Research Outputs: Investigating the economic activity of undergraduate students using administrative data

This is early research using administrative data to investigate the economic activity of undergraduate students in England and Wales for the academic year ending 2015. These Research Outputs are not official statistics.

In this section

1. Disclaimer
2. New analysis using administrative data
3. Things you need to know about this release
4. Other data sources on the economic activity of students
5. How did we determine whether a student was employed?
6. What do the outputs show?
7. Conclusions and next steps
8. Feedback
9. Annex A: Comparison with existing data sources on the economic activity of students
10. Annex B: Differences between these Research Outputs and the Student Income and Expenditure Survey (SIES)

Disclaimer

These Research Outputs are not official statistics. Rather they are published as outputs from research into an Administrative Data Census approach. These outputs must not be reproduced without this disclaimer and warning note and should not be used for policy- or decision-making.

Back to table of contents

New analysis using administrative data

The Administrative Data Census Project is working to assess whether the government-stated ambition that "censuses after 2021 will be conducted using other sources of data" can be met. We're aiming to produce population estimates, household estimates and population and housing characteristics using a combination of administrative and survey data. This is to meet demands for improved population statistics and as a possible alternative to the census.

Economic activity is a high-priority topic for users of census statistics. This article contains early research towards producing economic activity within an Administrative Data Census and demonstrates the type of analysis possible by linking different administrative datasets. The research is working towards providing information at a small-area level around the main policy areas such as economic activity and the labour market.

We're publishing this research to ask for feedback on our methodology and ideas for improvement. We're also particularly interested to hear about the breakdown of economic activity you need. Please send your feedback to Admin.Data.Census.Project@ons.gov.uk.

The 2001 and 2011 censuses included questions that enabled the classification of an individual's economic activity, providing important insights into the labour market in England and Wales. Individuals were classified as economically active or economically inactive. Economically inactive individuals were then further classified by their reason for inactivity, while economically active individuals were classified as employed, self-employed or unemployed. Economically active full-time students were classified separately from other active individuals.

The main administrative data sources needed to produce outputs on economic activity are income and benefits data from the Department for Work and Pensions (DWP) and HM Revenue and Customs (HMRC). However, current administrative data acquired by the Office for National Statistics (ONS) are not sufficient to develop a complete picture of economic activity for the whole population. This is because we don't have access to detailed breakdowns of Pay As You Earn (PAYE) data (which would enable the separation of income from pensions and employment) or information from Self Assessment returns (which includes information on the self-employed). More information about these data sources can be found in the Income and Benefits Data Source overview.

Due to the age distribution of undergraduate students and the age distribution of the self-employed, we expect these limitations to have a smaller impact on the student population. Therefore, to start researching our ability to produce estimates of economic activity from administrative data, this paper looks at undergraduate student employment only. By linking Higher Education Statistics Agency (HESA) student record data to HMRC's PAYE data, we've produced analysis on the employment status of undergraduate students in England and Wales for the academic year ending 2015 at local authority level.

This research has demonstrated the potential to produce statistics on economic activity at a subnational level from administrative data. It also demonstrates the potential for an Administrative Data Census to produce multivariate analyses – in this case, on economic activity. For example, cross-tabulations of the percentage of students in employment by variables such as mode of study and ethnicity were produced.

Back to table of contents

Things you need to know about this release

- These Research Outputs are not official statistics; they are published to demonstrate the type of analysis possible using administrative data and to ask for your feedback.
- This research is working towards providing information around the main policy areas such as economic activity and the labour market from an Administrative Data Census.
- The analysis was restricted to undergraduate students aged 16 and over who were not in their first or last year of study during the academic year ending 2015; this subset of students may not be representative of the whole student population (more information can be found in section 5).

