ending March 2012 and 215,000 higher in year ending March 2016 than those estimated by LTIM data. The most likely driver for this difference is the uncertainty of these migrants in their intentions to move to the UK when responding to the IPS. This suggests that differences between IPS figures and RAPID may be related to RAPID being based on actual observed patterns of behaviour. For example, in the IPS where their intentions were to stay in the UK less than 12 months (and therefore would be recorded as a short-term migrant), they may have stayed longer than this and therefore could have been recorded as a long-term migrant. Where estimates from RAPID data are based on actual observed patterns of behaviour of non-UK nationals after arriving in the UK and therefore are not reliant on migrants knowing their intended length of stay.

The increase in the number of arrivals in the RAPID and LTIM data estimates between the year ending March 2013 and the year ending March 2016 are mainly driven by a higher number of EU14 and EU2 nationals arriving in the UK. For EU2 nationals, a large increase in the estimated number of arrivals from RAPID data can be seen after the lifting of transitional controls restricting their access to the UK labour market in January 2014. From the year ending March 2015, EU2 nationals make up around half of the total difference between the estimated number of long-term EU arrivals in RAPID and LTIM data. After the year ending March 2016, estimates from RAPID data show a decrease in EU arrivals, consistent with the LTIM trend.

Data by nationality group can be found in the [accompanying dataset](#).

Figure 2: RAPID data show consistently higher estimates of long-term EU arrivals compared with LTIM data

EU long-term international immigration, UK, year ending March 2012 to year ending March 2020

Notes:

1. RAPID data has been adjusted (see Section 3 for further detail on the adjustments).

2. The final two years of RAPID data represented by a dashed line are provisional (see Section 3 for further detail).

3. The LTIM data represent our current best estimates and preliminary adjustments have been made based on Department for Work and Pensions (DWP) and Home Office data. Different types of lines have been used to represent where adjustments have and have not been applied. For LTIM estimates solid lines indicate adjustments have been applied (see note 2). Dashed lines indicate no adjustment has been made yet because of data availability.

4. Preliminary adjusted immigration and net migration LTIM estimates for EU8 citizens have been produced for the year ending December 2009 to the year ending March 2016.

5. Those aged under 16 years have been removed from the IPS component of LTIM data to align with the population covered by estimates from RAPID.

6. Confidence intervals are based on IPS estimates.

7. There are small differences between the annual time periods covered by each data source. Estimates from RAPID are tax years (ending 6 April) and LTIM estimates are year ending 31 March. Since these are broadly comparable time periods and for the purpose of clarity, for each data source we refer to a common annual period of 'year ending March'.

Download the data

[.xlsx](#)

When looking at estimates from RAPID data of long-term outflow alongside the LTIM estimates, both sources are showing similar trends in the departures of EU nationals. Between year ending March 2013 and year ending March 2020, both sources show an increase in the number of departures in this timeframe (Figure 3).

While the two sources are not directly comparable because of the methodological differences in measuring long-term migration, estimates from RAPID data suggest they are between 36,000 higher in the year ending March 2012 and 132,000 higher in the year ending March 2016 than those previously estimated by LTIM data. Until the year ending March 2016, the difference between the RAPID and LITM data estimates is mainly driven by RAPID estimating higher outflow of EU8 nationals. From year ending March 2016, the difference between the RAPID and LTIM data estimates is driven by both EU8 and EU2 nationals.

Estimates from RAPID data show a higher number of long-term departures for EU nationals than LTIM, however, it is important to put this in the context of the number of arrivals estimated by each data source. Estimates from RAPID data show a higher level of inflow, therefore the resident EU migrant population who could outflow is higher. Another likely cause of the difference between these two sources is uncertainty of these migrants in their intentions to move from the UK, where RAPID data estimates are based on actual observed patterns of behaviour.

Figure 3: RAPID estimates a higher number of long-term departures for EU nationals than LTIM

EU long-term international emigration, UK, year ending March 2012 to year ending March 2020

Notes:

1. RAPID data has been adjusted (see Section 3 for further detail on the adjustments).

2. The final two years of RAPID data represented by a dashed line are provisional (see Section 3 for further detail).

3. The LTIM data represent our current best estimates and preliminary adjustments have been made based on Department for Work and Pensions (DWP) and Home Office data.

