Update to associations between characteristics of students (ABCS)

How do outcomes differ when accounting for multiple student characteristics?

Reference OfS 2020.54

Enquiries to official.statistics@officeforstudents.org.uk

Publication date 26 November 2020

Contents

Summary

1. Association between characteristics of students (ABCS) is a set of analyses that seeks to better understand how outcomes vary for groups of students holding different sets of characteristics. We define groups of students by looking at a set of characteristics so that we can determine the effect of not just one characteristic on an outcome, but the effect of multiple characteristics. This report is an update to the 2019 publication Association between characteristics of students.

2. This update includes some changes to the methodology as we continue to develop the ABCS measure. We have incorporated additional characteristics in the ABCS continuation measure and developed a new ABCS continuation measure for student groups who are studying part-time: ABCS PT continuation.

3. ABCS continues to be an experimental official statistic. As such, we are looking for feedback regarding any improvements that could be made to the methodology or the presentation of the ABCS measure. We are also keen to understand how you might use these measures. Please get in touch with us to let us know your thoughts and feedback

Introduction

4. Associations between characteristics of students (ABCS) is a set of analyses which aim to improve our understanding of the outcome different groups of young people are likely to experience across the student lifecycle. At the moment, it focuses on access to higher education for young people and continuation in higher education for students of all ages.

5. We have used statistical modelling[1] to calculate modelled access and continuation rates for different student groups. These student groups are defined by a combination of all the characteristics included in the model. For this update, we have then used these modelled access and continuation rates to split the various student groups into quintiles.

6. Those student groups with the lowest modelled rates will be in the lowest access or continuation quintiles and those with the highest will be in the highest access and continuation quintiles. As well as looking at the quintile for each complete student group (defined by combining all of the characteristics for each measure), the interactive dashboards[2] allow users to examine how a particular characteristic or combination of characteristics is distributed across the quintiles. For example, it is possible to see the proportion of those in student groups that include only females in each of the quintiles, or who are both Chinese and from a POLAR4[3] quintile 1 area. In this way, we clearly see how those who share a single, or even multiple, characteristics might not have the same experience when it comes to access to, or continuation in, higher education.

7. Since the ABCS quintiles are defined using modelled access and continuation rates, they group students by their likelihood of accessing or continuing in higher education based on the factors that are included in the statistical models (see the individual Access, continuation for full-time students and continuation for part-time student sections for details of the characteristics included). In selecting the factors for use in these models we are looking for characteristics that should not influence the outcome in question, but where there is evidence that the outcomes for groups within these characteristics differ. For example, there is no reason why a student's ethnicity should have an impact on the likelihood of them continuing into the second year of their course. However, our analysis of continuation rates shows that black students have lower continuation rates than students from any other ethnic background[4]. Conversely, we know that the level 3 qualifications obtained by young people will have an impact on the likelihood of a young person entering higher education (for example, a student with three A-levels at A* is more likely to go into higher education than a student with three Es). However, prior attainment will not be included in the model because this is a justifiable – or valid – relationship.

---

[1] These models are binary logistic regression models with two-level interactions. The interactions to include have been determined using stepwise regression. See Annex A for details.

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

[3] For details about POLAR4 visit: www.officeforstudents.org.uk/data-and-analysis/young-participation-byarea/

[4] See the continuation tab of the Access and Participation dashboard here: https://www.officeforstudents.org.uk/data-and-analysis/access-and-participation-data-dashboard/

8. These measures were published as an experimental statistic in September 2019. Since then, we have been developing them further, looking at additional characteristics that could be included, and producing an additional continuation measure for students who are studying part-time. ABCS now includes three measures:

    a. ABCS access – this measure looks at the access rates of young people in England.

    b. ABCS FT continuation – this measure looks at the continuation rates for full-time students.

    c. ABCS PT continuation – this measure looks at the continuation rates for part-time students.

9. In this report we describe these three ABCS measure and give examples of the findings for each of them. However, we encourage users to examine our interactive dashboard to explore the different combinations of characteristics for these three measures. We also set out the changes that we have made to the original methodology, including an update to the statistical modelling approach and an explanation of the change from access and continuation groups to access and continuation quintiles. We detail the four new characteristics that have been included in ABCS FT continuation: care experience, free school meal eligibility, socio-economic classification (NS-SEC) and parental higher education.

