

Office for
Students



Exploring student outcomes

**Differences in continuation, completion
and progression between students at
English higher education providers**

Reference OfS 2022.07

Enquiries to Stanley Rudkin at official.statistics@officeforstudents.org.uk

Publication date 20 January 2022

Contents

| | |
|---|-----------|
| Overview | 3 |
| Rationale | 3 |
| Interpretation of data charts | 5 |
| Annex A: Methodology | 11 |
| Populations and outcomes..... | 11 |
| Split indicators | 12 |
| Statistical modelling | 13 |
| Choice of model factors and groups..... | 13 |
| Estimation of percentage point differences in outcomes after controlling for other factors..... | 16 |
| Annex B: Definitions | 20 |
| Annex C: List of abbreviations | 33 |

Overview

1. This document outlines the interpretation and methodology behind the Office for Students (OfS) analysis: 'Exploring student outcomes'.¹
2. We analyse rates of continuation (the proportion of students who continued their studies into a second year); completion (the proportion of students who can be tracked through to the end of their qualification)²; and progression (the proportion of graduates in employment or further study 15 months after graduation) for various populations of students at English higher education providers.
3. For each of these student outcomes, we take the percentage point difference between two groups of students, across seven student characteristics, or 'split indicators': age group, disability type, ethnicity, sex, Index of Multiple Deprivation (IMD) quintiles, Tracking Underrepresentation by Area (TUNDRA) and free school meals eligibility.
4. We then use statistical modelling to control for a range of factors related to student characteristics which also correlate with these outcomes. This indicates the extent to which observed differences in outcomes simply reflect other underlying factors which vary across student groups.
5. It is important to note that this analysis can only account for factors which we can observe in the data. It is possible that there remain unobserved factors correlated both with the student characteristic of interest and with the outcome itself. These unobserved factors could introduce statistical bias. For this reason, users should not interpret the estimates in this release as reflecting causal relationships.

We are actively seeking feedback regarding these experimental statistics.³ If you have any queries or suggestions, please contact Stanley Rudkin at official.statistics@officeforstudents.org.uk.

Rationale

6. At the OfS, we want every student, whatever their background, to have a fulfilling experience of higher education that enriches their lives and careers. One of the ways in which we can assess how well we are meeting that aim is by carrying out analysis that looks at student outcomes across different characteristics.

¹ The related dashboard on exploring student outcomes is at www.officeforstudents.org.uk/data-and-analysis/student-outcomes-and-experiences-data-dashboards/.

² Our January 2022 consultation on constructing student outcome and experience indicators for use in OfS regulation (www.officeforstudents.org.uk/publications/student-outcomes-and-teaching-excellence-consultations/outcome-and-experience-data/) proposes two possible measures of completion outcomes. This document reports analysis of the first of those options, referred to as the cohort-tracking method, which measures the proportion of students who can be tracked through to the end of their qualification.

³ For more information, see <https://osr.statisticsauthority.gov.uk/policies/official-and-national-statistics-policies/experimental-statistics/>.

7. In addition to this, the OfS has issued consultations about the assessment of registration condition B3 and the construction of student outcome and experience indicators to be used in that assessment.⁴ Therefore, as well as supporting us to assess progress against our primary regulatory objectives, this document also provides supporting information for the consultation about B3. It should aid stakeholders in understanding aspects of the approach we have proposed for setting numerical thresholds for student outcomes. The definitions of student outcome measures used in this analysis are aligned with our consultation proposals for B3 assessments and the evidence that they will draw upon, and remain subject to change upon conclusion of the consultation exercise.⁵
8. Through the indicators consultation, we have proposed to construct the following student outcomes data indicators⁶:
 - a. A **continuation indicator** that measures the percentage of students that continue in the study of a higher education qualification (or have gained a qualification) after one year (two years for part-time students).
 - b. A **completion indicator** that measures the percentage of students that complete a higher education qualification, which we propose to base on either tracking individual entrants, or by calculating the proportion of students likely to complete the qualification they started, through up to six subsequent years of study.
 - c. A **progression indicator** constructed from the Graduate Outcomes survey data that measures progression to managerial or professional employment, or further study, 15 months after a higher education qualification has been awarded.
9. This analysis seeks to identify factors associated with continuation, completion⁷ and progression, and better understand the extent to which differences in these outcomes can be accounted for by other underlying factors which vary across student groups.
10. Where differences in student outcomes reduce after controlling for other factors, this should not be interpreted as them being explained or excused by other underlying differences in student characteristics. Rather, these cases simply indicate where multiple characteristics may be relevant in determining the outcome.

⁴ Both of these consultations are available at www.officeforstudents.org.uk/outcomes-and-excellence/.

⁵ For the purposes of ensuring a consistent modelling approach across each student outcome, graduates with unknown Standard Occupational Classification (SOC) codes for their employment activity were removed from the analysis of progression rates. In our proposed definition for the progression indicator (IPEMPINDNUM), these graduates are instead assigned a weighted value (see IPEMPSOCWEIGHT) derived from the population of graduates at the provider with the graduate's mode of study and broad level of study who reported being employed, with no other positive outcomes. In this analysis, however, the Bernoulli distributional assumption in our modelling approach requires the outcome to be binary (equal to 0 or 1). These removed observations never represented more than 3 per cent of the population under consideration.

⁶ For more details on the data indicators, see the consultation at www.officeforstudents.org.uk/publications/student-outcomes-and-teaching-excellence-consultations/student-outcomes-data-indicators/.

⁷ This document reports analysis of completion outcomes as defined by the cohort-tracking method, which measures the proportion of students who can be tracked through to the end of their course.

11. Once again, it is important to note that this analysis can only account for factors which we can observe in the data. It is possible that there remain unobserved factors correlated both with the student characteristic of interest and with the outcome itself. These unobserved factors could introduce statistical bias.

Interpretation of data charts

12. The charts associated with this release⁸ present percentage point differences in continuation, completion and progression rates between one group of students and a chosen reference group, after controlling for a given set of factors.
13. These differences are reported across seven 'split indicators': age group, disability type, ethnicity, sex, Index of Multiple Deprivation (IMD) quintiles, Tracking Underrepresentation by Area (TUNDRA) and free school meals eligibility.
14. The relevant 'reference group' is shown above each chart in the descriptive title. This indicates the group of students against which the outcomes of other student groups are compared. By default, this is the largest category within the characteristic chosen in the filter above the charts. The exceptions to this are for TUNDRA, IMD and free school meals eligibility, where the group with the highest average positive outcomes offers a more meaningful comparison than the 'Unknown' group, which is sometimes larger.
15. The bar labelled 'no other factors (actual differences)', shows the actual difference in outcomes between the chosen student group and the reference group, averaged over the four most recent years of available data (or two years for progression).
16. Each of the bars beneath this indicate the estimated difference in continuation, completion or progression rates after controlling for a given set of factors through a statistical model. The broad group of factors included in the model are shown on the left-hand side and additional bars can be added by adjusting the filter under the heading 'Controlling for:'. Hovering over each bar shows a more detailed list of the factors being controlled for, alongside a summary of the data.
17. By default, the bars at the bottom of the charts represent the estimated difference in student outcomes after controlling for all factors.
18. The bars may either be negative or positive. A negative bar indicates that the student group selected in the filter has a lower positive outcome rate than the reference group, after controlling for the factors shown on the left-hand side. A positive bar indicates that the student group selected in the filters has a higher positive outcome rate than the reference group, after controlling for the factors shown on the left-hand side. In most cases, the reference group has a higher positive outcome rate, so the bars are negative.
19. After controlling for other factors, the bars can do one of three things, compared with the 'actual difference' shown at the top of the charts:

- a. **If the bar gets smaller**, this suggests the 'actual difference' is overstating the extent of the relationship between the characteristic and the outcome. The way in which other

⁸ Available at www.officeforstudents.org.uk/data-and-analysis/student-outcomes-and-experiences-data-dashboards/exploring-student-outcomes-dashboard/.

factors differ between the two student groups is contributing to (at least some of) the actual difference in outcomes. See paragraph 22 for an example.

