Developing methods to estimate international migration of EU nationals using the Home Office Borders and Immigration (HOBI) data and replace the existing Registration and Population Interactions Database (RAPID).

Executive Summary

We have been exploring the use of Home Office Borders and Immigration data to estimate the migration of EU nationals. Our ambition is to move to this new data source for our headline estimates of EU long-term migration published in November 2024.

Questions for MARP

1. Does the panel think we have provided sufficient evidence that the Home Office Borders and Immigration data is a more suitable data source?
2. Can you provide feedback on the methods developed for those with EU settled status?
3. Does the panel feel there is sufficient evidence to support moving to these new methods and if not what further evidence would be required?
4. Can you provide any feedback on further future research we should consider?

Background

The International Passenger Survey (IPS) was used to estimate the migration of EU nationals until March 2020. These estimates were based on an individual's intention to move to or from the UK and had its challenges as many EU nationals changed their plans, and their actual behaviour no longer reflected their intentions. For EU nationals this led to us undercounting immigration. While adjustments were introduced to account for changing intentions the IPS was recognised as being stretched beyond its original purpose.

In May 2022 we moved to using administrative data to estimate the migration of EU nationals. This was based on the [Registration and Population Interactions Database (RAPID).](#) This contains data for everyone with a National Insurance Number (NINo) and their interactions with the tax and benefits system since 2008. We identified people who had recently moved to the UK (using their NINo registration date) and looked for sustained interactions with the tax and benefits systems to indicate a long-term immigration into the UK. When we saw evidence of these interactions stopping for a sustained period this was assumed to be an emigration. As these data are based on an individual's actual behaviour rather than their intention it provided a better basis for estimating migration.

Since January 2021, with the introduction of the new immigration system, EU nationals moving to the UK have either needed a visa or for those already living in the UK in January 2021 needed to apply to the EU Settlement Scheme. Therefore, we have been progressing research using the Home Office Borders and Immigration (HOBI) data which combines visa and travel information. We already use this data source for [estimating the migration of non-EU nationals.](#)

Data Sources: coverage and quality

The Home Office Borders and Immigration (HOBI) data is 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 UK, together with data relating to their immigration status, such as periods of leave (visas) granted. Using this travel information, we can estimate how long an individual has been in the UK and based on this make an assumption about whether their stay is long-term.

The current data source used for EU nationals, the Registration and Population Interactions Database is a population stock dataset from which we use activity starting or ending to infer a migration flow into and out of the UK. RAPID has limited coverage for populations who do not interact with these systems. For example.

- Children under the age of 16.
- International students who do not work alongside their studies, or only work for part of the year.

- Those living in the UK but not working (for example due to caring responsibilities, with no entitlement to benefits).

Population coverage of HOBI

HOBI has better population coverage than our current data source RAPID. Table 1 shows some of the limitations of RAPID and how HOBI covers these same populations.

Table 1: Population coverage of RAPID and HOBI.

RAPID	HOBI
We use interactions with the tax and benefits system to infer an individual's presence in the UK. It is possible that some of these interactions occur while the individual is abroad (for example working abroad during the COVID-19 pandemic).	Uses travel information to show when an individual is in the UK and out of the UK.
Does not include students who don't work alongside their studies. We make an adjustment using data from the Higher Education Statistics agency to fill the gap.	Theoretically includes all international students regardless of whether they work
Does not include children under the age of 16. We rely on insights from the IPS to fill this gap.	Includes EU nationals of all ages coming to and from the UK. Both children arriving on a visa and those with EUSS.
Does not include other populations who don't interact with the tax and benefits system. Or emigrates individuals who stop interacting.	Includes all EU nationals coming to and from the UK, regardless of their interactions with the tax and benefits systems.

Timeliness of HOBI

HOBI provides more timely data than is available through RAPID. HOBI is updated quarterly and is provided around 2 months after the end of the reference period. RAPID is tax year data set and is updated annually around 4-6 months after the end of the tax year.

Estimates of International Migration are published for the YE June and the YE December. Therefore, we must apply temporal disaggregation to RAPID based estimates to project forward 3 months for YE June and 9 months for YE December estimates. By moving to HOBI we remove the need to project forward estimates of migration as the record level data is available.

Data Quality of HOBI

We have completed exploratory analysis to assess the differences between the new EU Home Office datasets and the existing non-EU data already provided and used for our headline estimates of non-EU migration. We have checked important quality components, over a time series that has enabled us to understand how data quality changed over time. When looking at data missingness and extreme values the quality of the data was comparable between the current non-EU data extract and the new EU data extract. In the section below we talk about missingness in the travel data. We also checked the number of cases included in the EU HOBI data compared to published data from the Home Office on the number of visas granted. For more information on the quality components explored see Annex 1.

Limitations of HOBI

While HOBI contains information for all non-UK nationals travelling into and out of the UK there are three limitations to this data source;

1. Travel through the Common Travel Area[1] is not included.

---

[1] The Common Travel Area is a long-standing arrangement between the YK, the Crown Dependencies (Bailiwick of Jersey, Bailiwick of Guernsey and the Isle of Man) and Ireland that pre-dates both British and Irish membership of the EU and is not dependent on it.

2. Dual Nationals/ Dual passport holders. Particularly anyone with a British passport as travel on their British Passport will not be included.
3. Irish Nationals do not need permission to come and live in the UK under CTA arrangements, they do not require a visa or EUSS so are not included.
4. There is no information about the geography (Local Authority) where migrants move to.
5. Irregular migration is not included. We use alternative sources to tell us about those who go on to claim Asylum. This mainly impacts non-EU migration.

Therefore, RAPID remains an important alternative source of data on international migration. It is an important component of producing migration estimates at a Local Authority (LA) level because RAPID contains information on where international migrants are moving to or from. RAPID contains information for Irish Nationals, who still need to apply for a NINo to work or claim benefits. RAPID will also remain an important source to quality assure our headline estimates of migration from HOBI.

