Annual Survey of Hours and Earnings, Office for National Statistics.
Last updated October 2017

ANNUAL SURVEY OF HOURS AND EARNINGS (ASHE) DATASET NOTES

INDEX

Contact information

If you cannot find what you are looking for or require further help or information about ASHE microdata please contact us:
By email at: earnings@ons.gov.uk
By telephone on: +44 (0)1633 456120

Changes to the 2016/2017 datasets

The contracted out of state pension variable (**cosp**) is no longer available from 2016.

New variables have been added to the datasets from 2016. The variables are as follows:

ppstart - this variable provides the start date of the employee's pay period. It takes the form **DDMMYY**.

hrpayx - this is the derived hourly rate used in the low pay calculation (see **Annex E**), excluding overtime and shift premium pay.

nlpflag - the 'not low pay' flag indicates employees that have a derived hourly rate below the minimum wage but are not classed as low paid as, legally, employees are only entitled to the new rate for pay reference periods that start on or after 1 April.

pcflag (2017 only) – the 'payroll calculator' flag indicates employees who are paid according to a payroll calculation (weekly hours x stated hourly rate x 52 weeks)/12 months) which differs slightly from the calculation stated on the ASHE questionnaire for calculating the paid hours worked. This results in their derived hourly rate coming out below the NLW/NMW.

autopen – this variable identifies whether an employee has been automatically enrolled into a workplace pension by their organisation.

The following geography variables have changed from 2016:
English Clinical Commissioning Groups (**wccg** and **hccg**)
Scottish Health Board/Welsh Local Health Board/English NHS area teams (**whlth** and **hhlth**)
Local Enterprise Partnerships (**wleps1**, **wleps2**, **hleps1** and **hleps2**)

The following geography variables were added from 2016 (and made available back to 2012/2013):
National Assembly for Wales Constituencies (**wwac** and **hwac**)
National Assembly for Wales Electoral Regions (**wwer** and **hwer**)
Scottish Parliamentary Constituencies (**wspc** and **hspc**)
Scottish Parliamentary Regions (**wspr** and **hspr**)

Updates to the Information Document - October 2017

Geography names/codes (with the exception of region) are no longer included in the annexes but instead provided as separate Excel lookups.

Five new annexes have been added which provide the SAS syntax we use for the following derivations:
Medians and percentiles (**Annex C**)
Mean hourly earnings (**Annex D**)
Proportion of jobs with hourly earnings below the NLW/NMW (**Annex E**)
Proportion of jobs with hourly earnings below the living wage (**Annex F**)
Change in median earnings for those in "continuous employment" (**Annex G**)

DISCONTINUITIES

There are three discontinuities in the ASHE datasets resulting from changes to the ASHE methodology. In 2004 supplementary information was included in the ASHE for the first time and in 2006 data from respondents with 'special arrangements' were treated as an extra stratum and occupations were coded using an automatic coding tool. In 2011

the Standard Occupational Classification 2010 (SOC 2010) replaced the Standard Occupational Classification 2000 (SOC 2000).

For continuity with the back-series, two datasets have been generated at the time of these discontinuities (for 2004, 2006 and 2011). The 1997 to 2004 series of datasets, excluding supplementary information, is identified by **numstrata = 1**. The 2004 to 2006 series of datasets, which includes supplementary information, is identified by **numstrata = 3**. The series for 2006 and later years is identified by **numstrata = 4**.

Since the SOC forms part of the methodology by which ASHE data are weighted to produce estimates for the UK, two full sets of provisional results were produced for 2011. The dataset 1997- 2011, on a SOC 2000 basis provides a time-series from 1997-2011 (2011 data is provisional). The 2011 revised data is only available on a SOC2010 basis. This marks the beginning of a new time series, meaning that care should be taken when making comparisons with earlier years.

From the revised 2012 dataset the geography codes are based on the 2011 Census output areas and the codes used are the official ONS 9-digit geography codes.

FILTERS AND WEIGHTS

The ASHE published tables apply filters, documented as footnotes, to both weekly (which also includes hourly) and annual pay variables.

The weekly filter is employees on adult rates whose pay for the pay period was not affected by absence. Additionally, employees who do not have a valid work region and who are less than 16 years old are filtered out because the age and region variables are required for weighting:
adr = 1 AND lop = 2 AND wgor > 0 AND age > 15

The annual filter is employees on adult rates who have been in the same job for more than one year. Additionally, employees who do not have a valid work region and who are less than 16 years old are filtered out. Employees with missing or zero annual gross salaries are also filtered out:
adr = 1 AND sjd = 1 AND agp > 0 AND wgor > 0 AND age > 15

In order to replicate published ASHE estimates, the weighting variable **'calwght'** should be used. The only exception to this is when estimating numbers of jobs with pay below the NLW/NMW or living wage, for which **'lpcalwght'** should be used.

JOB COUNTS

Weighted figures for numbers of jobs provide a broad idea of the numbers of employee jobs within specified domains. However, these are indicative only and they should not be considered accurate estimates of numbers of jobs. Caution should be applied when using these numbers.

