

Summary Quality Report for Annual Survey of Hours and Earnings (ASHE)

1 Introduction

This report is part of a rolling programme of quality reports being introduced by the Office for National Statistics (ONS). The full programme of work being carried out on [Statistical Quality](#)¹ is available on the National Statistics website. Summary Quality Reports are overview notes which pull together key qualitative information on the various dimensions of quality as well as providing a summary of methods used to compile the output.

This report relates to the [Annual Survey of Hours and Earnings \(ASHE\)](#)², conducted by ONS. ASHE provides information about the levels, distribution and make-up of earnings and paid hours for employees in all industries and occupations. [ASHE](#)² replaced the New Earnings Survey (NES) from 2004. [ASHE](#)² brought improvements to the coverage of employees, imputation for item non-response and the weighting of earnings estimates.

2 Summary of quality

2.1 Relevance

The degree to which the statistical product meets user needs for both coverage and content.

[ASHE](#)² measures the average level and distribution of earnings and paid hours for employee jobs. Estimates are presented by a number of break-downs including gender, occupation, industry and region.

What it measures	Estimates of the average level and distribution of earnings and paid hours for employee jobs.
Frequency	Annual
Sample Size (numbers used)	175,000 employees
Periods available	2004-2009 (ASHE ² inc supplementary surveys) 1997-2004 ASHE ² excl supplementary surveys) 1970-2003 (NES unweighted estimates)
Sample Frame	HM Revenue and Customs Pay As You Earn (HMRC PAYE) Register
Sample Design	All employees whose National Insurance Number ends in a particular pair of digits are selected, ONS treats this as a 1 per cent simple random sample.
Weighting	Employee jobs weighted together based on classes defined by occupation, region, age and gender, to population totals taken from the Labour Force Survey (LFS) ³ .
Imputation	Donor imputation carried out for missing data items on partial returns. No non-response imputation carried out. More detail is provided in Section 3 of this report.
Outliers	No outlier treatment. All validated values are used in estimates.

The survey is a sample survey of employee jobs, although information is collected from employers. It is based on a 1 per cent random sample of jobs on the HMRC PAYE register. It covers all employee jobs in all industries and occupations across the whole of the UK. It has a reference date in April, asking about individuals who are employees at that time. This reference date changes each year depending on when Easter falls.

[ASHE](#)² analyses for weekly and hourly earnings relate to employees on adult rates whose earnings for the survey pay period were not affected by absence. They do not include the earnings of those who did not work a full week, and whose earnings were reduced because of, for example sickness.

[ASHE](#)² analyses for annual earnings relate to employees on adult rates of pay who have been in the same job for more than one year.

As [ASHE](#)² is a survey of employee jobs, it does not cover the self-employed or any jobs within the armed forces. Given the survey reference date in April, the survey does not fully cover certain types of seasonal work, for example employees taken on for only summer or winter work.

Estimates by a range of different breakdowns covering gender, occupation, industry, age and geography are produced from the survey. However, because information is collected from the employer, information about social variables that are not held by the employer, such as ethnic background, household composition or level of disability cannot be produced.

[ASHE](#)² abides by the principal coding schemes used for classifying occupation and industry. Occupation is coded using the [Standard Occupational Classification \(SOC\) 2000](#)⁴ and industry using the [Standard Industrial Classification \(SIC\) 2007](#)⁵.

[ASHE](#)² uses the HMRC PAYE register as a sampling frame. Consequently a number of employee jobs are not available for selection. This happens when no employee in the business earns above the threshold requiring the business to register for PAYE. Although this means that a number of part-time employees with low earnings are not available for selection, investigations have shown that these employees are paid similar amounts to other part-time employees working in the same occupations for businesses that are above the PAYE threshold.

One of the main uses of [ASHE](#)² is to estimate the number of jobs falling below the [National Minimum Wage \(NMW\)](#)⁶. There are a number of circumstances under which an employee would not be subject to [NMW](#)⁶ legislation; for example if the employee is an apprentice or where accommodation is received as a benefit in kind. [ASHE](#)² does not collect information to identify where these circumstances apply, and so the survey does not estimate non-compliance with [NMW](#)⁶.