- b. **If the bar gets larger**, this suggests the 'actual difference' is understating the extent of the relationship between the characteristic and the outcome. In other words, because students in this group tend to hold other characteristics which are also associated with the outcome in question, but in the *opposite* direction to the observed difference, this masks the true extent of the relationship between the characteristic and the outcome. See paragraph 26 for an example.
- c. **If the bar does not change much**, such that the error bars still overlap with those of the 'actual difference' at the top, this means there is little evidence to suggest that the combination of factors on the left hand side are making any net contribution to the difference in outcomes between the two student groups.⁹ It is possible that different factors within the group on the left hand side are related with the outcome in different ways.

20. In all these cases, there may still be some other factor that is not observed or captured in the data, which is influencing the size and direction of the difference that remains.

21. In addition to this, the differences are estimated from statistical models with uncertainty, which is indicated by 95 per cent confidence intervals to the right-hand side of the charts.¹⁰ These indicate how much (observable) uncertainty there is around a given statistic; we would expect the true value of the statistic to lie between the intervals 95 per cent of the time, given the data in front of us. In other words, we are 95 per cent 'confident' that the true value lies between the two intervals. The 95 per cent significance level was primarily chosen to be illustrative of the observable statistical uncertainty. It also provides a tolerance of 'Type II' errors that suits our uses on this occasion, based on our expert judgement.¹¹

Examples

22. As an example, consider the difference in continuation rates of male (88.2 per cent) and female (90.9 per cent) UK-domiciled full-time undergraduates, which amounts to 2.7 percentage points.

23. Our analysis shows that after controlling for differences in the Level 3 qualifications held by male and female students on entry to their course (such as A-levels), the difference in continuation rates reduces to an estimated 2.4 percentage points. In other words, while differences in the entry qualifications of male and female students account for some of the difference in continuation rates, there remains a difference of 2.4 percentage points even after controlling for this.

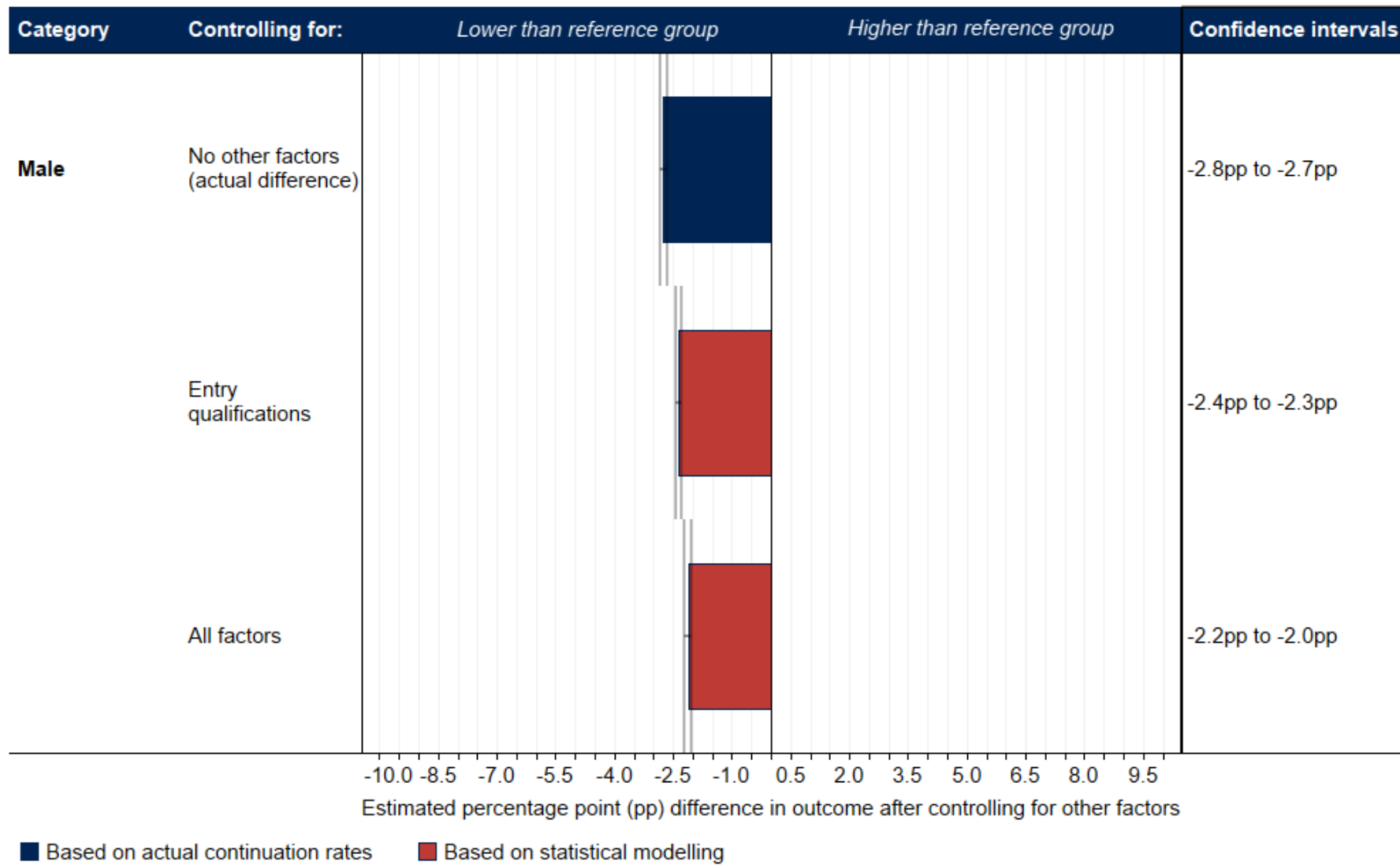
⁹ This comparison of confidence intervals does not represent a formal test of statistical significance.

¹⁰ For more information on how these confidence intervals are calculated, see Annex A.

¹¹ In statistical hypothesis testing, a Type II error occurs when one accepts a null hypothesis that is actually false. For example, under the null hypothesis that a given difference in continuation rates is equal to zero, we would be making a Type II error if we concluded, given the data in front of us, that the difference in continuation rates was equal to zero, but the true value was in fact different from zero. Selecting a higher level of statistical significance means a Type II error is less likely to occur.

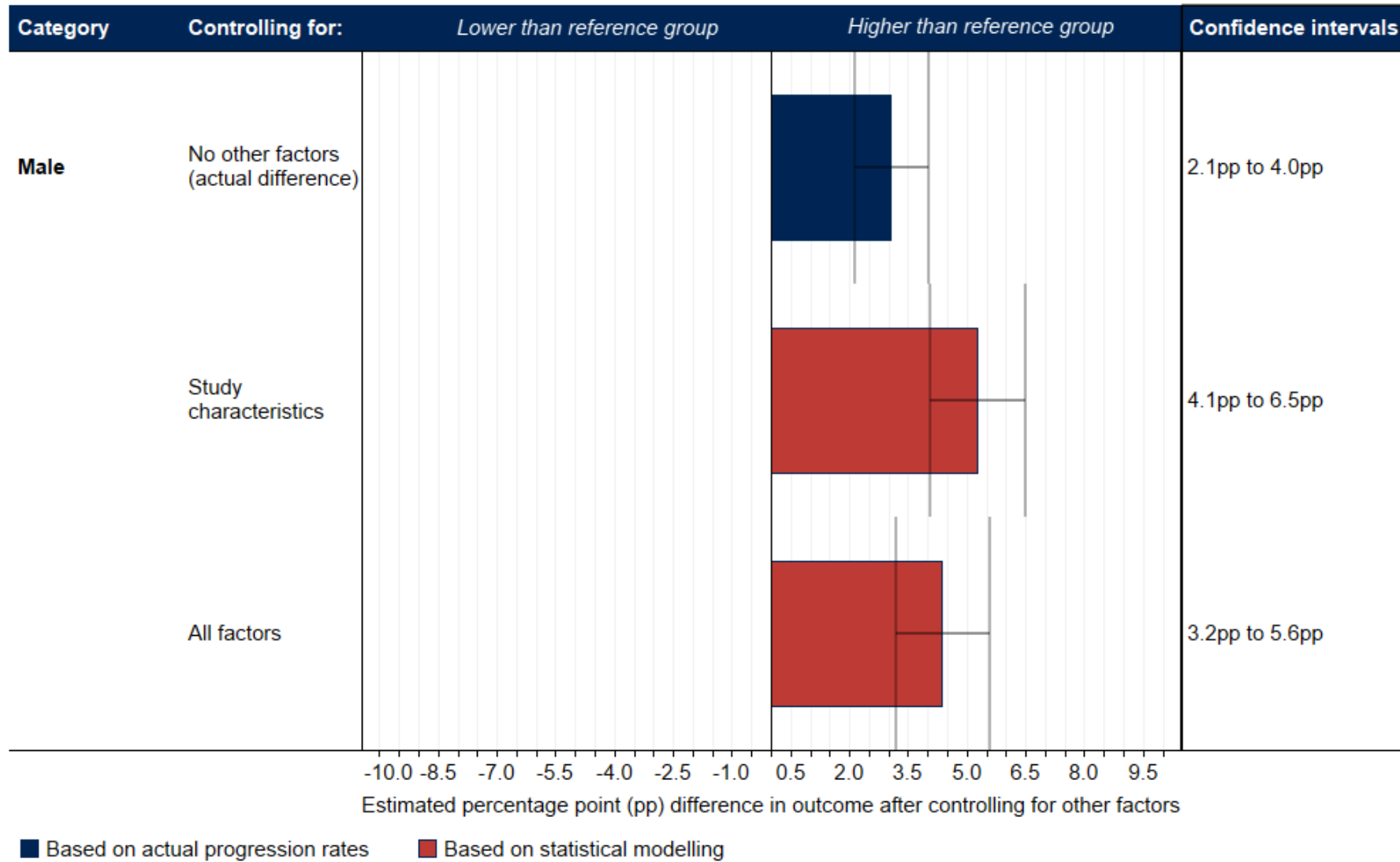
24. In fact, after controlling for all other observed factors, from equality characteristics to the types of study undertaken, while the difference does reduce further, it remains at an estimated 2.1 percentage points. This remaining difference is determined by a combination of:
- a. genuine differences in the current student experience between male and female students, where one group faces barriers to continuation that the other does not; and
 - b. other unobserved factors that differ between male and female students, which are also related to the likelihood of continuation, but have not been controlled for in the statistical model.

