

# Methods for measuring international migration using RAPID administrative data

A summary of the current methods in using the administrative data source the Registration and Population Interaction Database (RAPID) from the Department for Work and Pensions in estimating international migration.

Contact:  
Mike James  
pop.info@ons.gov.uk  
+44 (0) 1329 444661

Release date:  
16 April 2021

Next release:  
To be announced

## Table of contents

1. [Introduction](#)
2. [Overview of RAPID](#)
3. [Methods for using RAPID to measure international migration](#)
4. [Coverage of RAPID](#)
5. [Timeliness of estimates from RAPID](#)
6. [Adjustments to estimates from RAPID](#)
7. [Comparing estimates from RAPID to IPS estimates to assess the quality in measuring migration](#)
8. [Data](#)
9. [Glossary](#)
10. [Related links](#)

# 1 . Introduction

This report summarises the methods used to create admin-based migration estimates from the Registration and Population Interaction Database (RAPID). It is important that the information and research presented on these pages be read alongside the outputs to aid interpretation and avoid misunderstanding.

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. 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 ABME are likely to change and evolve as we explore these data sources in more detail.

## 2 . Overview of RAPID

The Registration and Population Interaction Database (RAPID) is created by the Department for Work and Pensions (DWP) to provide 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 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 the one that ended in 2011 to the most recent tax year available (currently the one that ended in 2020).

The following DWP, HMRC and local authority datasets currently feed into creating RAPID:

- Customer Information System<sup>1</sup>
- Migrant Worker Scan
- Pay As You Earn (PAYE) (Employments and occupational and private pensions)
- Self-Assessment
- State Pension
- Housing Benefit
- Child Benefit
- Income Support
- Pension Credit
- Tax Credits
- Universal Credit
- Personal Independence Payments
- Job Seekers Allowance
- Employment Support Allowance
- Attendance Allowance
- Disability Living Allowance
- Invalid Care Allowance
- Industrial Injuries Disablement Benefit
- Incapacity Benefit
- Severe Disablement Benefit
- Passported Incapacity Benefit
- Maternity Allowance
- Bereavement Benefit
- Widows Benefits

## **Notes for: Overview of RAPID**

1. Maintains changes in name, address, marriage (where necessary), dates of birth and death and NINo registration details.

### 3 . Methods for using RAPID to measure international migration

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.

To estimate international migration to and from the UK using Registration and Population Interaction Database (RAPID) data, the Department for Work and Pensions (DWP) and the Office for National Statistics (ONS) worked together to develop a methodology using the annual tax year summary datasets to create a version of RAPID including only Non-UK nationals (RAPID Migration Dataset). There are two main steps in the creation of the RAPID migration dataset which are important in understanding the methodology for identifying patterns of migration (both long-term and short term).

Firstly, use of information from the Migrant Worker Scan (MWS) - RAPID includes data from the MWS which identifies all non-UK nationals registering for a National Insurance number (NINo) from 1975 onwards. This gives us further information including, self-reported date of first arrival, NINo registration date, nationality at registration and previous country of residence. To ensure the most accurate information from the MWS is included DWP processed all MWS files from 2011 onwards to extract the earliest self-reported date of arrival.

Secondly, creating rules of residency in the UK - RAPID contains a combination of information on geographical location and activity with the underlying datasets as a proxy for inferring if someone was resident in the UK at any point during each tax year<sup>1</sup>. For someone to be classed as resident in the tax year they must have at least one week of interactions with tax or benefit systems. In addition, they must also have at least one geo-flag<sup>2</sup> showing they have a UK address. For those on State Pension, Bereavement Benefit or Widows Benefit this must not be paid abroad in a frozen rate country<sup>3</sup>. Non-UK nationals who have registered for a NINo will be counted as resident from the tax year of their self-reported date of arrival.

Using all this information DWP can create the RAPID migration dataset by removing UK nationals and determining whether non-UK national records within the dataset are either short-term or long-term. For each non-UK national record, the RAPID migration dataset contains one row for each tax year since first arrival in the UK, containing a summary of the activities for employment (including self-employment), DWP and HM Revenue and Customs (HMRC) benefits, housing benefit and pensions as well as the self-reported date of first arrival in the UK. This longitudinal data allows us to assess patterns of interactions over time. Records are then categorised as either long-term or short-term by looking for patterns of interactions with the tax and benefits system. Our analysis has so far focused on long-term migration.

#### Identifying long-term migrants in the RAPID data

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 into the UK. To determine 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” within DWP and HMRC datasets. There are a series of steps taken in processing the data which form the main components in determining if arrivals are long-term. The main components to this processing are:

- identifying migrants - migrants are identified from the MWS where any person registering for a NINo with a non-British nationality at the point of registration will be included
- identifying first arrivals and registrations - first arrival and NINo registration date are obtained from the MWS
- amalgamate all tax year datasets into a single dataset - in order to estimate a person's activity over time, all separate tax year data files in RAPID are amalgamated into a single dataset holding all tax years from 2010, or for those arriving after 2010, the tax year of first arrival
- activity - RAPID captures interactions with DWP, HMRC and local authority systems and calculates the length of each interaction in the tax year (the number of weeks); these interactions therefore show that that person is "active" within the source systems and we therefore use this to show "activity" within the administrative data

To estimate a person's total activity in the year, activities that cannot occur at the same time, such as employment and an out of work benefit, are added together. Other interactions may be allowed to overlap and therefore the maximum activity value is used. This gives a total number of weeks as an estimate of a person's total activity in the tax year. The number of weeks between the first arrival date and NINo registration date are also calculated. During this period the person did not have a NINo and therefore there are unlikely to be activities on the source systems covering this period.

Using all this information we can make an estimation of whether arrivals to the UK are long-term or short-term based on their activity profile in the RAPID data. All our research using administrative data so far has shown that people's lives are complex, therefore we have created multiple categories of long-term interactions to account for this complexity.

We have 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. It is important to note that the information contained within the RAPID data only specifies the total number of weeks of interactions within each tax year and does not specify that this activity is continuous, however, we have assumed the total activity measured to be sufficient to indicate long-term presence in the UK. These two categories make up the largest proportion of long-term arrivals in the RAPID data (over 90%). We have also included two further categories that expand on this definition of long-term activity, in order to reflect the complexity of people's lives, although it is important to note that each these groups only make up a small proportion of arrivals.

Category 1: 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.

- it is assumed that the 52 weeks of activity over the two-year period is consecutive, however, the RAPID data does not show when in the tax year this activity was; even if this activity is not consecutive it is assumed that the person is resident across that whole time period
- Category 1 arrivals make up 80% of all non-UK long-term arrivals in the RAPID data

Category 2: 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.

