

Article

# Young people's earnings progression and geographic mobility, England and Wales: tax year ending 2012 to tax year ending 2016

Young people's earnings progression with comparisons by various personal characteristics, and the effect of geographic mobility on earnings growth.

Contact:  
Ruth Davies  
policy.evidence.analysis@ons.  
gov.uk  
+44 (0)1633 651827

Release date:  
23 October 2018

Next release:  
To be announced

## Table of contents

1. [Disclaimer](#)
2. [Main points](#)
3. [Things you need to know about this release](#)
4. [What is included in the dataset](#)
5. [Earnings growth higher for younger people](#)
6. [Young people most likely to experience relative earnings progression](#)
7. [Sex and ethnicity are important factors in earnings progression](#)
8. [More than one in five young people on relative low annual pay progressed to higher earnings](#)
9. [Higher skilled young people more likely to move](#)
10. [Geographic mobility associated with higher earnings growth](#)
11. [Quality and methodology](#)
12. [Authors and acknowledgements](#)

# 1 . Disclaimer

These Research Outputs are experimental statistics on earnings progression. This analysis is based on a feasibility study exploring the use of administrative data linked to the 2011 Census.

It is important that the research presented here be read alongside the quality and methodology information to aid interpretation and avoid misunderstanding. These outputs must not be reproduced without this disclaimer and warning note.

## 2 . Main points

- Young people aged 16 to 24 years had median earnings growth of 28.7% from tax year ending (TYE) 2015 to TYE 2016. This is higher than all other age groups and compares with 6.8% for those aged 25 to 29 years. This is in part due to transitioning from part-time to full-time work.
- Of the 5.2 million young people aged 16 to 29 years who had earnings in both TYE 2015 and TYE 2016, there were 16% that experienced progression of at least two earnings deciles.
- For young people, aged 16 to 29 years, the level and growth of median annual earnings were generally lower for women than men, whether they were degree- or school-educated, or had no qualifications.
- Young people's earnings progression is related to geographic mobility. The highest average annual growth in earnings was 22% for those moving to London between TYE 2012 and TYE 2016, compared with 7% for those that did not move local authority or moved elsewhere.
- City regions tended to attract young people from nearby local authorities; for example, 33% of movers to London were from the South East and 36% of movers to Greater Manchester were from the North West.

## 3 . Things you need to know about this release

In its role as the largest producer of independent official statistics in the UK, Office for National Statistics (ONS) continues to provide the data and analysis that helps to understand how people in the UK experience life. This information has traditionally come from surveys or the census, however, these sources often cannot provide enough detail or timeliness. For this reason, ONS is developing its use of administrative and linked data and this article provides research into how census-linked datasets might be utilised in the future.

This analysis has been produced in collaboration with the [ONS Centre for Equalities and Inclusion](#). The aim of this centre is to work with others to ensure that the right data are available to address the main social and policy questions about fairness and equality in society.

The results presented in this article should be treated as experimental statistics. They are research outputs and whilst these new data have the potential to give us far better insight into some of the factors affecting earnings progression it is accepted that many circumstances are not captured in the data. Any disparities between groups in the level and growth of earnings presented in this research may be due to other characteristics other than the one being directly measured and compared.

This analysis only covers earnings progression. Changes in occupation or improvement to conditions of work, which may be considered progression, are not analysed here as this information is not currently available on the dataset.

To understand the overall picture of earnings progression and the relative importance of factors affecting progression, both descriptive statistics and [regression analysis](#) are included. Unless otherwise stated, commentary refers to findings from the descriptive analysis. It should also be noted that whilst regression analysis can indicate a relationship between factors, it does not imply causality.

## 4 . What is included in the dataset

Data from the 2011 Census have been linked to earnings and benefits information from Department for Work and Pensions (DWP) and Her Majesty's Revenue and Customs (HMRC) between tax year ending (TYE) 2012 and TYE 2016. This longitudinal dataset lets individual cases be matched over time allowing analysis of how earnings change. Individuals are contained in the dataset if they were present on the 2011 Census, resident in England or Wales and on the HMRC Pay As You Earn (PAYE) datasets at least once between TYE 2012 and TYE 2016. Linking PAYE to the census provides new insights by bringing personal and household characteristics together with earnings information.

Personal and household characteristics used in the analysis: sex, age, ethnicity, qualifications, occupation and household type are taken from the 2011 Census. Local authority geography for individuals is taken from the DWP Customer Information System to analyse the impact of moving on earnings progression.

PAYE contains individuals' total money earned or paid (including employments as an employee and occupational and personal pensions in payment) during the tax year. This analysis focuses on earnings progression and for this purpose the dataset was restricted to those aged 16 to 64 years in each year. This resulted in 28 million individual cases being included in this research. The scale of the dataset allowed for multiple disaggregations that would not be possible using traditional survey data.

