

Data Explained

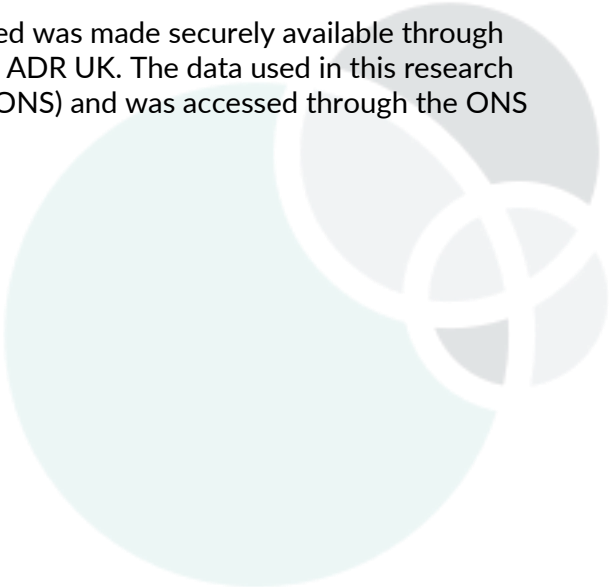
Annual Survey of Hours and Earnings linked to 2011 Census – England and Wales

Author: John Forth, Alex Bryson, Van Phan, Felix Ritchie, Carl Singleton, Lucy Stokes and Damian Whittard

Date: July 2023

This Data Explained summarises experiences and learning from working with the Annual Survey of Hours and Earnings (ASHE) linked to 2011 Census dataset. This publication is intended to help guide future researchers using this data and to provide feedback into future dataset development and documentation.

The administrative data discussed in this Data Explained was made securely available through the [Wage & Employment Dynamics](#) project funded by ADR UK. The data used in this research project comes from the Office for National Statistics (ONS) and was accessed through the ONS Secure Research Service.

A large, light teal decorative graphic in the bottom right corner, consisting of a large circle partially overlapping a white circle, mirroring the ADR ENGLAND logo.

Introduction

The ASHE dataset is an annual survey based on a 1% sample of employee jobs and conducted by the ONS. This de-identified dataset contains information on employees' earnings, paid hours, occupation and employer. To expand the number of de-identified, personal characteristics that can be observed for employees, the ASHE dataset has been linked to the 2011 Census for England and Wales.

The data linking process used employee records from ASHE in 2010, 2011 and 2012. ONS sought to identify these employees in the 2011 Census for England and Wales by matching on a set of common identifying fields. Where a robust link was made, a range of personal characteristics, including the employee's educational qualifications, country of birth, ethnicity, religion, and disability status, have been matched onto ASHE from the census. These census data items are only available for individuals in either ASHE 2010, 2011 or 2012, who also participated in the 2011 Census, and for whom a robust match could be made on a set of common identifying fields. However, the data items are available for these matched individuals in any year that they appear in ASHE from 2001-2018.

The linked ASHE - 2011 Census dataset allows researchers to examine the pay and career progression of a cohort of individuals who were in employment around the time of the 2011 Census. This data is linked and made available under the provisions of the 2017 Digital Economy Act, which provides a legal gateway for researchers to access government data in a secure way.

Information contained in the ASHE – 2011 Census linked dataset

- Personal characteristics:** Age, gender, residential location, country of birth, English language proficiency, ethnicity, religion, marital or civil-partnership status, health and disability, educational and vocational qualifications, household circumstances.
- Job characteristics:** Paid hours, usual hours, occupation.
- Employer characteristics:** Size (employees), industry, legal status, workplace location.

Office for National Statistics

ADRUK
Data-driven change

How is the data collected

The Annual Survey of Hours and Earnings (ASHE) is based on a 1% sample of employee jobs, taken from His Majesty's Revenue and Customs (HMRC) Pay As You Earn (PAYE) records. Employees are selected into the sample by virtue of the last two digits of their National Insurance number. The survey is completed by their employer, who receives the survey request from the ONS in April. Approximately 150,000 returns are completed each year.