DESCRIPTION OF VARIABLES

Name	Description	Comments
Year	Year	Year for which the survey was conducted. The survey reference date is always in April. No missing values.
NumStrata	Stratification of dataset identifier	1 = 1997-2004 series excluding supplementary information 3 = 2004-2006 series including supplementary information 4 = 2006- series with special arrangements as new stratum and occupations coded using automatic coding tool (ACTR).
Piden	Personal identifier	Random identifier created to identify multiple occurrences of the same person within the dataset. This allows comparison over time.
Sex	Sex	1 = Male 2 = Female No missing values.

Age	Age	From 1997-2003 the data was collected from the New Earnings Survey which used 1 January as the reference date for age. Since the introduction of ASHE in 2004, age is as at the survey reference date in April. The dataset only contains people aged 16 and over at the survey reference date. No missing values.
Serno	ONS Serial Number	Uniquely identifies records within each year within the ONS datasets.
Ft	Full / part time marker	1 = Full time 2 = Part time Full-time employees are defined as those who work more than 30 paid hours per week or those in teaching professions working 25 paid hours or more per week. No missing values.
Pt	Permanent / temporary marker	1 = Permanent 2 = Temporary / casual
Adr	Adult rate marker	1 = Full adult rate 2 = Trainee / junior rates No missing values.
Lop	Loss of pay marker	1 = Earnings were affected by absence 2 = Earnings were not affected by absence No missing values
Sjd	Same job marker	Has the employee worked in the same job for more than one year 1 = Same job 2 = Not same job
Djob	Double job marker	Indicates people with more than one job 0 = one job 1 = more than one job
Mjob	Main job marker	Indicates the main job, but only for people with more than one job. 0 = Not main job 1 = Main job
Empsta	Employee start date	Month and year in which employee started working for the organisation, MMYYYY
Colag	Collective agreement	Whether the employee's pay is set with reference to a collective agreement. See **Annex A**
App	Apprentice marker	Is the employee an apprentice? 0 = Missing 1= Yes 2 = No Available from 2013
Appdate	Apprenticeship start date	Format 'MMYYYY' Available from 2013
Sic03	Industry	Between 1997 and 2008 industry is based on Standard Industrial Classification 2003. Link to SIC archive
Sic07	Industry	From 2008 industry is based on Standard Industrial Classification 2007
Occ90	Occupation	1997 – 2001 occupation based on Standard Occupational Classification 1990. Vol 1 Page 25
Occ00	Occupation	Occupation based on Standard Occupational Classification 2000. (2002 -2011) NB Occ00 codes present for 1997-2001 are imputed for weighting purposes only and should NOT be used for tabulation or analysis.

Occ10	Occupation	2011 onwards occupation is based on Standard Occupational Classification 2010. Vol 1 Page 12 NB: 2011 coded to both SOC2000 and SOC2010
Sano	Special arrangement number	ONS number allocated to contributors with special arrangements for responding to ASHE.
Sernol	Serial number	ONS serial number for jobs in previous year. If same job appeared in previous year's datset, then sernol gives previous year's serial number of that particular job
Payp	Pay period	Type of pay period that original returned data was based on. NB. all pay and paid hour variables in the dataset are already converted to weekly averages 1 = One week 2 = Two weeks 3 = Four weeks 4 = Calendar month
Bhr	Basic paid hours	Basic weekly paid hours worked Missing values of Bhr are imputed, unless lop = 1
Ovhrs	Paid overtime hours	Weekly paid overtime hours worked during the reference period
Thrs	Total paid hours	Total paid hours worked during the reference period (Bhr + Ovhrs)
Bpay	Basic Pay	Basic Weekly Earnings Missing values of Bpay are imputed, unless lop = 1
Bpayinc	Basic Pay including Other Pay (from 2004)	Basic Pay including Other Pay Missing values of Bpay are imputed, unless lop = 1
Othpay	Other Pay	Pay received in the pay period for other reasons
Gpay	Gross pay	Gross weekly earnings for the reference period 1997-2003, 2004 strata1 definition (Bpay + Ipin + Ipop + Sppay + Ovpay) 2004 definition (Bpay + Ipin + Sppay + Ovpay + imputed Othpay) Current (from 2005) definition (Bpay + Ipayin + Sppay +Ovpay + Othpay)
Gpox	Gross pay excluding overtime	Gross weekly earnings excluding overtime for the reference period
Ovpay	Overtime pay	Weekly overtime pay for the pay period
Ipay	Incentive pay paid this pay period (1997-2004 back series)	Total incentive pay paid in this pay period, including any relating to another pay period
Ipin	Incentive pay paid in this pay period (1997-2004 back series)	Incentive pay paid in this pay period that relates to this pay period
Ipop	Incentive pay paid outside this pay period (1997-2004 back series)	Incentive pay that relates to this pay period, where the amount was paid outside this pay period
Ipayall	All incentive pay paid in this pay period	Total incentive pay paid in this pay period, including any relating to another pay period
Ipayin	Incentive pay paid in this period relating to the pay period	Incentive pay that relates to this pay period
Sppay	Shift and premium payments	Additional premium payments during the pay period for shift work and night or weekend work not treated as overtime
He	Hourly earnings	Hourly earnings for the reference period (Gpay / Thrs)
Hexo	Hourly earnings excluding overtime	Hourly earnings for the reference period, excluding overtime (Gpox / Bhr)
Tpay	Stated total pay	Total pay as stated by respondent
Hpay	Stated hourly rate of pay	Hourly rate of pay as stated by respondent
Agp	Annual gross pay	Gross annual earnings paid for the tax year ending 5 April of the reference year