Figure 1: Estimated difference in continuation rates between male and female UK-domiciled full-time undergraduates after controlling for other factors



25. It is important to note that this method can only account for observed factors which are included in the model. For example, if female students are generally more motivated than male students (which we cannot observe in the data) and if more motivated students are more likely to continue with their studies, then the difference between male and female students after controlling for all other factors would still not represent the true effect of sex on continuation rates – some of the difference would simply reflect the effect of motivation on continuation instead.
26. In some cases, differences in outcomes may increase after controlling for other factors. The interpretation is slightly different here. For example, consider the difference in progression rates between male (81.9 per cent) and female (78.8 per cent) UK-domiciled part-time undergraduates, which amounts to 3.1 percentage points.
27. Our analysis shows that, after controlling for underlying differences in the study characteristics of male and female part-time students, such as their subject of study, this difference in progression rates rises to an estimated 5.3 percentage points.
28. In other words, despite the fact that male part-time students tend to study courses which are associated with lower progression rates, they remain more likely to progress into managerial or professional employment, or further study, 15 months after graduation. This suggests that the observed 3.1 percentage point gap is understating the true extent to which being female is associated with lower progression rates for part-time students.

Figure 2: Estimated difference in progression rates between male and female UK-domiciled part-time undergraduates after controlling for other factors



Annex A: Methodology

Populations and outcomes

1. There are eight student populations for which three student outcomes (continuation, completion and progression¹²) are analysed in this publication.¹³ These populations are given by all possible combinations of the three categories below:
 - a. **Student domicile:** UK-domiciled or non-UK domiciled
 - b. **Level of study:** undergraduate or postgraduate
 - c. **Mode of study:** full-time or part-time.
2. These population splits were chosen because there are structural differences in the availability of data and in the relationships between student outcomes and student characteristics across these groups.
3. For each population and outcome, we have aggregated the four most recent years of available data (or two years for progression, using the available Graduate Outcomes survey data) to maximise statistical power and borrow strength across years. This aggregation of years is aligned with the approach taken in our calculation of indicators of student outcomes.
4. The exact years differ according to modes of study and the student outcome under consideration:
 - a. For **full-time continuation**, both the year of entry and the following year are needed to determine whether an individual continued into their second year of study. Given that 2019-20 is the most recent year available, we considered entrants between 2015-16 and 2018-19.
 - b. **Part-time continuation** is measured in the third year of study to account for longer courses. Given 2019-20 is the most recent year of data available, we analyse part-time entrants between 2014-15 and 2017-18.
 - c. For **full-time completion**, for which four years of data are needed to track students through to the end of their course, we considered entrants between 2012-13 and 2015-16.

¹² Progression rates are not analysed for non-UK domiciled graduates. This is due to the increased risk of unobserved non-response bias arising from consistently lower response rates in the Graduate Outcomes survey amongst non-UK domiciled graduates. In addition, visa rules around employment and further study (see <https://www.gov.uk/graduate-visa>) may mean that survey responses for non-UK domiciled graduates who remain in the UK are unrepresentative of the wider population.

¹³ The student outcomes analysed here are those within scope of our current B3 consultation (see www.officeforstudents.org.uk/publications/student-outcomes-and-teaching-excellence-consultations/student-outcomes-data-indicators/). In future, this analysis could be extended to other measures of the student lifecycle, such as access and attainment.

- d. For **part-time completion**, for which six years of data are needed to track students through to the end of their course, we considered entrants between 2010-11 and 2013-14.
 - e. For both full-time and part-time **progression**, we considered graduates from the 2017-18 and 2018-19 cohorts, using all the available data from the Graduate Outcomes survey to identify the activities of graduates 15 months after graduation.¹⁴
5. Detailed population restrictions for each student outcome and definitions of the outcomes themselves can be found in our indicators consultation document.¹⁵
6. To avoid so-called ‘complete separation’ of data points and other instability in the statistical models described below, we excluded student data for providers where either there were fewer than 30 students registered in total, or there was no variation in the outcome. This amounted to excluding between 360 and 1,790 students from each of the various populations and outcomes in this analysis.

Split indicators

7. This analysis reports differences in outcomes across the following student characteristics or ‘split indicators’:
- a. Age on entry
 - b. Disability type
 - c. Ethnicity
 - d. Index of Multiple Deprivation (IMD)
 - e. Tracking underrepresentation by area (TUNDRA)
 - f. Sex
 - g. Free school meals (FSM) eligibility.
29. These characteristics were chosen because they broadly align with the ‘split indicators’ that we have proposed to use in the assessment of registration condition B3 and in the Teaching Excellence and Student Outcomes Framework (TEF) in our in our indicators consultation document.¹⁶ Although this analysis is reported at a more granular level across these split indicators.
30. In other analyses, we additionally include Associations Between Characteristics of Students (ABCS) measures as split indicators.¹⁷ However, given that ABCS measures are themselves

¹⁴ See <https://www.graduateoutcomes.ac.uk/>.

¹⁵ Available at www.officeforstudents.org.uk/publications/student-outcomes-and-teaching-excellence-consultations/student-outcomes-data-indicators/.

¹⁶ Available at www.officeforstudents.org.uk/publications/student-outcomes-and-teaching-excellence-consultations/student-outcomes-data-indicators/.

¹⁷ See www.officeforstudents.org.uk/data-and-analysis/associations-between-characteristics-of-students/.

derived from statistical modelling of student outcomes given a set of student characteristics, they are not included in this analysis to avoid duplication of information in the fitting of the statistical models.

Statistical modelling

31. After calculating the actual percentage point differences in outcomes across the split indicators listed above, we then used statistical modelling to estimate these same differences once again, this time controlling for underlying differences in other student characteristics and information about the higher education course undertaken.
32. All factors were modelled as fixed effects using binary logistic regression, as in Equation 1 below.¹⁸

Equation 1

$$\textit{Student outcome}_i \sim \textit{Bernoulli}(N, \pi_i)$$

where N is the number of observations in the modelling population and π_i is the probability of a positive outcome for observation i .

$$\textit{logit}(\pi_i) = \beta_0 + \tilde{\beta}_1 X_{1,i} + \dots + \tilde{\beta}_k X_{k,i}$$

Note: Each $\tilde{\beta}$ denotes a vector of fixed effects coefficients (of different sizes), the elements of which relate to an attribute within the associated explanatory variable, $X_1 \dots X_k$, which denotes a vector of dummy variables. There are k explanatory variables in the model.