ABCS Access

10. For ABCS access, 'access' is defined as participation in higher education at age 18 or 19. The measure uses data from DfE's National Pupil Database (NPD) from the year in which pupils obtain their key stage four (KS4) qualifications – most commonly GCSEs. We then look for these individuals in higher education records two or three years later – those with a matching record have accessed higher education. This can be participation at any level of higher education and any mode of study. The only restriction placed on the higher education instance is that it must be two or three years after the completion of KS4.

11. Aside from the differences described in the 'changes to methodology' section, the other change that we have made to the ABCS access measure is that the five-year time series has been rolled forward a year. Therefore, the data included now is for pupils who received their KS4 qualifications from summer 2011 to the summer of 2015 and entrants to higher education from 2014-15 to 2018-19.

12. The access measure looks at six characteristics: ethnicity, FSM eligibility, IDACI, IMD, POLAR4 and sex. As in the original ABCS access model, we have used these characteristics in a statistical model to calculate predicted access rates for every possible combination of categories within these six variables. Details of the model, the interactions included, and the model results can be found in Annex C.

13. We have then used the predicted access rates to split the students included in the modelling in to five quintiles. We have calculated the quintiles in such a way that no student groups will be split across quintiles, and if there are student groups with the same predicted access rates, these will also not be split across quintiles. Table 1 shows the number and proportions of students in each quintile, as well as the mean, minimum and maximum predicted access rate.

Table 1: description of ABCS access quintiles 2020

Access quintile	Number of students	Proportion of students	Mean modelled access rate	Minimum modelled access rate	Maximum modelled access rate
Quintile 1	553,090	20.0%	16.8%	0.1%	24.6%
Quintile 2	556,030	20.1%	31.0%	24.6%	36.3%
Quintile 3	547,955	19.8%	41.3%	36.3%	45.7%
Quintile 4	555,730	20.1%	51.0%	45.7%	56.8%
Quintile 5	553,845	20.0%	64.8%	56.8%	95.2%

14. These quintiles allow us to explore the groups of young people who are least likely (access quintile 1) and most likely (access quintile 5) to access higher education. Each student group, made up of a particular combination of categories of the six characteristics used in the statistical model, is assigned to a quintile. However, the interactive dashboards[5] allow users to look not only at which quintile is assigned to each student group, but what the distribution is across the quintiles when you select only one characteristic, or two – all the way up to the

[5] See www.officeforstudents.org.uk/data-and-analysis/associations-between-characteristics-of-students/access-to-higher-education/

complete set of six. There are only two groups of students whose access quintile is defined entirely by a single characteristic. The first is those whose ethnicity is Gypsy/Roma or Irish Traveller – 100 per cent of whom are in access quintile 1 (those least likely to access higher education). This means that their ethnicity is having a very strong impact on the likelihood of them accessing higher education, regardless of their other characteristics. In contrast 100 per cent of those whose ethnicity is Chinese are in access quintile 5.

15. For all other student groups, combining more than one characteristic shows how considering multiple characteristics tells us more about the likelihood of a student group accessing higher education. For example, when looking at all student groups where the ethnicity is black-Caribbean, 25 per cent are found in access quintile 5. When you restrict this to black-Caribbean ethnicity and only those who were eligible for FSM, none are found in access quintile 5.This shows that while some black Caribbean groups are highly likely to enter higher education, this is not the case for those who are also eligible from FSM. Similarly, if you look at student groups where the ethnicity is black Caribbean and sex is female, 49 per cent are in access quintile 5; however, for black Caribbean male groups, none are found in access quintile 5. This indicates that there is a difference in the likelihood of accessing higher education for black Caribbean males and females.

16. Another example: looking at student groups where IDACI is restricted to just quintile 1 areas, 45 per cent are found in access quintile 1. If we then look at those student groups where IDACI is restricted to quintile 1 and sex is restricted to female, this decreases to 35 per cent. However, if we instead restrict to males, then this increases to 55 per cent. That is, student groups from IDACI quintile 1 areas are not likely to access higher education, and they are even less likely to do so if they are male.

Continuation for full-time students

17. ABCS FT continuation is a measure that assesses the likelihood of different student groups continuing in their full-time higher education studies. For full-time students, continuation measures the proportion of students who qualify, transfer or continue into a second year of study[6]. In particular, this measure looks at continuation for UK domiciled, full-time undergraduate students studying at English higher education providers.