Missing travel data

The Home Office Borders and Immigration data matches an individual's visa information to their travel events. There are some coverage issues with the travel data included. For example, some routes or some nationalities do not have full coverage.

Travel data through air routes appears to be more complete than travel through sea or rail routes (Eurostar/ LeShuttle). This means nationalities that are more likely to use these routes, such as those from; France, Belgium, Netherlands are more likely to have missing travel data.

Looking at the completeness of travel data over time shows that we have seen an improvement in the arrival date however departures are less well covered and appear to show and increasing proportion of missing information. Overall, 9% of arrivals are missing and 4% of departures are missing.

Imputing missing travel data

We have developed a few options for imputing missing travel events. These have initially focussed on simple options for imputing the missing travel dates. Where we see two arrivals with no departure in the middle, or two departures with no arrival in the middle we have used imputation to make an assessment about when that departure or arrival may have been. It is of course possible that we are missing some travel events entirely (for example missing both and arrival and subsequent departure), but it is not possible to measure the scale of this.

The default option could be to keep rows with missing data and accepting the missingness. To do complete case analysis rows with missing data can be removed. However, this is removing travel data that is available. Therefore, we explored three options to impute the missing dates.

1. Longest date: adding a date that makes an individual's stay in the UK the maximum length based on available travel.
2. Shortest date: adding a date that makes their stay in the UK 1 day.
3. Middle date: adding a date in the middle of the available travel dates.

We have assessed these imputation options to understand the impact for estimating the migration of those with EU settled status. This shows the different imputation options had a large impact on the estimates produced. To help assess which imputation option was most suitable we created a test dataset from records with complete travel data, we replicated the missingness by removing 9% of arrival dates and 4% of departure dates. This suggests that using the longest date is the most suitable option and produces results closest to the "truth". Therefore, this is the imputation option we have taken forward for the rest of our analysis. See Annex 2 for more details on the results of the imputation work.

We recognise that whilst choosing the longest date is the most appropriate based on current travel behaviour, this behaviour could change over time. Therefore, we will continue to monitor whether this is the most appropriate solution in the future. In addition, we are going to work with experts in the Methodology and Quality Division to understand the impact of the different imputation options on creating uncertainty estimates.

Methods to estimate Long-Term International Migration using HOBI

EU visa holders

We have been able to produce estimates of long-term international migration for EU visa holders, from the year ending (YE) December 2021 to the YE December 2023, from HOBI data. These estimates have been produced by applying the same First Arrival/Last Departure (FA/LD) method that was developed for non-EU nationals. This methodology has been developed in collaboration with Home Office colleagues and has been endorsed through our Migration Statistics Expert Group.

For immigration, this method uses an individual's first arrival and last departure dates to approximate their length of stay in the UK within the period for which they have a valid long-term visa. Individuals whose stay lasts 12 months or more are classified as long-term immigrants. Using this methodology, the provisional estimate for total long-term immigration for EU visa holders for the year ending December 2023 was approximately 30,000 (Figure 1). Compared with Statistics published by the Home Office, this was 65% of all visas granted to EU nationals became long-term immigrants.

Figure 1: Number of EU nationals immigrating long-term to the UK, between the YE December 2021 and the YE December 2023
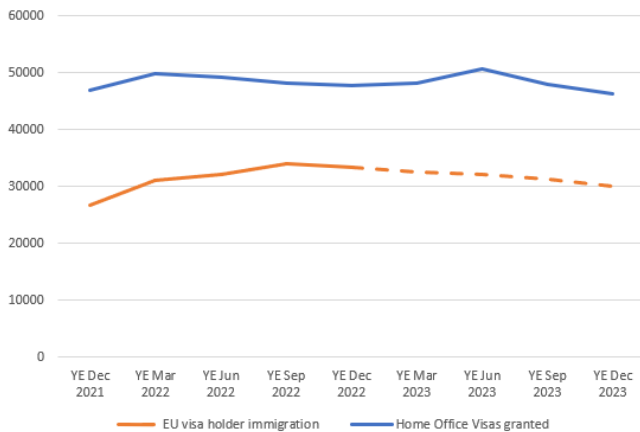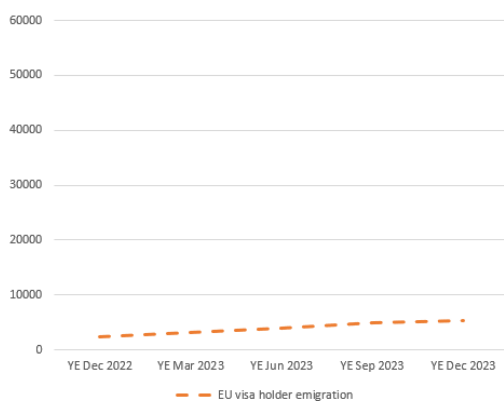

Notes:

1. Estimates for the most recent 12 months are provisional. Where we don't yet have travel data to say whether those who arrived in the most recent year will stay long term we have to make assumptions.

For emigration, we identify previous long-term immigrants with a last departure from the UK during the reference period and record them as long-term emigrants if they do not return to the UK within 12 months, or if they only return for a short-term stay.

Figure 2: Number of EU nationals emigrating long-term from the UK, between the YE December 2022 and the YE December 2023

Notes:

Estimates for the most recent 12 months are provisional. Where we don't yet have travel data to say whether those who arrived in the most recent year will stay long term we have to make assumptions

Emigration estimates remain low as EU nationals have only required a visa to move to the UK since January 2021. Not enough time has passed for many individuals to reach the end of their visas and leave the UK. For example, if a person arrived in January 2021, they would need to remain in the UK for a period of 365 days (and be confirmed as a long-term immigrant), and then subsequently leave for another period of 365 days before being considered a long-term emigrant.