- Students on sandwich courses were included in the analysis but were not necessarily on their sandwich year during the academic year ending 2015.
- Self-employed students were not categorised as employed in this analysis; therefore, the term “employment” used throughout this publication refers to employment processed through the Pay As You Earn (PAYE) system only.
- Students classified as employed in this analysis were in employment at some time during the tax years ending 2015 or 2016; therefore, it’s not possible to determine which academic year or years they were in employment.
- It’s not possible in this analysis to distinguish between students who were in employment during the holidays and students who were in employment during term-time.
- The term “gross income” used throughout the publication refers to income from Pay As You Earn data only; other income, such as income taxed via Self Assessment and benefits, are not included in the income amount.

Back to table of contents

Other data sources on the economic activity of students

As discussed in section 2, the 2001 Census and 2011 Census published information on the economic activity of students in England and Wales. The Office for National Statistics (ONS) Labour Force Survey is the other main data source on economic activity with results published in the UK Labour market bulletins. These bulletins contain the official estimates of employment, unemployment and economic activity for the UK and should be used for all analysis and decision-making. The economic activity status of 16- to 24-year-olds is classified separately for individuals in full-time education and those not in full-time education.

The Department for Education (DfE), Welsh Government and the Department for Business, Innovation and Skills (BIS – now the Department for Business, Energy and Industrial Strategy) have also previously published data on the economic activity of students.

BIS and Welsh Government released results from the Student Income and Expenditure Survey (SIES) for English-domiciled students (academic year ending 2012) and Welsh-domiciled students (academic year ending 2015) respectively.

DfE are also releasing a series of experimental statistics on the employment and earnings outcomes of higher education graduates. These experimental statistics are produced using the longitudinal educational outcomes (LEO) dataset that combines administrative data sources from DfE, Higher Education Statistics Agency (HESA), Department for Work and Pensions (DWP) and HM Revenue and Customs (HMRC). The experimental statistics released so far look at employment and earnings after graduation. The administrative data sources used in LEO differ from the administrative data used in these Research Outputs. Unlike LEO, the administrative data sources we used are a subset of income and benefit data provided to ONS by DWP and HMRC for feasibility research.

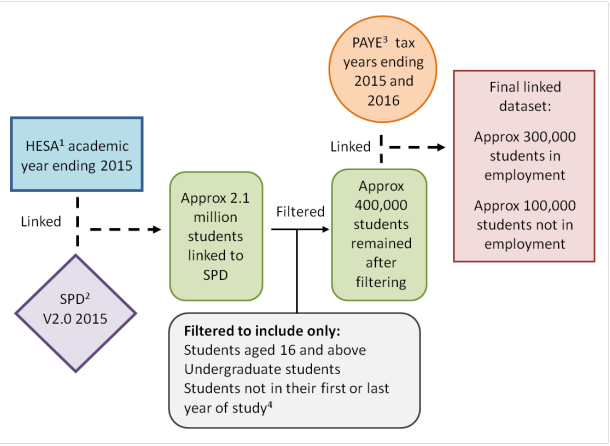
The data sources discussed here are either survey based or focus on graduate outcomes. None combine administrative data sources with the aim of producing multivariate analysis of economic activity at the small-area level. For a more detailed comparison of data sources on the economic activity of students, see Annex A.

Back to table of contents

How did we determine whether a student was employed?

For this research, we have linked Higher Education Statistics Agency (HESA) data for the academic year ending 2015 to Statistical Population Dataset[1] (SPD) V2.0 for 2015 via a unique identifier (Figure 1). For more information on SPD creation see the Administrative Data Research Reports. Students were only included in the analysis if they were present on SPD V2.0 2015. For more information on students not included in SPD V2.0, see the Methodology of Statistical Population Dataset V2.0 report.

Figure 1: Methodology to determine whether a student is in employment



Source: Office for National Statistics

Notes:

1. Higher Education Statistics Agency.
2. A Statistical Population Dataset (SPD) is a single, coherent dataset that forms the basis for estimating the size of the resident population. It's produced by linking records across multiple administrative data sources and applying a set of inclusion and distribution rules.
3. Pay As You Earn.
4. If a student dropped out of their course before the expected last year, they will have been included.

