Long-Term International Migration estimates and their reliability and accuracy

Understanding the reliability and accuracy of Long-Term International Migration estimates, including sampling distribution, sampling error, statistical significance and confidence intervals.

Contact:
Mike James
MigStatsUnit@ons.gov.uk
+44 (0) 3000672165

Release date:
27 February 2020

Next release:
To be announced

Table of contents

1 . Reliability of Long-Term International Migration estimates

Estimates of Long-Term International Migration (LTIM) are derived from the International Passenger Survey (IPS). This survey collects information about passengers entering and leaving the UK and has been running continuously since 1961.

For more information on the data source see the International migration statistics first time user guide. For more information on the method used to calculate LTIM estimates see the Long-Term International Migration estimates methodology.

Estimates of Long-Term International Migration (LTIM) are:

- the best estimates available at this time

- based on a consistent definition since 1991

- produced in accordance with the Code of Practice for Statistics

To see how these estimates of LTIM measure against the Code of Practice Pillar of Quality, please see the Long-Term International Migration QMI.

Latest development of migration statistics

We are transforming our migration statistics, making use of all available data to provide a richer and deeper understanding of migration. Since July 2018, we have been integrating outcomes from this work into the Migration Statistics Quarterly Report (MSQR), making preliminary adjustments to the headline LTIM measures.

We published our first set of preliminary adjustments in the August 2019 MSQR to produce our best possible assessment of migration trends.

As other new data sources become available, we will continue to refine our adjustments further and reflect this in our migration statistics. We plan to develop the preliminary adjustments, and will inform users of these when possible, with final adjustments expected to be developed by summer 2020.

While we go through this transformation journey, the MSQR has been reclassified as Experimental Statistics to support this period of development and innovation to meet the public good. This also allows us to gain a good understanding of the quality of the statistics, including their accuracy and reliability.

We are improving our understanding of the coherence of migration data sources, including the International Passenger Survey (IPS) and Annual Population Survey (APS) alongside other administrative data sources.

Previous changes to the IPS to improve quality

The quality of international migration estimates improved following the introduction of fundamental changes to the International Passenger Survey (IPS). These changes included:

- redesigning the sample design at the beginning of 2009 to make the survey more focused on identifying migrants

- rebalancing the IPS interviewer resource away from Heathrow and towards other routes

- introducing a new IPS processing system early in 2009 to enable improvements to be made to the IPS weighting methodology

Further information is available in International Passenger Survey: quality information in relation to migration flows .

2 . Sampling distribution and the normal distribution

The different possible samples of passengers that could have been selected by the International Passenger Survey (IPS) can be used to produce a sampling distribution for the figure we are trying to estimate.

For example, if we are estimating immigration within a particular year, one sample may produce an estimate of 500,000, another may have resulted in an estimate of 515,000 and another may have produced an estimate of 490,000. If we could take a lot of samples like this and plot the estimates from each sample, we would produce a chart of the sampling distribution of our estimate.
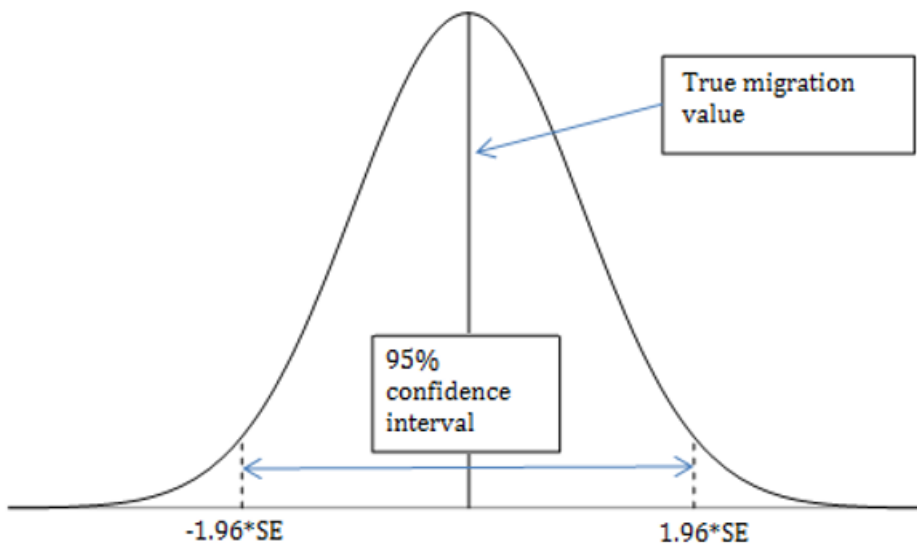
Assuming that the estimation method we use to produce the estimate for each sample is unbiased, the shape of the plot would follow the widely recognised normal distribution, where the most likely estimates of the true value are centred towards the middle, and the least likely estimates are at the "tail ends".

In practice, to estimate the true value for a specific population of, for example, the number of immigrants, we take one sample and produce a single estimate. We assume that the sampling distribution of our estimate would follow approximately a normal distribution, centred on the true value, and we can use a statistical formula to calculate the standard error around the estimate. This is a measure of the accuracy of the estimate.