4. Those aged under 16 years have been removed from the IPS component of LTIM data to align with the population covered by estimates from RAPID.

5. Confidence intervals are based on IPS estimates.

6. There are small differences between the annual time periods covered by each data source. Estimates from RAPID are tax years (ending 6th April) and LTIM estimates are year ending 31 March. Since these are broadly comparable time periods and for the purpose of clarity, for each data source we refer to a common annual period of 'year ending March'.

Download the data

.xlsx

When looking at estimates from RAPID data of long-term net migration alongside the LTIM estimates both sources are showing similar trends across the whole time series, where EU net migration has fallen since peak levels between the year ending March 2015 and the year ending March 2016 (Figure 4).

While the two sources are not directly comparable because of the methodological differences in measuring long-term migration, estimates from RAPID data suggest they are between 67,000 higher in the year ending March 2015 and 117,000 higher in the year ending March 2017 than those previously estimated by LTIM.

Up to the year ending March 2014, EU net-migration estimates from RAPID data are driven by a higher number of arrivals of EU8 and EU14 nationals than is estimated by LTIM data. After the year ending March 2014, estimates from RAPID data of EU net migration are driven by a higher number of arrivals of EU2. This is because of the employment restrictions placed on EU2 nationals being lifted in January 2014. For the year ending March 2019, EU net migration estimates from RAPID data is higher than the surrounding years; this is driven by EU14 nationals where both inflow increases and outflow decreases for the first time since the year ending March 2012.

The most likely cause of the difference between the two sources is the impact of EU nationals having a high degree of uncertainty in their intentions to move to and from the UK. This makes it much more challenging to measure migration using methods such as the IPS data, where our estimates are dependent on the information people give us about for how long they intend to move to and from the UK. However, administrative data sources such as in RAPID can estimate the actual observed patterns of behaviour of these migrants after they arrive in the UK.

Our previous research highlighted the uncertainty of EU8 nationals in their intentions to move to and from the UK and we therefore applied preliminary adjustments to the IPS estimates to address this. However, at that point there was not enough evidence around the migration patterns of EU14 or EU2 nationals to adjust these groups. Our research based on RAPID data suggests that the IPS data were not adequately capturing the migration patterns of these nationals.

Figure 4: RAPID estimates higher EU net migration figures than LTIM

EU long-term international net migration, UK, year ending March 2012 to year ending March 2020

Notes:

1. RAPID data has been adjusted (see Section 3 for further detail on the adjustments).

2. The final two years of RAPID data represented by a dashed line are provisional (see Section 3 for further detail).

3. The LTIM data represent our current best estimates and preliminary adjustments have been made based on Department for Work and Pensions (DWP) and Home Office data. Different types of lines have been used to represent where adjustments have and have not been applied. For LTIM estimates solid lines indicate adjustments have been applied (see note 2). Dashed lines indicate no adjustment has been made yet because of data availability.

4. Preliminary adjusted immigration and net migration LTIM estimates for EU8 citizens have been produced for the year ending December 2009 to the year ending March 2016.

5. Those aged under 16 years have been removed from the IPS component of LTIM data to align with the population covered by estimates from RAPID.

6. Confidence intervals are based on IPS estimates.

7. There are small differences between the annual time periods covered by each data source. Estimates from RAPID are tax years (ending 6 April) and LTIM estimates are year ending 31 March. Since these are broadly comparable time periods and for the purpose of clarity, for each data source we refer to a common annual period of 'year ending March'.

Download the data

.xlsx

Understanding non-EU migration data in RAPID

For non-EU nationals, both RAPID and LTIM data show a similar trend in the number of arrivals into the UK (Figure 5). While the two sources are not directly comparable because of the methodological differences in measuring long-term migration, estimates from RAPID data suggest they are between 4,000 lower in the year ending March 2014 and 52,000 lower in the year ending March 2018 than those previously estimated by LTIM data.