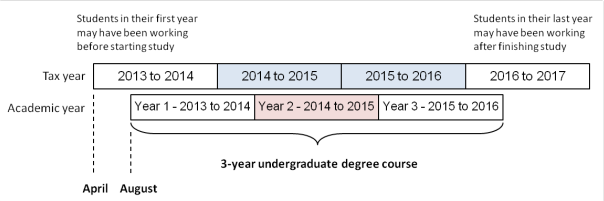
Download this image

.png

The linked HESA SPD dataset was filtered to include only undergraduate students aged 16 and over. Age referred to age at 30 June 2015. Further filtering was then applied due to the discrepancy in dates between the tax year in Pay As You Earn (PAYE) data and the academic year in HESA. While the academic year begins in August, the tax year begins in April (Figure 2).

If a student was present in the PAYE data for the tax year that overlaps with the beginning of their course, they may have only been in employment prior to starting their studies. Similarly, if a student was present in the PAYE data for the tax year that overlaps with the end of their course, they may have only been in employment after finishing their studies (Figure 2). Therefore, students in their first or last year of study were removed prior to analysis. However, if a student dropped out of their course before the expected last year, they will have been included. Students on an industrial placement year or studying abroad were included in the analysis. It is important to note that the subset of undergraduate students included in this analysis may not be representative of the whole student population.

Figure 2: The discrepancy between the academic year in Higher Education Statistics Agency data and the tax year in Pay As You Earn data



Source: Office for National Statistics

Download this image

.png

Students that remained after filtering (approximately 400,000) were linked to the PAYE data for the tax year ending 2015 and the tax year ending 2016 via a unique identifier.

It's not possible within the PAYE data that we've been supplied to distinguish between individuals in employment and individuals in receipt of occupational or personal pensions. However, due to the age distribution of undergraduate students, the proportion in receipt of occupational or personal pensions is likely to be small. Therefore, if a student was present within the PAYE data for either tax year with an income amount greater than £0, we assumed that they were in employment (approximately 300,000). If a student was not found in either PAYE tax year with an income amount greater than £0, we assumed that they were not in employment (approximately 100,000).

When a student was found within both PAYE tax years, an average amount was taken to calculate their annual PAYE income amount. For information on why a student may not have income recorded in the administrative data, please see the Income Research Output publication.

Term-time addresses recorded in the HESA data were used for geographical analysis of the findings.

Notes for: How did we determine whether a student was employed?

1. An SPD is a single, coherent dataset that forms the basis for estimating the population. It is produced by linking records across multiple administrative data sources and applying a set of inclusion and distribution rules.

Back to table of contents

What do the outputs show?

Overall, 73% of undergraduate students in England and Wales who were not in their first or last year of study during the academic year ending 2015 were in employment at some stage during the tax years ending 2015 and 2016. A higher proportion of female students were in employment (76%) than male students (69%). These figures include employment during holidays or term time. As the

employment could have occurred at any time during the tax years ending 2015 and 2016, it's not possible in this analysis to determine which academic year or years they were in employment. You can find more information on the limitations of this analysis in section 3 and section 5.

Data from the Labour Force Survey (LFS) for the period January to March 2015 showed that 31.5% of 18- to 24-year-olds in full-time education were in employment (Table A06). During the summer months, from June to August 2015, this figure rose to 39.5% (Table A06). Definitional differences between the LFS and our Research Outputs largely explain why these figures are lower than our estimate of 71% for full-time students.

To identify employments, the LFS questionnaire asks an individual if they undertook paid work during a reference week and then classifies an individual by their main employment. Our Research Outputs classify a student as in employment if they had any earnings processed through the Pay As You Earn (PAYE) system at any point during either tax year (see section 5 for more details on the methodology).