At the 95% confidence level, which is a widely accepted level, we would expect the confidence interval to contain the true value 95 times out of 100. Equivalently, we can say that there would be a 1 in 20 chance that the true value would lie outside of the range of the 95% confidence interval.

As Figure 1 illustrates, 95% of the estimates would lie within 1.96 multiplied by the true standard deviation of the sampling distribution. This also works the other way round, so we can say that for 95% of random samples taken, our estimate will be no more than 1.96 multiplied by the standard error of that estimate away from the true value that we are trying to estimate. Using this knowledge, we can calculate a confidence interval around our estimate.
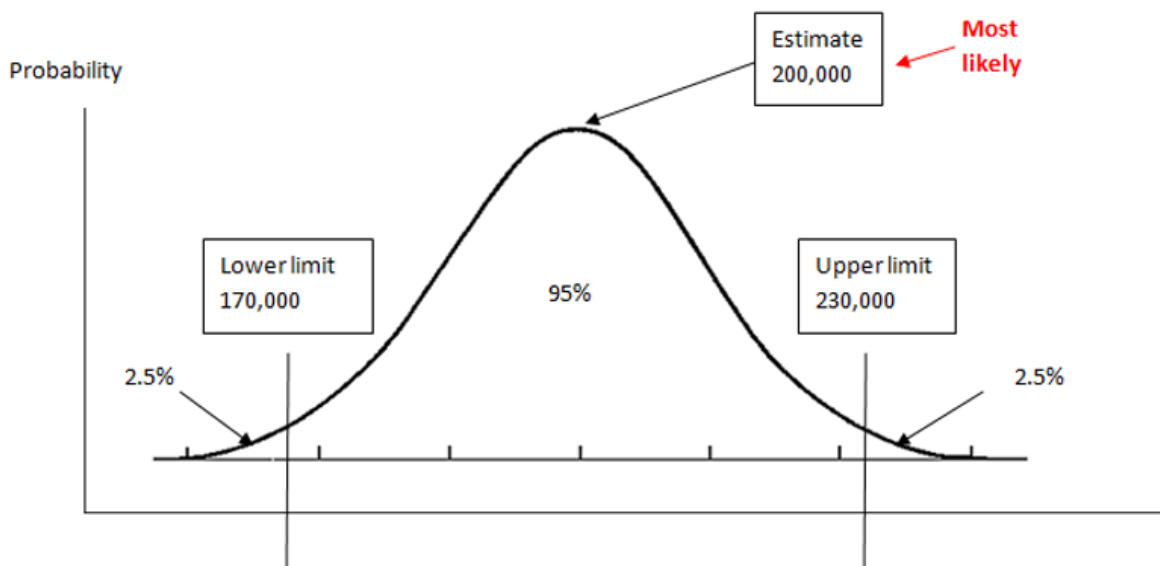
Figure 1: Normal distribution curve

Source: Office for National Statistics

3 . Accuracy and confidence intervals

Confidence intervals are indicators of the extent to which the estimate may differ from the true value. The larger the confidence interval, the less precise the estimate is. The central value within the confidence interval is the best estimate of the true value. The confidence interval around the estimate captures the uncertainty of the estimate and gives an interval within which we can say that there is a high probability that the true value lies.

A confidence interval is not a range in which any figure is equally likely, but the most likely true figure is the estimate itself (Figure 2). However, when interpreting these confidence intervals and statistically significant changes, you should be aware that there is no method for quantifying the error associated with the non-survey components of Long-Term International Migration (LTIM), which are unlikely to be random.

Figure 2: Explaining uncertainty with confidence intervals

Source: Office for National Statistics

The published estimate 200,000 is the most likely figure in the range of plus or minus 30,000.

Users are advised to be cautious when making inferences from estimates with relatively large confidence intervals. Confidence intervals become larger (meaning there is more uncertainty) for more detailed estimates (such as citizenship by reason for migration). This is because the number of people in the sample who have these specific characteristics (for example, EU8 citizens arriving to study) is smaller than the number of people sampled in higher-level categories (such as the total number of EU citizens arriving to study).

Where possible, it is better to use the highest-level breakdown of data available. For immigration and emigration estimates where the lower confidence interval is below zero, users should assume the estimate is above zero.

Users are advised that estimates that have the preliminary adjustments applied to them do not currently have confidence intervals.

4 . Sampling and non-sampling error

Details of the possible effects of sampling error on the migration estimates by various characteristics are given in Table 1.02 of the Long-Term International Migration (LTIM) tables. Entries in this table show that estimates based on the sampling of passengers on certain routes have proportionately larger confidence intervals associated with them. As such, generally speaking, the reliability of the estimate increases in proportion to the size of the estimate.

Sampling error arises because of the variability that occurs by chance when a sample, rather than an entire population, is surveyed; sampling error results because not every migrant who enters or leaves the UK is interviewed. Sampling errors are determined both by the sample design and the sample size. Sampling error may sometimes present misleading changes as a result of the random selection of those included in the sample.

Non-sampling error is all error that is not sampling error. The challenge with non-sampling error is that it is difficult to directly calculate a numerical measure of its effect. This makes it hard to incorporate when analysing results.