18. As noted above, for this update we have included four additional characteristics to the ABCS FT continuation measure, meaning there are now 12 characteristics in the measure: age, care experience, disability, ethnicity, FSM eligibility, IDACI, IMD, local or distance learner[7], NS-SEC, parental education, POLAR4 and sex.

19. We have used a statistical model to calculate predicted continuation rates for each of the student groups. We have included all the characteristics as main effects and used an automated approach (stepwise) to determine which of the two-way interactions should be included. Details of the model, the interactions included, and the model results can be found in Annex D.

20. We have then used these predicted continuation rates to split the students into quintiles. The creation of the quintiles ensures that no student groups are split across quintiles, and no student groups with the same predicted continuation rates are split across quintiles. This means that each quintile does not contain exactly the same number of students. Table 2 shows the number and proportion of students in each quintile, as well as the mean, minimum and maximum predicted continuation rate in each quintile.

Table 2: description of ABCS FT continuation quintiles 2020

FT continuation quintile	Number of students	Proportion of students	Mean modelled FT continuation rate	Minimum modelled FT continuation rate	Maximum modelled FT continuation rate
Quintile 1	404,910	20.0%	80.6%	32.7%	86.1%
Quintile 2	404,890	20.0%	88.4%	86.1%	90.4%
Quintile 3	405,000	20.0%	91.8%	90.4%	93.2%
Quintile 4	404,835	20.0%	94.4%	93.2%	95.5%
Quintile 5	404,925	20.0%	96.7%	95.5%	99.9%

21. Similarly to ABCS access, these quintiles allow us to explore the groups of students who are least likely (FT continuation quintile 1) and most likely (FT continuation quintile 5) to continue into the second year of their full-time undergraduate courses. Each student group, made up of a particular combination of the categories within the 12 characteristics used in the statistical

[6] For details of how we calculate full-time continuation rates, see the OfS access and participation data methodology and rebuild instructions available from: https://www.officeforstudents.org.uk/data-and-analysis/institutional-performance-measures/technical-documentation/

[7] See paragraph 59 for definitions of local and distance learners.

model, is assigned to a quintile. However, the interactive dashboard[8] allow users to look at which quintile is assigned to each student group, and the distribution across the quintiles when you select only one characteristic, or two – all the way up to the complete set of 12. Unlike ABCS access, there are no groups of students whose FT continuation quintile is defined entirely by a single characteristic.

22. Combining more than one characteristic shows how considering multiple characteristics tells us more about the likelihood of a student group continuing in higher education. For example, for those student groups where care status is restricted to only those who are care experienced, 54 per cent are in FT continuation quintile 1 (least likely to continue). However, for student groups restricted to those who are both care experienced and ethnicity is restricted to Chinese, 41 per cent are in FT continuation quintile 5 (most likely to continue). This shows that the likelihood of continuing is higher for those who are care experienced and Chinese than the average for all those who are care experienced.

23. Another example: for restricting to student groups where age is 21-25 at the start of their course, 47 per cent are in FT continuation quintile 1. If we then look at student groups where age is restricted to 21-25 and local or distance learner is restricted to those who were neither distance nor local learners (those who are studying at a provider and away from home), this decreases to 35 per cent. This reduces again to only 23 per cent once we restrict the student groups further to include only female students. This shows that the likelihood of 21-25 year olds continuing their studies is higher for those who were neither local or distance learners, and is higher again for 21-25 year olds who are neither local or distance learners and are female.

[8] See www.officeforstudents.org.uk/data-and-analysis/associations-between-characteristics-of-students/continuing-in-full-time-higher-education/

Continuation for part-time students

24. As part of this update, we are introducing a new measure: ABCS PT continuation. Continuation is measured differently for students studying part-time than for those studying full-time: those who have qualified, transferred or continue into the third year of their studies are said to have continued[9].

25. Because of this difference in definition, the most recent available year of data for part-time continuation is for students who started their course in 2016-17, and so the five-year time series we use is from 2012-13 to 2016-17. The measure includes UK-domiciled students on part-time undergraduate courses at higher education providers in England.