Anipay	Annual incentive pay	Component of gross annual earnings that comes from incentive payments
Bikfilter	Benefits in kind filter	Did the employee receive any benefits in kind in the tax year ending 5 April? For example, a company car or subsidised housing. 1 = Yes 2 = No 0 = Missing
Bik	Annual value of benefits in kind	Annual value of benefits in kind for the tax year ending 5 April
Ppfilter	Pension provision filter	Was the employee a member of any pension scheme run or facilitated by their organisation 1 = Yes 2 = No
Pens	Pension category for 1997-2004 back series	Pension provision of the employee within the company 1 = Contracted out salary related scheme only 2 = Contracted out money purchase scheme only 3 = Not contracted out salary related scheme 4 = Not contracted out money purchase scheme 5 = Contracted out salary related scheme and a not contracted out occupational pension 6 = Contracted out money purchase scheme and a not contracted out occupational pension 7 = Group personal pension arrangement only and contracted out of SERPS 8 = Group personal pension arrangement only and not contracted out of SERPS 9 = None of the above
Tpen	Type of pension scheme – 2005-2012	1 = Defined benefit scheme 2 = Defined contribution scheme 3 = Group personal pension scheme 4 = Stakeholder pension 5 = Pension category unknown 0 = No pension with employer
Tpen	Type of pension scheme – from 2013	1 = Defined Benefit scheme 2 = Defined Contribution scheme 3 = Group Personal Pension scheme 4 = Group Stakeholder Pension 5 = Group Self Invested Personal Pension (SIPP) 6 = National Employment Savings Trust (NEST) 7 = Pension category unknown 0 = No pension with employer
Spens	Stakeholder pension marker for 1997-2004 back series	1 = Employee has stakeholder pension 2 = Employee does not have stakeholder pension
Spayd	Stakeholder payment marker for 1997-2004 back series	1 = Stakeholder pension paid through payroll deductions 2 = Stakeholder pension not paid through payroll deductions
Penpay	Pensionable pay - post 2004	The amount of employee's weekly pay that was pensionable
Ownpay	Employee contributions - post 2004	Weekly employee's pension contributions
Compay	Employer contributions - post 2004	Weekly employer's pension contributions
Ownperc	Employee's percentage pension contribution - post 2004	The employee's contributions made in the pay period as a percentage of pensionable pay
Comperc	Employer's percentage pension contribution - post 2004	The employer's contributions made in the pay period as a percentage of pensionable pay
Cosp	Contracted Out of State Second Pension – no longer available	1 = Yes 2 = No

	from 2016	0 = Missing . = Not available
Autopen	Automatically enrolled in a pension scheme - from 2016	1 = Yes 2 = No 0 = Missing
Salsac	Salary sacrifice (from 2013)	Were employee contributions made through a salary sacrifice arrangement 1 = Yes 2 = No 0 = Missing
Idbrsta	Inter-Departmental Business Register (IDBR) legal status	Legal status of the enterprise on the IDBR 1 = Private company 2 = Sole proprietor 3 = Partnership 4 = Public corporation & nationalised industries 5 = Central government 6 = Local authority 7 = Non-profit body or mutual association
Idbrnemp	IDBR employment	Number of employees in the enterprise on the IDBR
Luref	Local unit reference	Number generated to indicate multiple occurrences of the same local unit for disclosure checking purposes.
Alday	Annual leave (days)	Annual leave (days)
Pubpriv	Public Private 1997-2013	1 = Public (legal status 4, 5 and 6) 2 = Private (legal status 1, 2 and 3) 0 = Unclassified (includes legal status 7)
Pubpriv	Public Private from 2014	1 = Public (legal status 4, 5 and 6) 2 = Private (legal status 1, 2 and 3) 3 = Not for profit (legal status 7) 0 = Unclassified
Stratum	Relates to the way in which the data was collected from 2004	0= Special arrangements for large employers 1= Main despatch (employee in the same job between January (initial sample date) and April (survey reference date)) 2= Employee changed job between January and April 3= New employee job between January and April
Calwght	Calibration weight	Weights used to calibrate ASHE returns to job totals from the Labour Force Survey based on a combination of dweight and gweight
Dweight	Calibration weight	Based on stratum
Gweight	Calibration weight	Based on LFS population totals
Lpcalwght	Low pay calibration weight	Used for low pay analyses. Cases with loss of pay due to absence (or in industries T or U or unclassified) are firstly removed before calculating weights that sum to the number of jobs in the labour market (based on a combination of lpdweight and lpgweight)
Lpdweight	Low pay calibration weight	Based on stratum
Lpgweight	Low pay calibration weight	Based on LFS population totals
Ppstart	Pay period start date	This variable provides the start date of the employee's pay period. It takes the form DDMMYY.
Hrpayx	The derived hourly rate used in the low pay calculation	The derived hourly rate is hourly pay excluding overtime and shift premium pay.
Nlpflag	The 'not low pay' flag	See **Annex E** for the derivation. 1 = Not low pay 0 = Other
Pcflag	Payroll calculator flag	Payroll calculator flag 1 = Subject to a different payroll calculation 0 = Not identified as being subject to a different payroll calculation

