



ADR
NORTHERN IRELAND

Earnings and Employees Study (EES) 2011

Information Session

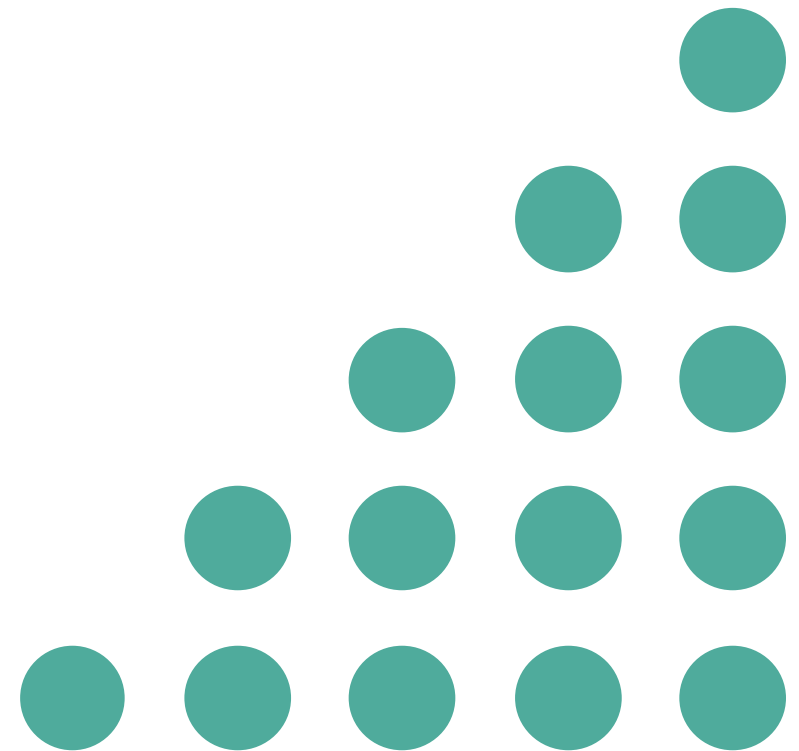
Research Support Unit, NISRA
Date: 30th November 2021



Overview of the Earnings and Employees Study (EES) 2011

- Introduction to the EES
- Linkage
- Quality assurance
- Applying for the data
- Metadata

Introduction to the EES



Census and Annual Survey of Hours and Earnings

Annual Survey of Hours and Earnings (ASHE)

ASHE provides information about the levels, distribution and make-up of earnings and hours paid for employees by gender and work pattern (full-time and part-time). Estimates are available for various breakdowns including industries, occupations, geographies and age groups within the UK.

The ASHE is collected under the Statistics of Trade and Employment (NI) Order 1988.

Census

The Census collects information every 10 years about people and households in Northern Ireland. It is used by central and local government, health authorities and many other organisations to plan and provide future services.

Censuses in Northern Ireland are governed by the Census Act (Northern Ireland) 1969.

