

## FREQUENTLY ASKED QUESTIONS

### **Why has there been a change to the earnings statistics from the New Earnings Survey (NES) to the Annual Survey of Hours and Earnings (ASHE)?**

The NES was designed to meet the policy needs of the 1970's and has changed little over the past thirty years. As part of the programme of National Statistics Quality Reviews, a Review of Distribution of Earnings Statistics was carried out. The review made recommendations regarding earnings statistics to improve their fitness for purpose for today's requirements. The changes that have been carried out as part of moving from the NES to ASHE are consistent with those recommendations.

A detailed account of the changes and the reasons can be found in the Review of Distribution of Earnings Statistics.

[http://www.statistics.gov.uk/methods\\_quality/quality\\_review/downloads/DOER\\_Final\\_Report.doc](http://www.statistics.gov.uk/methods_quality/quality_review/downloads/DOER_Final_Report.doc)

### **What are the changes?**

Firstly there have been improvements to the coverage of the survey. NES was mainly based on a 1% sample of employees on the Inland Revenue PAYE register for February. While this is still the main basis of the ASHE survey, this sample is supplemented by additional samples drawn from the Inland Revenue PAYE register in April, to cover employees that have either moved into the job market or changed jobs between the time of selection and the survey date, and drawn from the Inter Departmental Business Register for businesses registered for VAT but not registered for PAYE, to cover business who do not have employees above the PAYE threshold.

We have also changed the age statistic to reflect people's age at the survey reference date in April, rather than using 1st January as was the case with NES.

However the changes that have the biggest impact are the introduction of imputation and weighting procedures for ASHE and the change to focusing on the median as the headline figure for earnings.

### **What is weighting?**

Because some types of respondents are more or less likely to respond than others, the proportions of the different types within the returned dataset is not always representative of the proportions of those different types within the general population. Therefore different weights are applied to different types of respondents, so that when the weights are added together, each type will have a proportion consistent with their proportions in the general population. For example, if respondents of type A are generally poor responders, each record of type A will need a relatively large weight to so that they can collectively represent all the type A people in the population. The general population proportions used by ASHE to calculate its weights come from the Labour Force Survey (LFS) and the types are determined by classifying people by age group, sex, occupation and a regional split.

## **What is imputation?**

Sometimes respondents return forms with only some of the variables completed. This means that different variables have different levels of response within the returned dataset. In turn this would mean that the proportions of each variable returned for each job type are different and therefore the weights required for each variable would be different.

Instead of using different weights for different variables we have imputed for the missing values so that the same weights can be used for all the variables.

The imputation method used for ASHE is donor imputation. To impute for a missing value in a record, this method looks for another record with similar characteristics and uses the value from the donor record for the missing variable. This ensures that the distribution of each variable within the imputed data set is similar to the distribution in the un-imputed dataset.

## **What is the impact of introducing weighting and imputation?**

The biggest impact on ASHE comes from the introduction of weighting. This compensates for types of job that are under represented in the ASHE dataset due to poor response. When compared with the LFS, the job-types that are under represented tend to be males, tend to be working in London and the South East and tend to be in Standard Occupational Classification (SOC) 2000 major groups 1 to 3. Therefore these jobs receive larger weights. However these jobs also tend to have higher earnings. Consequently weighting tends to increase all estimates to compensate for the under representation of these high earning job types.

Any domain that has a high concentration of these under represented high earning job types will tend to have a larger increase in the earning estimates than those domains with a lower concentration. Hence, for males the increase in estimates of earnings is more than the increase for females. Notably this affects hourly pay excluding overtime, which is used in the calculation of the Office for National Statistics (ONS) preferred measure of the gender pay gap. The estimate of hourly pay for males is increased more than the estimate for females, which widens the estimate of the gap between male and female hourly pay.

Also estimates of the level of earnings for people working in London are increased more than estimates for other regions. This widens the estimate of the difference in pay between London and other regions of the UK.

## **Why move to focussing on medians rather than means?**

The mean and the median measure different things and can both be appropriate measures depending on what the user is trying to measure. The mean measures the average amount earned by individuals within a domain. In a skew distribution, such as earnings, this measure is susceptible to small numbers of very high earners. The median measures the amount earned by the typical individual within a domain. Since the majority of users seem to be interested in the amount that the typical individual earns, this makes the median a more appropriate measure to focus on than the mean.

## **What is going to be published and when?**

An article giving more detail regarding the changes and their impact, including tables of summary statistics has been published on 15th October. The article focuses on the effect of the methodology changes on historical data. Because in previous releases we have always focused on the mean as the principle measure of earnings this article looks at the impact on the mean figures.

Also at this time tables of data for key domains became available on the National Statistics website for the back series from 1998-2003. These domains cover:

- Male / female
- Full / part time
- Occupation groups (2 digit SOC 2000 for 2002 and 2003, 2 digit SOC1990 for 1998-2001)
- Industry groups (2 digit Standard Industrial Classification (SIC) 2003)
- Age Groups
- Geography by place of work (government office region, county and local authority)
- Public / Private sector

On 28th October, 2004 ASHE results were released, along with revised data for 2003. The release covered all the key domains. We also released data on a comparable coverage basis to the NES results, i.e. excluding the new supplementary surveys. However there will be no data released based on NES methodology, i.e. without weighting and imputation.