This is mainly driven by estimates from RAPID data showing a lower number of arrivals of Asian nationals than LTIM data, whereas the two sources show a very similar number of arrivals of migrants from the rest of the world. Data by nationality group can be found in the accompanying dataset.

Previous analysis of IPS data shows us that over half of arrivals of Asian nationals are arriving for formal study. Therefore, it is likely that the difference between these two sources is explained by the measurement of students where RAPID data do not adequately capture the volume of students arriving in the UK long term.

Figure 5: RAPID estimates a lower number of non-EU arrivals than LTIM

Non-EU long-term international immigration, UK, year ending March 2012 to year ending March 2020

Notes:

1. RAPID data has been adjusted (see Section 3 for further detail on the adjustments).

2. The final two years of RAPID data represented by a dashed line are provisional (see Section 3 for further detail).

3. The LTIM data represent our current best estimates and preliminary adjustments have been made based on Department for Work and Pensions (DWP) and Home Office data.

4. Those aged under 16 years have been removed from the IPS component of LTIM data to align with the population covered by estimates from RAPID.

5. Confidence intervals are based on the IPS data.

6. The IPS data component of LTIM data in year ending March 2017 and the year ending March 2020 saw changes in the number of non-EU citizens arriving to study which were not reflected in the most comparable Home Office student visa data. Therefore, the datapoints were revised to allow users to make more meaningful comparisons. It was not possible however to calculate confidence intervals for the revised estimates.

7. There are small differences between the annual time periods covered by each data source. Estimates from RAPID are tax years (ending 6 April) and LTIM estimates are year ending 31 March. Since these are broadly comparable time periods and for the purpose of clarity, for each data source we refer to a common annual period of 'year ending March'.

Download the data

.xlsx

When looking at estimates from RAPID of long-term outflow alongside the LTIM estimates, both sources are showing similar trends in the departures of non-EU nationals. Between the year ending March 2013 and the year ending March 2020, both sources show a decrease in the number of departures. From the year ending March 2017, the IPS trend is broadly stable where estimates from RAPID show a continued decrease. However, it is important to note that over this same time period the RAPID trend falls within the confidence intervals of the LTIM estimates (Figure 6).

Figure 6: RAPID estimates a higher number of non-EU long-term departures than LTIM

Non-EU long-term international emigration, UK, year ending March 2012 to year ending March 2020

Notes:

1. RAPID data has been adjusted (see Section 3 for further detail on the adjustments).

2. The final two years of RAPID data represented by a dashed line are provisional (see Section 3 for further detail).

3. The LTIM data represent our current best estimates and preliminary adjustments have been made based on Department for Work and Pensions (DWP) and Home Office data. Different types of lines have been used to represent where adjustments have and have not been applied. For LTIM estimates solid lines indicate adjustments have been applied (see note 2). Dashed lines indicate no adjustment has been made yet because of data availability.

4. The preliminary student adjustment to non-EU emigration and net migration LTIM data was applied from 2012/13 onwards.

5. Those aged under 16 years have been removed from the IPS component of LTIM data to align with the population covered by estimates from RAPID.

6. Confidence intervals are based on IPS estimates.

7. There are small differences between the annual time periods covered by each data source. Estimates from RAPID are tax years (ending 6 April) and LTIM estimates are year ending 31 March. Since these are broadly comparable time periods and for the purpose of clarity, for each data source we refer to a common annual period of 'year ending March'.

Download the data

.xlsx

When looking at estimates from RAPID of long-term net migration alongside the LTIM estimates, both sources are showing similar trends across the whole time series, where non-EU net migration has increased since the year ending March 2013 (Figure 7).

While the two sources are not directly comparable because of the methodological differences in measuring long-term migration, estimates of non-EU net migration from RAPID data suggest they are between 20,000 lower in the year March ending 2014, and 60,000 lower in the year ending March 2018 than those previously estimated by LTIM. This is mainly driven by the lower inflow estimates from RAPID data across the time series.

Our research has shown that one of the main drivers for the lower estimates of non-EU arrivals from RAPID data is the measurement of students migrating to and from the UK. This is because students who do not hold a NINo and students who do hold a NINo but do not work long term alongside their studies will not be included in RAPID data as a long-term migrant.