33. It is possible that the relationship between a given characteristic and outcome varies according to other characteristics; for example, the relationship between ethnicity and continuation may vary depending on the sex of a student. Interaction terms offer a way of controlling for these relationships. In any future updates, sensitivity analyses could be undertaken to understand whether the inclusion of interaction terms would improve the precision of model estimates or shed light on the extent to which relationships between student characteristics and outcomes vary according to other characteristics.

Choice of model factors and groups

34. From the outset, we considered as comprehensive a list of factors as possible as candidates for inclusion in this analysis. They are listed below. Definitions of factors which were ultimately included in the statistical models can be found in Annex B.
35. One of the factors considered for inclusion in the statistical models is the higher education provider where the student is registered. There are other characteristics of a provider which might correlate with the outcomes of its students, such as the number of students registered, the 'mission group' of the provider or the average tariff scores of its entrants. The decision was taken not to control for this information when modelling student outcomes, as it is not our

¹⁸ While a random intercept could have been used to control for the provider at which a student is registered, as we have done in other analyses (e.g. 'Data analysis of unconditional offers: Update' at <https://www.officeforstudents.org.uk/publications/data-analysis-of-unconditional-offers-update/>), since we only want to control for variation in student outcomes at different providers, and fixed effects are less computationally intensive, a fixed effects approach was favoured for this analysis.

intention to explain or excuse student outcomes through characteristics which are within the provider's control.

Candidate factors for inclusion in statistical models

- Age on entry
- Participation of Local Areas (POLAR4)
- Tracking Underrepresentation by Area (TUNDRA)
- Index of Multiple Deprivation (2019)
- Income Deprivation Affecting Children Index (2019)
- Disability
- Degree awarding body
- Local/distance learning
- Entry qualifications
- Ethnicity
- Level of study (or qualification level)
- Sex
- Subject of study
- Socioeconomic background
- Parental education
- Care experience
- Sandwich courses
- Length of course
- Foundation years
- Free school meal eligibility
- Region of study
- School type
- Household Residual Income (HRI)
- Estrangement
- Home region
- Higher education provider
- Sexual orientation
- Religion or belief
- Gender identity
- Degree classification (progression only)
- Year of study
- Graduate Outcomes quintiles

36. All factors above were considered as candidates for inclusion in statistical modelling through exploratory analysis of their relationship with each student outcome within each population. The following section sets out the formal criteria against which these factors were assessed to determine whether they should be included as fixed effects in the final statistical models.

Criteria for inclusion in statistical models

37. Sufficient data quality and coverage

- a. We first determined coverage using the same approach as in our report: 'Equality, diversity and student characteristics data',¹⁹ which applies the OfS data quality framework to determine whether data quality and coverage are sufficient for reporting data on various student characteristics.²⁰ These decisions are summarised in Table B1 of this report.
- b. Having done this, we then ensured the proportion of observations that are unknown (after linking and suppression) did not exceed approximately 60 per cent. We chose this approximate threshold because, in our expert judgement, it strikes an appropriate balance between retaining as much useful information as possible and ensuring that we do not rely on data that is unrepresentative of the underlying population. Nonetheless, the final decision to exclude a factor depended also on the other criteria listed below; for example, we might still include a certain factor if 61 per cent of records were missing data, but there was nonetheless a strong relationship with the outcome and no reason to suspect the available data was misrepresentative of the wider population.

¹⁹ See www.officeforstudents.org.uk/publications/equality-diversity-and-student-characteristics-data-2010-11-to-2019-20/.

²⁰ See 'Annex A: Data quality framework – a method for assessing the quality of student characteristic data' at www.officeforstudents.org.uk/publications/differences-in-student-outcomes-further-characteristics/.

38. Correlation with the outcome

- a. We considered the extent to which the factor was statistically significantly related to the outcome, which we specified should hold at least at the 90 per cent confidence level.²¹ We chose a lower level of statistical significance than that used to calculate the illustrative confidence intervals described from paragraph 52 onwards. This was to ensure that we were not excluding factors which, when later modelled alongside other factors, were found to have a highly statistically significant relationship with the outcome in question.
- b. Where the statistical significance was marginal, we further considered the magnitude of the differences in outcomes between student groups as well as the number of students in each group.
- c. On a small number of occasions, the split indicators themselves were not statistically significantly related with the student outcome under consideration. Where the criteria for inclusion in the model was otherwise met, it was decided to still include these factors in the statistical models regardless, since a finding of statistical non-significance remains of interest.

39. No strong correlation with other factors

- a. We ensured that no factor was strongly correlated with any other (multicollinearity), as this would mean the effect of each factor on the outcome could not be disentangled from one another, resulting in high statistical uncertainty around model estimates. We define 'strong multicollinearity' by a variance inflation factor (VIF) greater than or equal to 10.²²

40. Graduate Outcomes response rates (progression only)

- a. The Higher Education Statistics Agency (HESA) found no evidence that the Graduate Outcomes survey would benefit from weighting to correct non-response bias, for both the 2017-18 and 2018-19 graduate cohorts.²³ Nonetheless, given the range of characteristics and the granularity of some student groups in our analysis, we also considered whether Graduate Outcomes response rates were sufficiently high across each category within a given factor. This applies only to progression rates, since this is the only outcome which uses data from the Graduate Outcomes survey.

²¹ This was evaluated by running a model of the same format as Equation 1, with only a single factor included as a fixed effect. A p-value was then calculated against the null hypothesis that the model provides no better fit than an 'intercept-only' model. In other words, we checked whether the factor could improve prediction of student outcomes, such that it is better than random guessing.

²² We used this relatively high threshold partially because of the size of our data, which mitigates the increased standard errors resulting from multicollinearity. Also, our statistical inference is concerned only with the 'split indicators' (the characteristics across which differences in outcomes are calculated), so inflated standard errors around other factors, which we are only controlling for, are of less concern.

²³ See <https://www.hesa.ac.uk/blog/21-06-2021/graduate-outcomes-no-need-to-weight>.

41. These criteria have some parallels with the benchmarking principles²⁴ proposed for benchmarking the indicators that will inform our regulation of student outcomes and the future TEF scheme.

Groups of factors in statistical models

42. The factors which were judged to have met the criteria above were then sorted into groups of similar characteristics. These are:

- a. Equality characteristics
- b. Area-based measures
- c. Entry qualifications
- d. Study characteristics
- e. Provider
- f. All factors above except provider
- g. All factors above.

43. For each student population and outcome, each of these groups of factors were then included as fixed effects in separate models. This allows us to identify the types of factors which appear to contribute most to observed differences in student outcomes.

44. The exact composition of these groups for each population and outcome can be found by hovering over the bars in the dashboards and in the datafiles associated with this release. Definitions of these factors can be found in Annex B.

45. It should be noted that some groups of factors occasionally contain only one characteristic where the criteria for inclusion was met.

Estimation of percentage point differences in outcomes after controlling for other factors

46. We estimate differences in outcomes after controlling for other factors using the following method.

47. Firstly, the statistical models are used to predict the outcome for each student, where all of the categories within one of the 'split indicators' described above are set to be equal to the reference group.²⁵ In other words, we estimate the likelihood of a positive outcome for each student, as if they belonged to the reference group instead of the value actually recorded in the data, while all their other characteristics remain unchanged.

48. Averages of these predictions are then taken at the sector level to obtain model estimated rates of positive outcomes, both before and after the characteristic is set to the reference

²⁴ Available at Annex D of the indicators consultation: www.officeforstudents.org.uk/publications/student-outcomes-and-teaching-excellence-consultations/student-outcomes-data-indicators/.

²⁵ By default, the reference group is chosen to be the largest category within the characteristic in question. The exceptions to this are for TUNDRA, IMD and free school meals eligibility, where the group with the highest average positive outcomes offers a more meaningful comparison than the 'Unknown' group, which is sometimes larger.

group. The difference between these two averages then represents an estimated difference in student outcomes after controlling for underlying differences in the other factors included in the model.