It should be noted that because the dataset records annual earnings, rather than hourly earnings, it is not possible to identify whether an increase in pay is due to an increase in hours, or a substantive rise in hourly pay. No information on earnings from self-employment is included in the dataset.

Only anonymised data has been used in this analysis and results are shown at an aggregated level, so individuals are never and can never be identified.

## 5 . Earnings growth higher for younger people

The cohort of young people, aged 16 to 24 years in 2011, had an average of 16.4% annual earnings growth over the period, tax year ending (TYE) 2012 to TYE 2016. This was higher than any other age group and the overall population aged 16 to 64 years experienced on average 1.4% annual earnings growth over the same period.

A change in annual earnings could be due to changing hours or level of hourly pay. It would be expected that younger people experienced higher earnings growth than older age groups as they may be moving from combining casual or part-time work and education into full-time, more formal employment. Although this analysis used annual earnings and it is not possible to determine if an individual increase is due to increasing hours or a rise in hourly pay, it is known that [more jobs done by those aged 16 to 21 years are part-time](#) than any other age group.

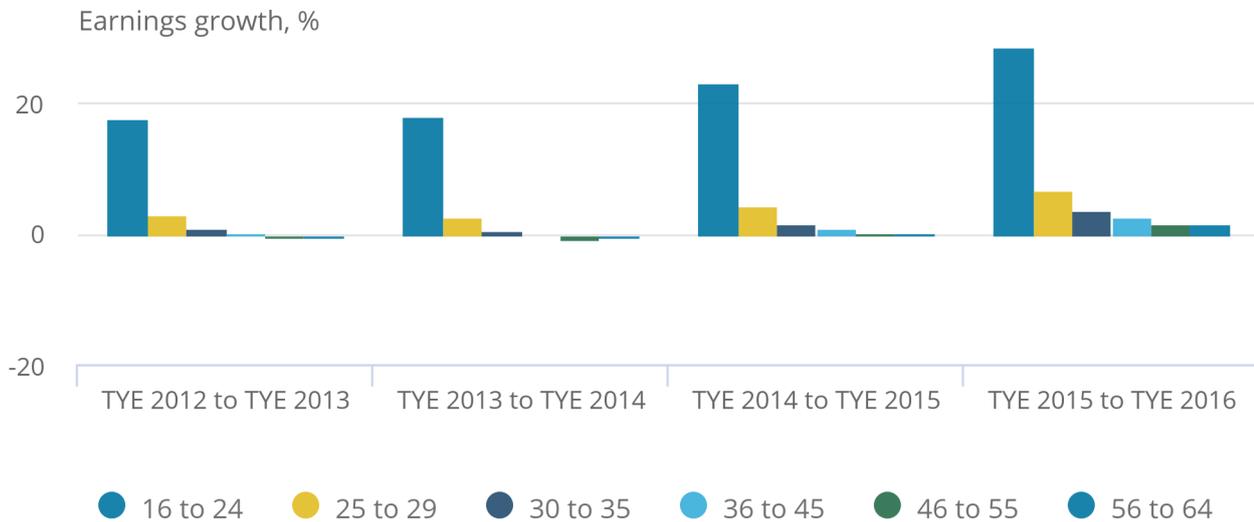
In addition, the year-on-year real median earnings growth for those aged 16 to 24 years in each year shows an increase from 17.6% to 28.7% between TYE 2012 and TYE 2016 (Figure 1). Again, this is the highest for all ages, although the annual median growth for the 25 to 29 years age group in each year more than doubled from 3.1% to 6.8%.

**Figure 1: Real median earnings growth rate by age group, tax year ending 2012 to tax year ending 2016**

England and Wales

Figure 1: Real median earnings growth rate by age group, tax year ending 2012 to tax year ending 2016

England and Wales



Source: Office for National Statistics and Department for Work and Pensions

Notes:

1. The earnings growth rate is calculated for those with PAYE earnings in both periods, and is adjusted for the effects of inflation using the Consumer Prices Index including owner occupiers' housing costs (CPIH).
2. Age groups are based on the age of people in the calendar year common to the period covered, that is, age in 2012 for earnings growth over the period tax year ending (TYE) 2012 to TYE 2013.

Comparing these growth rates with official statistics published from the Annual Survey of Hours and Earnings (ASHE) reveals a higher level of earnings growth but a similar trend. These differences may be due to this analysis only including individuals aged 16 to 64 years living in England or Wales present on the 2011 Census with a PAYE record in at least one year between TYE 2012 and TYE 2016. [Analysis of those continuously employed between two years](#) on ASHE has also previously shown higher growth rates than for those not in the same job for more than 12 months. Furthermore, this analysis calculates the median growth rate of individuals whereas ASHE analysis calculates growth in the distribution median. Further information is in the [Quality and Methodology](#) section at the end of the article.

