

International passenger survey methodology

International Passenger Survey data are collected via face-to-face interviews with passengers passing through ports and on routes into and out of the UK.

Contact:
Charlie Culwick
pop.info@ons.gov.uk
+44 1329 444661

Release date:
6 October 2023

Next release:
To be announced

Table of contents

1. [Background](#)
2. [Sampling](#)
3. [Weighting](#)
4. [Imputation](#)
5. [Seasonal adjustment](#)
6. [Related links](#)
7. [Cite this methodology](#)

1 . Background

Survey data are collected on the International Passenger Survey (IPS) via voluntary face-to-face interviews with passengers passing through ports and on routes into and out of the UK. This is a continuous survey (conducted on 362 days a year).

In 2022, just under 285,000 passengers responded, representing about 0.15% of travellers. Data collected are then used to produce estimates of visits to the UK by visitors from overseas, visits overseas by UK residents, and the amounts spent on visits.

2 . Sampling

A multi-stage sampling design is employed that involves first sampling a port or route on a given day and within a given period of the day (this is referred to as a “shift”). Shifts can be periods of time, for example a morning shift at airports and the Eurostar terminal at St Pancras Railway Station. Within each shift, passengers are selected for interview in a systematic way. This works by choosing passengers at fixed intervals, for example every 20th passenger that crosses an imaginary line set by the interviewer is interviewed.

Alternatively, a shift can be a selected ferry crossing at seaports and trains crossing to France via the Channel Tunnel. Within these shifts, certain passengers are systematically chosen for interview at fixed intervals from a random start; for example, every 20th passenger on a Dover ferry will be interviewed starting on the upper deck. On Channel Tunnel trains and seaports where dockside interviewing is carried out, each passenger in a vehicle is counted and, for example, every 20th person is selected.

A specific fixed interval is applied on each shift, which means that each shift can have a different interval dependent on the number of passengers travelling during the shift. The greater the number of passengers, the larger the interval. This gives the interviewers enough time to complete the interview with a passenger before they must begin their next interview. Each of the passengers selected for interview are asked questions relating to travel and tourism.

A contact can result in one of three response outcomes (complete, partial, and minimum) or a non-response. The three response outcomes are:

- complete interview – all the questions applicable to the contact are answered
- partial interview – core questions are answered but responses to other questions are missing and so are subsequently imputed where possible or designated as “don’t know” answers
- minimum interview – only a few valid answers have been provided on nationality and residency and they are included in the data for weighting and supplementary information only

The non-response outcome can take two forms:

- an interviewer attempts to contact the respondent but no interview takes place; this includes non-contacts (the interviewer could not approach the respondent because they were speaking on a mobile phone, or for onboard ferry collection, eating in a restaurant); refusals to take part; and ineligible (including on-duty military or embassy personnel, merchant navy personnel, airline crew or unaccompanied school children)
- no interviewer was available to contact the next identified respondent

Non-responses are recorded for weighting purposes including some ineligible – in ideal circumstances, only those who are eligible for interview would be counted. However, at some ports the conditions are such that, inevitably, some people who are not eligible for interview are included in the count.

The classification of response determines how each contact is treated in the weighting procedure.

3 . Weighting

The basis of the weighting of International Passenger Survey (IPS) data is that the total set of respondents interviewed at a port or route is weighted or scaled up and calibrated to passenger traffic known to have passed through that port or route in the period in question. This allows estimates of totals from the survey data, such as total expenditure and the total number of passengers travelling for holidays.

The known passenger traffic administrative data is provided to the IPS research team by the Civil Aviation Authority (CAA), Department for Transport, Eurostar, Eurotunnel, Heathrow Airport Holdings Ltd (formerly British Airports Authority) and several individual airports.

The weighting approach incorporates several stages that take account of all passengers selected for interview.

Weighting is conducted separately for each port or route and direction of travel combination, employing the same principles at each one.

The weighting stages are listed in order of application.

Stage 1: design weighting

A design weight is employed to account for the probability of sampling a passenger using the sampling rate.

The calculation compares the number of shifts or crossings sampled (at each port or route and direction of travel combination) with the number of shifts or crossings that could have been sampled for that combination in the sampled period.

In addition, it also accounts for the sampling rate. For example, in a case where a contact was sampled at a port with the following details:

- 10 shifts were run in the period across a calendar quarter on a weekday
- 100 shifts could have been run in the period the contact was sampled
- a sampling rate of 20 (that is, every 20th passenger was selected)

The design weight for this contact would be 200, calculated as:

$$(100/10) * 20$$

This example, which is also used in the following weighting stages, is just for explanatory purposes and uses figures to help explain the process.