Non-sampling error is best understood by referring to examples that apply to the International Passenger Survey (IPS).

The first non-sampling error may be the result of non-response. Bias will occur when passengers who do not respond to the survey have different characteristics to those who do respond.

A further source of bias may arise from contacts deliberately concealing their migration intentions from the interviewers. In addition, the question that determines whether the contact is a migrant or not and their length of stay is based on intentions and not actual behaviour. Measurement errors could therefore be introduced if there is a discrepancy between those intending to migrate, but who subsequently stay less than a year and those not intending to migrate, but who stay for a year or more.

For those contacts identified by the IPS as migrants, the level of non-response is very low for most characteristics. Latest details of survey non-response can be found in Table 1.03.

International Passenger Survey: quality information in relation to migration flows provides an overview of the quality and reliability of the IPS in relation to producing estimates of migration flows.

5 . Measuring change – statistical significance

Changes in the estimates from the International Passenger Survey (IPS) from one period to the next may occur simply by chance. In other words, the change may be because of which individuals were selected to answer the survey and may not represent any real-world change in migration.

We are able to measure whether this is likely to be the case using standard statistical tests. These tests examine the difference between two estimates, and calculate a confidence interval of the difference. The usual standard is to carry out these tests at the 5% level of "statistical significance".
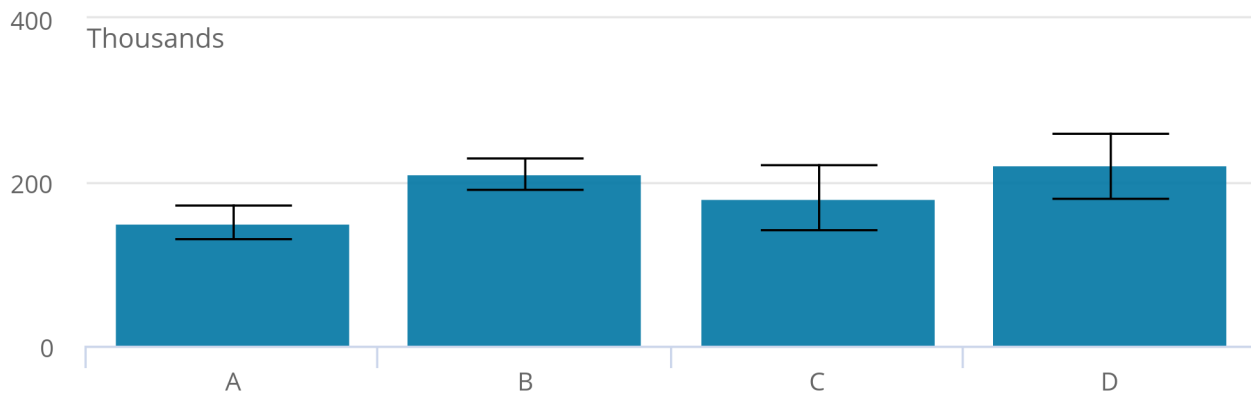
When we report on statistical significance, we provide an assessment of how likely it is that we would see results as unusual as these if the true value of the population remained unchanged. The phrase "statistically significant at the 5% level" indicates that, if the true value of the population remained unchanged, a result like this would occur less than 5% of the time.

When comparing two estimates, a t-test is performed, which results in the calculation of a 95% confidence interval for the difference between these estimates. If this interval excludes the value zero, then we can conclude that the difference is very likely to be a real difference in migration figures and not a result of sampling variation.

A quick method of identifying if the difference between two estimates is statistically significant is to determine if there is an overlap of their confidence intervals. If they do not overlap, then the differences can be described as statistically significant. However, if they do overlap, then a t-test should be performed to determine statistical significance. For example, there is a significant difference between datasets A and B, but may not be between C and D (Figure 3).

Figure 3: Is there a statistically significant difference?

Figure 3: Is there a statistically significant difference?

Source: Office for National Statistics

A t-test ascertains if the difference between two estimates is statistically significant, that is, if it were repeated with a different sample, the difference would occur 19 out of 20 times.

This test divides the difference of the estimates by the square root of the sum of the squared standard errors. The resulting t-value needs to be greater than 1.96 to be 95% certain that the estimates are different. It can also be used to create a confidence interval around the difference. It calculates the standard error of the difference directly from using the difference between the two individual standard errors.

All main statistical software packages have the functionality required to perform a t-test. If you need assistance with identifying whether the difference between two international migration estimates is statistically significant then please contact migstatsunit@ons.gov.uk.

6 . Related links

Long-Term International Migration estimates methodology
Methodology | Last updated 27 February 2020
In-depth look at methodology of both Long-Term International Migration (LTIM) and the International
Passenger Survey (IPS), including assumptions made for variables, and changes over time.

Long-Term International Migration QMI
Methodology | Last updated 28 November 2019
Quality and methodology information for Long-Term International Migration estimates, detailing the strengths
and limitations of the data, methods used and data users and uses.

International migration – terms, definitions and frequently asked questions
Methodology | Last updated 28 January 2019
Useful terms and definitions for long-term and short-term international migration.