Since ASHE is completed by employers using data from their payroll records, it provides highly accurate information on employees' earnings. However, the survey dataset contains few personal characteristics for the employee. The 2011 Census for England and Wales has three specific attributes that make it suitable as a potential donor of such information for ASHE sample members.

First, it is linkable to ASHE: the original datasets for ASHE and the 2011 Census held by ONS share a set of common fields that can be used to link individual records with a high degree of confidence. Second, the reference date is close to that of ASHE: the 2011 ASHE survey has a reference date of 13 April 2011, while the 2011 Census has a reference date of 27 March 2011. Third, the 2011 Census is comprehensive: by definition, one would expect almost all employees working in England and Wales on 13 April 2011 to be present in the dataset.

The process of linking ASHE to the 2011 Census for England and Wales proceeded in two phases. In phase 1 of the linking process, the ONS sought to establish links between ASHE and the 2011 Census by matching on an employee's name, sex, age, and residential postcode. Robust links were made for 62% of the employee job records in the 2011 ASHE belonging to employees resident in England and Wales. In many cases where a robust link could not be made, this was due to the absence of name data in the ASHE dataset. In phase 2 of the linkage process, the ONS expanded the list of linkage variables to include the employee's occupation, industry sector and workplace location. By including cases that match on sex, age, residential postcode, workplace location and occupation/industry, the ONS were able to increase the match rate to 74%.

The second edition of ASHE- 2011 Census dataset (released in May 2023) contains the full set of linked records arising from the two phases of the linking process.

Key variables

The ASHE- 2011 Census dataset contains four sets of variables:

ASHE variables: The dataset contains all variables from the ASHE microdata made available in the ONS Secure Research Service, with the exception of variables relating to residential geography below the level of local authority district. The removal of ASHE residential geography below this level is intended to maintain the confidentiality of census respondents.

Census variables: The dataset contains 317 variables taken from the 2011 Census dataset held in the ONS Secure Research Service. All Census variables in the ASHE-2011 Census dataset have the name suffix `_census`. Variables are labelled, but values remain unlabelled; code frames are provided in the core documentation library for 2011 Census.

Linkage variables: The dataset contains some variables that document aspects of the linkage process. A selection are described in the table below. All have the name suffix `_link`; variables and values are labelled. These variables are provided so that researchers can examine the sensitivity of their results to the inclusion or exclusion of cases with differing levels of match quality (e.g., selecting only those cases with a match score of 1.0).

Variable name	Type	Description
<code>anycensus</code>	Numeric	Dummy variable identifying ASHE records that have been linked to 2011 Census records
<code>desc_link</code>	String	Identifies the method by which the ASHE and 2011 Census records were linked for this case
<code>score_link</code>	Numeric	Overall match score (0-1)

Weighting variables: A weighting variable (`acew11_wt`) has been compiled to ensure that estimates based on those ASHE job records that could be linked to the 2011 Census are representative of all jobs held by employees resident in England and Wales. This weight variable is observed for all cases belonging to employees residing in England and Wales ($1 \leq hgor \leq 10$) and has been derived in each year of the data from 2001 to 2018.

What can the data be used for

This dataset will help to fill a crucial evidence gap in information about the experiences of individuals in the labour market. Importantly, for example, ASHE - 2011 Census will enable researchers to understand the underlying causes of intersectional pay gaps and provide a deeper understanding of how demographic factors have informed labour market transitions.

The dataset could help answer a range of research questions to inform policy, such as:

- What role do employers play in wage inequality?
- What are the returns on investment from education?
- What is the relationship between migration and the labour market?
- How does wage progression differ depending on characteristics such as gender, disability, or ethnicity?
- How do personal characteristics such as household structure affect wage progression?

Existing research or examples of previous research

The *ASHE – 2011 Census User Guide* (Forth et al, 2022) provides some illustrations of the analytical capabilities of the ASHE – 2011 Census linked dataset. Firstly, it illustrates the wage premium experienced by those with degree-level qualifications, and the wage penalties experienced by those with a disability and those born outside the UK. Secondly, it illustrates how the share of earnings variance that can be explained by observable characteristics increases after expanding the list of explanatory variables beyond those present in ASHE to include personal characteristics taken from the 2011 Census.