Adjustments

We use the UN definition of a long-term international migrant and therefore need to wait for 12 months of travel data to confirm people's long-term migration status. There is a user need for more timely statistics, so we use assumptions about people's expected length of stay or departure based on past trends to produce adjustments, which provides provisional estimates for more recent periods. These adjustments have been developed and refined for non-EU nationals and we have applied these same methods to EU nationals.

The first adjustment applied to the HOBI data is the early leavers adjustment. An early leaver is defined as an individual who arrives on a long-term visa but does not stay for at least 12 months and therefore does not become a long-term migrant. For those who arrived in the last 12 months, we use past behaviour to estimate how many will leave before that point and remove them from our estimates. This adjustment is broken down by reason for migration and covers those who arrived or departed between January 2023 and December 2023. An example of this can be seen below where we have compared the early leavers adjustment applied to work-related and study-related visas for non-EU & EU visa holders, as they are the largest contributors to migration into the U.K.

Figure 3: Comparison of non-EU & EU early leavers adjustment, for Work and Study-related visas, Year ending June 2023 to Year ending December 2023
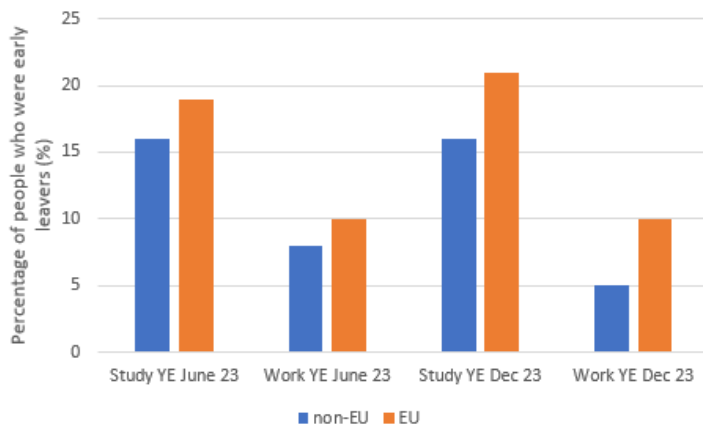
An additional adjustment we make is the emigration re-arrival adjustment. This reflects someone who left the UK within the last 12 months but has returned and therefore does not count as a long-term emigrant. We need to make an adjustment to account for those for whom we do not have 12 months of travel data. In the year ending December 2023, this adjustment was approximately 3% for EU nationals compared with 5% for non-EU nationals.

Irish Nationals do not require a visa or to have obtained settled status to enter the UK. Therefore, we must consider alternative sources to account for this, such as RAPID from which we are able to produce estimates or Irish nationals. It is assumed that the economic behaviour of Irish nationals is similar to that of other EU nationals, therefore we propose using estimates produced from RAPID, for Irish nationals, to inform a proportional adjustment which will be applied to HOBI estimates. For example, if we can see that Irish nationals contribute to 5% of overall EU immigration in RAPID then we will apply this same proportion to increase our HOBI immigration estimates for EU nationals.

RAPID is currently the only available data source to tell us about the migration of Irish Nationals and we acknowledge that RAPID will have both under coverage (students and those not working) and over coverage (those living in the Republic of Ireland but working in Northern Ireland). Therefore, we have a programme of work with the Northern Ireland Statistics and Research Agency (NISRA) and the Irish Central Statistics Office (CSO) to share data on the movement of Irish nationals between the Republic of Ireland and Northern Ireland which will further refine this adjustment in the future.

Future research & limitations

As EU nationals have only required visas since 2021, we only have two full years of data to show what proportion of arrivals on long-term visas end up staying long-term. However, as we receive updated extracts, we will be able to understand whether this behaviour is changing over time and include more data periods to inform these adjustments. In addition, we are further developing the early leavers adjustment to use more information such as nationality, age and sex in addition to visa type to better understand which characteristics inform a person's likelihood to stay long-term or leave before staying 12 months.

There are also some inconsistencies between statistics published by the HO showing the number of visas granted and our own long-term migration estimates when broken down by reason for migration. Research is currently ongoing to establish the cause of this issue, which we believe to lie within the allocation of visas to the correct subcategory (e.g work, study, family etc). We aim to have this issue resolved for our November 2024 publication.

EU Settled Status

Unlike EU nationals who are given a visa to move to the UK those with EU Settled Status require a different approach. Initial research using the "first arrival, last departure" (FALD) method has shown it was not suitable for estimating migration events for this sub-group of the population. Most people who are entitled to EU settled status or indefinite leave were residents in the UK before their application. This makes our FALD method (which assumes a visa is granted before arrival into the UK) unsuitable for this leave type, as most of those applying will already be residents in the UK and should not be included in our long-term immigration estimates. The last departure approach makes assumptions around an absence from the UK relative to the time remaining on a period of leave. Those with indefinite leave do not have a time limit on the amount of time they can reside in the UK.

There are two types of EUSS:

- Settled: Have been resident[2] for 5 years and have no limits on their length of stay in the UK.
- Pre-settled: Have not been resident in the UK for 5 years but were resident on 31st December 2020[3] or are joining a family member of someone from the EU, Switzerland, Norway, Iceland or Liechtenstein who was living here by 31 December 2020, and has settled or pre-settled status, and you joined them in the UK on or after 1 April 2021

A set of assumptions are applied to those with EUSS.

1. For those with full settled status we assume that they arrived at least 5 years prior to their EUSS application date (the rules state individuals must have been living in the UK for 5 years to be eligible for EUSS).
2. For those with pre-settled status we assume that they arrived at the time of their application date.