Earnings and Employees Study (EES) 2011

Description

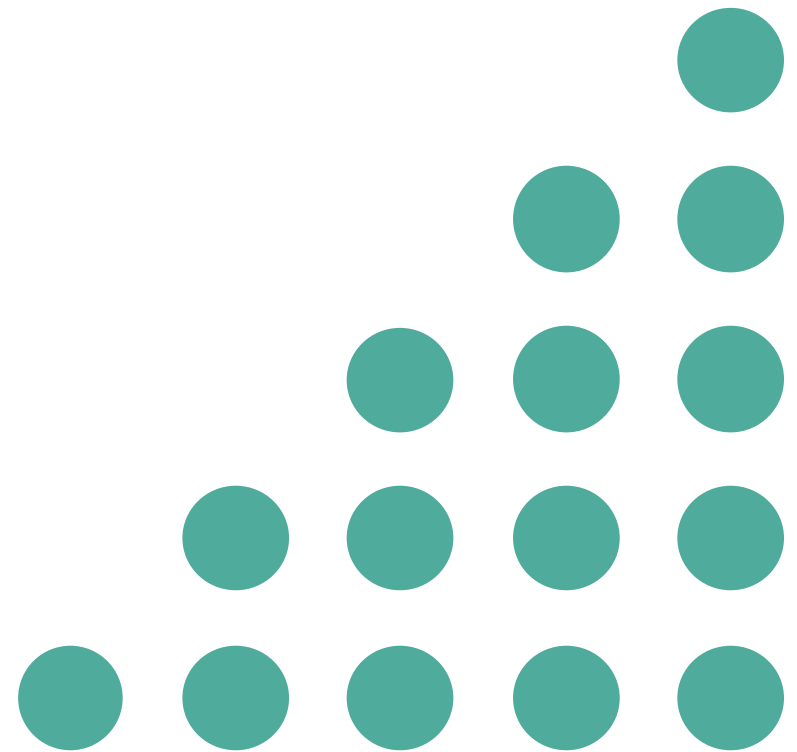
- The EES 2011 links together variables from the Annual Survey of Hours and Earnings (ASHE) 2011 with variables from the Census of Population and Housing 2011, and Capital Value data from the Land and Property Services.
- The dataset takes the form of one pre-linked table.

Sample

- The ASHE sample is comprised of 5,770 distinct ASHE records, approximately 1% of all employees in NI who were covered by Pay As You Earn (PAYE) schemes.
- These records are linked to 5,528 distinct Census records. The difference in the number of distinct ASHE and distinct Census records is because several individuals within the dataset have more than one job.



Linkage



Data sources

DATASET 1	
DATA SUPPLIER:	Economic and Labour Market Statistics Branch
DESCRIPTION:	Annual Survey of Hours and Earnings (ASHE) - a UK wide business survey that provides estimates on hourly, weekly and annual earnings for employee jobs. The ASHE is a 1% sample taken from the HM Revenue and Customs Pay as You Earn System.
NUMBER OF RECORDS:	5,770 records for 2011 5,528 individuals for 2011
PRE-MATCHING CHECKS	
<ol style="list-style-type: none">1. The ASHE has a very high response rate and there are no imputed records in the data.2. Individuals selected for ASHE 2011 were eligible for matching. This data was supplemented with 2012-2018 ASHE data for those individuals selected for 2011 only, in order to maximise the response rate.3. Information available to the TTP for linkage purposes:<ul style="list-style-type: none">• Name information which includes surname, forename initial and mid-name initial,• Home and work postcode,• Sex and• Date of birth.	

Data sources

DATASET 2	
DATA SUPPLIER:	Census Office
DESCRIPTION:	Extract from the 2011 Census (entire enumerated population aged 16 and over including those both economically active and economically inactive)
NUMBER OF RECORDS:	1,365,124
PRE-MATCHING CHECKS	
<ol style="list-style-type: none">1. The data did not include those individuals who were imputed due to no response which accounted for 5% of the total population. <i>Match rates have not been adjusted for imputation as the data only included enumerated individuals.</i>2. All 1,365,124 records contained enough data to enable them to be available for linkage.3. Information available to the TTP for linkage purposes :<ul style="list-style-type: none">• Name information which includes forename, mid-name and surname,• Address information at the time of the Census,• Sex and• Date of birth.	

Linkage method

'Rules based' or deterministic linkage approach using Match-Keys

ID	Forename	Surname	Sex	DOB	Postcode	Match-key 1: Forename, Surname, Sex, DOB, Postcode
Dataset 1	John	Smith	M	01/01/1990	BT6 XXX	JOHNSMITHM01011990BT6XXX

ID	Forename	Surname	Sex	DOB	Postcode	Match-key 1: Forename, Surname, Sex, DOB, Postcode
Dataset 2	John	Smith	M	01/01/1990	BT6 XXX	JOHNSMITHM01011990BT6XXX

- A single match-key alone cannot resolve all of the differences that occur between data sources, hence the need for multiple Match-Keys.
- A series of Match-Keys have been developed.
- The strongest level of linking is exact linking, which links pairs of records that are identical on all linking fields.
- Typically, the Match-Keys are processed in a stepwise manner starting with the most exact match-key and working down to the last match-key.