- in most cases the period between arrival and registration is less than 52 weeks, however in some instances there can be a [long lag between arrival and registration](#) and there are many reasons someone may not need a NINo when they first arrive, for example, if they are a student, but then later on register for a NINo in order to work
- Category 2 arrivals make up 10% of all non-UK long-term arrivals in the RAPID data

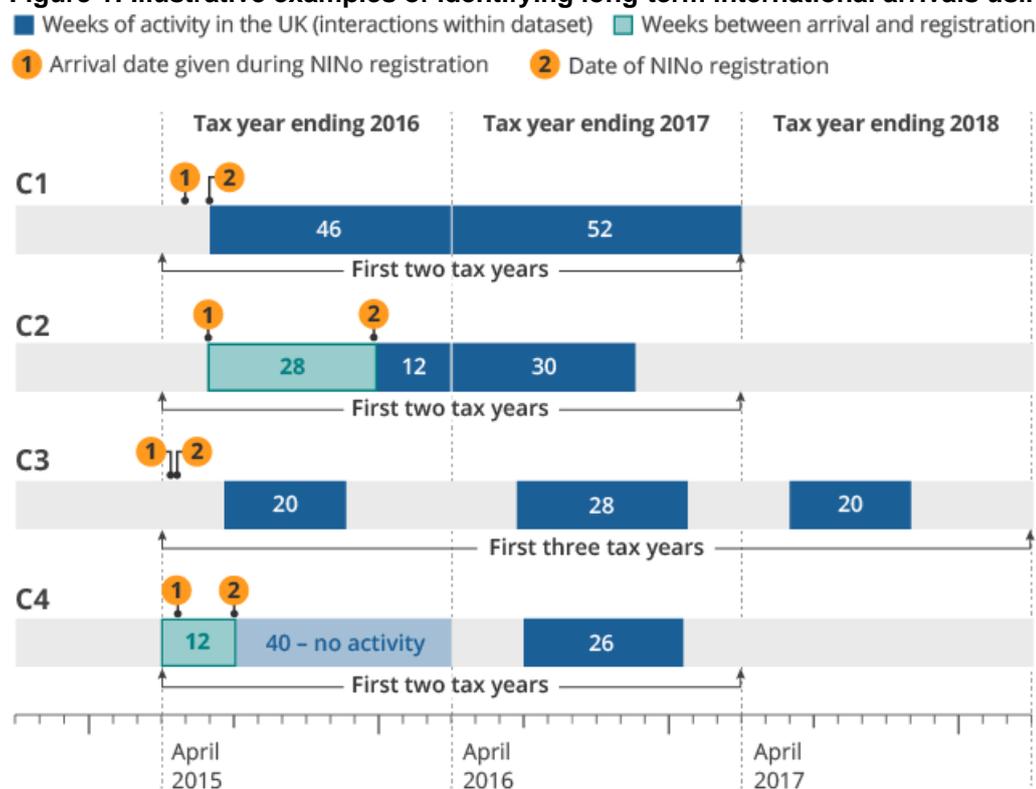
Category 3: activity occurred in three consecutive years from registration, (where registration is counted as an activity), and where the 52-week activity criteria is not met. However, the presence of activity across multiple tax years suggests they are a resident long-term.

- Category 3 arrivals make up 9% of all non-UK long-term arrivals in the RAPID data

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 1 is over 52 weeks in total, where there must be at least one week of activity in the registration year plus 1.

- it is assumed that the person is resident from registration; there are many reasons why someone may not start showing activity straight after registration for example those who are looking for a job
- Category 4 arrivals make up 1% of all non-UK long-term arrivals in the RAPID data

**Figure 1: Illustrative examples of identifying long-term international arrivals using RAPID data**



Source: Department for Work and Pensions – Registration and Population Interaction Database (RAPID)

**Notes:**

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

These categories are applied in a hierarchical order, and once categorised, an arrival is not re-categorised. If someone does not fit into any of these four categories, they are assumed to be a short-term migrant.

## **Outflow**

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 data, either through claiming benefits or through their earnings or pension. Therefore, to measure long-term emigration we need to determine individuals who no longer have activity in the RAPID 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 data are counted as a long-term emigrant<sup>4</sup>.

## **Re-arrival measure**

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. For example, a non-UK national arrived in 2004, was issued a NINo, worked in the UK for 2 years before leaving then subsequently re-arrived in the UK in 2015 and started working again. In this scenario RAPID would classify them as a re-arrival of a non-UK national in 2015.

## **Data limitations of RAPID in measuring long-term migration**

### **NINo registrations via Home Office visa application route**

Since October 2018, the Home Office has been allocating NINos (via DWP) as part of the visa application process for Tier 2 and Tier 5 visas for non-EU nationals. This means the NINo is issued before the person arrives in the UK, and it is possible that they may never arrive. Caution needs to be applied to the current methodology of counting the period between registration and arrival for non-EU nationals. As of January 2021, this is likely to be also the case for EU visa applications, therefore we will be investigating the impact of the change to the NINo registration process in future iterations of this methodology.

### **Self-employment**

The available self-employment data that feeds into RAPID are incomplete for tax years up to 2013, partially complete for the tax year ending 2014, and incomplete for the most recent tax year (ending 2020). Where someone's only interactions with the sources that feed into RAPID are through self-employment this would lead to migrants appearing to have no activity. However, it is likely this activity has continued but the data are not yet available. The methodology applied here uses recorded self-employment data as a proxy measure for the years that are not fully populated. Where a self-employment record is present this person is allocated 52 weeks of activity in the tax year. When the self-employment data is available in the following year, RAPID will be updated with this activity.

### **Benefit activities**

People arriving in the UK from non-EU countries have no access to income related benefits for the first 5 years or until they have acquired indefinite leave to remain. This will affect certain migrant footprints which may have an impact on whether they are determined to be a long-term migrant or not. Since January 2021, this may also apply to newly arriving migrants from the EU.

- Inactivity whilst resident – there are some instances where a period of inactivity may be expected and may not mean someone has left the UK. For example, when the only interactions within RAPID are from child benefit and this comes to an end because the youngest child has reached the age where child benefit stops. Residency rules have been created to identify this type of scenario. These rules keep both the parent and child resident and stop a departure being generated.

## Measuring migration during the coronavirus pandemic

The latest RAPID data covers the period April 2019 to March 2020 therefore does not cover the onset of the coronavirus (COVID-19) pandemic. As part of our future development of this work we will be considering how we can use this data source to understand migration patterns during the coronavirus pandemic.

### Notes for: Methods for using RAPID to measure international migration

1. This will include both long-term and short-term residents at this stage.
2. RAPID contains three geo-flag indicators: first address in the tax year, last address in the tax year and longest address in the tax year. This information comes from DWP's Customer Information System.
3. The UK State Pension is payable overseas but is not increased ("uprated") annually unless there is a legal requirement to do so, for example, where there is a relevant reciprocal social security agreement between the UK and the person's country of residence. The same rules apply for Bereavement Benefit and Widows Benefit. For more information see [Frozen Overseas Pensions](#).
4. There are some instances where inactivity may not be an indicator of emigration. For example, where a child benefit claim ends due to the child reaching the age where the benefit is no longer payable. RAPID applies rules in these instances to keep both the child and the adult in receipt of the child benefit claim as resident to stop them appearing as an outflow in the data. For the child this is rolled forward until the child is 20 years old or has another activity in the dataset. For the adult in receipt of the child benefit claim the rule is rolled forward until retirement age or has another activity.

## 4 . Coverage of RAPID

As the Registration and Population Interaction Database (RAPID) covers everyone with a National Insurance number (NINo) it includes UK nationals and migrants from EU and non-EU countries. Anyone 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 due to the wide range of data sources included, however, there are some populations where activities with the source datasets are less well covered, where a NINo would not have been issued and some populations that have been removed for the purpose of measuring international migration at this time.

### UK nationals

As RAPID contains a record for everyone with a NINo it includes both UK and non-UK nationals. In the creation of the RAPID migration dataset we currently remove UK nationals. Exploring the use of RAPID to estimate the migration patterns of UK nationals is an important part of our future development of admin-based migration estimates (ABMEs).

## Migrants under the age of 16 years

Migrant children under 16 years of age are not included in RAPID data. Children who arrive into the UK do not need to register for a NINo in the same way as adults. They are not recorded or captured by the Migrant Worker Scan (MWS) (which identifies non-UK nationals registering for a NINo from 1975 onwards). Whilst Child Benefit data are contained within RAPID, it does not provide any evidence of the nationality of the child and is not suitable for the analysis of migration into or out of the UK. Therefore, any records relating to those under the age of 16 years have been removed from the dataset and the RAPID data presented in the accompanying report concentrates on people age 16 years and over. As part of the future development of ABMEs we will be evaluating alternative data sources for estimating the migration patterns of those under the age of 16 years.