While we have adjusted estimates from RAPID using further administrative data to better capture the volume of student migration to and from the UK, we have taken a cautious approach to the adjustment at this stage. It is therefore likely that the adjustment does not fully address this coverage gap and that some student migrants are still missing from the estimates. Further understanding the migration patterns of students will form an important part of our future research to improve our admin-based migration estimates (See Section 9: Future developments for more detail).

Figure 7: RAPID estimates lower net migration than LTIM

Non-EU long-term international net migration, UK, year ending March 2012 to year ending March 2020

Notes:

1. RAPID data has been adjusted (see Section 3 for further detail on the adjustments).

2. The final two years of RAPID data represented by a dashed line are provisional (see Section 3 for further detail).

3. The LTIM data represent our current best estimates and preliminary adjustments have been made based on Department for Work and Pensions (DWP) and Home Office data. Different types of lines have been used to represent where adjustments have and have not been applied. For LTIM estimates solid lines indicate adjustments have been applied (see note 2). Dashed lines indicate no adjustment has been made yet because of data availability.

4. The preliminary student adjustment to non-EU emigration and net migration LTIM data was applied from 2012/13 onwards.

5. Those aged under 16 years have been removed from the IPS component of LTIM data to align with the population covered by estimates from RAPID.

6. Confidence intervals are based on IPS estimates.

7. The IPS data component of LTIM data in year ending March 2017 and the year ending March 2020 saw changes in the number of non-EU citizens arriving to study which were not reflected in the most comparable Home Office student visa data. Therefore, the datapoints were revised to allow users to make more meaningful comparisons. It was not possible however to calculate confidence intervals for the revised estimates.

8. There are small differences between the annual time periods covered by each data source. Estimates from RAPID are tax years (ending 6 April) and LTIM estimates are year ending 31 March. Since these are broadly comparable time periods and for the purpose of clarity, for each data source we refer to a common annual period of 'year ending March'.

Download the data

.xlsx

Understanding non-UK migration data in RAPID

Our work so far has focused on separately understanding what is driving the trends in the administrative data for EU and non-EU nationals. It is important that we further our understanding of the coverage of RAPID and develop our adjustments in each group before making a final assessment of the migration patterns of non-UK nationals as a whole.

Looking at EU and non-EU nationals combined (that is, all non-UK nationals) the levels and trends of net migration are similar when comparing estimates from RAPID data alongside LTIM estimates (Figure 8). This is because of the differences we see between RAPID and LTIM data for EU and non-EU nationals offsetting each other. The estimate from RAPID data for EU net migration is higher than the LTIM data estimate, while the estimate from RAPID data for non-EU net migration is lower than the LTIM estimate.

Figure 8: For the non-UK population, RAPID estimates and LTIM estimates show similar levels of net migration

Non-UK long-term international net migration, UK, year ending March 2012 to year ending March 2020

Notes:

1. RAPID data has been adjusted (see Section 3 for further detail on the adjustments).

2. The final two years of RAPID data represented by a dashed line are provisional (see Section 3 for further detail).

3. The LTIM data represent our current best estimates and preliminary adjustments have been made based on Department for Work and Pensions (DWP) and Home Office data. Different types of lines have been used to represent where adjustments have and have not been applied. For LTIM estimates solid lines indicate adjustments have been applied (see note 2). Dashed lines indicate no adjustment has been made yet because of data availability.

4. Preliminary adjusted immigration and net migration LTIM estimates for EU8 citizens have been produced for the year ending December 2009 to the year ending March 2016.

5. The preliminary student adjustment to non-EU emigration and net migration LTIM data was applied from 2012/13 onwards.

6. Those aged under 16 years have been removed from the IPS component of LTIM data to align with the population covered by estimates from RAPID.

7. Confidence intervals are based on IPS estimates.

8. The IPS data component of LTIM data in year ending March 2017 and the year ending March 2020 saw changes in the number of non-EU citizens arriving to study which were not reflected in the most comparable Home Office student visa data. Therefore, the datapoints were revised to allow users to make more meaningful comparisons. It was not possible however to calculate confidence intervals for the revised estimates.