Furthermore, the LFS figures include 18- to 24-year-olds in full-time education while our figures only include undergraduate students registered at government-funded higher education institutions. Despite these definitional differences, it's important to explore the potential to produce statistics from administrative data, as surveys, such as the LFS, have limited scope for multivariate analysis at small-area level. However, there are some similarities between the LFS figures and our results. For example, both show higher employment levels for female students than male students.

Results of the 2011 Census for student employment were also lower than these outputs show. Of all full-time students aged 16 to 74, 29% were in employment during the week prior to the 2011 Census.

Definitional differences are again likely to explain these different findings. For example, similarly to the LFS, the census only includes employment during the reference week prior to census. Additionally, all full-time students aged 16 to 74 are included in the census findings, as opposed to just undergraduate students.

Results of the Student Income and Expenditure Survey (SIES) for English-domiciled students (academic year ending 2012) and Welsh-domiciled students (academic year ending 2015) are closer to our estimate of student employment. The SIES for English and Welsh-domiciled students combined for the academic year ending 2012 found that 70% of students undertook some form of paid work during either the academic year or the previous summer vacation[1]. For a description of the main methodological differences between the SIES and these Research Outputs, see Annex B.

Our analysis showed that at local authority level in England and Wales, the percentage of undergraduate students who were not in their first or last year of study during the academic year ending 2015 and who were in employment at some stage during the tax years ending 2015 and 2016, ranged from 36% to 95%. Local authorities refer to the term-time address recorded on the Higher Education Statistics Agency (HESA) data for a student. Results for all local authorities in England and Wales can be found in the accompanying data download.

We'd like feedback on the definition and breakdown of economic activity that you need at small-area level. Please send your feedback to Admin.Data.Census.Project@ons.gov.uk.

Employment by mode of study and age

Student employment in England and Wales varied by mode of study (for example, part-time, full-time, sandwich) for the tax years ending 2015 and 2016. A greater percentage of all part-time students were in employment at some stage during the two tax years (83%) compared with full-time students (71%).

The SIES for English-domiciled students (academic year ending 2012) and Welsh-domiciled students (academic year ending 2015) also found a greater percentage of part-time students to be in employment than full-time students. For English and Welsh-domiciled students combined, during the academic year ending 2012, 66% of full-time students undertook some form of paid work during either the academic year or the previous summer vacation, compared with 87% of part-time students[1]. Key methodological differences between the SIES and our analysis may explain differences in the exact findings. You can find more information on these differences in Annex B.

Our analysis also demonstrated that 80% of sandwich students were in employment at some stage during the tax years ending 2015 and 2016. It's the nature of sandwich courses that the majority of students undertake a year in industry (or elsewhere) during their second or third year of study. Therefore, higher employment levels for sandwich students were expected. It's important to note that sandwich students included in this analysis were not necessarily on their year in industry (or elsewhere) during the academic year ending 2015.

Older students were also more likely to be in employment than younger students. For example, 76% of students aged 41 and above were in employment during the tax years ending 2015 and 2016 compared with 66% of students aged 16 to 19.

Higher Education Statistics Agency (HESA) data for the UK showed a relationship between mode of study and age for student enrolments during the academic year ending 2015. While 57% of students enrolling on a part-time course were aged 30 and over, only 6% of full-time students were (Table 6a and 6b).

We therefore explored the relationship between student age and employment broken down by mode of study (Figure 3). After doing so, mode of study appeared to be a bigger driver of student employment patterns than age. In all age groups except 41 and over, full-time students were the least likely to be in employment. At ages 16- to 22-years, sandwich students were the most likely to be in employment. This reflects the fact that most students on sandwich courses undertake their year in industry (or elsewhere) during their second or third year. From the age of 23 and above, part-time students were the most likely to be in employment.