26. The starting place for which characteristics to include in this measure was the characteristics used in the original ABCS FT continuation model: age, disability, ethnicity, IDACI, IMD, local or distance learners, POLAR4 and sex. In all cases, we have found that, as for full-time continuation, there is a relationship between all these characteristics and part-time continuation. However, there is a large proportion (88.2 per cent) of students who are 21 or over at the start of their course, meaning there are a substantial number of students whose POLAR4 quintile is not calculated. For this reason, we have chosen to omit POLAR4 from this measure. As discussed in the 'Using additional characteristics in ABCS PT continuation' section (page 13), none of the additional characteristics for ABCS FT continuation can be included in the part-time measure.

27. As for the other measures, we used a statistical model to calculate predicted continuation rates for each of the student groups. We included all the characteristics as main effects and used an automated approach (stepwise) to determine which of the two-way interactions should be included. Details of the model, the interactions included, and the model results can be found in Annex E.

28. We have then used these predicted continuation rates to split the students into quintiles. The creation of the quintiles ensures that no student groups are split across quintiles, and no student groups with the same predicted continuation rates are split across quintiles. This means that each quintile does not contain exactly 20 per cent of the population. Table 3 shows the number and proportion of students in each quintile, as well as the mean, minimum and maximum predicted continuation rate in each quintile.

[9] For details of how we calculate part-time continuation rates, see the OfS access and participation data methodology and rebuild instructions available from: www.officeforstudents.org.uk/data-and-analysis/institutional-performance-measures/technical-documentation/

Table 3: description of ABCS PT continuation quintiles 2020

PT continuation quintile	Number of students	Proportion of students	Mean modelled PT continuation rate	Minimum modelled PT continuation rate	Maximum modelled PT continuation rate
Quintile 1	92,190	20.0%	50.8%	0.4%	54.7%
Quintile 2	92,745	20.2%	58.1%	54.7%	60.5%
Quintile 3	90,900	19.8%	63.3%	60.5%	67.3%
Quintile 4	91,815	20.0%	69.2%	67.3%	72.3%
Quintile 5	92,420	20.1%	78.1%	72.3%	99.8%

29. As for the other two ABCS measures, these quintiles allow us to explore the groups of students who are least likely (PT continuation quintile 1) and most likely (PT continuation quintile 5) to continue into the third year of their part-time undergraduate courses. Each student group, made up of a particular combination of the seven characteristics used in the statistical model, is assigned to a quintile. However, the interactive dashboards[10] allow users to look at which quintile is assigned to each student group, and the distribution across the quintiles when you select only one characteristic, or two – all the way up to the complete set of seven. Like ABCS FT continuation, there are no groups of students whose PT continuation quintile is defined entirely by a single characteristic.

30. Combining more than one characteristic shows how considering multiple characteristics tells us more about the likelihood of a student group continuing in higher education. For example, only 18 per cent of females are found in PT continuation quintile 1, but when we look at student groups where the sex is restricted to female and learner type is restricted to distance learners, this increases to 37 per cent – indicating that being a distance learner has a negative impact on the likelihood of female part-time learners continuing their studies.

31. Another example: 30 per cent of those from IMD quintile 1 areas are in PT continuation quintile 1, while 13 per cent are in PT continuation quintile 5. When we look at only males from IMD quintile 1 areas, the proportion in PT continuation quintile 1 increases to 44 per cent, but the proportion in PT continuation quintile 5 also increases to 24 per cent. This highlights the complexity of understanding how different student groups' continuation rates vary for some in IMD quintile 1 depending on other characteristics. Being male increases their risk of not continuing for some student groups who are also from IMD quintile 1 areas, but for other males from IMD quintile 1 areas, it decreases the risk. We can add other characteristics to explore which ones are having an impact on likelihood of continuing. For example, if we further narrow down the group to only those from IMD quintile 1 areas who are male and are neither local or distance learners, only 1 per cent are in PT continuation quintile 1.

[10] See www.officeforstudents.org.uk/data-and-analysis/associations-between-characteristics-of-students/continuing-in-part-time-higher-education/

Changes to the methodology

Additional characteristics

32. In June 2020, the OfS published a report titled Differences in student outcomes: further characteristics,[11] which explored how outcomes differ for characteristics that had not previously been explored by the OfS. Following on from this work, we have considered the possibility of including some of these measure in the ABCS modelling. A number of these variables are not available on the National Pupil Database (NPD) – the base dataset for ABCS access – and so we focused on ABCS continuation. The report showed that for continuation of full-time students, there was sufficient data, and evidence of difference in continuation, for the following: free school meal (FSM) eligibility, parental higher education, national statistic – socio-economic classification (NS-SEC) and care experience.