## Students

Migrants who come to the UK with the sole purpose of studying may not be included in RAPID data. As RAPID contains records for anyone with a NINo, students who do not hold a NINo will not be included in RAPID data at all. Any students who do hold a NINo will be included in RAPID, however may not be identified as resident in the UK if they do not undertake any activity that verifies residency, for example, some form of work alongside their studies. In addition, students who do work alongside their studies may not have enough activity to class them as a long-term migrant if they only work during term time. See the adjustments to estimates of RAPID section for information on our preliminary method to account for the coverage of students in RAPID data.

## Economically inactive

There are some instances where non-UK nationals could become inactive in the dataset whilst still living in the UK. For example, someone who was previously working but stops working in order to look after children or a family member. In some instances they could be eligible to claim [Child Benefit](#), however, some migrants do not have [recourse to public funds](#) so would be ineligible to claim other types of benefit. In these instances, if there was no activity for a whole tax year the rules in the data would show this person to outflow from the UK. In addition any non-UK nationals who move to the UK with their spouse or other family members but do not work or claim benefits will either be excluded from the dataset (if they do not have a NINo), or will be included but not determined to be a long-term migrant (if they have a NINo but have no activity).

## Migrants who gain UK citizenship

In addition to the gaps in RAPID data, there are also some differences in how RAPID categorises records as non-UK nationals compared with other data sources such as the International Passenger Survey (IPS) and Home Office border data. RAPID identifies non-UK nationals 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 classed as a migrant. RAPID does not currently include any data on applications for UK citizenship or indefinite leave to remain and therefore does not update nationality for these people. Consequently, RAPID classifies everyone under their nationality at registration even if they have subsequently gained UK citizenship. Therefore, these people would be counted as outflow of a non-UK national even though they have subsequently gained UK citizenship. See the adjustments to estimates of RAPID section for information on our preliminary method to account for migrants who gain UK citizenship in RAPID.

## Differences in the coverage of migrant populations between RAPID data and Long-Term International Migration (LTIM) estimates derived from IPS

There are some differences in the coverage of specific populations when comparing RAPID data to the IPS data. These are detailed below including information on how we have improved the comparability of estimates presented in our accompanying report.

## UK Nationals

RAPID contains information for everyone with a National Insurance Number (NINo), therefore will also include UK Nationals. However, in creating the RAPID migration dataset UK Nationals are removed.

The IPS estimates migration for UK and non-UK Nationals. For the purpose of analysis presented here and in the main report, UK Nationals have been removed.

## Migrants under the age of 16 years

Migrant children under 16 years of age are not separately identifiable in RAPID data. 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. Child benefit records enable the linking of a parent (or carer) to a child, however, it is not known from the records where the child was born or the nationality of the child or other parent. If the linked parent is a migrant, then it might be that the child is also a migrant, although this will not always be the case. Therefore, these records have been excluded for the estimation of migration.

Children under 16 years are eligible to be interviewed on the IPS. If the sampled person is under 16, where possible the interview is carried out with the child after having first received permission from a parent, guardian or responsible adult travelling with them (for example, a teacher if they are on a school trip). If the child is too young to complete the interview themselves, proxy information is collected from the parent, guardian or responsible adult, wherever possible. For the purpose of analysis presented here and in the main report those under the age of 16 have been removed from the IPS element of the long-term international migration (LTIM) estimates.

## Asylum seekers

Asylum seekers will not be included in RAPID data until their asylum application has been granted. At the point of asylum being granted they will be allocated a NINo and can therefore interact with the benefits or employment systems.

Persons arriving as asylum seekers or as refugees on resettlement schemes are identified and removed from IPS data. However, one of the components of LTIM is an adjustment for asylum seekers and refugees based on Home Office administrative records.

## Dual nationals

As previously stated, RAPID identifies non-UK nationals using the MWS. The MWS only includes people who present themselves at the point of registration as non-British. It therefore excludes people who have never been in the UK previously but who have dual nationality and present themselves at registration with British documentation.

Citizenship is taken from the passport shown at the time of the IPS interview, or (if this is not available) taken as stated by respondent. Therefore, dual nationals will be recorded by whichever nationality is on the passport they present.

## Migrants accompanying or joining family or friends

Some spouses of migrants may not yet have registered for a NINo, because they have not interacted with anything that requires a NINo. For example, anyone who is living on their partner's income may not be in the RAPID data. Alternatively, those who previously have worked but stop working, for example to look after family and are supported by their partners income would appear in RAPID data but would not be classed as resident in the year if there is no income generating activity.

The IPS asks for a person's main reason for migration. One of the options is accompanying or joining friends or family.

## Migrants of pension age

Migrants claiming pension from their home state will not be captured in RAPID data. If the person is resident in the UK and not registered for NI or benefit purposes they will not be in the RAPID data. RAPID will not count people as resident if their only activity and income is derived from investment income, as this is not included in the self-assessment data extract used in RAPID.

The IPS sampling frame is based on all arrivals into the UK regardless of their reason for migration and age

## 5 . Timeliness of estimates from RAPID

In addition to the population coverage gaps in the Registration and Population Interaction Database (RAPID) there is also the challenge of the timeliness of the data. 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 after arrival and consequently, will not be present in the administrative data until they do. Analysis comparing date of registration and date of self-reported arrival suggests that around 80% of migrants arrive in the UK and register for a National Insurance number (NINo) in the same tax year, around 17% register in the following tax year and around 3% take over two years to register for a NINo. There are many reasons why someone may not register for a NINo immediately for example, if you have a residence permit that allows you to live in the UK, it may include the condition that you have [no recourse to public funds](#). If so, it means you will not be able to claim most benefits, tax credits or housing assistance that are paid by the state. Alternatively, if you come to the UK to study you may not need a NINo unless you decide to work alongside your studies, or your university course involves a placement within the workforce.

We also need to take account of the time needed between someone interacting with the administrative data sources and being categorised as a long-term migrant in the dataset. As this administrative data source is based on interactions with administrative systems, it takes at least 12 months of activity to have taken place before a long-term measure of duration can be identified and, in most cases, more than 12 months of data are needed. As RAPID is an annual dataset, if someone arrives in the UK halfway through a tax year, we will need data covering two tax years in order to identify at least 52 weeks of activity to categorise someone as a long-term migrant. See the adjustments to estimates of RAPID section for information on our adjustment to account for the time needed to assess whether an arrival is long-term in RAPID.

In addition to this there are several timeliness characteristics of the administrative data that generates the RAPID dataset. RAPID provides an annualised view of the activities in the tax year and is generated once the tax year is completed. This initial dataset is a provisional assessment of some of the activities as it can take time for information to be submitted by employers or individuals and recorded on some administrative systems. RAPID is refreshed six months after the end of the tax year to create a final version of the data, which captures any information that was submitted to systems late or was missing from the provisional version<sup>1</sup>.

## Notes for: Timeliness of estimates from RAPID

1. The deadline for submitting self-assessment returns is the January after the end of the tax year (for example in January 2021 returns were due for the 2019/20 tax year). Therefore, when RAPID is refreshed 6 months after the end of the tax year final self-assessment data is not available for the latest tax year. We therefore roll forward information from the previous tax year as outlined in Section 3. In addition, as part of this refresh six months after the end of the tax year we include all update to the self-assessment data, including late submissions from previous tax years.

## 6 . Adjustments to estimates from RAPID

To address the coverage gaps identified and take account of the time needed to assess whether people are long-term migrants, we have applied a series of adjustments to the estimates derived from the Registration and Population Interaction Database (RAPID) using other available administrative data. These are preliminary adjustments and as we further our understanding of the RAPID data and as other new data sources become available, we will continue to refine these adjustments and will reflect these in future research outputs.