As well as port, route and direction, this weight is calculated separately for weekday or weekend, and morning, evening, or night weighting strata.

Stage 2: non-response weighting

A non-response weight factor is employed to take account of contacts selected for interview, but who were subsequently not actually interviewed, either because it was not possible to contact them, or they refused to participate.

The weight is applied using weighting strata for each port or crossing, direction of travel (arrivals or departures), and whether the interview took place during the week or at the weekend.

It involves uplifting “completes and partials” and “minimum” cases by a factor calculated as:

- the sum of the design weights applied to all “complete and partial”, “minimum” and “non-response” records
- divided by the sum of “completes and partials” and “minimums” at that port or route and direction of travel combination

For example, using the case study described in Stage 1 where the design weight was 200 for a weekday shift, if 100 passengers were interviewed and another 20 were non-responses, this would mean that the design-weighted estimate of total passengers would be:

$$200 * 120 = 24,000$$

Then the design-weighted estimate from interviewed passengers would be:

$$200 * 100 = 20,000$$

This non-response weight would be:

$$24,000/20,000 = 1.2$$

The non-response weight essentially increases the total weight applied to a strata with high non-response relative to a strata with lower non-response.

Stage 3: weight factor for minimum responses

A weight factor is applied for discarding minimum respondents.

Minimum interviews are discarded in this step of the weighting, with other cases weighted up to compensate.

Minimum interviews do not have sufficient data to be included in the data for analysis.

However, the purpose of applying this weight is that it is possible that the profile of minimums might be skewed to certain nationalities or residents of certain countries (for example, because of language difficulties, meaning that only minimal information is provided to the interviewer).

This weighting step works to the same principle as the non-response weight. It uses port or route and direction of travel as weighting strata.

In the example given for stage 1 and 2, if among the 100 passengers interviewed, 10 were minimum interviews and they were all from Asia, while the number of passengers with full or partial interviews from Asia was 20, then the minimum weight for these cases would be calculated by:

$$200 * 1.2 * 30/200 * 1.2 * 20 = 1.5$$

Simplified, this is the total passengers from Asia divided by the minimum response from Asia, that is, 30/20= 1.5. The minimum weight essentially increases the total weight of those strata with high numbers of minimums relative to those with lower numbers; the minimum interviews are then discarded from the dataset used for analysis.

Stage 4: administrative data weighting

The data is weighted to the total number of passengers in the period using administrative data provided by CAA and other bodies. Here the population (that is, passenger traffic) for the ports and routes covered by the sampling are used to weight the data.

The population excludes interlining passengers (those neither entering nor leaving the UK from this port, that is, simply changing international flights) and out-of-hours traffic (that is, arriving or departing outside the hours covered by the IPS interviewing at that port).

The weight is applied at each port or route, and direction of travel combination.

Taking again the above example, the weekday and weekend shifts are pooled and then the sampled traffic weight is calculated as:

Administrative total number of passengers divided by stage 1,2 and 3 weighted survey data.

So, if the administrative data were 25,000 in our example and the design-weighted estimate of total passengers were 24,000 (from Stage 2 above), then the sampled traffic weight would be:

$$25,000/24,000 = 1.04$$

Stage 5: under coverage weighting

Weighting for sample under coverage. This extends the population weighting described in Stage 4 to compensate for not covering certain ports and times of day (out-of-hours traffic) in the survey sample.

The weight uses port or route, and direction of travel as weighting strata and incorporates the region of the world that the traffic has come from or has gone to.

The weight reflects the fact that flights to and from some parts of the world are more likely than others to arrive or take off at night when no interviewing is conducted at airports.

Using the example in stages 1 to 4, if administrative data were 3,000 passengers that had travelled out of the hours sampled by the IPS, all of them were from Asian flights and in the weighted data so far, then the number of passengers from Asian flights is:

$$200 * 1.2 * 1.5 * 1.04 * 20 = 7,488$$

This would mean the unsampled traffic weight for these passengers would be:

$$(3000 + 7488)/7488 = 1.4$$

Stage 6: observed imbalance weighting

Weighting for observed imbalance. This step is used to correct an observed imbalance between the number of non-migrants entering and leaving the UK.

These weights are applied as a series of fixed factors, relating to direction of travel, port or route, and country or residence.