[ASHE](#)² is also used as the main data source for the quadrennial [European Structure of Earnings Survey](#)⁷.

2.2 Accuracy

The closeness between an estimated result and the (unknown) true value.

Estimates from this survey are subject to various sources of error. Total error consists of two elements, the sampling error and the non-sampling error.

Sampling error

This occurs because estimates are based on a sample rather than a census. [ASHE](#)² estimates this error through coefficients of variation (cv) which are published alongside all [ASHE results](#)⁸. The cv is the ratio of the standard error of an estimate to the estimate itself, expressed as a percentage. The smaller the cv the higher the quality of the estimate.

In published tables [ASHE](#)² uses colour coding as a quick reference guide to the cv of the estimates.

Key	
	cv <= 5%
	cv > 5% and <= 10%
	cv > 10% and <= 20%
x = unreliable	cv > 20% or unavailable

Non-sampling error

Non-sampling errors are not easy to quantify and include errors of coverage, measurement, processing and non-response. [ASHE²](#) minimises any non-response bias by imputing for item non-response and weighting for unit non-response.

Various procedures are in place to minimise errors in returned data. Returns undergo a range of checks which include validation against previous returns and expected values, selective editing and re-contacting businesses for verification. Similar checks are also made at the aggregate level for key results.

Revisions

Results are published in the November following the survey reference date. Revised results are then published one year later alongside the following year's provisional results. The revised results take account of late returns to the survey and amendments to data resulting from validating returns to the current year's survey.

Revisions are usually quite small, with revision at the UK level typically around 0.1 per cent. However estimates for domains with smaller sample sizes are susceptible to larger revisions.

2.3 Timeliness and Punctuality

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

The survey reference date for [ASHE²](#) is in April of each year. The [ASHE Statistical Bulletin⁹](#) is published in November of the same year, covering a number of headline statistics and information from the survey. [ASHE²](#) has consistently met target publication deadlines.

The [National Statistics Release Calendar¹⁰](#), available on the National statistics website, provides twelve months advance notice of releases. In the unlikely event of a change to the pre-announced release schedule, public attention will be drawn to the change. The reasons for the change will be explained fully at the same time, as set out in the [Code of Practice for Official Statistics¹¹](#).

2.4 Accessibility and Clarity

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

[ASHE results⁸](#) are published as a number of standard web tables, in Excel format. Due to the volume of data, these tables are made available on the National Statistics website by the end of the calendar year. Additionally, all the tables are available from the Earnings Helpdesk following the publication of the first wave of tables.

The Earnings Helpdesk team can be contacted by telephone: 01633 456120 or via email: earnings@ons.gov.uk.

2.5 Comparability

The degree to which data can be compared over time and domain.

An earnings survey has been carried out each year since 1970. However, over the years there have been a number of changes to coverage, methodology and classification conventions used for the survey.

[ASHE results⁸](#) for 1997-2003 have been produced by reworking data provided for the NES using [ASHE methodology¹²](#).

In 2004 additional supplementary surveys were introduced to improve the coverage of [ASHE²](#). These focus on employees that have started a new job between the original [ASHE²](#) sample extract being produced in January and the survey reference date in April. Results for 2004-2006 are produced including these new supplementary surveys.

Data for 2004 have also been produced excluding the new supplementary surveys so that it is comparable with the data for 2003 and earlier.

In 2006 [ASHE](#)² moved to the ONS standard for geographic areas using Output Areas (OAs) as building blocks to higher level geographic areas. This improves comparability with other ONS geographic results and allows further geographic analysis to be produced.

In addition to this, from 2006 the [LFS](#)³ moved from using seasonal quarters to calendar quarters. As [ASHE](#)² uses [LFS](#)³ data in the calculation of aggregation weights, it was necessary to use data from [LFS](#)³ Quarter 2 rather than [LFS](#)³ Spring quarter. A consistent back-series which takes into account both the new geographies and [LFS](#)³ weights has been produced going back to 2004.

