

# Annual population survey (APS) QMI

Designation: National Statistic

Contact:

[socialsurveys@ons.gsi.gov.uk](mailto:socialsurveys@ons.gsi.gov.uk)

Release date:

14 September 2012

Next release:

To be announced

## Table of contents

1. [Methodology background](#)
2. [Overview](#)
3. [Executive summary](#)
4. [Output quality](#)
5. [About the output](#)
6. [How the output is created](#)
7. [Validation and quality assurance](#)
8. [Concepts and definitions](#)
9. [Other information](#)
10. [Sources for further information or advice](#)

# 1 . Methodology background

<b>National Statistic</b>	
<b>Survey name</b>	Annual Population Survey (APS)
<b>Frequency</b>	Annual
<b>How compiled</b>	Sample based survey
<b>Geographic coverage</b>	UK
<b>Sample size</b>	320,000
<b>Last revised</b>	14 September 2012

## 2 . Overview

- the sample size is approximately 320,000 respondents
- has the largest coverage of any household survey and allows the generation of statistics for small geographical areas
- uses data from the Labour Force Survey (LFS)
- the data sets consist of 12 months of survey data and are broken down on a quarterly basis
- the first APS data set was published for the period January to December 2004

The [Annual Population Survey \(APS\)](#) is a continuous household survey, covering the UK. The topics covered include employment and unemployment, as well as housing, ethnicity, religion, health and education.

The purpose of the APS is to provide information on important social and socio-economic variables at local levels. The published statistics enable monitoring of estimates between censuses for a range of policy purposes and provide local area information for labour market estimates.

The LFS (and the APS) are the sources recommended for employment-related statistics, such as estimates of the number of people in employment or unemployed. They are also a unique source of comprehensive, coherent information about economic inactivity, as it separates information about people who want a job and those who do not.

The APS is not a stand-alone survey; it uses data combined from 2 waves of the main [Labour Force Survey \(LFS\)](#), collected on a local sample boost. These boosts are sponsored by the [Department for Pensions \(DWP\)](#), [Department for Business, Innovation and Skills \(BIS\)](#), the Welsh Government and the [Scottish Government](#). There are also many other central and local government users of the APS data.

## 3 . Executive summary

The Annual Population Survey (APS) is a continuous household survey, covering the UK, with the aim of providing estimates between censuses of main social and labour market variables at a local area level. The APS is not a stand-alone survey, but uses data combined from two waves of the main [Labour Force Survey \(LFS\)](#) with data collected on a local sample boost. Apart from employment and unemployment, the topics covered in the survey include housing, ethnicity, religion, health and education.

The datasets comprise 12 months of survey data and are disseminated quarterly. The achieved sample size is approximately 320,000 respondents.

This report contains the following sections:

- Output quality
- About the output
- How the output is created
- Validation and quality assurance
- Concepts and definitions
- Other information, relating to quality trade-offs and user needs
- Sources for further information or advice

## 4 . Output quality

This report provides a range of information that describes the quality of the output and details any points that should be noted when using the output.

We have developed [Guidelines for Measuring Statistical Quality](#); these are based upon the five European Statistical System (ESS) Quality Dimensions. This report addresses these quality dimensions and other important quality characteristics, which are:

- relevance
- timeliness and punctuality
- coherence and comparability
- accuracy
- output quality trade-offs
- assessment of user needs and perceptions
- accessibility and clarity

More information is provided about these quality dimensions in the following sections.

## 5 . About the output

### Relevance

(The degree to which the statistical outputs meet users' needs.)

The purpose of the Annual Population Survey (APS) is to provide information on important social and socio-economic variables at a local level. The data enable monitoring of main estimates between censuses for a range of policy purposes and provide local area information for labour market estimates.

The annual local area sample boosts for England, Wales and Scotland are sponsored by [Department for Work and Pensions \(DWP\)](#), [Department for Business, Energy and Industrial Strategy \(BEIS\)](#), [Welsh government \(WG\)](#) and [Scottish government \(SG\)](#). There are many other central and local government users of the APS data.