Miss_ind	Imputation indicator	Indicates which variable(s) (bpay, bhr, ovpay, ovhrs and/or agp) is missing and requires imputation
Imp	Successful imputation flag	Indicates if the variable(s) has been successfully imputed 1 = Record imputed . = Record not imputed
Wgor	Work region	See **Annex B**
Hgor	Home region	See **Annex B**
Warea	Work area	See geography lookups
Harea	Home area	See geography lookups
Wla	Work local authority	See geography lookups
Hla	Home local authority	See geography lookups
Wpc	Work parliamentary constituency	See geography lookups
Hpc	Home parliamentary constituency	See geography lookups
Wtec	Work training enterprise council	See geography lookups
Htec	Home training enterprise council	See geography lookups
Wttw	Work travel-to-work area	See geography lookups
Httw	Home travel-to work area	See geography lookups
Wha	Work Health Authority	See geography lookups
Hha	Home Health Authority	See geography lookups
Whlth	Work Health Geography	See geography lookups Introduced April 2013 but applied from 2012 revised
Hhlth	Home Health Geography	See geography lookups Introduced April 2013 but applied from 2012 revised
Wpct	Work Primary Care Trust	See geography lookups
Hpct	Home Primary Care Trust	See geography lookups
Wccg	Work Clinical Commissioning Group	See geography lookups (England only) Introduced April 2013 but applied from 2012 revised
Hccg	Home Clinical Commissioning Group	See geography lookups (England only) Introduced April 2013 but applied from 2012 revised
Wcoa	Work census output area	
Hcoa	Home census output area	
Wlsoa	Work lower super output area	
Hlsoa	Home lower super output area	
Wmsoa	Work middle super output area	
Hmsoa	Home middle super output area	
Wnuts 1	Work Nomenclature of Territorial Units for Statistics	See geography lookups
Wnuts2 and 3	Work Nomenclature of Territorial Units for Statistics	See geography lookups
Wnuts4	Work Nomenclature of Territorial Units for Statistics	See geography lookups
Hnuts 1	Home Nomenclature of Territorial Units for Statistics	See geography lookups
Hnuts2 and 3	Home Nomenclature of Territorial Units for Statistics	See geography lookups
Hnuts 4	Home Nomenclature of Territorial Units for Statistics	See geography lookups
Hlau1 2012-2013	Home Local Administrative Unit (formally NUTS level 4)	See geography lookups
Wlau1 2012-2013	Work Local Administrative Unit (formally NUTS level 4)	See geography lookups
Hlau1nat from 2014	Home Local Administrative Unit (formally NUTS level 4)	See geography lookups
Wlau1nat from 2014	Work Local Administrative Unit (formally NUTS level 4)	See geography lookups
Wleps1 and 2	Work local enterprise partnerships codes from 2012	See geography lookups

Hleps1 and 2	Home local enterprise partnerships codes from 2012	See geography lookups
Wwac	Work National Assembly for Wales Constituency code from 2012	See geography lookups
Hwac	Home National Assembly for Wales Constituency code from 2012	See geography lookups
Wwer	Work National Assembly for Wales Electoral Region codes from 2013	See geography lookups
Hwer	Home National Assembly for Wales Electoral Region codes from 2013	See geography lookups
Wspc	Work Scottish Parliamentary Constituency codes from 2012	See geography lookups
Hspc	Home Scottish Parliamentary Constituency codes from 2012	See geography lookups
Wspr	Work Scottish Parliamentary Region codes from 2013	See geography lookups
Hspr	Home Scottish Parliamentary Region codes from 2013	See geography lookups
Lgd14w	Work NI Local government districts	See geography lookups
Lgd14h	Home NI Local government districts	See geography lookups

Annex A

Collective Agreement 1997 - 2004 – colag

No.	Organisation	Agreement detail
465	British Broadcasting Corporation	National joint agreement
448	Local authorities' service - England & Wales	Local Authority Single Status National Agreement
442	Local authorities' service - England & Wales	Craft & associated employees JNC
410	Local authorities' service - Scotland	Local Authority Single Status Agreement
411	Local authorities' service - Scotland	Craft operatives JNC
420	London Regional Transport	Railways, general & operating grades
427	National Health Service	Professional & Technical Staff A Whitley Council
428	National Health Service	Professional & Technical Staff B Whitley Council
429	National Health Service	Hospital doctors & dentists
430	National Health Service	Other doctors & dentists
431	National Health Service	Administrative & clerical staffs Whitley Council
432	National Health Service	Nurses & midwifery staff
433	National Health Service	Ancillary staffs Whitley Council
434	National Health Service	Maintenance staff
463	National Health Service	Ambulance staff Whitley Council
456	Police & Fire Services	Police service (ranks below superintendent only)
457	Police & Fire Services	Fire services - operational ranks below station officer
458	Police & Fire Services	Fire services - operational ranks from station officer to senior divisional officer
459	Police & Fire Services	Fire services - control room & non-operational staff
415	Prison Service	Prison service - prison officers
412	Civil Service	Inland Revenue
491	Post Office	Royal Mail clerical
492	Post Office	Royal Mail supervisory / specialist
493	Post Office	Royal Mail engineering & allied
494	Post Office	Royal Mail manuals
495	Post Office	Counters clerical
496	Post Office	Parcel force manuals
435	Teaching - England & Wales	Primary, secondary & special schools
438	Teaching - England & Wales	Staff in sixth form colleges JNC
437	Teaching - Scotland	Primary & secondary schools SJNC
460	Universities (old)	Clinical academic staff
461	Universities (old)	Computer operatives
466	Universities (old)	Academic & academic related staff
467	Universities (old)	Clerical staff
468	Universities (old)	Technical staff
469	Universities (old)	Manual staff
470	Universities (new)	Administrative, professional, technical & clerical staff
471	Universities (new)	Lecturers
479	Universities (new)	Manual staff
401	Others (not included in other categories listed)	National/industry agreement
402	Others (not included in other categories listed)	District agreement
403	Others (not included in other categories listed)	Company agreement