Methods to estimate migration of those with EU Settled Status

---

[2] 5 years continuous residence means that for 5 years in a row, individuals much have been in the UK for at least 6 months in any 12-month period other than for exceptional circumstances.
Apply to the EU Settlement Scheme (settled and pre-settled status): What you'll get - GOV.UK (www.gov.uk)
[3] Individuals with pre-settled status must show continuous residence to be eligible to move to settled status, otherwise their pre-settled status will not be extended.

We have developed two methods to estimate long-term international migration of those with EU Settled Status.

1. Cumulative time spent in the UK
2. 12 months in 16 spent in or out of the UK.

These methods have been developed alongside expert advice and collaboration with the Home Office who have advised on the specific rules that apply to those with EUSS.

Cumulative time spent in the UK

Owing to the requirement of residency to be entitled to settled status, the approach to classifying someone as a long-term immigrant or emigrant must be reassessed. We start from a position where someone with settled status is likely a resident in the UK, and therefore, the first event we look for is a long-term emigration. We have travel data going back to 2015, which means we can assess travel patterns over several years.

Figure 4: Illustrative example of periods in and out of the UK
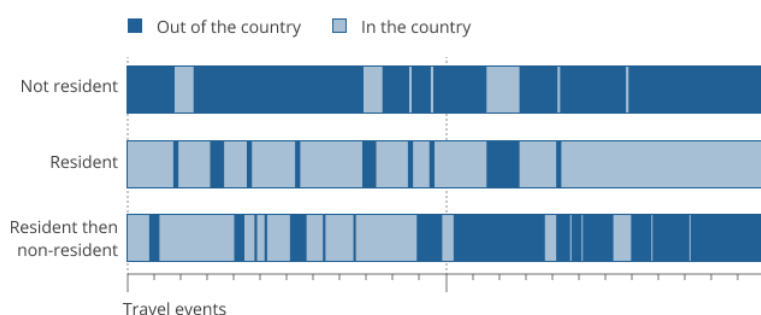

Figure 4 provides an illustrative example of three types of travel history we might see in the data. The first example shows evidence of a person who is not a resident in the UK but took several short trips to the UK. The second example shows someone who is a resident in the UK and took several trips outside of the UK. The third example shows someone that was a resident then transitioned to a non-resident. The third example also highlights one of the challenges that we face with categorising an emigration event and determining exactly when it occurred.

The method developed is based around travel patterns and the amount of time spent in and out of the country. The method has been designed to be flexible in terms of defining travel periods to assess the impact of using different definitions. The first event the method is looking for is an emigration event. This method uses three steps.

1. Using travel dates, calculate the length of stay in and out of the UK. This groups extended trips out of the UK into a single travel period; then assess the cumulative time in and out of the country.
   a. To group these extended trips we have assessed four different options, grouping trips of over; 31 days, 45 days, 60 days or 90 days.
   b. This grouping allows one shorter period outside of the UK in between the longer trips. For example if someone spent 100 days outside of the UK, then in their next trip only spent 7 days outside of the UK, followed by another 90 days all of these trips would be linked into one extended event.
2. Using these grouped periods; we identify those who spend more time out of the UK than in.
3. We then look at the total amount of time spent outside of the UK. If the total time out of the UK exceeds a set limit, they are classed as a long-term emigrant.
   a. We have assessed three different options for the total amount of time spent outside of the UK; 183 days, 270 days or 365 days.

Immigration events are identified using an inverse approach (identifying extended periods within the UK, which we link together to determine whether a person meets the immigration criteria).

This method does identify multiple immigration and emigration events. One of the final steps of the method is to ensure that we only have events occurring in the correct order, i.e. there cannot be two consecutive immigration events, without an intervening emigration event. We have also identified and removed those people who have migration events identified once they have obtained UK nationality.

12 months in 16

The 12 months in 16 method is a more simple and less flexible way of estimating emigration and immigration. This method starts with the default assumption that someone was not resident at the start of their travel data; and then the method will identify a first immigration date. We have travel data back to 2015, for those resident in the UK at the end of the EU transition period (1 January 2021) the method will have used their travel between 2015 and 2021 to assess they are resident in the UK.

This method then uses a rolling 16-month period in which is it looking for time spent in and out of the UK. If the time spent out of the UK exceeds 12 months in a 16-month period this individual will be emigrated. Conversely when the time spent in the UK exceeds 12 months in a 16-month period this individual will be re-immigrated.

This approach has been used by Stats New Zealand and the Australian Bureau of Statistics.

Figure 5: Illustrative example of 12/16 method for international migration
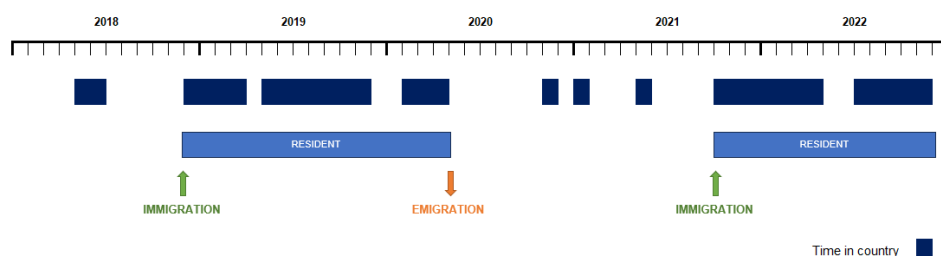
**Findings**

Cumulative method

When all the possible combinations of parameters are run through the method figure 6 shows that the trends between all the parameters are the same, but each parameter gives a different level of immigration of those with EU settled status. The different grouped stay lengths makes a larger impact on the immigration estimates. This suggests that when people immigrate, they have more travel to and from the UK.

Figure 6 shows the same trends for emigration. For emigration the different grouped stay lengths has a smaller impact on the estimates. This suggests that when someone emigrates, they have less travel to and from the UK.