## 6 . Young people most likely to experience relative earnings progression

Relative earnings growth is another important measure of progression as it can give an indication of social mobility. It is defined here by an individual moving up at least two deciles in the earnings distribution from the previous year. More information on deciles is in the [Quality and Methodology](#) section.

Young people were the most likely to experience relative earnings growth. There were nearly 3 million people aged 16 to 24 years with positive earnings in both tax year ending (TYE) 2015 and TYE 2016. Of these, 19% experienced progression of at least two earnings deciles (Figure 2), compared with 7.8% for the overall population aged 16 to 64 years.

Young people's higher relative earnings progression can be partially explained as they are over-represented in the lower earnings deciles. Of those aged 16 to 24 years, 46% were in deciles 1 and 2 in TYE 2016, which give the greatest scope to move up the distribution. This age group would also be expected to see large changes in annual earnings as they move from part-time to full-time work or into more formal careers after education.

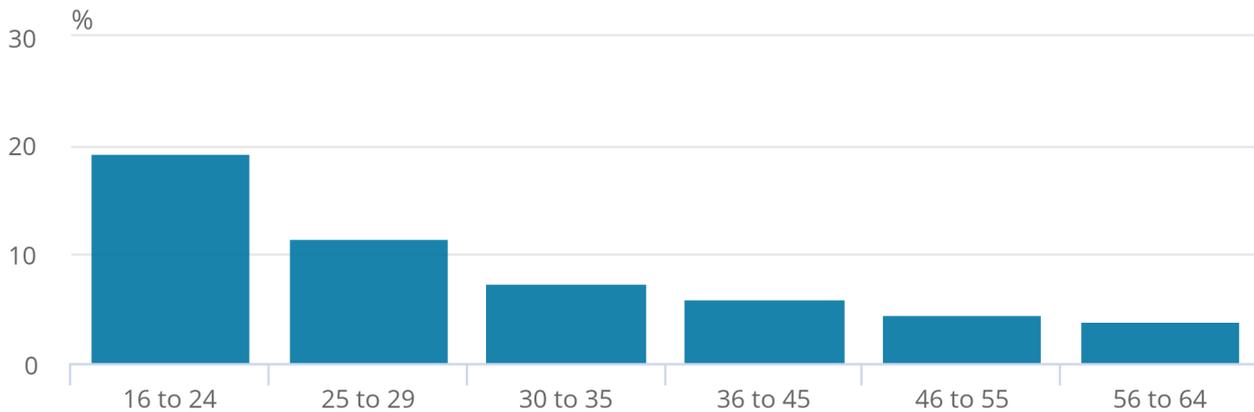
For the 25 to 29 years age group, 12% moved up at least two deciles (Figure 2), again higher than for older age groups. Over half (54%) of this age group were split between deciles 5 to 8. The proportions of individuals experiencing this level of decile progression were consistent in all year-on-year changes from TYE 2012.

**Figure 2: Percentage of those continuously employed with earnings decile progression by age group, tax year ending 2015 to tax year ending 2016**

England and Wales

Figure 2: Percentage of those continuously employed with earnings decile progression by age group, tax year ending 2015 to tax year ending 2016

England and Wales



Source: Office for National Statistics and Department for Work and Pensions

Notes:

1. The rate of in-employment progression is the proportion of employees in each age group who moved up at least two earnings deciles from tax year ending (TYE) 2015 to TYE 2016.
2. Age groups are based on the age of people in the calendar year common to the period covered, that is, age in 2015.
3. This analysis excludes those turning 65 during the reference period.

## 7 . Sex and ethnicity are important factors in earnings progression

It is expected that younger age groups experience higher annual earnings growth than older age groups. The size of this dataset, however, allows analysis into how young people with varying personal characteristics, living in different parts of England and Wales, experience earnings progression. This would not be otherwise possible with traditional surveys.

Generally, median earnings and growth in median earnings were lower for women compared with men. This is also true when accounting for qualification level. Men tend to have greater earnings progression when comparing those with degrees, those with school-level qualifications and those with no qualifications. As might be expected, median earnings and median earnings growth are higher for those with degree qualifications and lowest for those with no qualifications as identified on the 2011 Census.