49. This method is mathematically equivalent to transforming the relevant parameter estimate from the statistical model into a 'marginal effect' in percentage points, relative to the reference category. With categorical variables, this result is more straightforward to interpret than an 'average marginal effect'; the 'average' student will invariably sit between categories recorded in the data, which makes the result hypothetical and harder to interpret.
50. These model predictions cannot truly capture the outcome of a given student if they were in the reference group instead, because this is hypothetical. The extent to which these estimates identify a genuinely 'causal effect' will depend on the existence of other unobserved factors which correlate both with the outcome and with the characteristic of interest. For example, if female students are more motivated than male students (which we cannot observe in the data) and if more motivated students are more likely to continue with their studies, then the difference between male and female students after controlling for all other factors would still not represent the true effect of sex on continuation rates – some of the difference would simply reflect the effect of motivation on continuation instead. This is known as 'omitted variable bias'. In other words, our estimated differences in outcomes can only account for observed factors which are included in the model; there may be other confounding effects which we have not been able to control for.
51. All estimated differences and rates are reported as rounded to one decimal place, while the reported number of students within each category is rounded to the nearest five. Student numbers are suppressed where there are fewer than 23 students (prior to rounding) in the chosen category. These thresholds were chosen to retain as much information as possible, while ensuring that information about individuals cannot be identified from the data.

Confidence intervals

52. Confidence intervals indicate how much (observed) uncertainty there is around a given statistic; we would expect the true value of the statistic to lie between the intervals 95 per cent of the time, given the data in front of us. In other words, we are 95 per cent 'confident' that the true value lies between the intervals.
53. It is important to note that the 95 per cent significance level was primarily chosen to be illustrative of the observable statistical uncertainty. It also provides a tolerance of 'Type II' errors that suits our uses on this occasion, based on our expert judgement.²⁶

²⁶ In statistical hypothesis testing, a Type II error occurs when one accepts a null hypothesis that is actually false. For example, under the null hypothesis that a given difference in continuation rates is equal to zero, we would be making a Type II error if we concluded, given the data in front of us, that the difference in continuation rates was equal to zero, but the true value was in fact different from zero. Selecting a higher level of statistical significance means a Type II error is less likely to occur.

54. For our estimated differences in outcomes, 95 per cent confidence intervals are obtained by first adjusting the predicted probability for each student according to the standard error around the parameter estimate for the characteristic that has been set to the reference category, as described in the previous section. For example, the predicted probabilities for male students ‘as if they were female’ are adjusted by $\pm 1.96 \times \text{s. e.}(\hat{\beta}_{\text{Sex, Male}})$, where $\text{s. e.}(\hat{\beta}_{\text{Sex, Male}})$ is the standard error for the estimate of the fixed effect of being male, relative to female.
55. As before, averages are then taken at the sector level to obtain model estimated rates of positive outcomes where the characteristic is set to the reference group, but this time each of the predicted probabilities have been adjusted by $\pm 1.96 \times \text{s. e.}(\hat{\beta})$. Then the differences between the actual rate and the upper and lower estimates of the rate (where the characteristic was set to the reference group) indicate the intervals between which we are 95 per cent confident the true difference lies, after controlling for underlying differences in the other factors included in the model.
56. For the actual difference in outcomes, represented by the bar labelled ‘No other factors (actual difference)’ in the charts, confidence intervals are similarly calculated using standard errors from a statistical model where only the characteristic across which differences are being measured is included as a fixed effect. However, because the actual difference in outcomes is already known, rather than being estimated, these confidence intervals have a slightly different interpretation. Instead of indicating uncertainty arising from the statistical models, these intervals represent the sensitivity of the difference to the effects of random variation in the outcome being measured. A large confidence interval typically indicates that there are small numbers of students informing the measure and it is advisable to be cautious when interpreting these values.
57. When multiple statistics are calculated on a given topic, it is often expected that users will wish to make comparisons between those statistics. To the extent that those statistics include information about statistical uncertainty, that uncertainty can be underestimated depending on the nature of the multiple comparisons that are being made. For example, in the case of 95 per cent confidence intervals, the likelihood that the computed confidence interval includes the true value of underlying performance may be substantially lower than the intended 95 per cent if multiple comparisons are being made. To overcome this, adjustments can be made to the calculations to control the error or false discovery rates (such as the Bonferroni correction). To align with the proposals in our construction of student outcome and experience indicators to inform the assessment of condition B3 or the TEF, we have not made any such adjustments for multiple comparisons within this analysis. For more information on this proposal, see paragraph 29 of the document, ‘Description of statistical methods’.²⁷
58. While we have proposed not to adjust for multiple comparisons, we do ask users who wish to make multiple comparisons to exercise caution when making their judgements because of the higher risk of false discovery when using lower levels of statistical confidence.

²⁷ Available at www.officeforstudents.org.uk/publications/student-outcomes-and-teaching-excellence-consultations/student-outcomes-data-indicators/.

We are actively seeking feedback regarding these experimental statistics. If you have any queries or suggestions, please contact Stanley Rudkin at official.statistics@officeforstudents.org.uk.

Annex B: Definitions

1. The data in this release is drawn from the following sources:
 - a. the Higher Education Statistics Agency's (HESA's) student record
 - b. the HESA Alternative Provider (AP), or Student Alternative (SA), record
 - c. the Education and Skills Funding Agency's (ESFA's) individualised learner record (ILR)
 - d. the Student Loans Company (SLC) data
 - e. the Graduate Outcomes survey from HESA
 - f. and the National Pupil Database (NPD) provided by the Department for Education.²⁸
2. In addition, the Income Deprivation Affecting Children Index (IDACI) and Index of Multiple Deprivation (IMD) quintiles are produced by the Ministry of Housing, Communities and Local Government²⁹ and are merged with the ILR and HESA student records.
3. When designing model factors, we first tried to use the most detailed grouping available, since this provides the statistical models with the most nuance and information. However, we also balanced this with the fact that smaller groups could lead to model instability. For example, for subject information, we used level 2 of the Common Aggregation Hierarchy (CAH2), since it provided more detail than the CAH1 grouping, while the CAH3 grouping was found to introduce model instability.
4. For the characteristics in this release which are also reported in our 'Equality, diversity and student characteristics data',³⁰ the same coverage decisions were applied, as summarised in Table B1 below. The contextual information which informed those decisions can be found in Annex B of the 'Equality, diversity and student characteristics data' report.
5. For all other characteristics, which generally relate to course information as opposed to personal characteristics, the quality and coverage were deemed sufficient for inclusion in the statistical models in all cases, according to principles from the OfS data quality framework.³¹

²⁸ The DfE does not accept responsibility for any inferences or conclusions derived from the NPD data by third parties.

²⁹ See <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015>.

³⁰ See www.officeforstudents.org.uk/publications/equality-diversity-and-student-characteristics-data-2010-11-to-2019-20/.

³¹ See 'Annex A: Data quality framework – a method for assessing the quality of student characteristic data' at www.officeforstudents.org.uk/publications/differences-in-student-outcomes-further-characteristics/.

Table B1: Summary of population coverage of student characteristics from the report ‘Equality, diversity and student characteristics data’

| Characteristic | Data source(s) | First year available | Domicile | Level of study | Other |
|---|---------------------|--|--------------------------------|----------------|---|
| Age on entry | HESA and ILR | 2010-11 | All | All | - |
| Disability type | HESA and ILR | 2010-11 | All | All | - |
| Ethnicity | HESA and ILR | 2010-11 | UK | All | - |
| Sex | HESA and ILR | 2010-11 | All | All | - |
| Religion or belief | HESA Student & SA | 2017-18 | All | All | - |
| Sexual orientation | HESA Student & SA | 2015-16 | All | All | - |
| Parental education | HESA Student | 2012-13 | All | All | - |
| Free school meals eligibility | NPD | 2014-15 entrants, 2016-17 qualifiers, 2016-17 all students | UK | UG | Attended English schools, under 21 on entry |
| Participation of Local Areas (POLAR4) | OfS, HESA and ILR | 2010-11 | UK | UG | Under 21 on entry |
| Index of Multiple Deprivation (IMD) | MHCLG, HESA and ILR | 2010-11 | English | All | - |
| Care experience | HESA Student | 2014-15 entrants, 2016-17 qualifiers, 2016-17 all students | UK | UG | - |
| Household Residual Income (HRI) | SLC | 2011-12 | English, Welsh, Northern Irish | UG | Full-time |
| Estrangement | SLC | 2014-15 | English, Welsh, Northern Irish | UG | Under 25 in year |
| Socioeconomic background | HESA Student | 2015-16 | UK | UG | Full-time |
| Tracking underrepresentation by area (TUNDRA) | OfS, HESA and ILR | 2010-11 | English | UG | Under 21 on entry |

Note: UG = Undergraduate students only. Other acronyms defined above

6. Where students refused to provide information, or it is not applicable, or otherwise unknown, we have grouped students into a single 'Unknown' category, since these distinctions do not provide any further information for the purposes of understanding differences in student outcomes.