33. FSM eligibility[12] indicates whether the student was ever recorded as being eligible to receive free school meals in the six years prior to the March census date in their final year of key stage four (year 11). Students who have been eligible for FSM have been shown to have consistently lower continuation rates than those who were not, making this a characteristic of interest for ABCS continuation[13].

34. Parental higher education[14] is collected through the student's response to the question 'Do any of your parents[15] have any higher education qualifications, such as a degree, diploma or certificate of higher education?' Continuation rates are consistently higher for students for whom at least one parent holds a higher education qualification than for students whose parents do not hold a higher education qualification[16].

35. NS-SEC[17] classifies the socio-economic background of students. For those students who are under 21 at the start of their course NS-SEC is based on the occupation of their highest earning parent. For those students aged 21 or over at the start of their course NS-SEC is

[11] See www.officeforstudents.org.uk/publications/differences-in-student-outcomes-further-characteristics/

[12] FSM eligibility data comes from the Department for Education's National Pupil Database. The DfE does not accept responsibility for any inferences or conclusions derived from the NPD data by third parties.

[13] Annex C: free school meal eligibility. Available from: www.officeforstudents.org.uk/publications/differences-in-student-outcomes-further-characteristics/. In the differences in student outcomes report, FSM eligibility has been restricted to only those who were under 21 at the start of their course. This restriction is loosened for this report, and so we have rerun the analysis to ensure that we find the same relationship between FSM eligibility and full-time continuation, which we do.

[14] Data for this characteristic comes from the HESA variable pared. See www.hesa.ac.uk/collection/c19051/a/pared

[15] HESA defines parents as 'This includes natural parents, adoptive parents, step-parents or guardians who have brought you up.'

[16] Annex D: parental higher education. available from: www.officeforstudents.org.uk/publications/differences-in-student-outcomes-further-characteristics/

[17] Data or this characteristic comes from the HESA variable SEC. See www.hesa.ac.uk/collection/c19051/a/sec.

based on the occupation of the student themselves prior to higher education. Occupations are coded using the Standard Occupation Classification (SOC) and then grouped further into these categories[18]:

a. Higher managerial, administrative and professional occupations

b. Intermediate occupations

c. Routine and manual occupations

d. Never worked and long-term unemployed.

36. The Difference between outcomes: further characteristics report restricted the population for NS-SEC to under 21s only, which we have chosen not to do for ABCS in favour of keeping as much of the available data as possible. Therefore, we have looked at the relationship between NS-SEC and full-time continuation rate for all ages. We find the same relationship as for the under 21s[19]: the continuation rates differ between the groups, with 'higher managerial, administrative and professional' having the highest continuation rates and 'Never worked and long-term unemployed' having the lowest. These consistent differences across the groups indicate that this is another useful characteristic to include in ABCS continuation.

37. Care experience describes whether a student has ever spent time in the care of a local authority in England or Wales, or Health and Social Care Trust in Northern Ireland, or who has self-declared as in care for three months or more[20]. Analysis has shown that those who are care experienced are at a higher risk of not continuing on their course than those who are not care experienced[21].

Data quality of the additional characteristics

38. Although for all four of these characteristics we find good evidence for their use in the ABCS FT continuation measure, they are not without their problems. Data about FSM eligibility can only be used for pupils sitting their KS4 qualifications in the summer of 2010 onwards[22] for whom we have the appropriate FSM field on the NPD, and it is not used to identify students who previously studied at independent schools. In practice, this means that we have FSM eligibility in each year of data we use for those who had not attended independent schools and had started higher education in these academic years at these ages:

[18] See Section 7 of the ONS' webpage titled The National Statistics Socio-economic classification (NS-SEC) available from: www.ons.gov.uk/methodology/classificationsandstandards/otherclassifications/thenationalstatisticssocioeconomicclassificationnssecrebasedonsoc2010

[19] See Annex F available from www.officeforstudents.org.uk/publications/differences-in-student-outcomes-further-characteristics/

[20] The definition and reporting of care experience is complex. For details of the data used, see https://www.hesa.ac.uk/collection/c19051/a/careleaver

[21] Annex B: care experience. Available from https://www.officeforstudents.org.uk/publications/differences-in-student-outcomes-further-characteristics/

[22] See footnote 13

a. 2013-14: aged 18

b. 2014-15 aged 18 and 19

c. 2015-16: aged 18, 19 and 20

d. 2016-17: aged 18, 19, 20 and 21

e. 2017-18: aged 18, 19, 20, 21 and 22.