Subsequent research by the Wage & Employment Dynamics team (Phan et al, 2022) has used the ASHE – 2011 Census dataset to investigate the role of employer characteristics in explaining wage differences between white and ethnic minority employees. The employer has typically been an omitted variable in studies of ethnic wage gaps, which have tended to rely on self-reported wage and hours data from household surveys. The research finds that employer-specific wage effects account for sizeable parts of the estimated differences between the wages of white and ethnic minority workers at the mean and other points in the wage distribution. In household survey data, the role of such effects would have been attributed either to other worker characteristics that correlate with firm-specific pay rates, such as education levels, workplace location and occupations, or to the unexplained part of these wage gaps. The ASHE-2011 Census dataset has thus enabled researchers to obtain a richer understanding of how ethnicity affects one's earnings in the labour market.

Data limitations encountered

The ASHE – Census 2011 data includes a wide array of variables from the 2011 Census that describe the personal characteristics of the de-identified employees. However, it includes fewer variables to describe employees' family or household circumstances. The Wage & Employment Dynamics team are liaising with the ONS to add further variables to describe the economic activity of other members of employees' family or household; this is work in progress.

As the employee characteristics that are inherited from the census are necessarily self-reported at a single point in time, the ASHE-2011 Census dataset is best suited to the analysis of characteristics that are likely to take the same value whenever they are collected, such as a person's country of birth. Other characteristics, such as disability status, religion or even ethnicity may change over time. These characteristics are increasingly likely to be measured with error as one moves further away from 2011.

2011 Census data is only available for employees who were observed in ASHE 2010, 2011 or 2012. A person who was observed in earlier years of ASHE but who retired from paid employment in 2009 would then not have any 2011 Census information available. The same would be true of a person who appears in ASHE for the first time in 2013. The implication is that the availability of census information across the ASHE panel dataset depletes as one moves further away from the linkage years. For instance, only half (51%) of the employee jobs in ASHE 2016 held by employees residing in England and Wales have linked data from 2011 Census.

Suggested future data linkages

The Wage & Employment Dynamics team would like to see the ASHE data linked to the Census 2021 for England and Wales and the 2022 Census for Scotland.

References

Forth, J. Bryson, A. Phan, V. Ritchie, F. Singleton, C. Stokes, L. and Whittard, D. (2022). ASHE – Census 2011 Data Linkage: User Guide for Drop 2 of the ASHE – Census 2011 Dataset, Bristol: Wage and Employment Dynamics project.

Office for National Statistics (2023). Annual Survey of Hours and Earnings linked to 2011 Census - England and Wales, released 03 July 2023, ONS Secure Research Service Metadata Catalogue. DOI: <https://doi.org/10.57906/80f7-te97>

Phan, V. Singleton, C. Bryson, A. Phan, V. Ritchie, F. Stokes, L. and Whittard, D. (2022). Accounting for firms in ethnicity wage gaps throughout the earnings distribution, IZA Discussion Paper No. 15284.



Disclaimer

This work was produced using statistical data owned by ONS and accessed through the ONS Secure Research Service. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analyses of the statistical data. The work uses research datasets which may not exactly reproduce National Statistics aggregates.

Acknowledgements

The ASHE - 2011 Census linked dataset was produced as a collaboration between ADR UK, the ONS and the Wage & Employment Dynamics project team. The Wage & Employment Dynamics project (wagedynamics.com) is funded by ADR England, an investment of the Economic and Social Research Council (ESRC), part of UK Research and Innovation (UKRI). Grant number: ES/T013877/1.

The Wage & Employment Dynamics project team comprises Dr Damian Whittard, Prof Felix Ritchie and Dr Van Phan from the University of the West of England, Prof Alex Bryson from University College London, Dr John Forth from City, University of London, Dr Carl Singleton from the University of Reading and Lucy Stokes from the National Institute of Economic and Social Research.

Contact

Name: John Forth

Email: john.forth@city.ac.uk