Throughout this annex, more detail on variables with the prefix 'IP' can be found in the 'Core algorithms' document.³²

Continuation

7. For full-time students, continuation indicators are calculated and reported for each cohort of entrants to higher education courses between 2015-16 and 2018-19. Their continuation outcomes are measured as at a census point 1 year and 15 days after the date on which they commenced their studies.³³
8. For part-time students, continuation indicators are calculated and reported for each cohort of entrants to higher education courses between 2014-15 and 2017-18. Their continuation outcomes are measured as at a census point 2 years and 15 days after the date on which they commenced their studies.³⁴

Completion

9. Full-time completion indicators are calculated and reported for each cohort of entrants to higher education courses between 2012-13 and 2015-16. Their completion outcomes are measured as at a census point four years and 15 days after the date on which they commenced their studies.³⁵
10. Part-time completion indicators are calculated and reported for each cohort of entrants to higher education courses between 2010-11 and 2013-14. Their completion outcomes are measured as at a census point six years and 15 days after the date on which they commenced their studies.³⁶

Progression

11. Progression indicators are calculated and reported for each cohort of leavers from higher education courses between 2017-18 and 2018-19. Rates of progression to managerial and professional employment or further study beyond graduation are produced based on responses to the Graduate Outcomes survey.³⁷
12. Progression rates are not analysed for non-UK domiciled graduates. This is due to the increased risk of unobserved non-response bias arising from consistently lower response rates

³² Available at www.officeforstudents.org.uk/publications/student-outcomes-and-teaching-excellence-consultations/student-outcomes-data-indicators/.

³³ For more information, see proposal 5 of the indicators consultation.

³⁴ For more information, see proposal 5 of the indicators consultation.

³⁵ For more information, see proposal 6 of the indicators consultation.

³⁶ For more information, see proposal 6 of the indicators consultation.

³⁷ For more information, see proposal 7 of the indicators consultation.

in the Graduate Outcomes survey amongst non-UK domiciled graduates. In addition, visa rules³⁸ around employment and further study may mean that survey responses for non-UK domiciled graduates who remain in the UK are unrepresentative of the wider population.

13. For the purposes of ensuring a consistent modelling approach across each student outcome, graduates with unknown Standard Occupational Classification (SOC) codes for their employment activity were removed from the analysis of progression rates. In our proposed definition for the progression indicator (IPEMPINDNUM), these graduates are instead assigned a weighted value (see IPEMPSOCWEIGHT) derived from the population of graduates at the provider with the graduate's mode of study and broad level of study who reported being employed, with no other positive outcomes. In this analysis, however, the Bernoulli distributional assumption in our modelling approach requires the outcome to be binary (equal to 0 or 1). These removed observations never represented more than 3 per cent of the population under consideration.

Domicile

14. UK and non-UK domiciled students are analysed separately in this release. For more detail on the source of this information, see IPDOM in the 'OfS 2021 core algorithms' document.³⁹

Level of study

15. Undergraduate and postgraduate students are analysed separately in this release.
16. In addition to this, a more detailed grouping of level of study (defined by IPLEVEL and IPLEVELNUM) within undergraduate and postgraduate cohorts is included as a fixed effect in the statistical models. These groupings are listed below, with the associated IPLEVEL values in parentheses:
 - a. Undergraduate
 - i. First degree (DEG)
 - ii. Other qualifications with a postgraduate component (PUGD)
 - iii. Other undergraduate (Levels 5 and 6) (OUG, IPLEVELNUM= 5, 6)
 - iv. Other undergraduate (Level 4 and other) (OUG, IPLEVELNUM = 0, 4)
 - b. Postgraduate
 - v. Postgraduate research (PHD)
 - vi. Postgraduate taught masters (PGTM)
 - vii. PGCE

³⁸ See <https://www.gov.uk/graduate-visa>.

³⁹ Available at www.officeforstudents.org.uk/data-and-analysis/institutional-performance-measures/technical-documentation/.

viii. Other postgraduate. (OPGT, OPGR, PUGO).

Subject of study

17. In this analysis, subject codes from the Joint Academic Coding System (JACS) are mapped to Level 2 of version 1.3 of the Common Aggregation Hierarchy (CAH), which contains 35 subject groups.⁴⁰
18. In the statistical models, each subject has a different observation but the data is weighted such that each student instance is only counted once.

Mode of study

19. Mode of study is defined as at the start of the course. See IPSTARTMODE. Modes of study are grouped into full-time and part-time. Students with other modes of study are excluded from this analysis.

Age on entry

20. Age on entry refers to the age of the student on 31 August in the year they commence their studies (see IPSTARTAGEBAND). In this analysis, age is sorted into six groups.

Disability type

21. Disability information is self-reported by students at the point of starting their course. Providers are advised that this information should then be updated annually, but it is unclear how common this is in practice. See IPDISABLETYPE.

Ethnicity

22. Ethnicity information is provided by the student on the basis of their own self-assessment (see IPETHNICDETAIL). For this release, ethnicity information is categorised into 15 groups.
23. Ethnicity information is restricted to UK-domiciled students in this release, since its collection is restricted to UK-domiciled students on the HESA Student record.⁴¹

Sex

24. Sex is reported by the student as 'Male', 'Female', or 'Other sex'. See IPSEX.

Religion or belief

25. Religion or belief is collected on the HESA Student record⁴² and HESA AP record⁴³ only. It records the religious belief of the student, on the basis of their own self-assessment. The field is collected for all students, having become compulsory in 2017-18.

⁴⁰ See <https://www.hesa.ac.uk/innovation/hecos>.

⁴¹ See <https://www.hesa.ac.uk/collection/c19051/a/ethnic>.

⁴² See <https://www.hesa.ac.uk/collection/c19051/a/reblf>.

⁴³ See <https://www.hesa.ac.uk/collection/c19054/a/reblf>.

26. It was decided to use this data for all students from 2017-18 onwards, because of the changes in reporting brought about by the move to make the collection mandatory.

Sexual orientation

27. Sexual orientation information is collected on the HESA Student record⁴⁴ and HESA AP record⁴⁵ only. It records the sexual orientation of the student, on the basis of their own self-assessment. The field is optional for all students and is used from 2015-16 onwards.

Parental education

28. Parental higher education data is collected on the HESA Student record only, indicating whether one or more of the student's parents have a higher education qualification.⁴⁶ This data is primarily collected as part of Universities and Colleges Admissions Service (UCAS) applications. Providers are also expected to collect this information for eligible students who did not use UCAS.

29. It was decided to use this data for all students from 2012-13 onwards.

Free school meal eligibility

30. The free school meal (FSM) measure is based on the population of students matched to the Department for Education's National Pupil Database (NPD)⁴⁷ who were identified as having ever been eligible for FSM.

31. The NPD census for key stage 4 (KS4) covers pupils attending maintained (and independent) schools in England, and censuses for academic years from 2009-10 to the latest have been matched to HESA and ILR student records. Since pupils are generally 15 years old in their last year of KS4, 2014-15 is the earliest academic year that a full cohort of higher education entrants, aged under 21 on entry, can be tracked back to the NPD.

32. Consequently, FSM data is used for entrants to higher education from 2014-15 onwards for analyses of continuation and completion. For analysis of progression rates, the first entrant year used is 2016-17, since qualifiers are older and so fewer students can be linked to an NPD record in earlier years.

33. Coverage of FSM in this analysis is limited to those who attended mainstream funded schools in England, who were aged under 21 on entry to higher education, studying undergraduate courses, and were domiciled in the UK.

⁴⁴ See <https://www.hesa.ac.uk/collection/c19051/a/sexort>.

⁴⁵ See <https://www.hesa.ac.uk/collection/c19054/a/sexort>.

⁴⁶ See www.hesa.ac.uk/collection/c19051/a/pared.

⁴⁷ See <https://find-npd-data.education.gov.uk/>.