39. Because of this absence of data in some years, some other analyses (such as the difference in outcomes: further characteristics report) chooses to restrict use of FSM eligibility to only young students (those aged under 21 at the start of their course) and to only use the data from 2015-16 onwards. However, for ABCS we have chosen to maximise the amount of data used in the model, and so we use FSM eligibility for all five years of our time series, covering the population described above. This results in FSM eligibility being unknown for 36.6 per cent of students in our population in 2013-14, reducing to 26.3 per cent for 2017-18.

40. The application of our Data Quality Framework[23] as part of the Differences in student outcomes: further characteristics work has shown that the data for care experience is only of a sufficient quality for use from 2014-15 onwards and NS-SEC for 2015-16 onwards. Therefore, for these variables, there will be a high proportion of unknowns, since the data is missing for some years of our collections.

41. In addition to the challenge of data availability for all years, there is also disparity between the record types. ABCS uses four main sources of data: the NPD, the Individualised Learner Record (ILR) and Higher Education Statistics Agency (HESA) student and alternative provider (AP) student records. In paragraph 38 we discussed the impact of taking the FSM field from the NPD record on the availability of data. For parental higher education, NS-SEC and care experience variables we have a similar problem with data availability in that details of these characteristics are not collected on the ILR. Therefore, students whose data is returned to the ILR will not have any information for these characteristics.

42. Although this means that the data is incomplete for all the additional variables, we can still use the data we do have in the full-time continuation model. To avoid losing information for individuals whose data is missing for these additional characteristics by removing them from the model, we include ‘unknown’ categories for these variables in the model. The inclusion of these variables does not cause any problems in running the model and they are found to contribute to explaining the variation in full-time continuation rates.

Using additional characteristics in ABCS PT continuation

43. Having found evidence supporting the use of these characteristics in ABCS FT continuation, we have also explored their possible use in ABCS PT continuation. However, because the part-time continuation measure looks at continuation across two years, rather than one, the years of data available to use in the analysis lag a year behind those for full-time continuation. Where data for a characteristic is only available for the two or three most recent years of data, this

[23] Annex A: Data quality framework – a method for assessing the quality of student characteristic data. Available from: www.officeforstudents.org.uk/publications/differences-in-student-outcomes-further-characteristics/

means that we do not find sufficient data to include these characteristics in the part-time measure. This is the case for NS-SEC and care leaver.

44. For FSM eligibility, we found insufficient data because of the age profile. As discussed above, FSM eligibility is only available for younger students. Because part-time study typically attracts older students, there is a much higher proportion of missing data for FSM eligibility for part-time students than for full-time (30.3 per cent unknown for all full-time data compared with 89.6 per cent for part-time).

45. While parental education is available for the whole time series, there is still significant amounts of missing data for part-time continuation. For part-time continuation 60.2 per cent of records have missing or unknown parental education information, compared with 24.5 per cent for full-time continuation. This is partly due to a higher proportion of part-time students being recorded on the ILR than for full-time (10.9 per cent compared with 7 per cent), but also due to a much higher proportion of part-time students responding 'don't know' or not providing a response compared with full-time students (31 per cent of part-time students recorded on the HESA record for 11.5 per cent for full-time).

46. Given all the above, none of the four new characteristics are included in the new ABCS PT continuation measure.

Statistical modelling

47. The introduction of the new characteristics to the ABCS FT continuation measure causes a substantial increase in the computational time taken to model the data. As a result, we have reconsidered the modelling approach we take in selecting and including interaction terms. The original ABCS methodology used a more complex method than is usual for including interactions, where some levels of interactions between two categorical variables could be included while others were excluded. The more standard approach is to include either all levels of interactions between two categorical variables, or none of them. We have looked at this more standard approach, and a hybrid of the two approaches, to assess whether the choices made in how we select the interactions impacts on the model results. In doing so, we have concluded that there is nothing gained by using the more complex method, or the hybrid method, and so have changed our approach to the standard approach. Further details of this follow and full details of these assessment can be found in Annex A

48. Details of the original methodology can be found in the first ABCS report[24]. In summary, we used a binary logistic regression model with all main effects kept in the model and used stepwise modelling to select which two-way interactions to include. We created dummy variables for every possible combination of categories within every characteristic (for example, there was a dummy variable for male and distance learner, and for white and POLAR4 quintile 1). This meant that interactions could be present in the model for some categories within a characteristic, but not necessarily all of them. It was entering this vast number of variables, representing all possible combinations of categories within characteristics, that caused the model process to become inefficient.