404	Others (not included in other categories listed)	Establishment agreement
405	Others (not included in other categories listed)	National/industry agreement supplemented by district/company or establishment agreement

Collective Agreement 2005 and later – colag

No.	Agreement detail
1	National or industry agreement
2	Sub-national agreement
3	Organisational agreement
4	Workplace agreement
5	National or industry supplemented by a sub-national, organisational or workplace agreement

Annex B

Region – wgor hgor

1	North East
2	North West
3	Yorkshire and The Humber
4	East Midlands
5	West Midlands
6	South West
7	East
8	London
9	South East
10	Wales
11	Scotland
12	Northern Ireland

Annex C – Median and percentile code

The ASHE system uses a methodology which produces slightly different median values to those produced by most statistical packages. So that users can replicate our published estimates, a brief description of the method is set out below followed by the SAS syntax. Hopefully the code will make sense even if you are not a SAS user but please let us know if you have any questions.

```
/*In SAS (proc means) the steps are roughly:

1. Sorts data values in ascending order
2. Calculates the cumulative weight
3. Works out the median point by dividing the sum of the weights by two
4. The median value is then the value where the associated cumulative weight contains
the mid-point value

The ASHE system is similar, but attempts to model for the fact that earnings are
continuous rather than discrete. The steps taken are:

1. Sorts data values in ascending order
2. Calculates the cumulative weight
3. Works out the median point by dividing the sum of the weights by two
4. The median is calculated by interpolating between the two data values depending on
where the mid-point falls in relation to the cumulative weights

This is a rough explanation but this is why you may get slight differences between the
figures you have calculated and those published on the ONS website.

/*This part remains the same for every median you require, only the keep statements
need to be updated*/

%macro median(input=,filter=,class=num,var=,output=,pcl=);

%if &class=num %then %do;
data ashe;
      set &input(keep=lop adr ft calwght &var wgor gpay sjd agp sex); /*update keep
statement if you need to include extra breakdowns such as occupation in your analysis*/
      &filter
      num=1;
run;
%end;

%else %do;
data ashe;
      set &input(keep=lop adr ft calwght &var &class wgor gpay sjd agp sex); /*update
keep statement if you need to include extra breakdowns such as occupation in your
analysis*/
      &filter
run;
%end;

proc means data=ashe n nway noprint;
      class &class;
      var calwght;
      output out=agg (drop=_type_ _freq_) sum=;
run;

data agg;
      set agg;
      median=calwght*&pcl;
run;
```

```
proc sort data=ashe;by &class;run;
proc sort data=agg;by &class;run;

data ashe_2;
      merge ashe(keep=lop adr ft calwght &var &class in=a) agg(keep=&class median);
      by &class;
run;

proc sort data=ashe_2;by &class &var;run;

data ashe_3;
      retain count 0 count2;
      set ashe_2;
      by &class &var;
      count+calwght;
      if first.&class then count2=count-calwght;
run;

data ashe_4(drop=count2);
      format status $6.;
      set ashe_3;
      count=count-count2;
      if count le median then status='lower';
      if count gt median then status='higher';
run;

proc sort data=ashe_4;by &class descending status &var;run;

data ashe_5(drop=del rename=(&var=highest count=higher));
      retain lower lowest;
      set ashe_4;
      by &class descending status &var;
      if last.status and status='lower' then del=1;
      if first.status and status='higher' then del=1;
      if last.status then do;lower=count;lowest=&var;end;
      if del ne 1 then delete;
      if status='lower' then delete;
run;

proc sort data=ashe_5;by &class;run;

data &output(keep=&class &var);
      set ashe_5;
      if highest-lowest=0 or higher-lower=0 then &var=lowest;
      else &var=lowest+((highest-lowest)/(higher-lower))*(median-lower);
run;

%mend median;


libname ds16 'M:\Datasets\2016';

data ashe16;
/*this would be your ASHE microdata set*/
set ds16.asheuk16_imp_wgt;
/*any additional coding should be done here*/
* Create Occupations;
                  if Occ00 = . then do;
                  Occ1 = substr(put(Occ10,4.),1,1);
                  Occ2 = substr(put(Occ10,4.),1,2);
```

```
                Occ3 = substr(put(Occ10,4.),1,3);
                Occ4 = substr(put(Occ10,4.),1,4);
                Occ1_10 = substr(put(Occ10,4.),1,1);
                Occ2_10 = substr(put(Occ10,4.),1,2);
                Occ3_10 = substr(put(Occ10,4.),1,3);
                Occ4_10 = substr(put(Occ10,4.),1,4);
                Occ1_00 = .;
                Occ2_00 = .;
                Occ3_00 = .;
                Occ4_00 = .;
        end;
        else do;
                Occ1 = substr(put(Occ00,4.),1,1);
                Occ2 = substr(put(Occ00,4.),1,2);
                Occ3 = substr(put(Occ00,4.),1,3);
                Occ4 = substr(put(Occ00,4.),1,4);
                Occ1_00 = substr(put(Occ00,4.),1,1);
                Occ2_00 = substr(put(Occ00,4.),1,2);
                Occ3_00 = substr(put(Occ00,4.),1,3);
                Occ4_00 = substr(put(Occ00,4.),1,4);
                Occ1_10 = .;
                Occ2_10 = .;
                Occ3_10 = .;
                Occ4_10 = .;
        end;
run;

/*For the lines of code below, you'll need to make sure that you enter the correct
filter dependent on which variable (annual or weekly) you're using
Ensure you change the pcl field to reflect whether you're extracting the median or a
specific percentile*/
/*The variables entered in the class field is the breakdown you require*/

*median AGP by 1 digit Occupation;
%median(input=ashe16,filter=where adr eq 1 and sjd eq 1 and wgor gt 0 and agp gt
0;,class=Occ1,var=agp,output=ashe_16_occ1_agp_median, pcl=0.5);
*10th percentile AGP by 1 digit Occupation;
%median(input=ashe16,filter=where adr eq 1 and sjd eq 1 and wgor gt 0 and agp gt
0;,class=Occ1,var=agp,output=ashe_16_occ1_agp_pcl, pcl=0.1);
*median GPAY by 1 digit Occupation;
%median(input=ashe16,filter=where adr eq 1 and lop eq 2 and wgor gt
0;,class=Occ1,var=gpay,output=ashe_16_occ1_gpay_median, pcl=0.5);

*The code falls over when you include more than one variable in the class field,
therefore the additional variables are included in the filter and then the tables are
appended together; This is when you need to alter the keep statement for new or dropped
variables;
*median GPAY by 1 digit Occupation and sex;
%median(input=ashe16,filter=where adr eq 1 and lop eq 2 and wgor gt 0 and
sex=1;,class=Occ1,var=gpay,output=ashe_16_occ1_m_gpay_median, pcl=0.5);
%median(input=ashe16,filter=where adr eq 1 and lop eq 2 and wgor gt 0 and
sex=2;,class=Occ1,var=gpay,output=ashe_16_occ1_f_gpay_median, pcl=0.5);

*You will then need to merge these together to produce one table of all the medians you
require;
data ashe_16_occ1_sex_gpay_median;
set ashe_16_occ1_m_gpay_median;
sex=1;
run;
data ashe_16_occ1_f_gpay_median;
set ashe_16_occ1_f_gpay_median;
```

```
sex=2;
run;
proc append base=ashe_16_occ1_sex_gpay_median data=ashe_16_occ1_f_gpay_median;run;

/*We'd suggest trying to match a few of our published figures with the output from the
code first to make sure it works within your systems*/
```