We have also outlined the assumptions we have made for each of these adjustments, the limitations of the adjustments and any future improvements we will look to make.

For indicative numerical examples of each of the adjustments made to the estimates from RAPID see accompanying datasets.

The adjustments currently applied to the estimates from RAPID are based on the most appropriate and available administrative data and use previously published research. They have been applied to demonstrate how alternative administrative data can be used to fill the gaps in RAPID. However, 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.

### Student inflow

The aim of our student inflow and outflow adjustments is to account for the under-coverage of students in the dataset. As RAPID uses information on interactions with the benefits and earnings datasets to estimate migration into and out of the UK, any students who do not work alongside their studies will not be identified as a long-term migrant in the dataset.

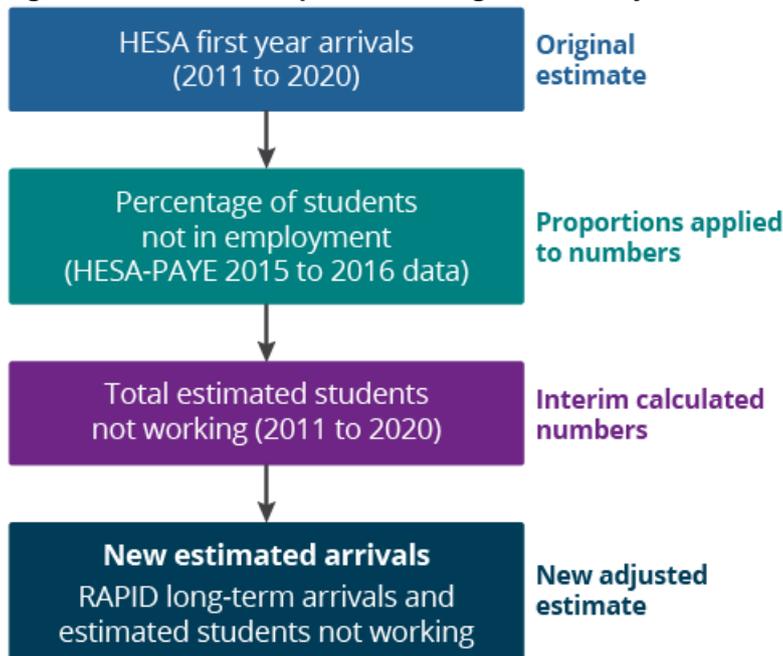
The student inflow adjustment uses data from Higher Education Statistics Agency (HESA) broken down into year of arrival, length of study and nationality group. We have [previously published research](#) linking together HESA with data from HM Revenue and Customs (HMRC) Pay as You Earn (PAYE) to inform us about employment and economic activity of international students in Higher Education. We have used this information to inform the adjustment to estimates from RAPID.

Using the HESA data, we calculate the number of first year enrolments for students studying undergraduate courses, by country of previous domicile. This allows us to estimate the total first-year student inflow. Here we only include students who are on a course that lasts over 1 year to ensure we are capturing students who fit the definition of a long-term migrant.

The HESA and PAYE research provides an estimate of the proportion of students who were not in any employment during either tax year overlapping with the academic year ending in 2016, therefore uses PAYE data covering the two tax years ending in 2016 and 2017 by nationality group. This research has only been completed for this time period, so we have applied this same proportion across the whole timeseries. By applying these proportions to the HESA first year inflow figure, we can estimate number of first year students not working, therefore the number of students who are likely to not be captured by RAPID.

Finally, to calculate the adjusted RAPID arrivals in each year, the total number of students not working is added on to the total RAPID figure. This is calculated for each year and for each nationality group (for example, EU2, Asia, Europe Other).

**Figure 2: Worked example calculating the new adjusted student arrivals**



Source: Office for National Statistics

## Student outflow

As we have applied a student inflow adjustment, it is important for net migration that an outflow adjustment is also made. For the student outflow adjustment, we have used:

- data from HESA, broken down into year of arrival, length of study and nationality group
- the Student Not Working figures calculated for the inflow adjustment
- Home Office Border data and the Longitudinal Educational Outcomes (LEO) data from the Department for Education

Firstly, using the HESA data on length of study we can calculate the proportions of first-year students who expect to leave the UK after a year, after three years (including students on courses two to three years in length), and after four years (including all students with course lengths of four years and over), for each nationality group and each inflow year. These proportions are applied to the relevant Students Not Working estimate calculated in the inflow adjustment. This estimates the number of students not working who are expected to have finished their course and therefore outflow for each year and each region between the academic year ending 2015 and the academic year ending 2020. As an example, to calculate the expected total outflow in the academic year ending 2017 we sum together the one-year outflow from the academic year ending 2016, the three-year outflow from the academic year ending 2014 and the four-year outflow from the academic year ending 2013<sup>1</sup>.

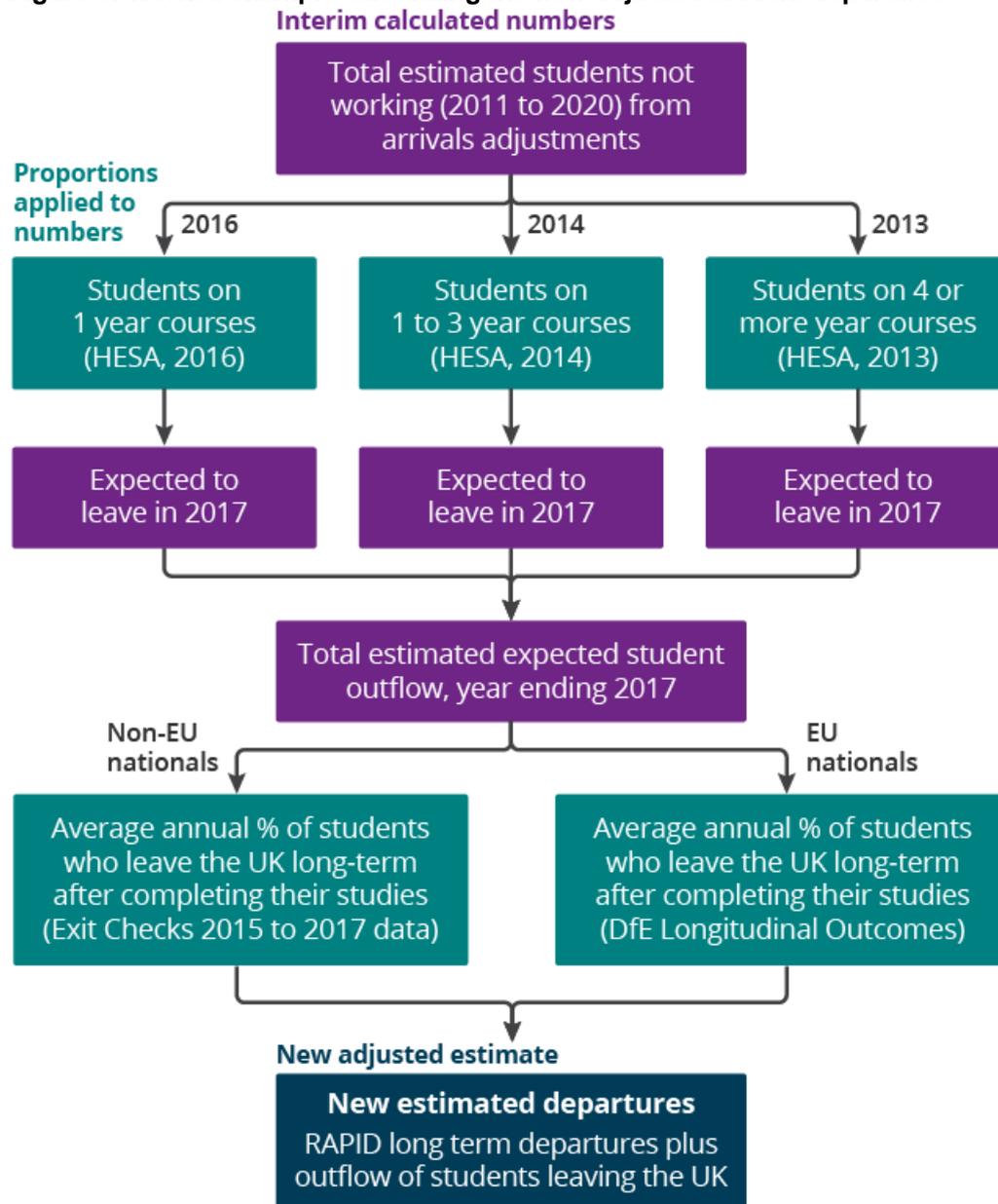
Previous research has shown that [not all students leave the UK at the end of their studies](#). Therefore, for the non-EU regions, we use Home Office Exit Checks data to identify the proportion of students who departed the UK long-term after completing their studies. This was done for both the year ending 7 April 2016 and year ending 7 April 2017 data, and an average of the two years was taken. This average proportion was then applied to the total number of students expected to outflow calculated above. This then provides a final estimate for the number of students who outflow in each non-EU region from the tax year ending 2015 to the tax year ending 2020.