The data are deposited at the [UK Data Archive](#) where they can be accessed by academic institutions and members of the public. Tables are available publicly on [NOMIS®](#), which uses the APS for estimates of employment, unemployment and inactivity rates down to county council and unitary authority or Parliamentary Constituency level. Estimates of educational attainment are also available on NOMIS® using APS data.

The APS has the largest coverage of any household survey and enables the generation of statistics for small geographical areas. Sampling errors are smaller compared with other social survey designs, because the APS has a single stage sample of addresses.

The sample design provides no guarantee of adequate coverage of any industry, as the survey is not industrially stratified and workers under 16 years of age are not covered. The coverage also omits communal establishments apart from NHS housing and students in halls of residence. Members of the armed forces are only included if they live in private accommodation.

### Timeliness and punctuality

(Timeliness refers to the lapse of time between publication and the period to which the data refer. Punctuality refers to the gap between planned and actual publication dates.)

The APS data are disseminated quarterly, with each dataset covering 12 months' data. Final data are released 3 months after the close of the survey period to which they relate.

For more details on related releases, the [GOV.UK release calendar](#) provides 12 months' advance notice of release dates. If there are any changes to the pre-announced release schedule, public attention will be drawn to the change and the reasons for the change will be explained fully at the same time, as set out in the [Code of Practice for Official Statistics](#).

## 6 . How the output is created

The sample frame for the survey in Great Britain is the [Royal Mail Postcode Address File](#) (PAF) and the National Health Service (NHS) communal accommodation list. Due to the very low population density in the far north of Scotland (north of the Caledonian Canal), telephone directories are used as sampling frames. A systematic sample is drawn each quarter from these three sampling bases and as the PAF is broken down geographically, the systematic sampling ensures that the sample is representative at regional level. In Northern Ireland, the [Rating and Valuation Lists](#) (which serve for the administration of land taxes) are used.

The resident population comprises persons who regard the sample address as their main address and also those who have lived in the dwelling for more than 6 consecutive months, even if they do not regard this as their principal dwelling. Persons absent for more than 6 months are not regarded as members of the resident population. A private household comprises one or more persons whose main residence is the same dwelling and /or who share at least one meal per day. Students living in halls of residence are sampled via the private households of their parents.

The APS survey year is divided into quarters of 13 weeks. From January 2006, the APS has been conducted on the basis of calendar quarters: January to March (Quarter 1), April to June (Quarter 2), July to September (Quarter 3) and October to December (Quarter 4). The APS design is not stratified.

The APS combines data from 4 successive quarters of the Labour Force Survey (LFS) with rolling-year data from the English, Welsh and Scottish Local Labour Force Survey (LLFS). Each quarter's LFS sample of private households is made up of 5 waves, each of approximately 12,000 households. Each cohort is interviewed in 5 successive quarters, so that in any one quarter, one cohort will be receiving their first interview (this is wave 1), another cohort their second (this is wave 2), and so on.

The APS data set is created by taking waves 1 and 5 from 4 successive quarters to obtain an annually representative sample of around 80,000 households. Over the period of the 4 quarters, waves 1 and 5 will never contain the same households to avoid the inclusion of responses from any household more than once in the dataset.

The APS sample has been increased by three annual boost samples. Respondents to the annual boosts are interviewed four times at yearly intervals and approximately one quarter of the sample is replaced each year. The first sample boost began in England (survey year ending 2001). It is combined with the LFS data from respondents in England to form the English local LFS. The second sample boost began in Wales in survey year ending 2002. It is combined with the LFS data from respondents in Wales to form the Welsh local LFS. The third sample boost began in Scotland in survey year ending 2004. It is combined with the LFS data from respondents in Scotland to form the Scottish local LFS. The achieved sample size is approximately 122,000 households (or 320,000 respondents) on each annual APS dataset.