9. There are small differences between the annual time periods covered by each data source. Estimates from RAPID are tax years (ending 6 April) and LTIM estimates are year ending 31 March. Since these are broadly comparable time periods and for the purpose of clarity, for each data source we refer to a common annual period of 'year ending March'.

Download the data

Notes for Results and analysis of RAPID data

1. While RAPID contains data for the tax year ending 2011, because of the methodology for calculating arrivals and re-arrivals, estimates of long-term migration can only be produced from the tax year ending 2012 onwards.

5 . Home Office border data and international migration

Data collected at the UK border by the Home Office have the potential to provide a direct measure of the movement of people in and out of the UK. Elements of these border data form part of a Home Office administrative dataset (also known as "exit checks"), which has been provided to the Office for National Statistics (ONS) for research purposes. Previous reports (from August 2017, July 2018 and January 2019) set out our journey to understanding these data more fully, including the caveats associated with using this complex dataset. Since 2016, the Home Office has also published an annual report on exit checks data.

Home Office border data currently only include nationals of non-European Economic Area (EEA) countries, and have only been collected from April 2015, allowing for migration estimates from the year ending 7 April 2017 onwards. Additionally, the data may not provide a complete account of border crossings as some people will enter or leave the UK via Ireland (Common Travel Area) and this is not captured in the border data.

In our previous research report, Exploring international migration concepts and definitions with Home Office administrative data published in February 2020, we proposed a method for how border data can be used to produce estimates of international immigration consistent with the UN long-term definition. As part of the development of admin-based migration estimates (ABMEs) we have built on this work to improve this method, apply it to a longer time series of data and assess how it could be developed further to provide measures of emigration and net migration in future.

Using border data to measure international immigration

The method looks at first arrival and last departure within a visa period as an approximation for length of stay in the UK. 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. Visits from non-visa nationals and those on long term visit visas are excluded.

The previous iteration of this method excluded records with an incomplete travel history other than first arrival or last departure. This was updated to include these records and focus on the first arrival and last departure travel events only.

The process of estimating long-term international immigration for any given 12-month reference period is illustrated in Figure 9, and uses a three-step process. The first step is identifying those people who have a visa period with a first arrival date within the reference period.

The second step is to use the time between the first arrival and last departure within a visa period to identify whether they have been resident in the country for 12 months or more (that is, whether they meet the usual residence threshold applied in the UN definition). This method means that short trips abroad over the course of an extended period of residence are excluded. If either the first arrival or last departure information is missing, then visa start or end dates are used as a proxy.

Lastly, the third step is to look at any previous visa period 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 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.

Figure 9: Illustrative example of identifying long-term international immigration using border data
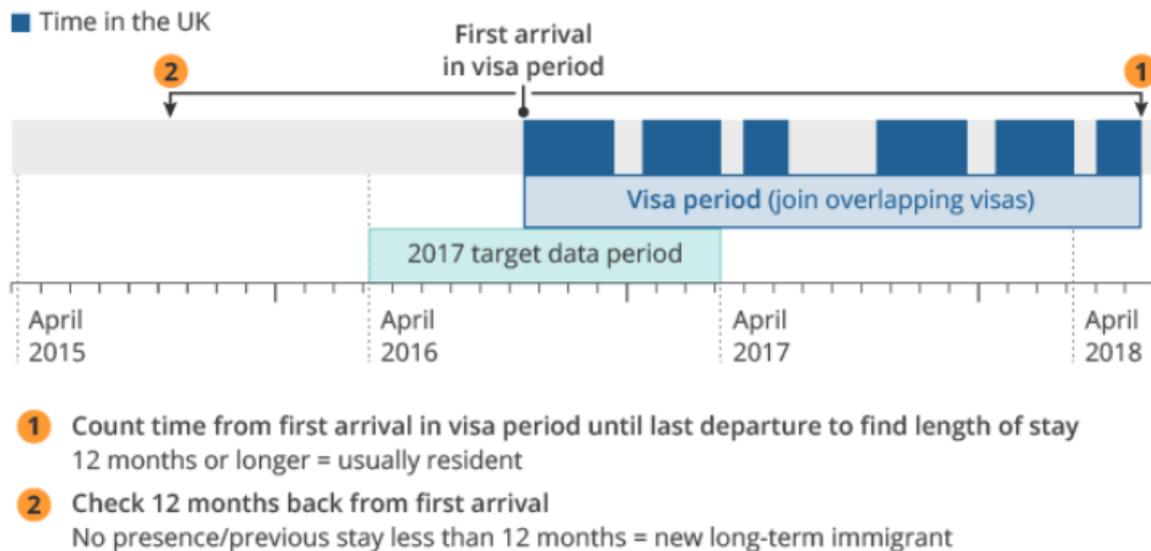
Source: Home Office