For both immigration and emigration estimates shown for the most recent year (2023) are not representative as not enough time has passed to be able to assess whether someone has immigrated or emigrated long-term based only on the travel data. We are developing methods to produce provisional estimates of immigration and emigration based on the proportion of total arrivals and departures of those with EU settled status who will become long-term, see Annex 3 for more information.

Figure 6: immigration and emigration of those with EUSS using cumulative time in UK method, all parameters, 2020 to 2023
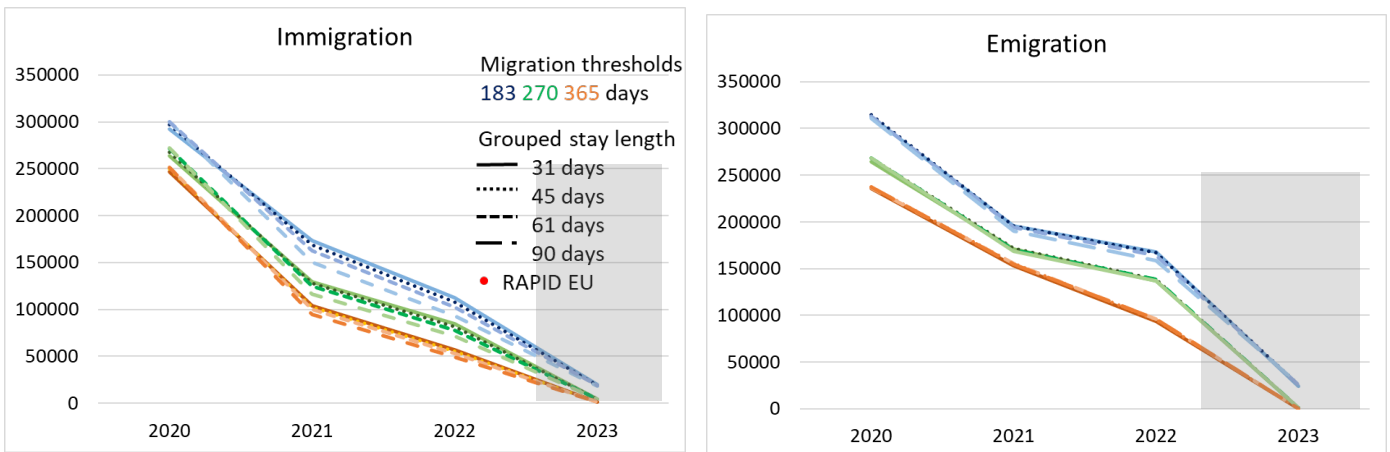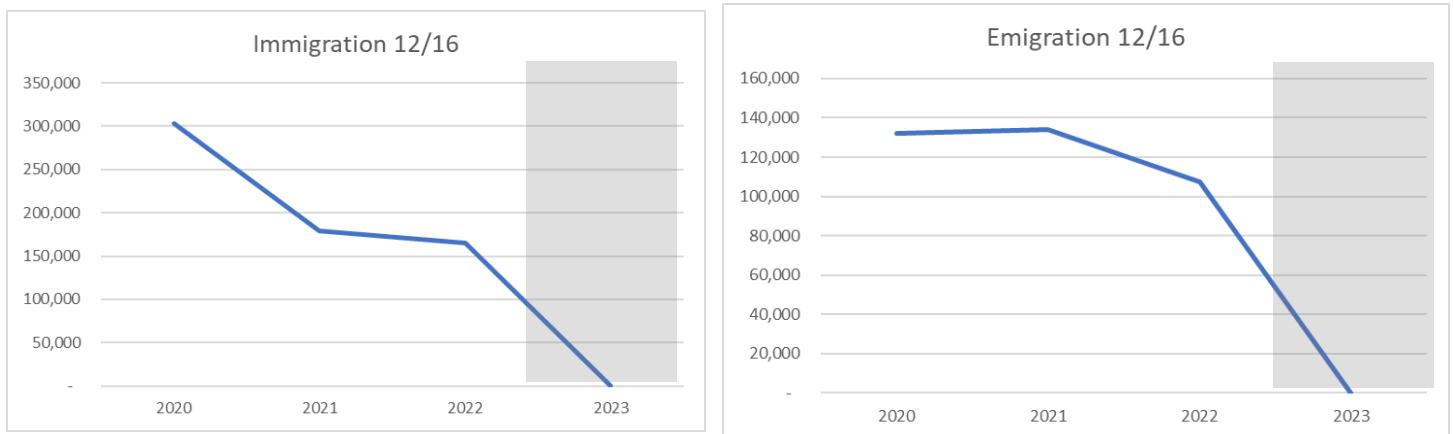

<u>12/16 method</u>

Figure 7 shows similar trends in immigration and emigration of those with EU settled status when using the 12/16 method to define long-term migration. Again like for the cumulative length of stay method estimates for the most recent yea (2023) are not representative and with the 12/16 method we will also need an alternative method for producing provisional estimates in the most recent year.

For both immigration and emigration estimates shown for the most recent year (2023) are not representative as not enough time has passed to be able to assess whether someone has immigrated or emigrated long-term based only on the travel data. We are developing methods to produce provisional estimates of immigration and emigration based on the proportion of total arrivals and departures of those with EU settled status who will become long-term, see Annex 3 for more information.