Two methodological changes took place during the collection and processing of the 2007 [ASHE](#)². The automatic coding tool, Automatic Coding by Text Recognition (ACTR) has been introduced for assigning Standard Occupational Classification (SOC) codes to [ASHE](#)² records. This replaces the current system whereby SOC codes are carried forward for employees who were in the same job as the previous year, or manually allocated by a team of coders if the employee was either not included in [ASHE](#)² for the previous year or has subsequently changed jobs. The second change is to treat large businesses who have a special arrangement (SA) in place with ONS to return their data electronically, as a separate stratum in the [ASHE](#)² weighting. These businesses have a response rate significantly higher than the returns from the paper questionnaires sent out in the original despatch. This meant that SA records were receiving a higher weight than they ideally ought to. Treating businesses with SA's as a separate stratum allows ONS to allocate more appropriate weights to them.

In March 2007, ONS released information on its statistical work priorities over the period 2007-2008. ONS announced that the sample size of the [ASHE](#)² was to be reduced by 20 per cent. The sample size was reduced for the 2007 and 2008 survey periods, with results based on approximately 140,000 returns, down from 175,000 in 2006.

The full 1 per cent sample has since been restored and 2009 [ASHE results](#)⁸ have been produced on this basis.

2.6 Coherence

The degree to which data that are derived from different sources or methods, but which refer to the same phenomenon, are similar.

The most comparable results to those obtained from [ASHE](#)² are figures from the [LFS](#)³. The [LFS](#)³ also collects information on hours and earnings, although it is collected from the employee by means of a household survey, rather than the data being supplied by the employer, as is the case with [ASHE](#)².

Consequently there are a number of important differences between [LFS](#)³ and [ASHE](#)². The data collected from [ASHE](#)² usually comes directly from company records. This means that [ASHE](#)² collects the actual amounts paid and the number of hours that the employee has been paid for. It classifies the employee's industry from the business perspective, which in the case of [ASHE](#)² is the main business of the enterprise. In addition, descriptions of individuals' occupations are provided by the business.

In contrast, [LFS](#)³ data is collected from the employee, or a member of the employee's household. This can lead to proxy and estimated responses, for example if someone responds on behalf of a partner who is not available at the time. The [LFS](#)³ also collects information on hours worked, rather than hours paid.

The monthly [Average Earnings Index](#)¹³, based on the [Monthly Wages and Salaries Survey](#)¹⁴ of 8,700 employers, provides information on changes in mean earnings for broad industrial sectors. No information is available on occupation, hours worked, and other characteristics of the workforce.

3 Summary of methods used to compile the output

Coverage

The [ASHE](#)² covers employee jobs across the whole of the UK. The survey covers employees from all industries and occupations, but does not cover the self employed or the armed forces. The survey reference date falls in April, hence the employees covered are only those that are active in that month, which can distort the numbers in some seasonal occupations.

Sample Frame and Design

The [ASHE](#)² sample is drawn from the HMRC PAYE register. All employees with National Insurance Numbers ending in a particular pair of digits are sampled. This can then be treated as a 1 per cent simple random sample of employees in all industries and occupations. This gives an original sample size of around 260,000.

The original sample is taken from the HMRC register in late January. However, the survey reference date is in April. Consequently a second extract is taken from the HMRC register in April that is used to identify people who have newly joined employers between those dates. These people are split between those on the April extract who have changed jobs since January and those who did not appear on the January extract at all.

The ONS has a special arrangement with some very large employers for them to provide electronic returns extracted from their employee records in April, selected on the same basis as the normal [ASHE](#)² sample.

Imputation

Since the introduction of weighting within [ASHE](#)², the problem of item non-response (ie where a questionnaire is returned by a respondent, but in an incomplete form) has become a significant issue when processing [ASHE](#)² data. A method of imputation, 'donor imputation', has been adopted. In this process, records with similar characteristics are sought to act as 'donors' for missing variables.