Donor imputation

- Donor imputation work was carried out to link the remaining unmatched ASHE individuals to Census individuals with similar characteristics.
- Followed the same deterministic linkage approach using Match-Keys as was used for the initial linkage. Using an iterative approach, the variables used to create the Match-Keys were:

Match-Key	Variables
1	3 digit SIC07, 4 digit SOC10, Gender, Age, Postcode Sector
2	2 digit SIC07, 4 digit SOC10, Gender, Age, Postcode Sector
3	3 digit SIC07, 4 digit SOC10, Gender, Age, Postcode District
4	2 digit SIC07, 4 digit SOC10, Gender, Age, Postcode District

- This process resulted in individuals in the ASHE dataset being linked to a pool of potential donor records in the 2011 Census dataset.
- This pool of individuals had similar characteristics to those in the ASHE dataset.
- Once the unmatched ASHE records had been linked to a pool of potential donors within the Census data a random record was selected to be the match for that ASHE individual.

Resulting match

- The resulting match was of high coverage and accuracy.
- The research dataset contains a fully linked data set of 5,528 ASHE individuals or 5,770 ASHE records as a small number of individuals have more than one job.

	Unadjusted match rate	Accuracy
ASHE 2011 to Census 2011	88.7%	100%

	Matched records	Donor records
ASHE 2011 to Census 2011	88.7%	11.3%

Quality assurance



Quality assurance

- Quality assurance practices were applied by the Research Support Unit (RSU) to the EES 2011 dataset to ensure the data was accurate, reliable, and representative:

Consistency within the data

Data for anomalies, unexpected values, and issues with aggregations

Consistency with published reports

Consistent with and representational of the whole population

Statistical Disclosure Control (SDC)

Industry and Occupation codes are deemed sensitive

Weighting

Information can be provided to researchers (on request) on the impact that weighting has when using different statistical packages.

Donor records comparison

No donor flags are included in the dataset and thus donor records will not be made known to the researchers or RSU.

Consistency with published reports

A key component of the QA process was to assess whether the EES data was both consistent with and representational of the whole population.

Consistency with ASHE dataset:

- Using weighted figures and through the application of appropriate filters, NISRA sought to replicate the headline statistics for hourly, weekly and annual earnings published in the ASHE 2011 Statistics Bulletin.
- As the EES 2011 is comprised of the full ASHE dataset (a 1% of all employees in NI who were covered by Pay As You Earn (PAYE) schemes), any inconsistency with the published report was minor, and fully explicable by the presence of donor records, the software used for the calculations, and the application of weighting.
- As such, NISRA noted that the Earnings and Employees dataset was both consistent with and representational of the ASHE sample.

Consistency with 2011 Census and NILS:

- Unweighted frequency distributions were calculated for each of the Person and Household variables within the EES dataset. RSU compared these findings to the Census 2011 summary report, NINIS, and the NILS.
- Bespoke frequencies were calculated from the published figures or NILS based on a demographic profile that best matched that of the EES (economically active (excluding self-employed and unemployed) and aged 16-74 years (where available)).
- Once again, NISRA noted that the linked dataset was consistent with and representational of the whole population.
- Where there were discrepancies between the Census 2011/NILS-Members 2011 and the Earnings and Employees sample distributions, this was due to bespoke aggregation of variables by Census or the unavailability of a published figure for a similar sample for comparison.

Consistency with published reports - ASHE

	ASHE-2011		Employees and Earnings Sample weighted*	
	Median	Mean	Median	Mean
Gpay (Gross pay - weekly)	354.5	423.2	354.6	423.2
Gpox (Gross pay excluding overtime - weekly)	343.1	408.7	343.3	408.7
Bpayinc (Basic Pay including other pay)	329.8	397.5	329.8	397.5

* Weight used = calwght, Adr (Adult rate marker) =1, Lop (Loss of pay marker) =2 & age >15