Figure 3: The percentage of undergraduate students in employment for each mode of study by age group, England and Wales

Tax years ending 2015 and 2016, men and women, aged 16 and over

Source: Office for National Statistics

Notes:

1. These Research Outputs are not official statistics on student employment.
2. Data source: Pay As You Earn (PAYE) data from HM Revenue and Customs and student record data from the Higher Education Statistics Agency (HESA). As the outputs are limited to these data sources, self-employed students are not classified as employed.
3. Students were only included in the analysis if they were present on SPD V2.0 2015. For further information on the methodology, see section 5.

4. Only undergraduate students aged 16 and over who were not in their first or last year of study during the academic year ending 2015 were included in the analysis. For further information on the methodology, see section 5.

Download this chart

Image .csv .xls

PAYE income[2] distribution of employed undergraduate students by age and mode of study

The PAYE income[2] distribution of undergraduate students in employment during the tax years ending 2015 and 2016 varied by age and mode of study. Part-time students generally earned more than sandwich students and full-time students (Figure 4) whilst older students generally earned more than younger students (Figure 5). These results are consistent with the SIES for English-domiciled students during the academic year ending 2012 and for Welsh-domiciled students during the academic year ending 2015.

Of the full-time students in employment during the tax years ending 2015 and 2016, 81% earned £5,000.00 or below annually compared with 25% of employed part-time students. Furthermore, 35% of employed part-time students earned over £20,000.00 annually compared with less than 1% of full-time students (Figure 4).

Figure 4: Annual PAYE income distribution of undergraduate students in employment by mode of study, England and Wales

Tax years ending 2015 and 2016, men and women, aged 16 and over

Source: Office for National Statistics

Notes:

1. These Research Outputs are not official statistics on student employment.
2. Data source: Pay As You Earn (PAYE) data from HM Revenue and Customs and student record data from the Higher Education Statistics Agency (HESA). As the outputs are limited to these data sources, self-employed students are not classified as employed.
3. Students were only included in the analysis if they were present on SPD V2.0 2015. For further information on the methodology, see section 5.
4. Only undergraduate students aged 16 and over who were not in their first or last year of study during the academic year ending 2015 were included in the analysis. For further information on the methodology, see section 5.
5. The term income refers to income processed through the Pay As You Earn system only. Income from other sources such as self-employment and benefits are not included in the income amount.
6. The PAYE income described here is gross nominal annual income before deductions such as National Insurance contributions and tax but after the deduction of pension contributions.

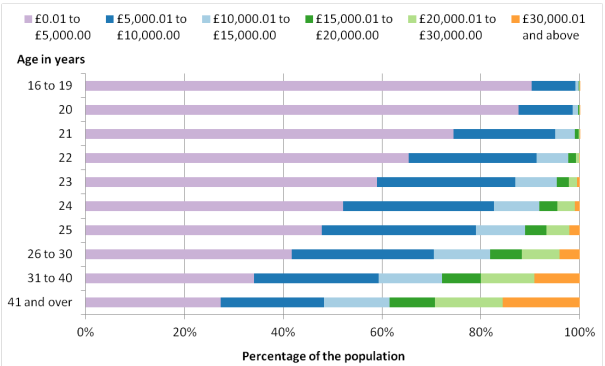
Download this image

.png .xls

Less than 2% of 20-year-old students in employment during the tax years ending 2015 and 2016 earned over £10,000.00 annually. This is compared with 41% of students in employment aged 31 to 40 (Figure 5).

Figure 5: Annual PAYE income distribution of undergraduate students in employment by age, England and Wales

Tax years ending 2015 and 2016, men and women, aged 16 and over



Source: Office for National Statistics

Notes:

1. These Research Outputs are not official statistics on student employment.
2. Data source: Pay As You Earn (PAYE) data from HM Revenue and Customs and student record data from the Higher Education Statistics Agency (HESA). As the outputs are limited to these data sources, self-employed students are not classified as employed.
3. Students were only included in the analysis if they were present on SPD V2.0 2015. For further information on the methodology, see section 5.
4. Only undergraduate students aged 16 and over who were not in their first or last year of study during the academic year ending 2015 were included in the analysis. For further information on the methodology, see section 5.
5. The term income refers to income processed through the Pay As You Earn system only. Income from other sources such as self-employment and benefits are not included in the income amount.
6. The PAYE income described here is gross nominal annual income before deductions such as National Insurance contributions and tax but after the deduction of pension contributions.