Annex D – Mean of hourly pay

For the majority of our pay variables the SAS proc means function can be used directly to calculate the mean of those pay variables.

For both hourly pay (**he**) and hourly pay excluding overtime (**hexo**), the calculation of the mean must be done slightly differently.

You must first calculate the sum of the pay and hours then use these to derive the mean of hourly pay/hourly pay excluding overtime.

The SAS code below can be used to replicate our published figures.

```
/*HE*/
proc means data=ashe17 noprint;
weight calwght;
var gpay thrs;
where adr=1 and lop=2 and wgor>0 and age>15;
output out=ashe17_he sum(gpay)= sum(thrs)= /autoname;
run;

data ashe17_he_mean;
set ashe17_he;
mean=gpay_sum/thrs_sum;
run;

/*HEXO*/
proc means data=ashe17 noprint;
weight calwght;
var gpox bhr;
where adr=1 and lop=2 and wgor>0 and age>15;
output out=ashe17_hexo sum(gpox)= sum(bhr)= /autoname;
run;

data ashe17_hexo_mean;
set ashe17_hexo;
mean=gpox_sum/bhr_sum;
run;
```

Annex E – Low pay derivations

Below is the SAS syntax we use to derive the variables used in the calculation of the number of jobs with hourly earnings below the NLW/NMW. Code has been provided for 2016 and 2017. Before the introduction of the NLW in 2016 the calculation was different. If you require the code for previous years please let us know.

hrpayx (derived hourly rate)
This is the derived hourly rate used in the low pay calculation. It is hourly pay excluding overtime and shift premium pay. It is derived as follows:

```
if bhr>0 then hrpay = 100*(gpox-sppay)/bhr;
hrpayx = round(hrpay*10000)/10000;
```

nlpflag ('not low pay' flag)
This identifies those employees whose pay period started before 1 April and whose derived rate is below the current NMW/NLW but who are paid at least the previous NMW/NLW rate and therefore are classified as not low paid for the purposes of low pay statistics. It is derived as follows:

/*2017*/

```
data ashe17nlpflag;
set ashe17;
appmonthchar = substr(PUT(appdate,6.),1,2);
appyearchar = substr(PUT(appdate,6.),3,4);
appmonth = INPUT(appmonthchar,2.);
appyear = INPUT(appyearchar,4.);
IF ((appmonth > 4 AND appmonth < 13 AND appyear = 2016) OR (appmonth > 0 AND appmonth < 5 AND appyear =
2017)) OR (app = 1 AND age = 16) THEN firstyear = 1;
ELSE firstyear = 0;
if ((age > 24 and app ne 1 and substr(ppstart,3,2)='03' and substr(ppstart,5,2)='17' and hrpayx ge 720 and hrpayx lt
750)
or (age > 20 and age < 25 and app ne 1 and substr(ppstart,3,2)='03' and substr(ppstart,5,2)='17' and hrpayx ge 695
and hrpayx lt 705)
or (age > 17 and age < 21 and app ne 1 and substr(ppstart,3,2)='03' and substr(ppstart,5,2)='17' and hrpayx ge 555
and hrpayx lt 560)
or (age < 18 and app ne 1 and substr(ppstart,3,2)='03' and substr(ppstart,5,2)='17' and hrpayx ge 400 and hrpayx lt
405)
or (age > 15 and age < 19 and app = 1 and substr(ppstart,3,2)='03' and substr(ppstart,5,2)='17' and hrpayx ge 340
and hrpayx lt 350)
or (age > 18 and app = 1 and firstyear = 1 and substr(ppstart,3,2)='03' and substr(ppstart,5,2)='17' and hrpayx ge 340
and hrpayx lt 350)
or (age > 18 and age < 21 and app = 1 and firstyear = 0 and substr(ppstart,3,2)='03' and substr(ppstart,5,2)='17' and
hrpayx ge 555 and hrpayx lt 560)
or (age > 20 and age < 25 and app = 1 and firstyear = 0 and substr(ppstart,3,2)='03' and substr(ppstart,5,2)='17' and
hrpayx ge 695 and hrpayx lt 705)
or (age > 24 and app = 1 and firstyear = 0 and substr(ppstart,3,2)='03' and substr(ppstart,5,2)='17' and hrpayx ge 720
and hrpayx lt 750)) then nlpflag=1;
else nlpflag = 0;
run;
```

/*2016*/

```
data ashe16nlpflag;
set ashe16;
if (age > 24 and substr(ppstart,3,2)='03' and substr(ppstart,5,2)='16' and hrpayx ge 670 and hrpayx lt 720) then
nlpflag=1;
else nlpflag = 0;
run;
```

lpmgx (low pay flag)
The code we used to produce the estimates for the number of jobs paid below the NMW/NLW (i.e. **lpmgx = 1**) in 2016 and 2017 is as follows:

```
/*2017*/

DATA lp2017;
        SET ashe17;
        WHERE age>15 and wgor>0 and lop=2 and lpcalwght NE . and hrpayx>0;
RUN;

DATA lp2017v2;
        SET lp2017;
        num=1;
        IF(bhr>0) THEN hrpay = 100*(gpox-sppay)/bhr;
        hrpayx = ROUND(hrpay*10000)/10000;                    /*Note - hrpayx already on dataset*/
        lpmgx = 0;
        IF age>15 and age<18 THEN agegroup=1;
        IF age>17 and age<21 THEN agegroup=2;
        IF age>20 and age<25 THEN agegroup=3;
        IF age>24 THEN agegroup=4;
        appmonthchar = substr(PUT(appdate,6.),1,2);
        appyearchar = substr(PUT(appdate,6.),3,4);
        appmonth = INPUT(appmonthchar,2.);
        appyear = INPUT(appyearchar,4.);
        IF ((appmonth > 4 AND appmonth < 13 AND appyear = 2016) OR (appmonth > 0 AND appmonth < 5 AND
      appyear = 2017)) OR (app = 1 AND age = 16) THEN firstyear = 1;
        ELSE firstyear = 0;
        IF((agegroup = 1 AND app ne 1 AND hrpayx < 405 AND hrpayx > 0) OR
          (agegroup = 2 AND app ne 1 AND hrpayx < 560 AND hrpayx > 0) OR
          (agegroup = 3 AND app ne 1 AND hrpayx < 705 AND hrpayx > 0) OR
          (agegroup = 4 AND app ne 1 AND hrpayx < 750 AND hrpayx > 0) OR
          (age > 15 AND age < 19 AND app = 1 AND hrpayx < 350 AND hrpayx > 0) OR
          (age > 18 AND app = 1 AND firstyear = 1 AND hrpayx < 350 AND hrpayx > 0) OR
          (age > 18 AND age < 21 AND app = 1 AND firstyear = 0 AND hrpayx < 560 AND hrpayx > 0) OR
          (agegroup = 3 AND app = 1 AND firstyear = 0 AND hrpayx < 705 AND hrpayx > 0) OR
          (agegroup = 4 AND app = 1 AND firstyear = 0 AND hrpayx < 750 AND hrpayx > 0)) THEN lpmgx = 1;
        IF nlpflag=1 THEN lpmgx = 0;
RUN;

/*2016*/

DATA lp2016;
        SET ashe16;
        WHERE age>15 and wgor>0 and lop=2 and lpcalwght NE . and hrpayx>0;
RUN;

DATA lp2016v2;
        SET lp2016;
        num=1;
        IF(bhr>0) THEN hrpay = 100*(gpox-sppay)/bhr;
        hrpayx = ROUND(hrpay*10000)/10000;                    /*Note - hrpayx already on dataset*/
        lpmgx = 0;
        IF age>15 and age<18 THEN agegroup=1;
        IF age>17 and age<21 THEN agegroup=2;
        IF age>20 and age<25 THEN agegroup=3;
        IF age>24 THEN agegroup=4;
        appmonthchar = substr(PUT(appdate,6.),1,2);
        appyearchar = substr(PUT(appdate,6.),3,4);
        appmonth = INPUT(appmonthchar,2.);
```

```
    appyear = INPUT(appyearchar,4.);
    IF ((appmonth > 4 AND appmonth < 13 AND appyear = 2015) OR (appmonth > 0 AND appmonth < 5 AND
  appyear = 2016)) OR (app = 1 AND age = 16) THEN firstyear = 1;
    ELSE firstyear = 0;
    IF((agegroup = 1 AND app ne 1 AND hrpayx < 387 AND hrpayx > 0) OR
      (agegroup = 2 AND app ne 1 AND hrpayx < 530 AND hrpayx > 0) OR
      (agegroup = 3 AND app ne 1 AND hrpayx < 670 AND hrpayx > 0) OR
      (agegroup = 4 AND app ne 1 AND hrpayx < 720 AND hrpayx > 0) OR
      (age > 15 AND age < 19 AND app = 1 AND hrpayx < 330 AND hrpayx > 0) OR
      (age > 18 AND app = 1 AND firstyear = 1 AND hrpayx < 330 AND hrpayx > 0) OR
      (age > 18 AND age < 21 AND app = 1 AND firstyear = 0 AND hrpayx < 530 AND hrpayx > 0) OR
      (agegroup = 3 AND app = 1 AND firstyear = 0 AND hrpayx < 670 AND hrpayx > 0) OR
      (agegroup = 4 AND app = 1 AND firstyear = 0 AND hrpayx < 720 AND hrpayx > 0)) THEN lpmgx = 1;
    IF nlpflag=1 THEN lpmgx = 0;
RUN;
```