Consistency with published reports – Census

Marital and Same Sex Civil Partnership Status

	Single (never married or in a same-sex civil partnership) (%)	Married or in a registered same-sex civil partnership (%)	Separated (but still legally married or still legally in a same-sex civil partnership) (%)	Divorced or formerly in a same-sex civil partnership which is now legally dissolved (%)	Widowed or surviving partner from a same-sex civil partnership (%)
Census 2011*	41.3	49.0	3.9	4.8	1.0
Earnings and Employees Sample Unweighted	37.6	52.6	3.7	4.9	1.3

*Source Table CT0306NI: MARITAL OR CIVIL PARTNERSHIP STATUS BY ECONOMIC ACTIVITY (economically active (excluding self-employed and unemployed); aged 16-64)

Consistency with published reports – Census

Census 2011 - Country of Birth - Five Cat

	Northern Ireland (%)	Republic of Ireland (%)	Other UK (%)	Other EU (%)	Other (%)
Census 2011*	86.0	2.1	5.4	4.0	2.6
Earnings and Employees Sample Unweighted	86.8	1.9	4.6	4.3	2.4

*Source Table DC2606NI: ECONOMIC ACTIVITY BY COUNTRY OF BIRTH BY AGE BY SEX (economically active (excluding self-employed and unemployed); aged 16-74)

Religion Brought Up In

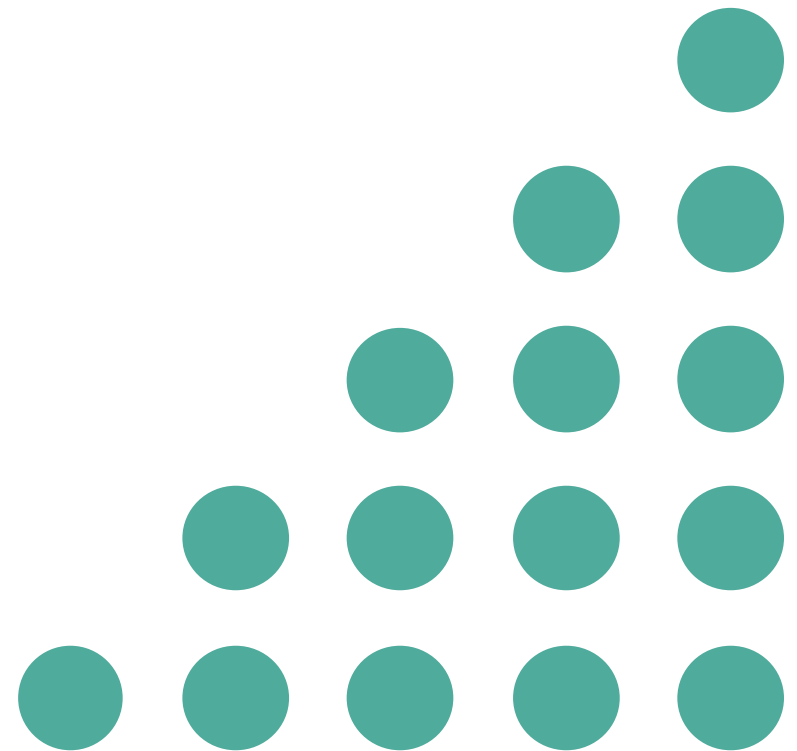
	Catholic (%)	Protestant and other Christian (including Christian related) (%)	Other Religions (%)	None (%)
Census 2011*	44.1	49.9	0.9	5.1
Earnings and Employees Sample Unweighted	45.1	49.6	0.9	4.5

*Source Table DC2617NI: ECONOMIC ACTIVITY BY RELIGION OR RELIGION BROUGHT UP IN BY AGE BY SEX (economically active (excluding self-employed and unemployed); aged 16-74)

Weight variables

- The use of weights within survey data is important to help ensure that the data is representative of the overall population.
- EES dataset includes weight variables, which have been calculated and used in the production of the ASHE published tables.
- When replicating the published ASHE figures it is important that the correct weight is used, the most frequently used weighting variable is 'calweight'.
- It is also important to include the correct filters.
- For more information on weights and filters please refer to the [ASHE dataset notes](#).