Download this image

.png .xls

Employment by ethnicity

Undergraduate student employment in England and Wales during the tax years ending 2015 and 2016 varied by ethnic group as recorded in HESA. Student employment also varied between UK-domiciled and non-UK-domiciled students. The term domicile refers to a student's normal residence prior to commencing their higher education study. Information on ethnic group was available for 76% of non-UK-domiciled students and 99% of UK-domiciled students in the HESA data.

Overall, 77% of UK-domiciled students in England and Wales were in employment at some stage during the tax years ending 2015 and 2016 compared with 30% of non-UK-domiciled students. Of the 11 ethnic groups, 10 also had higher employment for UK-domiciled students compared with non-UK-domiciled students (Figure 6).

For example, among the white population, 80% of UK-domiciled students were in employment compared with 39% of non-UK-domiciled students (Figure 6). Asian/Asian British: Bangladeshi was the only ethnic group where this pattern was not observed. However, as a small number of students were present within this ethnic group, this result should be interpreted with caution.

The ethnic group Asian/Asian British: Chinese had the lowest levels of employment amongst both UK-domiciled and non-UK-domiciled students. For UK-domiciled students, 53% were in employment at some stage during the tax years ending 2015 and 2016 and for non-UK-domiciled students 11% were in employment (Figure 6). The 2011 Census also showed that Asian/Asian British: Chinese was the ethnic group with the lowest percentage of students in employment.

It's important to note that postgraduate students were not included in this analysis and are likely to have had different patterns of employment across the ethnic groups.

Figure 6: Percentage of the undergraduate student population in employment by ethnic group for UK-domiciled and non-UK-domiciled students, England and Wales

Tax years ending 2015 and 2016, men and women, aged 16 and over

Source: Office for National Statistics

Notes:

1. These Research Outputs are not official statistics on student employment.
2. Data source: Pay As You Earn (PAYE) data from HM Revenue and Customs and student record data from the Higher Education Statistics Agency (HESA). As the outputs are limited to these data sources, self-employed students are not classified as employed.

3. Students were only included in the analysis if they were present on SPD V2.0 2015. For further information on the methodology, see section 5.
4. Only undergraduate students aged 16 and over who were not in their first or last year of study during the academic year ending 2015 were included in the analysis. For further information on the methodology, see section 5.
5. Ethnic groups marked with an asterisk have small counts (less than 100 in at least one group) and should be interpreted with caution.
6. The term domicile refers to a student's normal residence prior to commencing their Higher Education study. This may differ from their nationality.
7. Information on ethnic group was available for 76% of non-UK-domiciled students and 99% of UK-domiciled students in the HESA data.

Download this chart

Image .csv .xls

Notes for: What do the outputs show?

1. Analysis was performed on data held by the UK data service from the Student Income and Expenditure Survey, 2012.
2. The PAYE income described here is gross nominal annual income before deductions such as National Insurance contributions and tax but after the deduction of pension contributions. Where a student had PAYE income for both tax years an average amount of the two tax years was calculated. Please see section 5 for more details on the methodology.

Back to table of contents

Conclusions and next steps

This analysis has demonstrated the potential to produce multivariate statistics on economic activity at local authority level using administrative data. Greater access to administrative data in the future should enable us to produce multivariate statistics on economic activity for the whole population of England and Wales for small areas.

In particular, access to Self Assessment data from HM Revenue and Customs (HMRC) would enable us to identify the self-employed population. Access to more detailed Pay As You Earn (PAYE) data from HMRC would also help separate individuals in employment from individuals in receipt of an occupational pension. For students, this data would also help to separate term-time employment from employment during the holidays. Combined with benefits data from the Department for Work and Pensions (DWP) these data sources would help to complete the picture of economic activity.