Home Office Exit Checks data only covers non-EU nationalities. Therefore, for the EU regions, we used the [Graduate outcomes \(LEO\): 2017 to 2018](#), this provides information on employment and earnings outcomes of higher education graduates in England. We use this data to calculate the average proportion of EU students who departed long-term after graduation between the tax years ending 2015 and 2018. The LEO graduate outcomes data splits graduating students into the following groups looking at their outcomes a full tax year after graduation: unmatched; activity not captured; no sustained destination; sustained employment only; and further study. As we are estimating the proportion who are no longer in the UK, we have summed those who are unmatched, whose activity is not captured and those with no sustained destination. These groups were used as the likely reason for inactivity is that they have left the UK. For more information on the data matching and quality of this dataset please see the [Graduate Outcomes methodology report](#).

This average proportion was then applied to the total number of students expected to outflow calculated above. This then provides a final estimate for the number of students who outflow in each EU region from the tax year ending 2015 to the tax year ending 2020.

HESA arrivals data are only available from the academic year ending 2011, therefore it is only possible to calculate the subsequent departures from the year ending March 2015 onwards. In order to estimate outflow for the years prior to this we had to extrapolate the trends backwards. For each region, we then calculated the average proportion of international student inflow who outflowed in each year between the tax year ending 2015 and the tax year ending 2020. This average proportion was applied to the students not working figures between the tax year ending 2011 to the tax year ending 2014, where the HESA data are not fully complete in order to work through the same methodology. This gives us the estimated outflow figures for each region from the tax year ending 2011 to the tax year ending 2020.

**Figure 3: Worked example calculating the new adjusted student departures**



Source: Office for National Statistics

### **Provisional inflow in the latest tax years where not enough data are available to assess whether arrivals are long-term**

The aim of this adjustment is to account for the time needed in administrative data to assess whether an arrival is long-term. As RAPID is based on actual observed patterns of behaviour, it takes at least 12 months of activity to have taken place before a long-term measure of duration can be identified. And in most cases, more than 12 months of data are needed. Therefore, for the tax years ending in 2019 and 2020 not enough data is available to observe long-term interactions in the dataset.

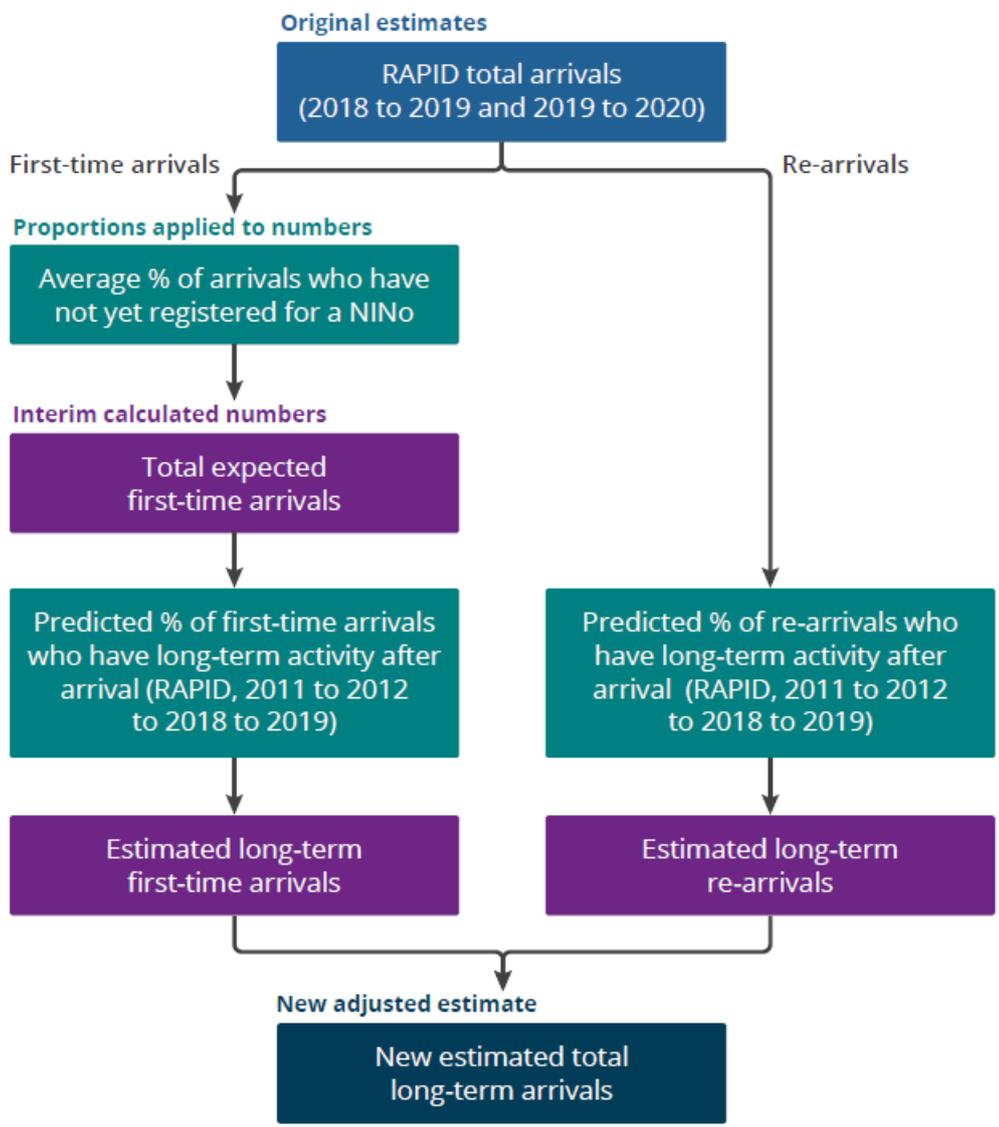
For this adjustment, the proportion of total arrivals that fall into each category in each year was calculated. The proportions for first arrivals and re-arrivals were separately calculated. For example, the proportion of first arrivals in each arrival category (C1, C2, C3, C4, short-term), and the proportion of re-arrivals in each re-arrival category (C1, C3, short-term). This was calculated for each year between tax years ending 2012 and 2018 (the last point final data are available). From these proportions, we calculated a three-year average to predict the proportion for the final two years.