In order to produce final estimates for a given reference period (which in Figure 9 is 8 April 2016 to 7 April 2017), the method requires data to be available for both 12 months prior to the start of the period (to identify any previous residence in the UK) and 12 months after the end of it (to identify stays of 12 months or more for all individuals with a first arrival in the period).

Provisional estimates 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.

The Home Office has provided updated annual datasets to support the development and testing of this new methodology. Currently the data used cover the time period 8 April 2015 to 7 April 2019, which allows the production of indicative estimates of long-term international immigration for three time points: 2016 to 2017, 2017 to 2018, and provisional data for 2018 to 2019. These indicative estimates are presented in Figure 10, alongside comparisons with estimates based on the most similar data available in the RAPID and IPS-based LTIM datasets.

Figure 10: Indicative immigration estimates for non-EEA nationals derived from Home Office border data compared with RAPID and IPS-based LTIM estimates

Non-EEA long-term international immigration, UK, year ending March 2017 to year ending March 2019

Notes:

1. Border data estimate for year ending 7 April 2019 is provisional.

2. The LTIM data represents the current best estimate of migration using International Passenger Survey (IPS) data and covers the year ending March for each time point.

3. Confidence intervals are based on the IPS data.

4. The IPS component of LTIM data in the year ending March 2017 saw changes in the number of non-EU citizens arriving to study which were not reflected in the most comparable Home Office student visa data. Therefore, the datapoints were revised to allow users to make more meaningful comparisons. It was not possible however to calculate confidence intervals for the revised estimates.

5. RAPID and LTIM data estimates cover non-EU population.

6. Those aged under 16 years of age are included in the IPS component of LTIM in Figure 10 to allow for a more like-to-like comparison with Home Office border data which covers this age group. Figures in Section 4 excluded those aged under 16 years of age in the IPS component of LTIM to allow for comparison with RAPID data which do not cover those aged under 16 years of age (see Section 3 for further detail on coverage of RAPID data).

7. RAPID data is adjusted (see Section 3 for further detail on the adjustments

8. There are small differences between the annual time periods covered by each data source. RAPID uses tax years (ending 6 April), LTIM estimates use year ending 31 March, while Home Office border data use year ending 7 April. Since these are broadly comparable time periods and for the purpose of clarity, for each data source we refer to a common annual period of 'year ending March.'

9. This border data methodology produces indicative figures that show a similar pattern and trend for non-EU inflow as RAPID data, but suggest a higher overall level of immigration (around 374,000 compared with RAPID data estimates of around 292,000). Border data cover a broader population than RAPID data, which may explain some of this difference. For example, those who travel on a family visa but are unable to work or claim benefits will be covered in the border data, but will not be captured by RAPID data. Also, there may be students not captured by the student adjustment to RAPID data. Differences in how migrants are defined and identified also contribute to the difference in these indicative figures. For example, at this stage the border data methodology also has the potential to over-estimate the level of migration by including some people in the estimate who would not in practice meet the UN definition of a long-term migrant.

Download the data

.xlsx

These figures are a useful starting point and indicate the potential for using border data to estimate immigration alongside other administrative data sources. However, it is important to note that there is uncertainty around these preliminary numbers based on border data. Further work is necessary to fully understand the complexities of the data, investigate what these data can tell us about particular groups of migrants (for example, students and workers) and to develop and improve the methodology used.