Figure 7: immigration and emigration of those with EUSS using 12/16 method, 2020 to 2023


<u>Comparing the two methods</u>

To make an assessment about which parameters might be most suitable we have made some comparisons between three of the parameters from the cumulative length of stay method and compared them to the 12/16 method. To do this we have taken a subset of the population for the 12/16 method to ensure we are comparing like for like. Figure 8 shows that using a cumulative length of stay of 183 days produces much higher immigration and emigration estimates. This is expected as 183 days could be including some short-term migrants as it doesn't line up to the definition of a long-term migrant. We also considered two other options of 270 days and 365 days. Whilst a cumulative length of stay of 365 days appears to line up best to the definition of long-term migration, when compared to the 12/16 method it produces lower immigration and emigration estimates. This is because the 12/16

method considers all stays in the UK, whereas the cumulative length of stay method only considers long-term stays and will remove repeated short stays in the UK. The cumulative length of stay of 270 days produces immigration and emigration estimates closest to the 12/16 method.

The cumulative length of stay method using a length of stay of 270 days is more timely than either the length of stay of 360 days or the 12/16 method so has the benefit of being able to produce more timely estimates of long-term migration.

Looking at emigration the 12/16 method has a challenge of sometimes flagging an emigration event earlier than it should. This is particularly evident during a shock event such as the COVID-19 pandemic. The 12/16 method shows a spike in emigration at the end of 2019, whereas the cumulative length of stay method shows this peak in Q1 2020 at the onset of the pandemic. This is because the 12/16 method identifies the first point at which someone spends more than 12 months in a 16 month period outside of the UK. Therefore, for individuals who travelled at the end of 2019 (perhaps to visit friends or family for Christmas) then returned to the UK before the start of 2010, the 12/16 method picks up this departure as the point at which they are spending more time out of the UK than in. Whereas we know that the onset of the COVID-19 pandemic happened during the first 3 months of 2020, and it is more likely this is when individuals emigrated.

Figure 8: emigration and re-immigration of those with EU settled status, comparing 12/16 method to three options from the cumulative length of stay method. Quarterly estimates, 2019 to 2023.
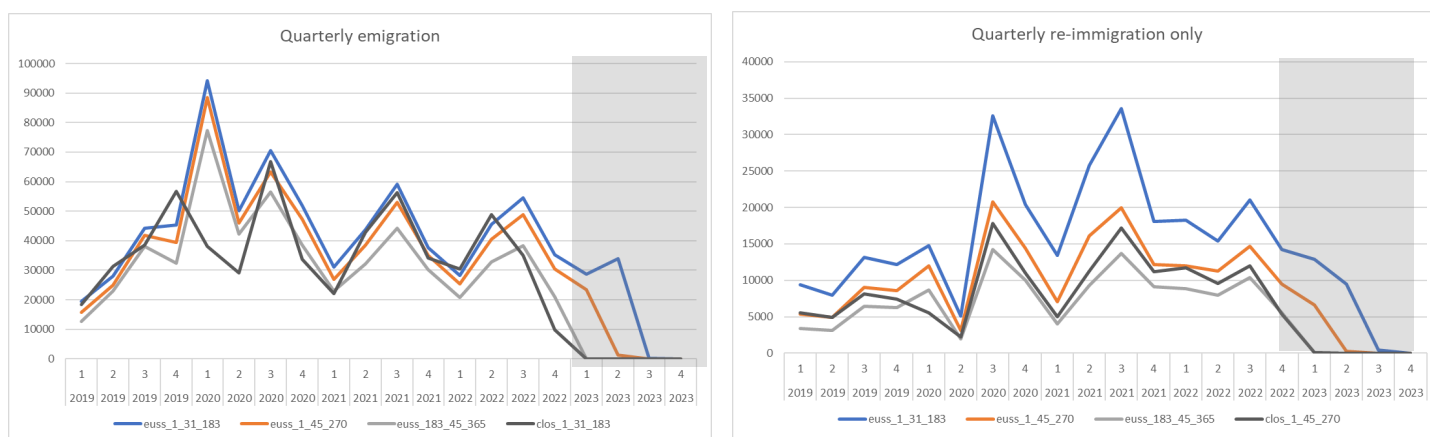
Conclusion

Our research shows that The Home Office Borders and Immigration (HOBI) data is a suitable data source for estimating long-term migration of EU nationals. We have developed methods to estimate the migration of both for EU visa holders and for those with EU settled status, and for those with EU settled status we have a few options from which to choose the most suitable method.

The HOBI data provides a more timely and better coverage data source for estimating long-term migration of EU nationals than the current methods based on RAPID.

Our programme of ongoing research and development will continue to refine these methods, particularly as we get a longer timeseries of data to see how behaviour may change over time. Our next steps include combining the migration estimates of EU visa holders and those with EU settled status to compare against alternative data sources including our current RAPID based estimates and estimates from the International Passenger Survey. We will work to understand any differences in estimates from each source to build further confidence in the estimates produced from HOBI.

The feedback from MARP will be crucial in guiding these improvements and achieving the standards required for official statistics designation.

Annex 1

An important quality component analysed was that of missingness. Variables that are used to produce migration estimates were isolated and comparisons were produced between those from non-EU countries & EU countries. For both non-EU & EU nationals there was no missingness for the variable "idnt_id", a unique identifier allocated to an individual across both the leave instance and visit histories datasets, used to link individuals across the different datasets to create accurate travel histories. Furthermore, within the leave instance datasets for both non-EU & EU nationals there was no missingness for the "primary_nat" and "drv_idnt_type" variables which are used to identify which countries people arriving into the U.K are from, and subsequently what type of leave they have obtained, for example did they have indefinite leave to remain in which case we would filter them out of our estimates.

Variables used to produce estimates that showed higher levels of missingness varied, some displaying higher levels for non-EU nationals than EU nationals, some displaying the opposite. For example, levels of missingness for the "crs_endorsement" and "cid_casetype" variables were higher for EU nationals than non-EU nationals. This was due to these two visa systems being in place before January 2021 (when EU nationals were required to obtain a visa following the U.K's transition out of the EU). Contrarily, the level of missingness for the "pcdp_casetype" variable was higher for non-EU nationals than EU nationals, due to most EU nationals entering the U.K through this system, including those on the EUSS who make up the largest proportion of records for this population of people. Further information, regarding missingness in variables used to produce migration estimates from HOBI data, is outlined in the table below.