Past comparisons carried out by the Office for National Statistics (ONS) of interviews collected at the start and end of visits showed an imbalance for certain nationals: more were recorded at the start of visits than at the end.

This method was introduced in 2019 and compares the results obtained in the arrivals and departures data. Where this provides robust evidence that the departures are under-recorded, an adjustment is applied to bring the departures figures in line with the proportion by country, among overseas residents, obtained in arrivals. Data using this method have been re-weighted back to 2009 when the imbalance was seen to increase.

For example, if the number of Japanese visitors were estimated as 1,000 in arrivals but only 500 in departures, this might be because these visitors are more likely to be non-responders in departures for some of the reasons listed earlier in this paper. Based on this, the imbalance weight for Japanese visitors would be calculated as:

$$1000/500 = 2$$

The other responders' weights would then be rescaled by the imbalance weight so the total sum of the weights sums to the overall traffic totals.

So, in our example, the total traffic was 28,000 passengers, of which 500 were Japanese. The imbalance for Japanese passengers has inflated their numbers to 1,000 so the imbalance weight for other passengers is then:

$$28,000/28,500 = 0.98$$

The adjustment is calculated by pooling all airports, seaports and the Channel Tunnel, as visitors can arrive via one entry point and leave by another, so only their overall totals have to balance.

Stage 7: final weighting

A final weight is calculated, which multiplies each of the weighting stages listed. In our example, the final weight for Japanese visitors is 1,048 from:

$$200 * 1.2 * 1.5 * 1.04 * 1.4 * 2$$

4 . Imputation

Where the responses for important items of interest are missing from the survey data for a partial interview, the values are imputed.

Imputation is applied to the following items:

- length of stay
- cost of fare (expressed in terms of cost of the single fare for the respondent)
- spend

For each of length of stay, cost of fare and spend, a value is calculated for the survey record that was missing the information.

The International Passenger Survey (IPS) employs a mean value within the class imputation procedure where the missing value is replaced with the average value for records with similar characteristics.

Matching variables

Length of stay imputation:

- country of visit for UK residents, country of residence for non-UK
- purpose of visit

Cost of fare:

- port in UK travelled to or from
- overseas port travelled to or from
- month of travel
- operator

Spend:

- country of visit for UK residents, country of residence for non-UK
- duration
- purpose of visit

Where the UK resident respondent has travelled on a package holiday, the cost of the fare is imputed and then deducted from the total cost of the package, and the residual cost (after removal of a percentage to cover travel agent fees) is assigned as their expenditure spent when on their trip.

Overseas residents staying in the UK are asked about their total expenditure in the UK.

This information is then imputed across the towns stayed in, proportionate to the length of stay in each one.

It is recognised that people tend to spend more when they stay in London than in other towns in the UK, and therefore an uplift index is calculated and applied to the spend allocated to London in cases where the respondent stayed in both London and other towns in the UK.

In cases where an overseas resident has not given details of all the towns in the UK that they stayed in, an uplift is applied to towns stayed in by similar records, using the same principles as outlined above for the imputation of stay, fares and spend.

5 . Seasonal adjustment

The number of visits and associated spending both have a clear seasonal pattern, with more taking place in the summer than in the winter. In addition, there are always additional peaks around other holiday periods such as Easter and half-term, which do not occur at the same time each year, and therefore the peaks occur at different times of the year.

Statistical techniques are used by the Office for National Statistics (ONS) with X-12-ARIMA software to produce seasonally adjusted figures. These figures show visits and spending with an estimate for the seasonal component removed. They allow more meaningful comparisons to be made between months and quarters of the year and help to illustrate underlying trends.

Because of the coronavirus (COVID-19) pandemic, no seasonal adjustments have been made to the data since Quarter 1 (January to March) 2020. We are working with the ONS Methodology Department to determine the right time to reintroduce seasonally adjusted data to the International Passenger Survey (IPS) publications.

6 . Related links

[International passenger survey QMI](#)

Methodology | Released 10 May 2023

Quality and Methodology Information for the International Passenger Survey detailing the strengths and limitations of the data, methods used, and data uses and users.

[International Passenger Survey methodological information](#)

Methodology | Released 24 March 2017

The International Passenger Survey background notes outline dataset definitions, sample methodology, geography, and purpose groupings.

[Travel trends: 2022](#)

Article | Released 26 May 2023

Annual estimates of completed international visits to and from the UK and earnings and expenditure associated with these visits.

7 . Cite this methodology

Office for National Statistics (ONS), released 6 October 2023, ONS website, content type, [International passenger survey methodology](#)