Time		Q2 2009	Q3 2009	Q4 2009	Q1 2010
LFS cases	cohort1	wave 5			
	cohort2	wave 4	wave 5		
	cohort3	wave 3	wave 4	wave 5	
	cohort4	wave 2	wave 3	wave 4	wave 5
	cohort5	wave 1	wave 2	wave 3	wave 4
	cohort6		wave 1	wave 2	wave 3
	cohort7			wave 1	wave 2
	cohort8				wave 1
	cohort9				
	cohort10				
	cohort11				
	cohort12				
APS (boost) cases	cohort a1	wave 4			
	cohort a2	wave 3			
	cohort a3	wave 2			
	cohort a4	wave 1			
	cohort a5				
	cohort a6				

Interviews in all waves are carried out either on a face-to-face basis with the help of laptops, known as Computer Assisted Personal Interviews (CAPI) or on the telephone, known as Computer Assisted Telephone Interviews (CATI). Information is collected using a software package called Blaise. Statistical software called SPSS is used to analyse the datasets. Further details of the methodology used in the LFS and APS can be found in the [Labour Force Survey User Guide Volume 1: LFS Background and methodology](#) Volume 1: LFS Background and methodology.

The APS datasets are weighted to reflect the size and composition of the general population, by using the most up-to-date official population data. Weighting factors take account of the design of the survey (which does not include communal establishments) and the composition of the local population by age and sex. The weights for other sample members are then adjusted to compensate for this.

Sample attrition is the term applied to respondents who begin the survey but subsequently drop out. It has been known for some time that these respondents tend to have different characteristics to those who remain in all waves of the survey, and can, therefore, result in sample attrition bias. The large sample size of the data sets helps to minimise sample attrition bias.

The APS datasets are reweighted back for 2 years, using the mid-year population estimates and are reweighted back for 10 years using the latest census population estimates.

Imputation is carried out to minimise non-response bias in estimates while simultaneously boosting the sample size. For person datasets, missing data are imputed by using roll-forward methods. For many variables, responses are rolled forward for wave 5 only. This means that, if a respondent cannot be contacted for wave 5, then the information that the respondent provided at their previous interview is used to impute a response for the current interview. This may give rise to error, for example, it is possible that the individual may have changed jobs.

To minimise non-response bias, where there are other individuals in the household who can answer on behalf of the absent respondent, proxy responses are also collected. Responses from the annual boosts are not rolled forward due to time elapsed since the response. For household datasets, missing data are imputed by using the "nearest neighbour" method.

[Statistical disclosure control](#) methodology is also applied to the datasets before release. This ensures that information attributable to an individual is not disclosed. [The Code of Practice for Official Statistics](#), and specifically Principle 5: Confidentiality sets out practices for how we protect data about an individual or organisation from being disclosed. The Principle includes a guarantee to survey respondents to "ensure that official statistics do not reveal the identity of an individual or organisation, or any private information relating to them".

On each quarterly APS dataset, the variable frequencies are compared with the previous period. This identifies at an early stage any significant discontinuities. All discontinuities judged significant are investigated to determine the reason for the discontinuity. Is it the product of questionnaire revision or processing error, derived variable revision or error or real world change? This process also ensures that the metadata associated with each variable are correct.

Specific main derived variables are checked in detail by extracting the underlying variables and recalculating in another application, then comparing the results with the values in the dataset. This ensures that the program used to calculate the derived variables is working correctly. On the assembled annual dataset, structural integrity checks are carried out to ensure data is included from the relevant period, outcome and sample. Also, time series checks are performed on main variables. Geographical consistency checks are carried out and all cases with missing geographical data are assigned missing values. Post-weighting checks are carried out to compare APS totals with source population totals and anomalies investigated and corrected.

## 7 . Validation and quality assurance

### Accuracy

(The degree of closeness between an estimate and the true value.)

As the Annual Population Survey (APS) is a sample survey, it provides estimates of population characteristics rather than exact measures. In principle, many random samples could be drawn and each would give different results, due to the fact that each sample would be made up of different people, who would give different answers to the questions asked. The spread of these results is the sampling variability. Confidence intervals are used to present the sampling variability. For example, with a 95% confidence interval, it is expected that in 95% of survey samples, the resulting confidence interval will contain the true value that would be obtained by surveying the whole population. The statistical methodology used to calculate the Labour Force Survey (LFS) sampling variability is published in the [LFS Quality and Methodology Information report](#).