Analysis of the highest and lowest earnings growth rates shows that, for those with a degree or higher in 2011, men of Black or Other ethnicity generally had some of the highest growth while White females had some of the lowest. However, the Black ethnic group typically had low initial median earnings, and consistently lower earnings over the period compared with other ethnicities. For example, men of Black ethnicity living in the North East region had the lowest median annual earnings for degree-educated 26- to 29-year-olds in tax year ending (TYE) 2012 of £12,565. This was over £10,000 less than the comparative figure for men of White ethnicity. Even though this group had the higher earnings growth of 63% compared with 27% over the five years, there remained a disparity of over £8,000 in TYE 2016.

The intersectional effects on earnings can be explored using the chart builder in Figure 3.

### **Figure 3: Comparison of young people's median annual earnings by personal characteristics, tax year ending 2012 to tax year ending 2016**

#### **England and Wales**

#### **Notes:**

1. This chart shows people aged 16 to 29 years with PAYE earnings in every period, tax year ending (TYE) 2012 to TYE 2016, who were in the same local authority across the period.
2. The population is restricted to non-movers to exclude the effect of moving on earnings growth, thereby allowing for fixed comparisons based on personal characteristics.
3. Degree educated - people with level 4 or above qualifications as recorded on 2011 Census.
4. School educated - people with below level 4 qualifications (including level 3, level 2, level 1, apprenticeships and other qualifications) as recorded on 2011 Census.
5. No qualifications - people with no formal educational qualifications as recorded on 2011 Census.
6. Any suppressions have occurred below a threshold of 20.

To exclude the effect of moving and allow comparisons based on various personal characteristics and location, the analysis in this section only included young people aged 16 to 29 years in 2011 that had not moved local authority over the period TYE 2012 to TYE 2016. It is recognised, however, that this analysis cannot take account of the variables that might affect an individual's level and growth of earnings that are not captured in the dataset. Disparities may be due to a wider variety of factors than the characteristics being directly compared.

In particular, the education information is taken from the 2011 Census and it is unknown whether an individual has gained additional qualifications or undertaken training in the subsequent years that will likely affect their earnings. The Department for Education has published research using the [Longitudinal Educational Outcomes](#) dataset that provides information on earnings from individuals that have attended higher education.

## 8 . More than one in five young people on relative low annual pay progressed to higher earnings

Relative low annual pay is defined as those with annual earnings of less than two-thirds of the aged 16 to 64 years employed populations' median earnings for each tax year. In tax year ending (TYE) 2012, this threshold was £10,307.

There were 2.2 million young people aged 16 to 24 years on relative low annual pay in TYE 2012, of which 24% moved to higher earnings and were consistently above the threshold from TYE 2014 to TYE 2016. Meanwhile, 42% cycled in and out of relative low annual pay over the period TYE 2014 to TYE 2016 and 21% remained on relative low annual pay throughout. The remaining 13% were no longer on the dataset in TYE 2016 so their progression is unknown.

For those in the 25 to 29 years age group, 21% consistently moved above the relative low annual pay threshold, whilst 31% cycled in and out of relative low annual pay and 26% remained below the threshold for all years.

When looking at the population aged 16 to 64 years on relative low annual pay in TYE 2012, using a logistic regression model to control for other factors, an increase in age decreases the likelihood of escaping.

Sex, ethnicity and equivalised household earnings in 2011 were further factors that affected the likelihood of people in the overall working age population escaping relative low annual pay. When holding other factors constant, women were less likely to escape relative low annual pay compared with men. Secondly, the odds of escaping relative low annual pay were 1.2 times higher for each 10% increase in annual household earnings.

Lastly, for those of Black ethnicity, the odds of consistently moving out of relative low annual pay by TYE 2016 were 1.6 times higher than for people of White ethnicity. However, those of White ethnicity were more likely to move out of relative low annual pay than people of Mixed, Asian or Other ethnicities.

## 9 . Higher skilled young people more likely to move

Younger age groups were more likely to have moved local authority, with those aged 18 to 29 years in 2011 accounting for 42% of employees aged 16 to 64 years that moved at least once between 2011 and 2015. Furthermore, holding other factors constant in a logistic regression model, the likelihood of moving decreases by year of age.

The highest proportion of movers are concentrated around London and other metropolitan areas for younger people aged 25 to 29 years. Young people living in the East of England and South East were most likely to move to London between 2011 to 2015, whilst those from Wales and the North East were least likely to move to London. This could reflect the disparity in living costs being a barrier to moving, or the geographical distance away from family and friends. Over half (57%) of young people moving to London are internal moves between local authorities in London, suggesting a highly mobile young population in the capital.

The pattern of movement becomes more evenly dispersed but with a higher concentration around London and the South East for those aged 30 to 45 years. For those aged 46 to 64 years, there are fewer movers, and an inverse pattern with the highest proportion of people in this group moving to coastal and rural areas (Figure 4).