Weighting

There are two processes involved in the weighting of responses for [ASHE](#)². The first allocates individual cases a design weight to adjust for non-response. For this purpose, responses are treated as being in one of four strata, depending on whether they were part of the original questionnaire despatch, one of the later supplementary surveys or have a special arrangement in place with ONS to return their data electronically.

For the second part of the weighting, the final file of responses is post-stratified to population estimates taken from the [LFS](#)³ in 108 post-strata. These post-strata are defined as a cross-classification of:

- occupation (9 groups) - major groups from [SOC 2000](#)⁴
- age-band (3 groups) - 16-21, 22-49, 50+
- gender (2 groups) - male and female
- region (2 groups) - London & South East and the rest of the UK

Statistical Disclosure

Statistical disclosure control methodology is applied to all outputs produced from the [ASHE](#)². This ensures that information attributable to an individual or individual organisation is not identifiable in any published outputs. The [Code of Practice for Official Statistics](#)¹¹, and specifically the Principle on Confidentiality set out practices for how we protect data from being disclosed. The Principle includes the statement that ONS outputs should "ensure that official statistics do not reveal the identity of an individual or organisation, or any private information relating to them, taking into account other relevant sources of information". More information can be found in [National Statistician's Guidance: Confidentiality of Official Statistics](#)¹⁵ and also on the [Statistical Disclosure Control Methodology](#)¹⁶ page of the ONS website.

Firstly, to protect individual earnings, a frequency count is applied suppressing all cells that are based on a count of less than three individuals. Secondly, to protect employer pay information, a dominance rule is applied, which uses the contribution from the largest employer and the overall standard error of the estimate to deduce whether information about the employer can be derived to any accuracy.

Given the nature and complexity of ASHE² outputs it is impossible to use a practical method to check for issues of secondary suppression. Instead ASHE² applies a policy where no sample counts are released, only weighted sample counts rounded to the nearest 1,000. This gives users enough information about the sample size for a cell for them to make quality inferences, without giving sufficient information to derive data by difference with any degree of certainty. Although in some circumstances a figure can be derived by difference, it would be impossible to tell how many individuals contributed to the figure.

4 References

	Title of Reference	Website location
1	Statistical Quality	http://www.statistics.gov.uk/about/data/methodology/quality/default.asp
2	Annual Survey of Hours and Earnings (ASHE)	http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=13101
3	Labour Force Survey (LFS)	http://www.statistics.gov.uk/StatBase/Source.asp?vlnk=358
4	Standard Occupational Classification (SOC) 2000	http://www.ons.gov.uk/about-statistics/classifications/current/SOC2000/index.html
5	Standard Industrial Classification (SIC) 2007	http://www.statistics.gov.uk/statbase/Product.asp?vlnk=14012
6	National Minimum Wage (NMW)	http://www.hmrc.gov.uk/nmw/
7	European Structure of Earnings Survey	http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/ses
8	ASHE Results	http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=13101&Pos=1&ColRank=1&Rank=192
9	ASHE Statistical Bulletin	http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=15313
10	National Statistics Release Calendar	http://www.statistics.gov.uk/ReleaseCalendar/currentreleases.asp
11	Code of Practice for Official Statistics	http://www.statistics.gov.uk/about/national_statistics/cop/default.asp
12	ASHE Methodology	http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=15237&Pos=2&ColRank=2&Rank=272
13	Average Earnings Index	http://www.ons.gov.uk/about-statistics/user-guidance/Im-guide/concepts/costs/about/aei/index.html
14	Monthly Wages and Salaries Survey	http://www.ons.gov.uk/about-statistics/user-guidance/Im-guide/sources/business-surveys/monthly-wages/index.html
15	National Statistician's Guidance: Confidentiality of Official Statistics	http://www.statisticsauthority.gov.uk/national-statistician/guidance/confidentiality-of-official-statistics.pdf
16	Statistical Disclosure Control Methodology	http://www.statistics.gov.uk/about/data/methodology/general_methodology/sdc.asp