Assessing the quality of border data-based estimates of immigration

As noted, the Home Office publishes annual reports on the quality of border data collected under the exit checks programme, improvements to data quality, and some statistical findings from the data. Quality indicators on the attributes of the data have been developed and are regularly reported on within these reports.

The ONS has also produced an [error framework for longitudinal administrative sources](#) using border data and we have used this approach to help identify the areas of the methodology for border data-based immigration estimates where further quality assurance or development is required.

6 . Data

[Estimating long-term international migration using RAPID](#)
Dataset | Released 16 April 2021
Data from our first iteration of our development of admin-based migration estimates (ABMEs) using the Registration and Population Interaction Database (RAPID).

7 . Glossary

Administrative data

Collections of data maintained for administrative reasons, for example, registrations, transactions, or record-keeping. They are used for operational purposes and their statistical use is secondary. These sources are typically managed by other government bodies.

Long-term international migration

"A person who moves to a country other than that of his or her usual residence for a period of at least a year (12 months), so that the country of destination effectively becomes his or her new country of usual residence."

RAPID

Registration and Population Interaction Database (RAPID) is a database created by the Department for Work and Pensions. It provides a single coherent view of interactions across the breadth of benefits and earnings datasets for anyone with a National Insurance number (NINo).

EU citizenship groups

EU estimates exclude British citizens. The following EU citizenship groups are used:

- EU14: citizens of countries that were EU members prior to 2004, for example, France, Germany and Spain

- EU8: citizens of Central and Eastern European countries that joined the EU in 2004, for example, Poland

- EU2: citizens of Bulgaria and Romania, which became EU members in 2007; between 2007 and 2013, these countries were subject to transitional controls restricting their access to the UK labour market; these restrictions were lifted on 1 January 2014

8 . Data sources and quality

In this first iteration of our development of admin-based migration estimates (ABMEs) we have presented analysis from the Registration and Population Interaction Database (RAPID). RAPID is a fully anonymised database created by the Department for Work and Pensions (DWP) for statistical purposes[1]. It provides a single coherent view of citizens' interactions across the breadth of systems in the DWP, HM Revenue and Customs (HMRC) and local authorities via Housing Benefit. RAPID data covers everyone with a National Insurance number (NINo) and for each person, the number of weeks of "activity" within these systems is summarised in each tax year from tax year ending 2011 to the most recent tax year available (currently tax year ending 2020).

The DWP developed the RAPID Migration Dataset to assist the Office for National Statistics (ONS) with the estimation of international migration to and from the UK. The RAPID Migration Dataset includes known first arrival dates and estimated departures and re-arrivals from the UK and categorises the arrivals as either short-term or long-term for each visit.

For more information on the creation of this dataset and the rules around categorising arrivals as long-term please see, Methods for using the administrative data source RAPID to measure international migration into and out of the UK.

Additionally, we have presented analysis using Home Office border data developed by the exit checks programme. The dataset used is derived from a linked database that combines data from Home Office systems to build travel histories that consist of an individual's travel in or out of the country, together with data relating to their immigration status, such as periods of leave granted. The data available for the ONS analysis cover travel events for nationals from non-EEA countries who had a valid period of leave between 8 April 2015 and 7 April 2019.

For more information on the methodology used to create this database, please see the Home Office user guide.

Notes for Data sources and quality

1. RAPID has been created to help generate insight that enables the DWP to formulate and improve employment and benefit policies so the department can react to changes in the UK economy and labour force.

9 . Future developments

Our admin-based migration estimates (ABME) development has so far focused on understanding the potential of two administrative data sources (Registration and Population Interaction Database (RAPID) and Home Office border data) to estimate migration into and out of the UK. We are making further steps in our journey towards a transformed migration statistics system based on administrative data. We will continue to ensure that the work to transform UK migration statistics aligns with our ongoing research to produce population statistics using administrative data. Migration statistics are an important component of estimating population change, and so we will ensure our research is integrated as we move towards a transformed population and migration statistics system in 2023.

We will be working closely with colleagues in the Home Office and the Department for Work and Pensions (DWP) and experts across the Government Statistical Service (GSS) to progress our analysis and understanding of both RAPID and Home Office border data. We are also collaborating closely with the National Records of Scotland (NRS) and the Northern Ireland Statistics and Research Agency (NISRA) to ensure that we can produce comprehensive UK migration statistics and produce statistics below the UK level.