### Figure 4: Proportion of residents that moved into local authority between 2011 and 2015 by age group

England and Wales

## Notes:

1. Each map shows the proportion of people in the specified age group in 2011 who moved into the local authority, relative to the age group's total population in each local authority.
2. Each of the age group maps uses the same classification. The classification used in the maps is based on a Jenks algorithm to define the breaks. This method seeks to reduce the variance of values within each break and maximise the variance between the values in different breaks.

It is not unexpected that younger people move more than older people. However, there was a difference amongst young employees depending on the type of job they did. Those aged 18 to 29 years and in lower-skilled occupations as recorded on the 2011 Census tended to be more restricted in where they moved and were more likely to move to London and city regions. Whereas young people in highly-skilled occupations moved to a wider range of local authorities, particularly across central and southern England (Figure 5).

It is known that [travel to work areas](#) for higher-skilled and higher-paid jobs are larger and the offer of higher pay can often be a motivation for moving for any employee regardless of age. For those in lower-skilled jobs, the National Minimum Wage and National Living Wage do not change by location, which may explain the tendency to move to city regions with denser labour markets.

Mover status is derived from the Department for Work and Pensions (DWP's) Customer Information System and refers to the local authority of the home address. This analysis does not specify the distance moved and the associated cost implications of this. Someone may move only a couple of streets away and be in a different local authority, whilst another move may cover many miles. In this analysis, both would be classified as movers.

It has not been possible to consider the difference in living costs between regions in this analysis. It recognised that an increase in annual earnings may not lead to an increase in living standards or overall financial well-being for an individual.

## **Figure 5: Proportion of young residents that moved into local authority between 2011 and 2015 by occupational skill level**

### **England and Wales**

## Notes:

1. Movers are defined as the proportion of young people, aged 18 to 29 years in 2011, with the specified skill level who have moved into each local authority between 2011 and 2015.
2. Skill level is based on Standard Occupational Classification (SOC) from 2011 Census. Corporate Managers and Directors (SOC 1100 to 1199) and Professional Occupations (SOC Major Group 2) are classified as High Skill. Other Managers and Proprietors (SOC 1200 to 1299), Associate Professional and Technical Occupations (SOC Major Group 3), and Skilled Trade Occupations (SOC Major Group 5) are classified as Upper Middle Skill. Administrative and Secretarial Occupations (SOC Major Group 4), Caring, Leisure and Other Service Occupations (SOC Major Group 6), Sales and Customer Service Occupations (SOC Major group 7), and Process, Plan and Machine Operatives are classified as Lower Middle Skill. Elementary Occupations (SOC Major Group 9) are classified as Low Skill.
3. Each of the skill level maps uses the same classification. The classification used in the maps above is based on a Jenks algorithm to define the breaks. This method seeks to reduce the variance of values within each break and maximize the variance between the values in different breaks.

## 10 . Geographic mobility associated with higher earnings growth

The [UK Industrial Strategy \(PDF, 8.5MB\)](#) identifies “People” and “Place” as two of the five foundations to boost earnings power throughout the UK. Ideally, all areas of the UK would provide equal opportunity for earnings growth, however, London consistently has the highest average earnings of any region, but also some of the highest living costs. This means that not everyone has the same opportunity to access the higher earnings available in London and it is important to understand how moving to other metropolitan areas affects earnings growth.

This analysis also considers earnings growth for those moving to the combined authority areas of Greater Manchester, West Midlands, West of England, Cambridgeshire and Peterborough, Sheffield, and in Wales, the Cardiff capital region (Figure 6).

The majority (81%) of young people, aged 18 to 29 years in 2011, did not move over the period 2011 to 2015. For those that did move 75% moved somewhere other than a city region.

City regions tended to attract people from nearby local authorities. For example, 33% of movers to London were from the South East, 36% of movers to Greater Manchester were from the North West, and 35% of movers to the West Midlands Combined Authority were from the West Midlands region (Figure 7).

Earnings growth in this section, refers to the average annual earnings growth over the period tax year ending (TYE) 2012 to TYE 2016. Greater London had at least four times more people moving there compared with the other city regions and the young people, aged 18 to 29 years making this move had the highest earnings growth, equivalent to an average of 21.8% per year over the period.

For Greater London, earnings growth tended to be higher for those moving from local authorities furthest away such as the North East and Wales, as the disparity in levels of median earnings is greatest (Figure 8).

When considering only those on relative low annual pay in the whole population aged 16 to 64 years in 2011, logistic regression analysis indicated that the initial region of residence did not have a significant impact on consistently moving out of relative low annual pay. However, whether the individual had moved during the period was significantly associated with moving out of relative low annual pay.