We completed further analysis of RAPID to identify the average length of time between arrival and registration where it was found around 17% of migrants register the year after arrival and around 3% register two years after arrival. Therefore, the most recent two tax years are adjusted to account for those who have arrived but not yet registered. This is applied to first time arrivals only as re-arrivals do not need to re-register for a NINo so this same delay will not apply.

The three-year average proportion calculated previously is then applied to the adjusted total number of arrivals for the latest two years. This gives us the total number of arrivals in each category which allows us to calculate the estimated number of projected long-term arrivals.

In future iterations of RAPID we will have a further tax year of data and therefore will be able to remove the adjustment and use the final RAPID data. For example, we have currently applied this adjustment to inflow for tax years ending in 2019 and 2020. When the RAPID migration dataset is available for the tax year ending 2021, we will be able to remove the adjustment for the tax year ending 2019 and use the actual observed interactions.

**Figure 4: Worked example calculating the provisional arrivals in the latest tax years**



Source: Office for National Statistics

## Provisional outflow in the latest tax year where no data is currently available

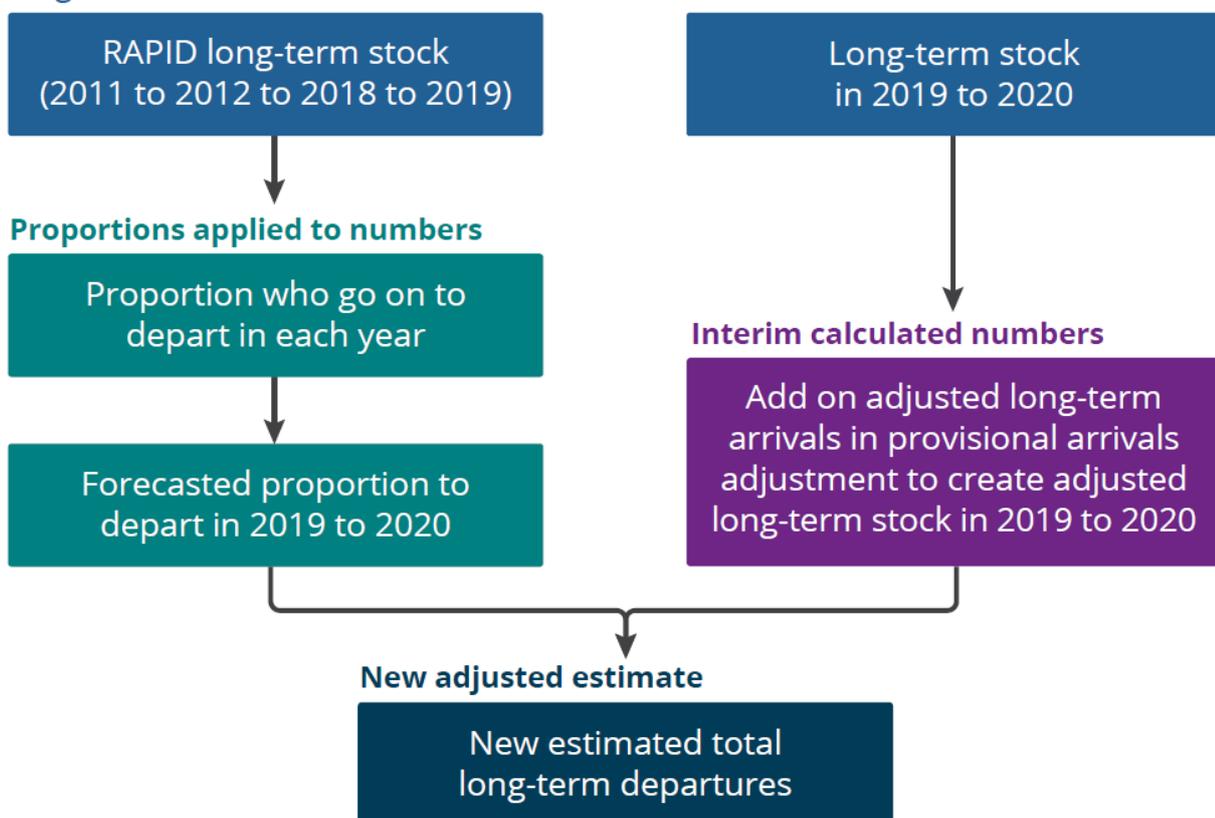
For this adjustment, for each year between the tax years ending 2012 and 2018 the proportion of long-term stock that go on to outflow is calculated using the long-term departures figure. Using the historical proportions calculated we calculate the predicted proportion for the final year.

We also adjust the total long-term stock by adding on the long-term inflow adjustment calculated in the provisional inflow adjustment. To this adjusted stock we then apply the forecasted proportions calculated above for the final year, which creates the new outflow estimate for the final year.

In future iterations of RAPID we will have a further tax year of data and therefore will be able to remove the adjustment and use the final RAPID data. For example, we have currently applied this adjustment to outflow for the tax year ending 2020. When the RAPID migration dataset for the tax year ending 2021 is available, we will be able to remove the adjustment for the tax year ending 2020 and use the estimates from RAPID.

**Figure 5: Worked example calculating the provisional departures in the latest tax year**

### Original estimate



Source: Office for National Statistics

## UK naturalisation adjustment

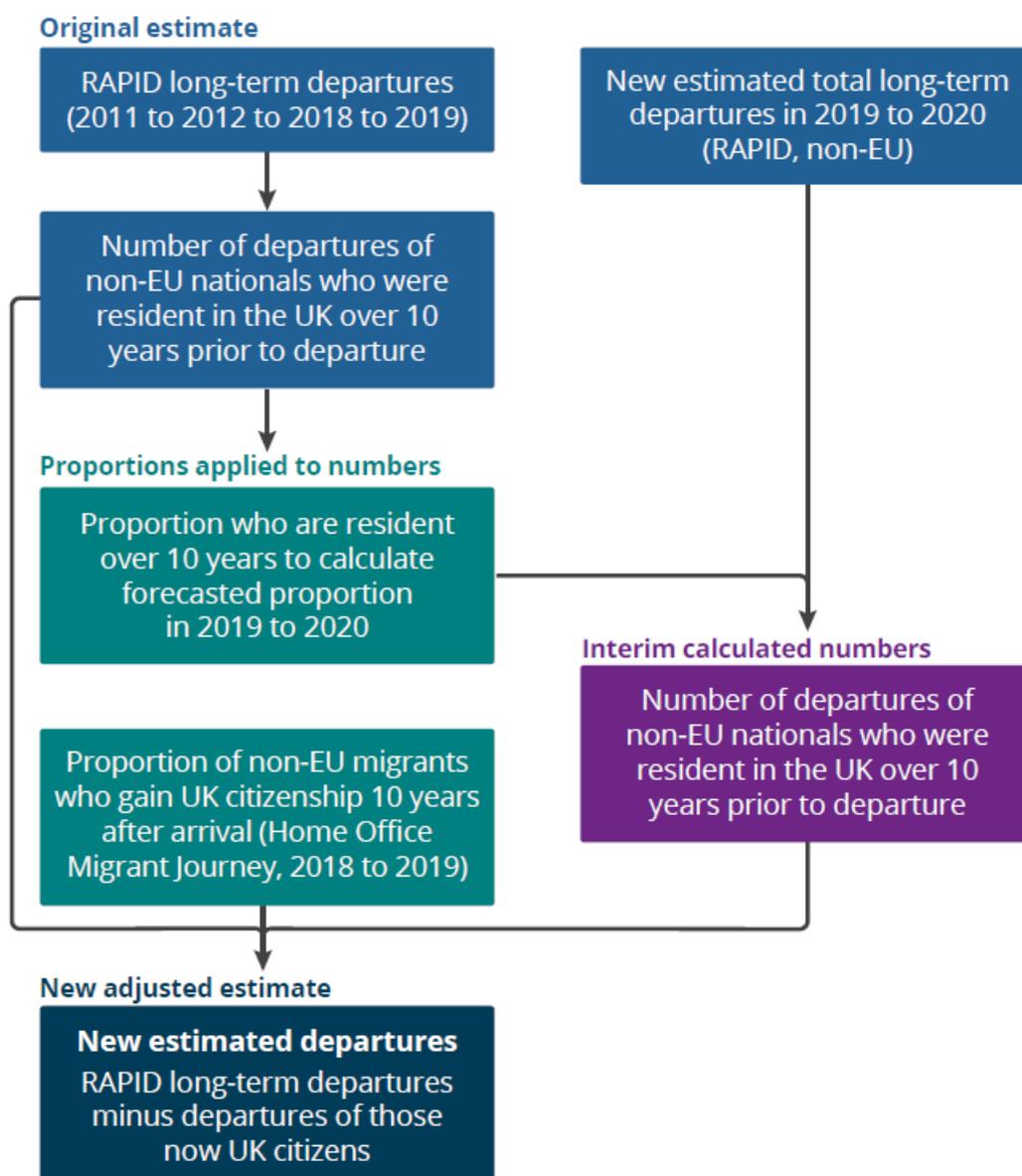
This adjustment uses both RAPID outflow data and data from the [Home Office Migrant Journey](#) data. The Home Office data provides information on the proportion of non-EU nationals that gain UK citizenship within a set period of time. We have focussed the adjustment on the outflow of non-EU nationals who have been resident in the UK for at least 10 years prior to outflow. Therefore, we have used the Home Office data to estimate the proportion of non-EU nationals who have gained UK citizenship within 10 years of their visa being issued.