The response rate indicates how many interviews were achieved as a proportion of those eligible for the survey at the first wave. The formula used is as follows:

RR equals (FR plus PR) divided by (FR plus PR plus OR plus CR plus RHQ plus NC plus RRI)

Where RR is response rate, FR is full response, PR is partial response, OR is outright refusal, CR is circumstantial refusal, RHQ is refusal to HQ, NC is non-contact, RRI is refusal to re-interview (applies to waves 2 to 5 only).

Full response indicates a household in which each household member has answered all applicable questions.

Partial response indicates a household in which questions were not completed because someone refused to be interviewed, refused part way through the questionnaire, or refused to let someone else answer on his or her behalf. However, at least one question block must have been completed.

Outright refusal indicates a household that refuses to respond to the survey and the interviewer feels that there is no chance of an interview at the current or in any future wave.

Circumstantial refusal indicates a household where the respondent refuses to respond because of a temporary circumstance (for example, going on holiday, too busy during the field period). A circumstantial refusal enables an interviewer to call back at the next wave.

Refusal to HQ indicates a household that contacts headquarters to refuse to participate in the survey in response to the advance letter.

Non-contact indicates an address that is occupied, but it has not been possible to contact any member of the household in the field period.

Refusal to re-interview indicates a household that takes part in the survey (at one or more of waves 1 to 4) but which, when asked to take part in the next wave (waves 2 to 5), refuses.

The LFS has to complete fieldwork to a tight timetable and interview as many of the sampled households as possible, which leaves limited time for recalls. LFS interviewers try to interview every adult (aged 16 and over) in each sampled household. However, when a household member is unavailable for interview, interviewers accept information by proxy from another responsible adult in the household. A proxy respondent is normally a person living with a partner who responds on their behalf or parents who respond on behalf of their adult offspring.

## **Comparability and coherence**

(Comparability is the degree to which data can be compared over time and domain, for example, geographic level. Coherence is the degree to which data that are derived from different sources or methods, but refer to the same topic, are similar.)

The first APS dataset was published for the period January to December 2004. It combines data from the Quarterly LFS waves 1 and 5 and the LLFS boosts for England, Wales and Scotland. From June 2006, the APS only comprises the LLFS cases and the LFS1 (waves 1 and 5).

The main LFS produces rolling quarterly output each month, but its main output is produced at the end of each calendar quarter (January to March, April to June, July to September, October to December). It is worth noting that, before 2006, the main LFS and LLFS outputs were produced at the end of each seasonal quarter (December to February, March to May, June to August, September to November).

The APS is one of a number of sources of data about the labour market. Some sources provide data that overlap with APS data on employment, unemployment and earnings. We have published guidance about the strengths and limitations of each source in relation to these topics and indicated which source is the most appropriate for different purposes.

The LFS (and APS) is the source we recommend for employment-related statistics such as estimates of the number of people in employment or unemployed. The LFS (and APS) is also a unique source of comprehensive, coherent information about economic inactivity, where it provides information separately about people who want a job and those who do not.

For estimates of change in earnings (for example, pay growth in the economy), the [Average weekly earnings \(AWE\)](#) is the most suitable source. It provides industry and whole economy information, but excludes small employers, the self-employed and government-supported trainees. Pay, commission, bonuses, overtime and pay award arrears are included, but redundancy payments and benefits in kind are excluded.

The [Annual Survey of Hours and Earnings \(ASHE\)](#) includes information about the levels, distribution and make-up of earnings and hours worked for employees in different occupations, industries, ages and regions. It should be used when the information required is not available from the Average Earnings Index (AEI) (such as for occupational groups, or regional analyses) and is the preferred source of the earnings of full-time employees, and of the average hourly earnings of all employees. The LFS (and APS) should be used when the information is not available from the AEI or from ASHE and is the preferred source of data about the earnings of part-time and low-paid employees.

## 8 . Concepts and definitions

(Concepts and definitions describe the legislation governing the output, and a description of the classifications used in the output.)

### **Economic activity**

Economically active people are those aged over 16 who are either in employment or International Labour Organisation (ILO) unemployed. This group of people are those active in the labour force.

### **Economic activity rate**

The economic activity rate is the number of people who are economically active as a percentage of the total population.

### **Economically inactive**

People who are neither in employment nor unemployed (on the ILO measure). This group includes people who are caring for their family or retired (as well as those aged under 16).