Participation of local areas (POLAR4)

34. The participation of local areas (POLAR) classification⁴⁸ groups areas across the UK based on the proportion of young people who participate in higher education. It looks at how likely young people are to participate in higher education across the UK and shows how this varies by area. POLAR classifies local areas into five groups – or quintiles – based on the proportion of 18-year-olds who enter higher education aged 18 or 19 years old. Quintile 1 shows the lowest rate of participation. Quintile 5 shows the highest rate of participation. In England it is calculated at middle-layer super output area (MSOA). See ‘IPPOLAR4’.

Index of Multiple Deprivation (IMD)

35. The Index of Multiple Deprivation (IMD) 2019 is a measure of deprivation for small areas within England. It is calculated at lower-layer super output area (LSOA) level and uses a number of different measures to determine levels of deprivation. It is produced by the Ministry of Housing, Communities and Local Government (MHCLG).⁴⁹ In this analysis, we group areas into quintiles (as opposed to deciles), where the most deprived areas are in quintile 1 and the least deprived are in quintile 5.

Income Deprivation Affecting Children Index (IDACI)

36. The Income Deprivation Affecting Children Index (IDACI) 2019 measures the proportion of children under the age of 16 in low income households for a particular area in England. It is calculated at lower-layer super output area (LSOA) level and is a supplementary measure to the Index of Multiple Deprivation (IMD), and therefore has been constructed for MHCLG by Oxford Consultants for Social Inclusion (OCSI).⁵⁰ In this analysis, we use IDACI quintiles, where the most deprived areas are in quintile 1 and the least deprived are in quintile 5. The measure is not calculated for pupils or students domiciled outside of England.

Care experience

37. Data on care experience is collected for entrants on the HESA Student record only.⁵¹

38. For English providers, care experience can be recorded as one of two categories. Firstly, data is recorded as part of the UCAS application via student self-disclosure. Providers are then able to verify care status and, where they confirm the student was in care, this is recorded as a different category which takes precedence over the UCAS category. Providers can choose to what extent they verify care status and they may use different verification criteria. For the purposes of this analysis these two categories are combined into one called ‘care experienced’.

39. Since collection of this data began in 2013-14 for entrants only, and the quality for entrants is deemed sufficient from 2014-15 onwards, it was decided that coverage for qualifiers and all students (for whom collection will be lagged) should begin from 2016-17 onwards.

⁴⁸ For more details of the POLAR measure, see www.officeforstudents.org.uk/data-and-analysis/young-participation-by-area/.

⁴⁹ See <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019>.

⁵⁰ See <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019>.

⁵¹ See <https://www.hesa.ac.uk/collection/c19051/a/careleaver>.

40. For providers in England, HESA collects care experience information for all Home/EU undergraduate and PGCE students. However, to maintain consistency and avoid misrepresenting the non-UK domiciled and postgraduate populations, it was decided to use data for UK-domiciled undergraduate students only.

Household Residual Income (HRI)

41. Household Residual Income (HRI) is assessed based on gross income (before tax and National Insurance) after accounting for certain costs such as dependants and pensions. It is based on Student Loans Company (SLC) data, and so is only available for those who have applied for income-dependent financial support. Further details of the collection of HRI are published by the SLC.⁵²

42. HRI data is collected on continuous scale so we have condensed the data into four groups for this analysis. The groups we have chosen are: £0, £1-£25,000, £25,001-£42,600 and £42,601 and over. These boundaries were determined based on values of HRI used when assigning loans and bursaries.

43. The assessment of HRI differs depending on the student's dependency status. If a student is dependent on their parents, then HRI is based on parental income. However, if the student is independent then HRI is based on their income (and their partner's income, if they have one). Dependency status is determined by the SLC. Examples of why a student may be considered independent include being 25 years old and over, not living with their parents, being married, having children, and being a care leaver. A full list of reasons is published by the SLC.⁵³ The background, experiences and outcomes of dependent and independent students are different, so we treat the HRI data for these two groups of students separately, creating separate measures for each.

44. The disclosure of HRI is not a compulsory part of a student loan application. If a student or their parents are aware that their HRI is at a level that means their student loan entitlement will not be increased by income assessment, or if they have other evidence of low income which means submitting income is not necessary, then their financial information will not be recorded. This leads to a reduction in disclosure rates, which could mean the data presented is not representative of the wider student population.

45. For some courses, such as nursing prior to 2017-18, a student may have an SLC record but receive finance that is not assigned using HRI. Again, these students would not be required to provide this information.

46. Income-dependent maintenance loans were only introduced for part-time students from 1 August 2018 onwards.⁵⁴ This means that, before this point, very few part-time students reported any HRI information. We are therefore waiting for a consistent time series to be established before using HRI data for part-time students. In addition, because part-time

⁵² See the document 'How you're assessed and paid', available at www.gov.uk/student-finance/new-fulltimestudents.

⁵³ See the document 'How you're assessed and paid', available at www.gov.uk/student-finance/new-fulltimestudents.

⁵⁴ See 'Student finance: how you're assessed and paid', page 25, available at <https://www.gov.uk/government/publications/student-finance-how-youre-assessed-and-paid>.

distance learners are only eligible for a maintenance loan if they are studying long distance because of a disability,⁵⁵ it may be that the data will still not be a fair reflection of the profile of part-time students, even once a time series is well established.

47. While HRI data is collected for those studying towards a Postgraduate Certificate in Education (PGCE), we do not use this information in this release, since these courses are not representative of the wider postgraduate population.
48. The HRI data in this release applies to students domiciled in England, Wales and Northern Ireland only. Students domiciled in Scotland are not included because they do not submit their financial information to the SLC as part of their student support applications.
49. It was decided to use data for English, Welsh and Northern Irish full-time undergraduates from 2011-12 onwards, separated measures according to dependency status.

Estrangement

50. Student estrangement is recorded by the SLC on a case-by-case basis. To be recorded as estranged, a student must be irreconcilably estranged from their living biological or adoptive parents for a considerable period of time (usually at least 12 months). This being the case, there will be students who are estranged but are not recorded as such, because they do not have an SLC record or have chosen not to declare their estrangement to the SLC. There may also be students who consider themselves to be estranged from their parents but do not fulfil the requirements of the SLC to be recorded as estranged.
51. Furthermore, the SLC only records estrangement for students under 25 and there will be students who are 25 and over that are estranged from their parents but not recorded as such in the data. For this reason, these statistics are limited to students who are under 25 as of 31 August in their academic year reported.
52. Estrangement status can change during a student's course, but if a student was recorded as estranged at any point during their time at the provider we include them as part of the estranged population.
53. The SLC data is a record of the estranged students in the current year and does not record students as not estranged. We have therefore assumed that all students with a linked SLC record who are not recorded as estranged are not estranged. This results in us calculating a disclosure rate of 100 per cent for students with an SLC record relating to their current provider when applying our data quality framework. As mentioned above, there may be cases where a student considers themselves to be estranged but they are not recorded as such with the SLC and will be included in the 'not estranged' population in these calculations.
54. For more information on our use of estrangement data from the SLC, see 'Annex G: Estrangement' to our report 'Differences in student outcomes – further characteristics'.⁵⁶ Note

⁵⁵ See 'Student finance: how you're assessed and paid', page 25, available at <https://www.gov.uk/government/publications/student-finance-how-youre-assessed-and-paid>.

⁵⁶ Available at www.officeforstudents.org.uk/publications/differences-in-student-outcomes-further-characteristics/.

that, while the methodology behind estrangement information is the same, the population coverage in that report differs from this release.

55. While estrangement data is collected for those studying towards a PGCE, we do not use this information in this release, since these courses are not representative of the wider postgraduate population.
56. The estrangement data in this release applies to students domiciled in England, Wales and Northern Ireland only. Students domiciled in Scotland are not included because of a lack of data availability. Student Finance England, Student Finance Wales and Student Finance Northern Ireland have the same criteria for a student to be considered estranged.
57. It was decided to use this data for English, Welsh and Northern Irish undergraduates aged under 25 in the academic year reported, from 2014-15 onwards.