A) Table of variables and their missingness percentage, within Home Office Borders & Immigration datasets, that are used to produce long-term migration estimates for non-EU and EU nationals

Variable	Definition	Non-EU (%)	EU (%)	Use in method	Impact
Idnt_id	Unique ID that is assigned to individuals by the system. Can be used to link across HOBI files.	0%	0%	Used to link individuals between datasets to correctly identify travel patterns linked to demographic information.	Able to successfully link records of individuals across leave instance and visit history files
Primary_nat	Three-digit code used to identify a person's country of origin before arriving in the U.K. This is identified by the passport they used upon entry into the U.K	0%	0%	Used to correctly identify an individual's nationality and ensure they are allocated to the correct set of estimates.	Able to correctly allocate individuals to appropriate estimates
Drv_status	The status for an individual's particular leave instance based on decisions about their	0%	0%	Used to filter out those whose value for this field was "in progress", "Expired(unused)" and "Issued(unused)" as these are not valid	Allows us to only create estimates of those who had a successful visa grant and used the visa to

	visa application and their travel movements into and out of the UK. Derived by the system.			grants of a visa.	travel.
Drv_idnt_type		0%	0%	Used to identify those who possess indefinite leave to remain and subsequently remove them from our estimates.	Able to remove individuals with the drv_idnt_type of 'E' & 'O' which signifies they have indefinite leave to remain.
Idnt_visit_seq	The chronological order of visits into the U.K	0.0%	0.0%	Used to identify an individual's first arrival and last departure, by sequence of travel	Able to identify with accuracy an individual's sequence of travel.
Arr_date	The arrival date of an individual	4.4%	11.8%	Used to identify an individual's arrival date into the U.K to create length of stay and inform an individual's first arrival.	If missing implicates the accuracy of someone's first arrival. When missing we impute the visa start date as an individual's first arrival.
Dep_date	The departure date of an individual	10.9%	7.8%	Used to identify an individual's departure date from the U.K to create their length of stay and inform an individual's last departure.	If missing implicates the accuracy of someone's last departure. When missing we impute the visa end date as an individual's last departure.
CRS_endorse ment	Variable used to identify the visa that someone has applied for through the Central referencing system. This system is older and is for those making visa applications from outside the U.K	72.9%	99.9%	Used to outline visa type for people who had a visa granted through this system, to identify those who have unmatched leave, those with Long-Term Visit visas and to allocate reason for migration.	Able to allocate visa types from this system to individuals. Missing ness is higher for EU nationals due to not needing a visa until January 2021 which this system pre-dates.

CID_casetype	Variable used to identify the visa that someone has applied for through the Case information database. This is used for those who are making applications from within the U.K.	85.9%	96.9%	Used to outline visa type for people who had a visa granted through this system, to identify those who have unmatched leave and to allocate reason for migration.	able to allocate visa types from this system to individuals. Missing ness is higher for EU nationals due to not needing a visa until January 2021 which this system pre-dates.
PCDP_casetype	Variable used to identify the visa that someone has applied for through the PCDP system. This is the newest visa system and can be used for applications made from within and outside of the U.K.	98.6%	85.8%	Used to outline visa type for people who had a visa granted through this system, to identify those who have unmatched leave and to allocate reason for migration.	Able to allocate visa types from this system to individuals. Missingness is lower for EU nationals due to this being the system that most will have a visa granted through.

Annex 2 – Imputation of missing travel dates

We produced 5 options of dealing with missing travel dates;

- ignore missing travel information (do nothing to the data).
- remove rows with missing travel
- imputing the smallest date (making the stay 1 day)
- imputing the largest date (making the stay in the UK the maximum possible length)
- imputing the middle date.

We ran these 5 options through the two methods developed to estimate long-term migration of those with EU Settled Status. Figures 2.1 and 2.2 show that the different imputation options have a large impact on the estimates, this means it is important we choose the most appropriate method for imputation.

We found that removing the rows of missing travel data produced very similar results to imputing the largest date, and that ignoring the missing travel information produces very similar results to imputing the smallest date. Therefore, we haven't shown these options on the charts below.

To assess which imputation option produces results closest to the "truth" we created a test dataset using records with complete travel data where we replicated the missing travel data. We removed 9% of arrival dates and 4% of departure dates at random; as this is the amount of missingness seen in the actual data. We then ran the 5 options through this test data to see which produced dates that were the closest to the actual dates removed. This is shown in Figure 2.1 and 2.2. This analysis found that of the 5 options we considered, using the largest date (making the individuals stays in the UK the maximum possible length) produced immigration and emigration estimates closest to the "truth" in the test dataset.