Further ASHE domains for 1998-2004 will become available over the following months.

ASHE data for 1992-1997 will be released at a later date once we are able to carry out adequate quality assurance on these years.

No data will be released for any periods prior to 1992 based on ASHE methodology.

## **How do I find the data I want in the tables?**

The various volumes that were previously published for the NES have been replaced by tables covering thirteen domains for each year for ASHE. The thirteen domain tables are:

- Table 1. All employees
- Table 2. Occupation (2 digit SOC 2000 for 2002 onwards, 2 digit SOC 1990 for 2001 and previous years)
- Table 3. Government Office Regions by Occupation
- Table 4. Industry (2 digit SIC 2003)
- Table 5. Government Office Regions by Industry
- Table 6. Age
- Table 7. Place of work by Local Authority
- Table 8. Place of residence by Local Authority
- Table 9. Place of work by Parliamentary Constituency
- Table 10. Place of residence by Parliamentary Constituency
- Table 11. Place of work by Travel-to-Work Area
- Table 12. Place of residence by Travel-to-Work Area
- Table 13. Public private sector

**NB** Initially only tables 1, 2, 4, 6, 7 and 13 have been released, with other to follow over the coming months.

Each of these domains there are tables covering eleven different variables:

Table .1	Weekly Pay: Gross
Table .2	Weekly Pay: Excluding Overtime
Table .3	Weekly Pay: Basic
Table .4	Overtime Pay
Table .5	Hourly Pay: Gross
Table .6	Hourly Pay: Excluding Overtime
Table .7	Annual Pay: Gross
Table .8	Annual Pay: Incentive
Table .9	Hours Worked: Total
Table .10	Hours Worked: Basic
Table .11	Hours Worked: Overtime

**NB** Overtime pay, Annual Pay: Incentive and Hours Worked: Overtime. Because a high percentage of respondents do not have non-zero values for these three variables, medians and percentiles have been presented as a distribution of non-zero values, whereas mean values are averages across all respondents.

For each of these variables the main table contains nine parts for sub-domains:

1. All employees
2. Male
3. Female
4. All Full-time
5. Male Full-time
6. Female Full-time
7. All Part-time
8. Male Part-time
9. Female Part-time

Finally, for each of these sub-tables, estimates are published for number of jobs, median and mean, with annual percentage change, deciles and quartiles.

So, if you want to find the median annual gross pay for a full-time employee in the construction industry in 2003; first you select the ASHE 2003 tables, then table 4 (industry), then 4.7 (Annual Gross Pay), then the Full-time tab, then scroll down to the Construction row.

### **What do the colours on the tables mean?**

In recent years, estimates from NES have been suppressed on quality grounds when the number of respondents contributing to an estimate has been less than 30 or the coefficient of variation (CV) of an estimate has exceeded 5%, for an estimate of a mean. These rules have also ensured that confidentiality has been maintained.

For ASHE the suppression for confidentiality reasons and issues regarding the quality of the estimates have been separated from one another. Confidentiality is maintained by applying

standard ONS criteria that require estimates to have a minimum of three respondents contributing, while ensuring that no one employer is dominant in their contribution to the estimate. Estimates that fail these criteria will be marked in the table as ".." disclosive.

The colours on the tables relate to the relative quality of the estimates in terms of CV. A CV of an estimate is the standard error of the estimate presented as a percentage of the mean. The lower the CV, the more accurate the estimate is considered to be. The new tables rank these CV's by band, which are colour coded.

CV  $\leq$  5% - Precise (previously these would have been the only estimates that ONS would have published)

5%  $<$  CV  $\leq$  10% - Reasonably precise

10%  $<$  CV  $\leq$  20% - Acceptable

20%  $<$  CV - Unreliable (these estimates are still not published, being considered too unreliable for practical purposes)

### **Where can I find more information?**

Further information can be found in:

The Review of Distribution of Earnings Statistics.

[http://www.statistics.gov.uk/methods\\_quality/quality\\_review/downloads/DOER\\_Final\\_Report.doc](http://www.statistics.gov.uk/methods_quality/quality_review/downloads/DOER_Final_Report.doc)

Methodology for the 2004 Annual Survey of Hours and Earnings

[http://www.statistics.gov.uk/articles/nojournal/ASHEMethod\\_article.pdf](http://www.statistics.gov.uk/articles/nojournal/ASHEMethod_article.pdf)

ASHE Results

<http://www.statistics.gov.uk/STATBASE/Product.asp?vlnk=13101>

Or contact:

Mark Williams

Office for National Statistics, Cardiff Road, Newport NP10 8XG

E-mail [earnings@ons.gov.uk](mailto:earnings@ons.gov.uk)

Tel. 01633 456728