Applying for the data



Accessing the data

Researchers should contact RSU in the first instance and will be asked to complete a one page proposal. Once this has been approved, researchers and their project will need to be accredited under the Digital Economy Act.

Accredited researchers

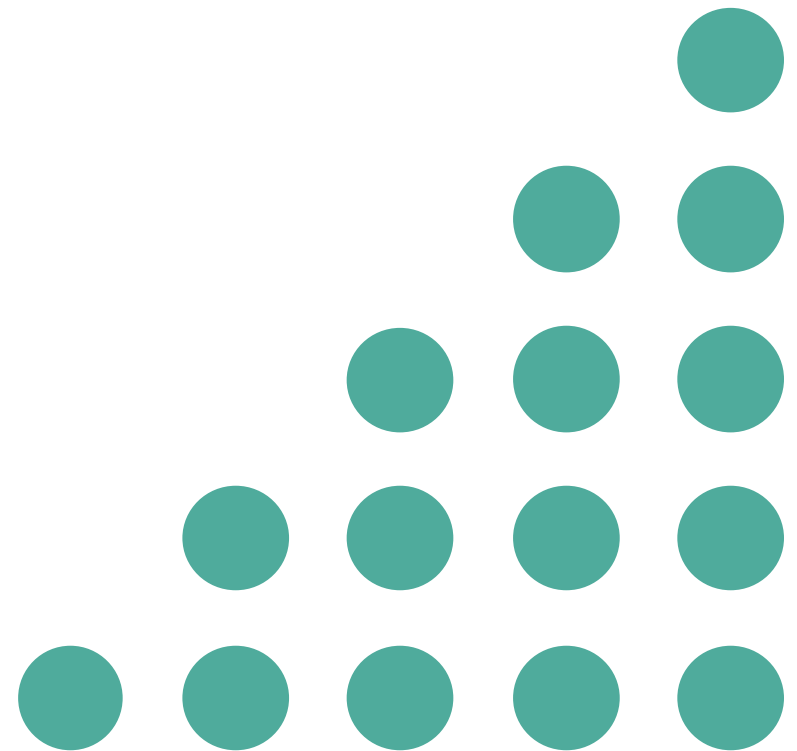
Access to the EES dataset is limited to researchers and research teams with [accredited researcher status](#) under the DEA. Information on how to apply for accreditation is available on the NISRA and ONS websites.

Accreditation of projects

RSU will organise an Project Sub Group meeting with research teams and the data providers. Researchers will be required to complete an application for approval by the UK Statistics Authority Research Accreditation Panel (RAP).

Data can be accessed in NISRA or on the ONS SRS subject to approval on a project by project basis.

Metadata



Metadata – ASHE variables

- Over 60 variables including:

Demographic variables	Sex Age-topcoded
Job markers	Full / part time marker Permanent / temporary marker Adult rate marker Loss of pay marker Same job marker Double job marker Main job marker
Industry and occupation	Sensitive variables. Researchers will be provided with 1 or 2 level codes. Consideration will be given to more detailed levels if justification is provided.
Hours worked	Pay period Basic paid hours Paid overtime hours Total paid hours

Metadata – ASHE variables

- Over 60 variables including:

Hours worked	Pay period Basic paid hours	Paid overtime hours Total paid hours
Pay details	Basic Pay Basic Pay including Other Pay Other Pay Gross pay - weekly Gross pay excluding overtime - weekly Overtime pay All incentive pay paid in this pay period Incentive pay paid in this period relating to the pay period	Shift and premium payments Hourly earnings Hourly earnings excluding overtime Stated total pay Stated hourly rate of pay Annual gross pay Annual incentive pay Benefits in kind filter Annual value of benefits in kind
Other pay variables	Pension provision filter Type of pension scheme – 2005-2012 Pensionable pay Employee contributions Employer contributions	Employee's percentage pension contribution Employer's percentage pension contribution

Metadata – ASHE variables

- Over 60 variables including:

Other variables	Employee start date Collective agreement IDBR employment Annual leave (days)	Public Private 1997-2013 Work NI Local government districts (2014) Home NI Local government districts (2014)
Weights	Calibration weight	Low pay calibration weight

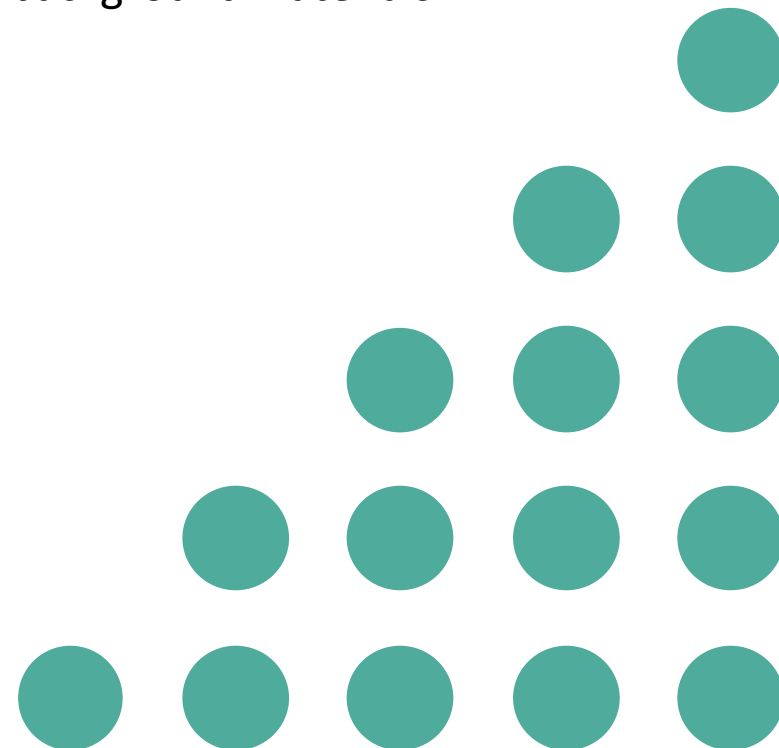
Metadata – Census variables

- Over 30 variables including:

<p>Person variables</p>	<p>Marital and Same Sex Civil Partnership Status Student Indicator Country of Birth - Five Cat Country of Migrant Origin Country of Previous Residence Passports National Identity Ethnicity Current Religion Religion – Brought Up In</p>	<p>Aggregation of Main Language Irish Language Ability Ulster-Scots Language Ability Disability Health Conditions Indicator Health Care Highest Level of Qualification Voluntary Work</p>
<p>Household variables</p>	<p>Household Size Alternative Household Composition Adaptions of Accommodation Tenure, Dwelling Number of Cars or Vans in Household Number of Adults in Household</p>	<p>Household Dependent Children Number of Carers in Household Deprived - Education Deprived - Health Deprived - Housing Deprived - Tenure</p>

Further information

Please visit the [ADRNI themed datasets](https://www.nisra.gov.uk/support/research-support/administrative-data-research-northern-ireland-adr-ni-themed-datasets) section of the NISRA website (<https://www.nisra.gov.uk/support/research-support/administrative-data-research-northern-ireland-adr-ni-themed-datasets>) for more information on the EES. There you will find links to the metadata and other background materials.





ADR NORTHERN IRELAND

- The Administrative Data Research Network takes privacy protection very seriously. All information that directly identifies individuals will be removed from the datasets by trusted third parties, before researchers get to see it.
- All researchers using the Network are trained and accredited to use sensitive data safely and ethically, they will only access the data via a secure environment, and all of their findings will be vetted to ensure they adhere to the strictest confidentiality standards.
- The help provided by the staff of the Administrative Data Research Network Northern Ireland (ADR-NI) and the Northern Ireland Statistics and Research Agency (NISRA) Research Support Unit is acknowledged. The ADR-NI is funded by the Economic and Research Council (ESRC). The Census and ASHE data have been supplied for the sole purpose of this project.

Find us at adruk.org or on
Twitter [@ADR_UK](https://twitter.com/ADR_UK)

Thank you