Using data from RAPID between the tax years ending 2012 and 2019 we can determine the number of departures who have been resident for 10 years prior to departure. For departures in the tax year ending 2020 we estimate this using the proportion of total departures who were resident in the UK for over 10 years prior to departure between the tax years ending 2012 to 2019. The historic proportions are then used to predict what proportion for the latest tax year. The provisional outflow adjustment shows the total estimated number of departures for the latest tax year, by applying the predicted proportion we can estimate the total number of departures who have been in the UK over 10 years prior to departure.

Using both the actual and predicted number of departures who have been resident in the UK over 10 years prior to departure we can apply the proportions from the Home Office Migrant Journey analysis showing the proportion who will have since gained UK citizenship.

The outflow of these UK citizens has been removed from the non-UK outflow total, however as we develop our approach for estimating international migration of UK nationals, we will include the migration patterns of this group.

**Figure 6: Worked example calculating the adjustment to account for the outflow of those who have since become British citizens**



Source: Office for National Statistics

## Limitations and Improvements to the adjustments

The preliminary adjustments applied to the estimates from RAPID are based on the most appropriate and available administrative data and using previously published research. They have been applied to demonstrate how alternative administrative data can be used to fill the gaps in RAPID. However, 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.

### Limitations of current methods used to adjust estimates from RAPID and areas for improvements

#### Student inflow:

##### Limitations

- the proportions of students not working as estimated by the HESA and PAYE case study are based on data from the end of the tax years 2016 and 2017; we have therefore applied this proportion across the timeseries and it is possible that this proportion changes over time
- as the HESA and PAYE research on the proportion of students not working alongside their studies was only completed for undergraduate students the adjustment has also only been applied to HESA data for undergraduate students

##### Areas for development

- understand the proportion of post graduate students who work alongside their studies
- understand how the proportion of students who work alongside their studies changes over time
- understand further the nature and duration of interactions of students and how this will impact on them being classed as a long-term migrant in RAPID

#### Student outflow:

##### Limitations

- this adjustment is based on data calculated for the student inflow adjustment therefore we are reliant on this methodology being correct
- the Home Office and Department of Education data estimates the proportion of students leaving at the end of their studies; we have assumed that this proportion is the same for working and non-working students
- it is possible that students who do not work alongside their studies are more likely to leave the UK at the end of their studies than those who do work
- the proportion of students working from the Home Office data are an average of the years ending 2016 and 2017 which we have applied across the timeseries. It is possible that this proportion changes over time
- the LEO data are for England higher education institutions only, we have applied this to students across the UK, however it is possible that international students in Scotland, Wales and Northern Ireland behave differently to those from England
- where we do not have HESA data available to estimate the number of students finishing their courses at the end of each year, we have applied proportions from later years

#### Areas for development:

- as further Home Office data becomes available, we can use this to understand how the proportion of students leaving at the end of their studies is changing over time
- understand the proportion of post graduate students who leave at the end of their studies

#### **Provisional inflow:**

##### Limitations

- we have assumed here that past relationships between arrivals and registrations are reflective of current behaviour
- we have also assumed that any proportions calculated are consistent over time and that past behaviour is reflective of current behaviour
- unknown impacts of EU exit and behaviour changes during the coronavirus pandemic on inflow adjustment methodology

#### Areas for development

- use the best possible evidence to adapt these assumptions over the period covering the onset of the coronavirus (COVID-19) pandemic where historic trends may not be indicative of recent behaviour

#### **Provisional outflow:**

##### Limitations

- we have assumed here that past relationships between arrivals and registrations are reflective of current behaviour
- we have used figures calculated for the provisional inflow adjustment so are assuming that this is correct and are taking on any assumptions made in that adjustment
- unknown impacts of EU exit and behaviour changes during the coronavirus (COVID-19) pandemic on inflow adjustment methodology in future years

#### Areas for development

- use the best possible evidence to adapt these assumptions over the period covering the onset of the coronavirus (COVID-19) pandemic where historic trends may not be indicative of recent behaviour

## UK citizenship:

#### Limitations

- we have assumed here that past relationships between arrivals and registrations are reflective of current behaviour
- we have assumed that the data from the Home Office Migrant Journey analysis is reflective of those who are outflowing
- the adjustment has only been applied to those who have been in the UK over 10 years prior to departure and some long-term migrants will gain UK citizenship earlier than this, therefore, the adjustment could be underestimating the size of this population
- we have assumed that the proportion of those who gain UK citizenship applies to those who are leaving the UK long-term, however it is possible that migrants who gain UK citizenship are less likely to outflow, therefore the adjustment could be overestimating the size of this population

#### Areas for development

- apply the adjustment for different lengths of stay prior to departure to account for those who gain UK citizenship prior to living in the UK for 10 years
- understand the propensity of individual nationalities to gain UK citizenship and therefore calculate the adjustment at a lower nationality level

## Notes for: Adjustments to estimates from RAPID

1. Whilst the academic year spans two tax years, we have assumed that the outflow will occur at the end of the academic year when the courses have finished. Therefore, for example, the outflow in the academic year ending 2015 has been applied to the outflow in the tax year ending 2016.

## 7 . Comparing estimates from RAPID to IPS estimates to assess the quality in measuring migration

Our report “developing our approach for producing admin-based migration estimates” presents high-level comparisons between estimates from the Registration and Population Interaction Database (RAPID) and previously published long-term international migration (LTIM) estimates. As part of building our understanding of RAPID and its measurement of long-term international migration, we completed further analysis comparing the findings to those from the International Passenger Survey (IPS). This includes demographic analysis of arrivals, departures and net migration for both EU and non-EU nationals. We also completed country level analysis of EU migration patterns to help understand what could be driving the trends seen in RAPID.

### Demographic analysis of RAPID

In this analysis, estimates from RAPID have not had any of the adjustments as described in the “adjustments to estimates from RAPID” section applied to them. At this stage the adjustments have only been calculated at the aggregate level by country group. Therefore, it is not possible to disaggregate this down by age or sex. Additionally, the IPS estimates presented alongside the RAPID estimates are not the adjusted LTIM estimates but those derived directly from the IPS.