Socioeconomic background

58. National Statistics socioeconomic classification (NS-SEC)⁵⁷ data is collected on the HESA student record and further details regarding its collection can be found on their website.⁵⁸ NS-SEC data is also available on the ILR; however, this data is not currently considered useable by the OfS due to quality concerns.⁵⁹
59. The collection of the NS-SEC data differs depending on the student's entry age. For students aged 21 and over at entry, it is based on their occupation prior to starting their course. For students under 21, it is based on the occupation of their parent, step-parent or guardian who earns the most.
60. It should be noted that collection of socioeconomic background data by HESA is limited to UK-domiciled full-time undergraduates who entered their higher education instance via a UCAS scheme, although providers are encouraged to provide this information for other full-time undergraduates as well. To ensure the consistency of data recording, the use of NS-SEC data in this analysis has been limited to only those students with a UCAS Application Scheme Code.⁶⁰ We cannot know whether providers reporting NS-SEC for students who did not use UCAS are doing so using the same collection method and NS-SEC mapping that is used by UCAS.
61. It was decided to use this data for full-time UK-domiciled undergraduates, who entered their higher education instance via a UCAS scheme, from 2015-16 onwards.

⁵⁷ See

<http://www.ons.gov.uk/methodology/classificationsandstandards/otherclassifications/thenationalstatistics socioeconomicclassificationnssecrebasedonsoc2010>.

⁵⁸ See www.hesa.ac.uk/collection/c19051/a/sec.

⁵⁹ See 'Annex F: Socioeconomic background' at www.officeforstudents.org.uk/publications/differences-in-student-outcomes-further-characteristics/.

⁶⁰ Defined by the HESA field 'UCASAPPID'. See <https://www.hesa.ac.uk/collection/c19051/a/ucasappid>.

Tracking underrepresentation by area (TUNDRA)

62. TUNDRA (tracking underrepresentation by area) is an area-based measure that uses tracking of state-funded mainstream school pupils in England to calculate young participation.⁶¹
63. TUNDRA classifies local areas across England into five equal groups – or quintiles – based on the proportion of 16-year-old state-funded mainstream school pupils who participate in higher education aged 18 or 19 years. Quintile 1 shows the lowest rate of participation. Quintile 5 shows the highest rate of participation.
64. TUNDRA is a different measure to POLAR4 because it focuses on the participation rate of state-funded mainstream school pupils and only applies to England.
65. In this release, we use Lower Super Output Areas (LSOAs) as the definition of a local area for TUNDRA quintiles.
66. Although TUNDRA is calculated using data from English state-funded mainstream schools only, TUNDRA quintiles exist for all local areas in England, which means they can be applied to any pupil with a known English postcode, even if they did not attend a state-funded mainstream school. For this reason, in this release, we apply TUNDRA quintiles to all young English undergraduates (from 2010-11 onwards).

Course length

67. This field indicates the expected length of the course on which a student is enrolled. While the actual length of course may be derived for some entrants from earlier years, this is not possible in more recent years where students are yet to complete their course. Therefore, for consistency, it was decided to use the expected length of course throughout. See IPCRSELGTH.

Home region

68. Home region is defined by the mapping of the home postcode of the student prior to their course (see IPPOSTCODE) to the relevant Nomenclature of Territorial Units for Statistics (NUTS) 1 region in the UK.⁶² For non-UK domiciled students, home region is split into two categories: 'European Union (EU)' and 'Neither EU nor UK'.

Entry qualifications

69. Entry qualifications are sourced from the ILR and HESA student and HESA AP records. For students with no previous higher education qualifications on entry to their course, entry qualifications are grouped in the same way as our 'Analysis of degree classifications over time',⁶³ while other students are grouped according to the field IPENTQUALGRP.

⁶¹ See www.officeforstudents.org.uk/data-and-analysis/young-participation-by-area/about-tundra/.

⁶² See <https://www.ons.gov.uk/methodology/geography/ukgeographies/eurostat>.

⁶³ See Table D1 in Annex D at www.officeforstudents.org.uk/publications/analysis-of-degree-classifications-over-time-changes-in-graduate-attainment-from-2010-11-to-2018-19/.

Awarding body

70. This field indicates whether the awarding body (see IPAWARDBOD) of the qualification is the same as the provider where the student is registered, or is either Edexcel or SQA, or whether validation arrangements are in place.

Degree classification

71. Degree classification (see 'IPDODEGCLASS') is only available for qualifiers on first degree courses or undergraduate courses with a postgraduate component, which means it can only be used for analysis of progression rates. Students who fall outside of this population have this field recorded as their level of study instead. For those within scope, this factor is split into four outcomes: 'First class honours', 'Upper second class honours', 'Other classifications of honours' and 'Unclassified'. Unclassified degree awards do not indicate failure to qualify, rather this indicates courses where degree classifications are not applicable, for example in medicine and dentistry.

Local and distance learning

72. This field categorises students into four groups: local learners, distance learners, neither local nor distance learners and unknown. It is defined by comparing home 'travel to work' areas with study 'travel to work' areas, which are calculated from home postcodes and study postcodes, respectively. See IPSTUDYLOCTYPE.

Foundation year

73. This field (IPFOUNDEAR) identifies those studying on an integrated foundation year. This is where completion of this first year of study gives the student access to a full degree programme in the following year.

Sandwich courses

74. This field identifies sandwich courses from information on the mode of study, on the HESA student record,⁶⁴ HESA AP student record,⁶⁵ and the ILR.⁶⁶ For HESA records, we identify courses as 'Sandwich' when any year of the course, even if this is not the current year, is expected to include a sandwich component. It should be noted, however, that some courses may be reported as including a sandwich component which may then not materialise (if the student fails to secure a placement, for example), or conversely, some courses reported as not including a sandwich component may then ultimately include one.

Graduate Outcomes quintiles

75. This field contains the quintile of the graduate's travel to work area, using information from the Graduate Outcomes survey. It is defined by IPGOTTWAQUINTILE. Quintile 1 indicates that the graduate lives in an area with the lowest rates of positive outcomes, whereas quintile 5

⁶⁴ See <https://www.hesa.ac.uk/collection/c19051/derived/xmode01>.

⁶⁵ See <https://www.hesa.ac.uk/collection/c19054/derived/xmode02>.

⁶⁶ See 'MODESTUD' at <https://www.gov.uk/government/publications/ilr-specification-validation-rules-and-appendices-2018-to-2019>.

indicates that the graduate lives in an area with the highest rates of positive outcomes. Further information on the methodology can be found in the OfS report 'A geography of employment and earnings'.⁶⁷

Region of study

76. Region of study is defined by the mapping of the study postcode of the student if they are not distance learning (see IPLOCPOSTCODE), or the home postcode prior to their course if they are distance learning (see IPPOSTCODE), to the relevant Nomenclature of Territorial Units for Statistics (NUTS) 1 region in the UK.⁶⁸

Provider

77. This field indicates the registering provider of the student (see IPUKPRNRC).

We are actively seeking feedback regarding these experimental statistics. If you have any queries or suggestions, please contact Stanley Rudkin at official.statistics@officeforstudents.org.uk.

⁶⁷ Available at www.officeforstudents.org.uk/publications/a-geography-of-employment-and-earnings/.

⁶⁸ See <https://www.ons.gov.uk/methodology/geography/ukgeographies/eurostat>.

Annex C: List of abbreviations

| | |
|----------------|--|
| ABCS | Associations Between Characteristics of Students |
| DfE | Department for Education |
| ESFA | Education and Skills Funding Agency |
| FSM | Free School Meals |
| HESA | Higher Education Statistics Agency |
| HESA AP | Higher Education Statistics Agency Alternative Provider record |
| HESA SA | Higher Education Statistics Agency Student Alternative record |
| HRI | Household Residual Income |
| IDACI | Income Deprivation Affecting Children Index |
| ILR | Individualised Learner Record |
| IMD | Index of Multiple Deprivation |
| LSOA | Lower Layer Super Output Area |
| MHCLG | Ministry of Housing, Communities and Local Government |
| MSOA | Middle Layer Super Output Area |
| NPD | National Pupil Database |
| NS-SEC | National Statistics socioeconomic classification |
| NUTS | Nomenclature of Territorial Units for Statistics |
| OfS | Office for Students |
| POLAR | Participation of Local Areas |
| SLC | Student Loans Company |
| TEF | Teaching Excellence and Student Outcomes Framework |
| TUNDRA | Tracking Underrepresentation by Area |
| UCAS | Universities and Colleges Admissions Service |
| UG | Undergraduate |



© The Office for Students copyright 2022

This publication is available under the Open Government Licence 3.0 except where it indicates that the copyright for images or text is owned elsewhere.

www.nationalarchives.gov.uk/doc/open-government-licence/version/3/