Annex F – Living wage derivations

ONS publishes estimates, on request, of proportions of employee jobs with earnings less than the living wage (as promoted by the Living Wage Foundation). In 2015 we consulted with key stakeholders and agreed on a single method for this calculation. Below is the SAS syntax we use to perform the calculation for 2016 and 2017.

```
libname ds16 'M:\Datasets\2016';
libname ds17 'M:\Datasets\2017';

/*2016*/

DATA lw16;
      SET ds16.asheuk16_imp_wgt;
      num=1;
      LW = 0;
      if wgor=8 and Hrpayx < 940 then LW = 1;
      if wgor ne 8 and Hrpayx < 825 then LW = 1;
      WHERE adr = 1 AND lop = 2 AND wgor > 0 AND age > 17 AND hrpayx ne . AND lpcalwght
ne .;
RUN;

proc means data = lw16 noprint;
class lw;
weight lpcalwght;
var num;
output out = lw16 (drop=_type_) sum=;
run;

/*2017*/

DATA lw17;
      SET ds17.asheuk17prov_imp_wgt;
      num=1;
      LW = 0;
      if wgor=8 and Hrpayx < 975 then LW = 1;
      if wgor ne 8 and Hrpayx < 845 then LW = 1;
      WHERE adr = 1 AND lop = 2 AND wgor > 0 AND age > 17 AND hrpayx ne . AND lpcalwght
ne .;
RUN;

proc means data = lw17 noprint;
class lw;
weight lpcalwght;
var num;
output out = lw17 (drop=_type_) sum=;
run;
```

Annex G – Continuous employment

In our ASHE bulletin we include a section on the change in median earnings for those employees in "continuous employment", i.e. in the same post for at least one year. There are a number of different ways that this could be calculated depending on the assumptions made. We use a paired year approach as, in the absence of longitudinal weights on ASHE, there is a greater risk that analysis over a longer time period could be biased due to sample attrition. Below is the SAS syntax we used to produce the 2016-2017 continuous employment datasets, on which the change in median earnings was calculated.

```
libname ds16 'M:\Datasets\2016';
libname ds17 'M:\Datasets\2017';

data a16orig;
set ds16.asheuk16_imp_wgt;
run;

proc sort data = a16orig;by serno;run;

data a17orig;
set ds17.asheuk17prov_imp_wgt;
run;

proc sort data = a17orig;by serno;run;

data a16 (keep=sernol ft adr lop);
set a16orig;
rename serno=sernol;
run;

proc sort data = a16;by sernol;run;

data a17 (keep=serno sernol ft sjd adr lop);
set a17orig;
run;

data a17v2;
set a17;
rename ft=ftt adr=adrt lop=lopt;
run;

proc sort data = a17v2;by sernol;run;

*Merge to get file of continuous employment from 2016 to 2017;

data d16_17;
merge a16(in=a) a17v2(in=b);
by sernol;
if a and b;
run;

data d16_17v2;
set d16_17;
where adrt=adr and lopt=lop and ftt=ft and sjd=1;
run;

data cont16(keep=sernol);
set d16_17v2;
run;

data cont17(keep=serno);
```

```
set d16_17v2;
run;

*Merge these files onto datasets to produce full datasets with continuous employment;

proc sort data = cont16;by sernol;run;
proc sort data = cont17;by serno;run;

data cont16;
set cont16;
rename sernol=serno;
run;

data cont17from16;
merge cont17 (in=a) a17orig (in=b);
by serno;
if a;
run;

data cont16v2;
set cont16;
run;

data cont16into17;
merge cont16v2 (in=a) a16orig (in=b);
by serno;
if a;
run;
```