### **Figure 6: Young people's average annual earnings growth by city region for movers and non-movers by selected local authority, tax year ending 2012 to tax year ending 2016**

England and Wales

Notes:

1. The chart shows the average annual median earnings growth, which is derived from the overall median earnings growth from tax year ending (TYE) 2012 to TYE 2016 (please refer to Notes page for further details). The earnings growth rate is calculated for those with PAYE earnings in both periods, and is adjusted for the effects of inflation using the Consumer Prices Index including owner occupiers' housing costs (CPIH).
2. Young people are those aged 18 to 29, based on the respondents age at the time of 2011 Census.
3. Mover status is based on whether or not an individual moved to one of the seven combined authorities, based on their local authority in 2015 compared with 2011. Moves within combined authorities are not included.
4. Local authorities with counts of below 10 are suppressed.

**Figure 7: Number of young movers by city region and number of non-movers from different local authorities, 2011 to 2015**

**England and Wales**

**Notes:**

1. This map shows the number of young people, 18 to 29, with PAYE earnings in tax year ending (TYE) 2012 and TYE 2016 based on their local authority in 2011, moving to one of the selected city regions by 2015.
2. Mover status is based on whether or not individuals moved to one of the seven combined authorities, based on their local authority in 2015 compared with 2011. Moves within combined authorities are not included.
3. Young people are those aged 18 to 29 years, based on the respondents age at the time of 2011 Census.
4. Local authorities with counts of below 10 are suppressed.
5. Each of the maps uses a unique classification. The classification used in the maps is based on a Jenks algorithm to define the breaks. This method seeks to reduce the variance of values within each break and maximise the variance between the values in different breaks.

**Figure 8: Young people's average annual earnings growth by city region for movers and non-movers from different local authorities, tax year ending 2012 to tax year ending 2016**

**England and Wales**

**Notes:**

1. This map shows the average annual median earnings growth, which is derived from the overall median earnings growth from tax year ending (TYE) 2012 to TYE 2016. The earnings growth rate is calculated for those with PAYE earnings in both periods, and is adjusted for the effects of inflation using the Consumer Prices Index including owner-occupiers' housing costs (CPIH).
2. Young people are those aged 18 to 29 years, based on the respondents age at the time of 2011 Census.
3. Mover status is based on whether or not individuals moved to one of the seven combined authorities, based on their local authority in 2015 compared with 2011. Moves within combined authorities are not included.
4. Local authorities with counts of below 10 are suppressed.
5. Each of the maps uses a unique classification. The classification used in the maps is based on a Jenks algorithm to define the breaks. This method seeks to reduce the variance of values within each break and maximise the variance between the values in different breaks.

## 11 . Quality and methodology

### Disclaimer

This work contains statistical data from Office for National Statistics (ONS) which is Crown Copyright. The use of the ONS statistical data in this work does not imply the endorsement of ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets, which may not exactly reproduce National Statistics aggregates.

### Information on datasets linked to 2011 Census: Benefits and Income Dataset (BIDs)

Research has previously been conducted to assess the [feasibility of including income questions](#) on the census. The decision was taken to not include such questions in 2011 due to the likely effect on response rates and potential introduction of bias in the results. There remains, however, a great level of interest in understanding how personal and household characteristics affect earnings and linking PAYE to the census can provide new insights.

PAYE – Pay As You Earn, which is administered by HM Revenue and Customs (HMRC), contains individuals' total money earned or paid (including employments as an employee, and occupational and personal pensions in payment) during the tax year.

The PAYE data cover the UK from the period 6 April 2011 onwards. In total, the dataset consists of between 36 and 41 million unique records per tax year with a single row per person per tax year. The total amount of PAYE pay is made up of the total amount earned per person during the tax year from pensions or through employment and it excludes any income from self-employment.

Most records show a positive value for earnings. It can, however, be negative if a person is due a tax rebate or zero if a person is receiving statutory sick pay or statutory maternity pay. Those with negative or zero PAYE earnings have been excluded from this analysis. The dataset will include people who are resident abroad, but get paid or receive their occupational pension from the UK, as well as people who may now be deceased.

This dataset population changes year-on-year with a similar trend to that of the Annual Survey of Hours and Earnings (ASHE) estimates of the number of jobs in England and Wales. The ASHE estimates for 2012 and 2016 are 18 million and 19 million respectively, while the working age population of the Census-Benefits and Income Dataset's data is 23 million for tax year ending (TYE) 2012 and 24 million for TYE 2016. ASHE only includes earnings for "employees who have been in the same job for more than a year". This differs to the earnings data used in this analysis, which includes all employees with earnings processed through PAYE. Some employees will have completed ad hoc or one-off periods of work during the year, resulting in low annual amounts, which will bring the median down.