Overall, looking at the number of arrivals and departures in RAPID by age group and sex, showed what we expect to see as a typical migrant population with a large proportion of arrivals falling between the ages of 16 to 30 years with fewer migrants arriving who are over the age of 30 years.

Demographic analysis of EU arrivals and departures show that the differences between RAPID and the IPS can be seen equally across both sexes and all age groups. This provides evidence that the higher estimates of migration seen in RAPID are not due to the over-coverage of any one demographic group.

For non-EU migrants, both previous research and our demographic analysis of RAPID compared with the IPS estimate shows us that [age groups that include significant numbers of students are not fully captured in RAPID](#), and this is likely to be an explanation as to why we are not seeing the same pattern in the RAPID data. When we look at the demographic breakdown of the two sources, the IPS is recording higher numbers of arrivals for both sexes for the student age population (aged 16 to 25 years), but similar for all other age groups. Further analysis suggests that this pattern is particularly driven by arrivals from Asia (rather than Rest of the World). This suggests that immigrating students, particularly from Asia, are not fully represented in the RAPID dataset. To account for this under-coverage of students in RAPID, it is necessary to create an inflow student adjustment of the RAPID data. Details of this adjustment can be found in the “adjustments to estimates from RAPID” section of this report.

For further breakdowns by EU and non-EU country groups see the accompanying datasets.

**Figure 7: For EU inflow in the year ending March 2018, estimates from both RAPID and the IPS show similar demographic profiles**

**Notes:**

1. Estimates from RAPID presented here have not had any of the adjustments as described in the “adjustments to estimates from RAPID” section applied to them.
2. The IPS estimates presented here are those derived directly from the IPS and not the long-term international migration (LTIM) estimates presented in the main report.
3. 95% confidence intervals have been presented for the IPS estimates. Whilst there is a level of uncertainty around the estimates from RAPID, it is not possible to measure this uncertainty currently.
4. 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](#)

**Figure 8: For EU outflow in the year ending March 2018, estimates from both RAPID and the IPS show similar demographic profiles**

Demographic analysis of EU long-term international emigration, UK, year ending March 2018

**Notes:**

1. Estimates from RAPID presented here have not had any of the adjustments as described in the “adjustments to estimates from RAPID” section applied to them.
2. The IPS estimates presented here are those derived directly from the IPS and not the long-term international migration (LTIM) estimates presented in the main report.
3. 95% confidence intervals have been presented for the IPS estimates. Whilst there is a level of uncertainty around the estimates from RAPID, it is not possible to measure this uncertainty currently.
4. 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](#)

**Figure 9: For EU net migration in the year ending March 2018, estimates from both RAPID and the IPS show similar demographic profiles**

Demographic analysis of EU long-term international net migration, UK, year ending March 2018

**Notes:**

1. Estimates from RAPID presented here have not had any of the adjustments as described in the “adjustments to estimates from RAPID” section applied to them.
2. The IPS estimates presented here are those derived directly from the IPS and not the long-term international migration (LTIM) estimates presented in the main report.
3. 95% confidence intervals have been presented for the IPS estimates. Whilst there is a level of uncertainty around the estimates from RAPID, it is not possible to measure this uncertainty currently.
4. 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](#)

**Figure 10: For non-EU inflow in the year ending March 2018, the IPS is recording higher numbers of arrivals at student ages**

Demographic analysis of Non-EU long-term international immigration, UK, year ending March 2018

**Notes:**

1. Estimates from RAPID presented here have not had any of the adjustments as described in the “adjustments to estimates from RAPID” section applied to them.
2. The IPS estimates presented here are those derived directly from the IPS and not the long-term international migration (LTIM) estimates presented in the main report.
3. 95% confidence intervals have been presented for the IPS estimates. Whilst there is a level of uncertainty around the estimates from RAPID, it is not possible to measure this uncertainty currently.
4. 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](#)

**Figure 11: For non-EU outflow in the year ending March 2018, RAPID is underestimating the departures of those of student age and shows higher outflows for those who are working age**

Demographic analysis of Non-EU long-term international emigration, UK, year ending March 2018

**Notes:**

1. Estimates from RAPID presented here have not had any of the adjustments as described in the “adjustments to estimates from RAPID” section applied to them.
2. The IPS estimates presented here are those derived directly from the IPS and not the long-term international migration (LTIM) estimates presented in the main report.
3. 95% confidence intervals have been presented for the IPS estimates. Whilst there is a level of uncertainty around the estimates from RAPID, it is not possible to measure this uncertainty currently.
4. 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](#)

## Country level analysis of EU immigration in RAPID

In the country comparisons undertaken here the RAPID estimates have not had any adjustments applied to them as described in Section 5. At this stage the adjustments have been applied using nationality groupings, and therefore it is not possible to disaggregate this to the individual country. Additionally, here the IPS estimates presented alongside the RAPID estimates are not the adjusted LTIM estimates but those derived directly from the IPS.

Here we have focused on EU arrivals only. As presented in the main report for EU nationals, estimates from RAPID are much higher than those based on the IPS. By disaggregating this down to country level it can help to understand if there is any bias in either the RAPID or IPS estimates in estimating migration from individual countries. We have not completed this analysis for non-EU nationals because we have not completed the student adjustment at an individual country level and the student adjustment has a much larger impact on the non-EU estimates than for the EU estimates and therefore estimates from RAPID at an individual country level would not be comparable to those based on the IPS.

Overall, looking at the proportional split of arrivals by country within the EU groups, the RAPID and IPS estimates present similar proportions across the data sources. This is the case across all the different EU groupings, for both the combined data for year ending March 2014, 2015 and 2016, as well as for the combined data for year ending March 2017, 2018 and 2019.

Additionally, for all the EU nationality groupings the main nationalities are the same for both the RAPID and IPS estimates across the different combined year periods. For EU14, the main nationalities are Italy and Spain; for EU8 the main nationality for immigrants is Poland; and then for EU2 and EU Other the main nationality is Romania and Cyprus respectively.

This analysis confirmed that the trends seen in RAPID are comparable to those estimated by the IPS.

### **Figure 12: Estimates of the proportion of arrivals from each EU country are similar in the two sources, RAPID and IPS**

Country comparisons of EU long-term international immigration, UK, combined year ending March 2014, 2015 and 2016

#### **Notes:**

1. Estimates from RAPID presented here have not had any of the adjustments as described in the “adjustments to estimates from RAPID” section applied to them.
2. The IPS estimates presented here are those derived directly from the IPS and not the long-term international migration (LTIM) estimates presented in the main report.
3. 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](#)

### **Figure 13: Estimates of the proportion of arrivals from each EU country are similar in the two sources, RAPID and IPS**

Country comparisons of EU long-term international immigration, UK, combined year ending March 2017, 2018 and 2019.

#### **Notes:**

1. Estimates from RAPID presented here have not had any of the adjustments as described in the “adjustments to estimates from RAPID” section applied to them.
2. The IPS estimates presented here are those derived directly from the IPS and not the long-term international migration (LTIM) estimates presented in the main report.
3. 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](#)

## **8 . Data**

[Analysis of RAPID including adjustment examples](#)

Dataset | Released 16 April 2021

Demographic analysis from our first iteration of our development of admin-based migration estimates (ABMEs) using the Registration and Population Interaction Database (RAPID) and numerical examples of the adjustments applied to the estimates.

## **9 . 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

## 10 . Related links

### [An evolving picture of UK population change](#)

Blog | Released 16 April 2021

### [International migration: developing our approach for producing admin-based migration estimates](#)

Article | Released 16 April 2021

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

### [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.