Figure 2.1: Impact of different imputation options using the cumulative length of stay method on test data compared to the true travel dates, 2019 to 2024.
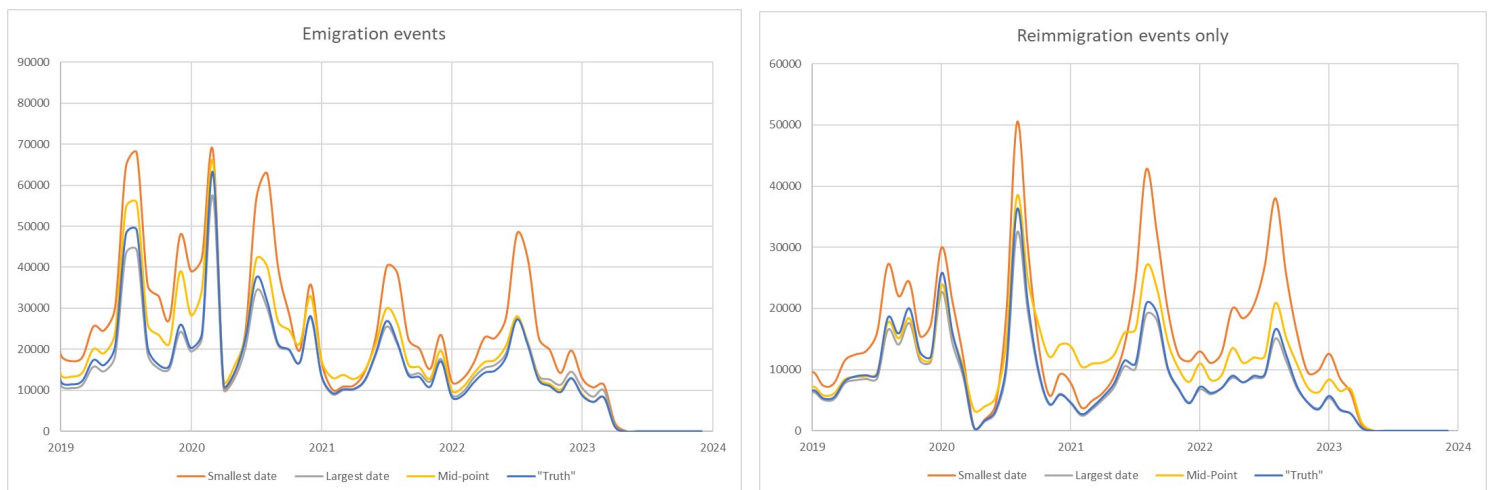
Figure 2.2: Impact of different imputation options using the 12/16 method on test data compared to the true travel dates, 2019 to 2024.
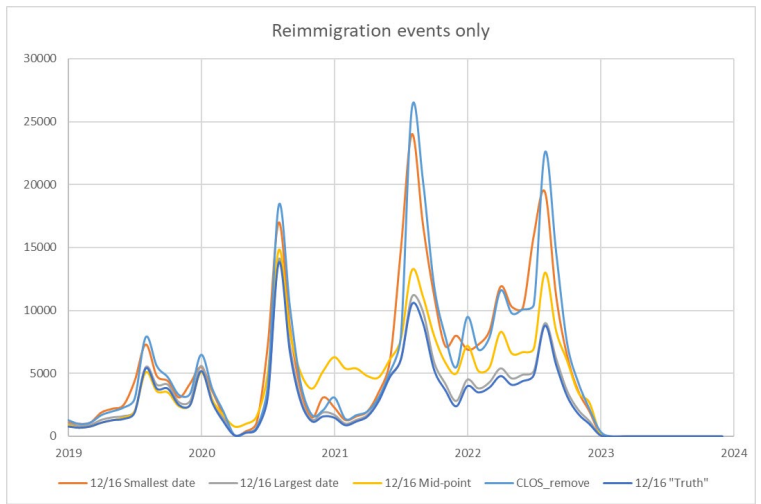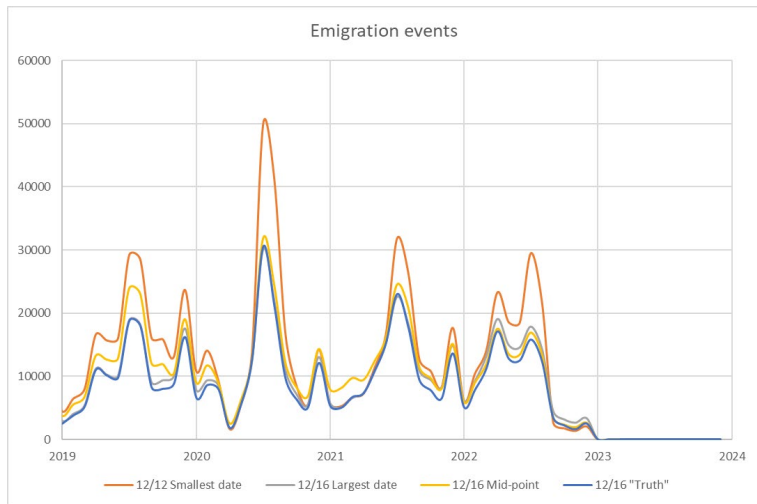
Annex 3 - Producing provisional estimates of long-term migration of those with EU settled status

The methods outlined above can be used to estimate emigration and immigration once enough time has passed for individuals to have spent enough time in or out of the UK to hit the qualifying limits. However, user needs require estimates of migration 5 months after the end of the reference period. For example, in November 2024 ONS will publish estimates of international migration up to the Year ending June 2024. Therefore, someone could leave the UK in February 2024 and will only have 4 months in which they are able to have spent time outside of the UK so will not hit these limits.

Therefore, like for visa holders where we apply the early leavers adjustment and emigration re-arrivals adjustment, for those with EU settled status we also need a method to produce provisional estimates of emigration and immigration.

Using data available from YE December 2022 we have identified those who have arrived during that year, that have any length of stay during the year and have spent less than 270 days outside of the UK, after their arrival. This gives a group of individuals who are 'potential LTIM immigration'. We then compared this to the actual immigration estimate for 2022, using the YE December 2023 data extract. This showed that around 93% of possible LTIM immigrants stayed in the UK long-term. We can therefore apply this proportion to possible LTIM immigrants in 2023 to estimate how many will go on to stay long-term. We have applied this same logic to emigration.

We currently only have EU HOBI data cuts available from YE December 2022 onwards. Therefore, we are not able to test this assumption over time. To further refine how we produce provisional estimates of migration for those with EU settled status we will monitor how this behaviour changes over time and include information such as nationality, age, and sex to further refine these weights.

These provisional estimates will be updated as more time passes and individuals are able to hit the qualifying limits.