Local authority and regional geography is taken from DWP's Customer Information System and is based on home address for each calendar year.

## Median earnings growth

The earnings growth rate is calculated for employees with PAYE earnings greater than zero in two consecutive tax years. Growth in earnings between the two tax years is calculated for each individual using the formula:

$$\text{growth} = (\text{year 2 earnings} - \text{year 1 earnings}) \div \text{year 1 earnings}$$

The median of all growth rates is then taken, to provide an estimate of "typical" earnings growth for employees between the two consecutive tax years. Earnings growth is adjusted for the effects of inflation using the Consumer Prices Index including owner occupiers' housing costs (CPIH).

This methodology differs to that normally produced using data from the Annual Survey of Hours and Earnings (ASHE). ASHE calculates the median earnings for a population in two consecutive tax years, then calculates the growth between these median levels of earnings. This method is designed to assess how the earnings of the "typical" employee have changed between two tax years, rather than the median of changes in pay produced in this analysis. Average growth rates analysis can be applied to ASHE data by restricting the data to people who have been employed in the same job for at least a year, producing similar results to those in this analysis.

## Average annual earnings growth

Average annual earnings growth rates are derived by annualising the overall earnings growth between tax year ending (TYE) 2012 to TYE 2016. This is calculated by converting the percentage growth into a decimal and then finding the nth root; in this case,  $n = 4$  as there are five tax years and therefore four separate annual growth rates. This average measure does not give the average of the four annual median earnings growth rates. For example, someone whose pay grew by 10% over the five years from TYE 2012 to TYE 2016 would have experienced an average annual increase of approximately 2.4% ( $1.1^{1/4} = 1.024..$ ).

## Relative low annual pay progression definitions

Individuals who started on relative low annual pay in TYE 2012 were categorised as follows based on their earnings over the period TYE 2012 to TYE 2016:

- escapers – those who earned above the relative low annual pay threshold in every year from TYE 2014 to TYE 2016 suggesting they have progressed onto higher wages
- stuck – those who were below the relative low annual pay threshold in every year between TYE 2013 to TYE 2016
- cyclers – those who fall between the first two categories, moving above the threshold for some but not all years between TYE 2013 and TYE 2016
- exited – those who are no longer present in the data at the end of the period and their progression is unknown

Although over a different time period, this replicates analysis done by the [Resolution Foundation](#) that used the ONS New Earnings Survey Panel Dataset and a panel version of the Annual Survey of Hours and Earnings between 2006 and 2016.

## Relative earnings progression using earnings deciles

Earnings deciles were created by putting the whole population aged 16 to 64 years into order from lowest to highest earnings, then splitting into 10 equal parts. For this analysis, earnings decile progression refers to an individual moving up at least two earnings deciles over a two-year period.

Categories of decile progression are based on data in the specified year (year two) in relation to the previous year (year one). Moving up two earnings deciles is an established measure of relative progression. Decile progression in this analysis is defined as:

- progressed, in work – those who had PAYE data greater than zero in both years and had moved up at least two deciles since the previous year
- progressed, into work – those who had PAYE data in the specified year but did not have PAYE data in year one
- not progressed, in work – those who had PAYE data greater than zero in both years, but had not moved up at least two deciles since the previous year
- not progressed, out of work – those had no PAYE data in the specified year, but did in the previous year

## City regions

The cities included in the geographic mobility analysis are the English combined authorities. The analysis has also been extended to include Cardiff Capital Region, however, this is not an official recognised geography whereas the English combined authorities are legal entities. Liverpool City Region and Tees Valley Combined Authority have been excluded as these produced small samples. These cities were selected in consultation with the ONS Centre for Subnational Analysis. For details on the city regions and their local authorities see Table 15 in the datasets.

## Logistic regression methodology

In addition to descriptive analysis, a logistic regression was conducted using the outcome variable: 'Whether the individual escaped relative low annual pay or not'. This was constructed from the relative low annual pay progression variable detailed previously, whereby the original variable was recoded into a binary between escaping or any other outcome. A second logistic regression on the likelihood of moving local authority was also conducted using the same methodology as detailed further in this methodology.

## Sampling for the regression models

A simple random sample stratified by region was taken for modelling. A sample size of 10,000 was used, after running the regression at various sample sizes between 1,000 and 15,000. This sampling test was to assess the impact on R<sup>2</sup> and AIC goodness of fit measures. A sample size of 10,000 also helped protect against “overfitting”, which can happen with large datasets. Overfitting means the model coefficients may only be relatable to the sample data used, rather than being generalisable to a wider population with similar characteristics. With this sample size, there was no issue with the minimum cell count, as all categories in each variable had a count of greater than 30.