Due to limitations of the analysis these results should not be used for policy- or decision-making at this stage. Self-employed students were not identified and some students present in the PAYE data will not have been in employment during the academic year ending 2015. This demonstrates the need for more administrative data to complete the picture of economic activity for both students and the whole population. Our UK labour market bulletins contain the official estimates of employment, unemployment and economic activity for the UK that should be used for policy- or decision-making.

We're conducting more research to produce administrative data-based population estimates and population characteristics. For further information on our overall aims, see our Annual Assessment.

Back to table of contents

Feedback

We're currently reviewing the potential to produce estimates on the other categories of economic activity. Therefore, it's important we understand your need for these statistics. We'd like to hear feedback on these Research Outputs and the methodology used to produce them, for example, whether the inclusion of students on an industrial placement year is appropriate for your use.

We're also interested to hear about the breakdown of economic activity that you need, for example, which categories at which geographic level and whether you need multivariate analysis. We plan to hold a user event on economic activity in 2018 to explore this in more detail. If you are interested in receiving more information about this event, or have any feedback after reading this release, please email Admin.Data.Census.Project@ons.gsi.gov.uk. Don't forget to include the title of the output in your response.

Back to table of contents

Annex A: Comparison with existing data sources on the economic activity of students

Annex A: Comparison with existing data sources on the economic activity of students

Publication	Producer	Status	Data source	Population coverage		Lowest geography level	Latest published time period
				Age	Geography		
Research Outputs	ONS	None	Linked administrative data	16 and above	England and Wales	Local authority	Academic year ending 2015 with tax years ending 2015 and 2016
UK labour market bulletins	ONS	National Statistics	Labour Force Survey	All ages Full-time education versus those not in full-time education is broken down for 16- to 24-year-olds	UK	Regional	June to August 2017
Student Income and Expenditure Survey - English-domiciled students	BIS	None	Student Income and Expenditure Survey	All ages	England and Wales	National	Academic year ending 2012
Student Income and Expenditure Survey - Welsh-domiciled students	Welsh Government	None	Student Income and Expenditure Survey	All ages	England and Wales	National	Academic year ending 2015
Graduate outcomes: longitudinal education outcomes (LEO) data[1]	DfE	Experimental Statistics	Linked administrative data	All ages	Graduates from English higher-education institutions	National	Graduating year ending 2013

Source: Office for National Statistics

Notes:

1. The longitudinal education outcomes (LEO) dataset is currently used to study graduate outcomes rather than the economic activity of students during their studies. Despite similarities in the methodology, the administrative data sources used in LEO differ from the administrative data used in these Research Outputs. Unlike LEO, the administrative data sources we use are a subset of income and benefit data provided to ONS by the Department for Work and Pensions (DWP) and HM Revenue and Customs (HMRC) for feasibility research.
2. ONS – Office for National Statistics.
3. BIS – Department for Business, Innovation and Skills.
4. DfE – Department for Education.

Download this image

.png

Back to table of contents

Annex B: Differences between these Research Outputs and the Student Income and Expenditure Survey (SIES)

Methodological differences between these Research Outputs and the Student Income and Expenditure Survey (SIES) may explain differences in the findings. Some of the main differences are that the SIES:

- includes students on any year of their course whereas these Research Outputs exclude students in their first or last year of study – as a result, students on a one- or two-year higher education course are included in the SIES but are excluded from these Research Outputs
- only includes students eligible for the student finance package during the academic year of study (for example, students completing a second degree that they are funding themselves are not included)
- only includes students who were domiciled in England or Wales for at least three years before starting study
- does not include students who were on a placement year or studying abroad at the time of the survey

Back to table of contents

You might also be interested in:

Administrative Data Census: annual assessments

Administrative Data Census Research Outputs

Related downloads

Research Outputs: Economic activity of undergraduate students