Our immediate next steps in the development of ABME will focus on refining our current methods, extending our use of RAPID data to include UK nationals and working closely with the Home Office as their new migration system is developed. We will provide users with updates of this development in autumn 2021.

Looking beyond this, we plan to publish a further ABME publication in early 2022 which will cover the period up to April 2021, reporting data beyond the point of the last International Passenger Survey (IPS) estimates and covering the onset of the coronavirus (COVID-19) pandemic. It will also include our latest progress in the development of ABME methods. See the transformation overview for the revised milestones on our transformation journey.

The next steps in further developing our approach to measuring international migration using RAPID data will focus on:

- continuing development of the adjustments made to RAPID data, particularly those designed to take account of the under coverage of the student population; this will include looking at options for linking administrative sources such as Higher Education Statistics Agency (HESA) and Pay as you Earn (PAYE) to better understand how students interact with administrative data on earnings

- where our adjustments are based on past trends, we will look for the best possible evidence to adapt these assumptions over the period covering the onset of the coronavirus pandemic, where historical trends may not be indicative of recent behaviour

- further understand the impact of migrants who gain UK citizenship after arriving in the UK

- extend our use of RAPID data to cover UK nationals, initially applying similar methods to those developed to derive flows of non-UK nationals

- explore methods for measuring migration of those aged under 16 years assessing all available data sources

- explore what RAPID data can tell us about the non-UK population total and compare this against previously published estimates

The next steps in further developing our approach to measuring international immigration using Home Office border data will focus on:

- building our understanding of border data and refining our methods for producing immigration estimates

- extending the time series as more data become available

- investigating more detailed breakdowns of the data by characteristics of migrants

- exploring similar methods to measure long-term emigration

We are also collaborating with the Home Office to better understand how the new immigration system will inform our overall understanding of international migration. Since January 2021, EEA nationals are included in Home Office travel and visa data, though British and Irish nationals will continue to be a coverage gap in the data. Together we are rapidly progressing work to understand the opportunities offered by these data.

We will also continue to progress the work to develop models for estimating international migration for the period impacted by the coronavirus pandemic as part of our wider strategy for ABMEs. This will build our understanding of a period of uncertainty in survey and administrative data for which historical trends alone cannot provide reliable estimates. The initial findings of this work are presented in more detail in Methods for using the administrative data source RAPID to measure international migration into and out of the UK.

As part of our longer-term development of ABMEs, we will investigate our options regarding the frequency of reporting migration statistics. This will include considering whether modelling approaches can be used to provide more timely insights and account for the time needed to assess whether activities in the administrative data indicate long-term migration. We are also exploring how multiple data sources (including RAPID and Home Office border data) could be integrated, considering their relative strengths and limitations, to provide the best possible estimates of international migration.

We welcome your feedback on this update and on our transformation journey. If you would like to contact us, please email us at pop.info@ons.gov.uk.

10 . Related links

Methods for using the administrative data source RAPID to measure international migration into and out of the UK
Article | Released 16 April 2021
A summary of the current methods in using the administrative data source the Registration and Populations Interactions Dataset (RAPID) from the Department for Work and Pensions in estimating international migration.

Transformation of the population and migration statistics system: overview
Article | Updated 16 April 2021
Latest update on our population and migration statistics transformation journey.

International migration and mobility: what's changed since the coronavirus pandemic
Article | Updated 26 November 2020
Bringing together a range of data sources on international migration and mobility patterns since the coronavirus (COVID-19) pandemic.

Home Office Immigration statistics quarterly release
Report | Updated 25 February 2021
The Home Office quarterly and annual statistics relating to those: coming to the UK, extending their stay, gaining citizenship, applying for asylum, and being detained or removed, as well as immigration for work, study and family reasons.

National Insurance numbers allocated to adult overseas nationals
Report | Updated 25 February 2021
The Department for Work and Pensions quarterly publication of National Insurance number allocations to adult overseas nationals entering the UK.