## Recoding independent variables

In some cases, it was necessary to recode or restructure independent variables for better interpretation in the model. Occupation, taken from the 2011 Census used the Standard Occupational Classification (SOC) 2010, was re-classified into four skill levels:

- Corporate Managers and Directors (SOC 1100 to 1199) and Professional Occupations (SOC Major Group 2) were classified as High Skill
- Other Managers and Proprietors (SOC 1200 to 1299), Associate Professional and Technical Occupations (SOC Major Group 3), and Skilled Trade Occupations (SOC Major Group 5) were classified as Upper Middle Skill
- Administrative and Secretarial Occupations (SOC Major Group 4), Caring, Leisure and Other Service Occupations (SOC Major Group 6), Sales and Customer Service Occupations (SOC Major group 7), and Process, Plan and Machine Operatives were classified as Lower Middle Skill
- Elementary Occupations (SOC Major Group 9) were classified as Low Skill

Household equivalised earnings were transformed into the logarithmic scale due to the skewed nature of earnings data.

## Logistic regression

Logistic regression was used to explore the relationship between an outcome variable and a range of explanatory variables, when the effect of each is isolated or held constant. As the outcome under assessment was categorical, logistic methods were used. Whilst it was possible to use multinomial regression, recoding the outcome variable to binary allowed better interpretation of the model.

## Procedure

The regression was conducted in SAS and used a backwards stepwise method to create the final model. All variables of interest were initially included. Variables with the highest p-values were gradually removed from the model, until only variables significant at the 0.05 level remained. The forward stepwise method was also tested and produced a similar model, though this method was not used for the final model as it allows for important variables to be missed when other variables are entered into the model first (termed “suppressor effects”).

## Variables

For information on the variables included in the final regression models please see dataset tables 22 and 23.

## Multicollinearity

Multicollinearity was accounted for in the model by running a correlation covariance matrix on the initial model's variables. Disability and health, ethnicity and nationality, household type, and age of youngest dependent child all correlated highly together above 0.7. For each pair, the former was chosen for the final model as they aided interpretation more so than the latter variable. There was some correlation between variables in the final model but not above 0.5.

## Goodness of fit

Goodness of fit tests describe how well a model fits, predicts and corresponds to the data from which it is generated. Firstly, the Likelihood Ratio Chi-Square Test, the Score Chi-Square Test and the Wald Chi-Square Test were all significant at the less than 0.001 level, meaning that at least one of the regression coefficients in the model was not equal to zero.

The further measurements used to assess model fit were the R-squared value, accuracy rate and Hosmer and Lemeshow test (H-L test).

The R-squared value for the escaping relative low annual pay model was 13% and indicates how well the logistic regression equation predicts the outcome with and without the independent variables. This is quite small, but it is acknowledged there are many factors that can contribute to a change in earnings that are not included in the dataset such as self-employment, new qualifications or training, and changes to family and household structure.

The accuracy rate tells us what percentage of responses a model predicts correctly; the final model had a 70% accuracy rate.

The Hosmer and Lemeshow test tests the null hypothesis that the model fits the data well. We accepted the null hypothesis as the p-value was 0.5377. See dataset table 23 for the equivalent goodness of fit measures for the model predicting whether people move local authority.

## Causality

Whilst regression analysis can indicate a relationship between factors, it does not imply causality. If a causal relationship between the independent and dependent variables is not indicated by existing knowledge, then the direction of causality is assumed to operate in either or both directions.

## Interpretation of results

The usual output from logistic regression is the odds ratio. This is obtained for each variable by exponentiating the given estimate. The odds ratio can be interpreted as: for a one-unit change in the predictor variable, the odds ratio for a positive outcome is expected to change by the respective coefficient, given the other variables in the model are held constant. For each individual category within a variable, the odds ratio is interpreted as the increase in probability of a positive outcome when compared with the reference category.

The interpretation differs for the household earnings variable as it was the log of this variable that was included in the model. Instead of a one-unit change in this variable, the value of the odds ratio is associated with a 10% increase in household earnings.

## 12 . Authors and acknowledgements

Ruth Davies, Bonang Lewis, Thomas Odell and Melanie Lewis, Office for National Statistics.

The authors would like to thank Tom Evans, Andrea Lacey, Chloe Lloyd, Vasileios Antonopoulos, Paola Serafino, Henry Lau, Jure Stabuc, and Lisa Jones alongside colleagues in Data as a Service, Research Support and Data Access, and Methodology (ONS), and contributors from the Race Disparity Unit and Department for Work and Pensions.