### **Employment**

People aged 16 or over who did some paid work in the reference week (whether as an employee or self-employed); those who had a job that they were temporarily away from (for example, on holiday); those on government-supported training and employment programmes and those doing unpaid family work (that is, working in a family business).

### **Employment rate**

The number of people in employment as a percentage of the population in that age group.

## ILO unemployment

The ILO measure of unemployment refers to people without a job who were able to start work in the 2 weeks following their Annual Population Survey (APS) interview and who had either looked for work in the 4 weeks prior to interview or were waiting to start a job they had already obtained. This measure is different to the Claimant Count measure of unemployment that is based solely on eligibility for benefits (that is, Jobseekers' Allowance). The ILO measure is generally accepted to be a more comprehensive measure and is usually higher than the claimant count for any given period, especially for women.

## ILO unemployment rate

The percentage of economically active people who are unemployed on the ILO measure. This refers to those aged 16 and over or those of working age.

For the range of international classifications used, please see the [Labour Force Survey \(LFS\) user guides](#) volume 5.

## 9 . Other information

### Output quality trade-offs

(Trade-offs are the extent to which different dimensions of quality are balanced against each other.)

Please see the Validation and quality assurance section.

### Assessment of user needs and perceptions

(The processes for finding out about users and uses, and their views on the statistical products.)

We use a number of groups to engage with users of labour market data:

- the Labour Market Topic Group for central government users
- the [Central and Local Government Information Partnership \(CLIP\)](#), to communicate with the local government sector
- the Labour Force Survey (LFS) Steering Group, which includes government users and the [Bank of England](#), to engage with central government users of the LFS
- the LFS User Group, which includes users from the local government sector
- academia and the media
- the Annual Population Survey is managed as a formal project, with internal and external data users represented at regular meetings

We run public consultations on the major decisions impacting on users. We consult the Labour Market Topic Group and LFS Steering Group on issues that we consider do not warrant full public consultation. We carry out [triennial reviews](#) of the surveys used for labour market statistics, and consult users as part of these reviews.

Our labour market statisticians keep a log of user engagement and monitor this to ensure that they are meeting users' needs.

## 10 . Sources for further information or advice

Accessibility and clarity (Accessibility is the ease with which users are able to access the data, also reflecting the format in which the data are available and the availability of supporting information. Clarity refers to the quality and sufficiency of the release details, illustrations and accompanying advice.)

The [UK Data Archive](#) at Essex University provides free access to the various Labour Force Survey (LFS) and Annual Population Survey (APS) datasets and can be contacted via the UK Data Archive3 website.

Tables are available publicly on [NOMIS®](#), which uses the APS for estimates of employment, unemployment and inactivity rates down to unitary authority and Parliamentary Constituency area. Estimates of educational qualifications are also available on NOMIS®.

The Office for National Statistics (ONS) Social Survey Data Access and Response Team provides APS data for a fee and can be contacted by phone: +44 (0)1633 455678. Tables using APS data can also be requested by emailing the ONS data service ([socialsurveys@ons.gov.uk](mailto:socialsurveys@ons.gov.uk)).

Labour market data, including data from the LFS, are published every month and include text, tables and charts. The data are widely available, generally free of charge, through a range of media. [Labour market statistical bulletins](#) and [Time series data](#) contained within the releases are available. The Labour market helpline can be contacted by email: [labour.market@ons.gov.uk](mailto:labour.market@ons.gov.uk).

Our recommended format for accessible content is a combination of HTML web pages for narrative, charts and graphs, with data being provided in usable formats such as CSV and Excel. The ONS website also offers users the option to download the narrative in PDF format. In some instances other software may be used, or may be available on request. Available formats for content published on the ONS website but not produced by ONS, or referenced on the ONS website but stored elsewhere, may vary. For further information please refer to the contact details at the beginning of this report.

- For information regarding conditions of access to data, please refer to the following links:
  - [Terms and conditions](#) (for data on the website)
  - [Copyright and reuse of published data](#)
  - [Pre-release access](#) (including conditions of access)
  - [Access to microdata](#) via the Virtual Microdata Laboratory
  - [Accessibility](#)