[24] See www.officeforstudents.org.uk/publications/associations-between-characteristics-of-students/

49. The model process we opted to use includes interactions between whole variables only – so, for example, the model includes the interaction between sex and local or distance learner, rather than the interactions between female and local, female and distance, female and neither, male and local, male and distance and male and neither. This amounts to the same interactions going into the model, but it prevents the stepwise method from removing some of the category-by-category interactions and not others – either all possible combinations of sex and local or distance learner are included, or none of them are. As a result, the final model will contain interactions between categories which are not statistically significant, but the overall interaction between the two characteristics will have been found to be statistically significant.

50. Because we want to maintain consistency across the ABCS measures, we also applied this modelling approach to the ABCS access and ABCS PT continuation measures. As well as testing the impact on the modelled full-time continuation rates, we have run similar comparisons for part-time continuation and access rates and concluded that using the new methodology does not cause substantial changes in the modelled rates when compared with the original methodology.

Grouping method

51. For the ABCS measures published in 2019, we used a methodology that aimed to create access and continuation groups that had genuine differences in their predicted outcome rates. This way, we could be certain that those in the first group really were those most at risk of poor outcomes, and that those in the second group really were less at risk than those in the first, and not that they just happened to be in group two because they were on a boundary.

52. However, this grouping method does not always give us the anticipated outcome of access and continuation groups with truly different rates. This is particularly the case for full-time continuation, where most student groups have very high continuation rates, and there is no obvious place to split these groups. This issue has been further exacerbated by the introduction of more characteristics to the ABCS FT continuation measure. It has also resulted in the grouping methodology becoming unstable – with very small changes in the predicted continuation rates from the statistical model leading to quite big changes in the make-up of the continuation groups. For example, the original method resulted in 13 per cent of students being in continuation group 2, whereas the new ABCS method resulted in there being 27 per cent in continuation group 2. Since the comparison of the models has shown that there are no big differences in predicted continuation rates for large student groups across the models, and the method that determines student groups does not use small student groups, we can be assured that these differences are caused by instability in the grouping methodology.

53. Because of this, we have chosen to use the more familiar approach of using quintiles[25]. These quintiles are created in such a way that students with the same predicted rates cannot be split

[25] In the case of ABCS access, the choice to publish quintiles, and the change to the new modelling methodology, means that ABCS access is now more similar in method to UCAS' MEM. However, the two measures continue to measure different outcomes: MEM looks at the proportion of students who are reported as confirmed and placed as a full-time undergraduate, while ABCS access look at the proportion of students who are reported as attending higher education at any level and studying as any mode. Additionally, not all providers are covered by UCAS, and so we will be looking at more providers than MEM does. OfS continues to be in communication with UCAS about the development of these measures.

across quintiles (which means the quintiles do not always contain exactly 20 per cent of the population, but it is always very close to that). This approach still enables us to identify those students most at risk from poor outcomes, but this stability will make the measures more useful as we continue to develop them further and include new years of data. Quintiles have been chosen to align with the presentation of other measures, such as POLAR (participation of local areas). However, we have the flexibility to create other groupings, such as deciles, as and when we discover uses for ABCS that require slightly different groupings.

54. When using the ABCS grouping methodology, not all student groups were assigned to the same ABCS group when using results from the three different modelling processes. Changing to using quintiles rather than ABCS groups has not completely resolved this issue. However, there are almost half as many student groups whose quintile changed between models compared with the number who changed continuation groups, and there were no student groups who moved more than one quintile between the three models.

55. This does not mean that we will not revisit the ABCS grouping methodology. As we continue to develop the ABCS measures, we will use the grouping methodology where we can identify groups in the data where predicted rates are sufficiently different to identify clear ABCS groups.

Definitional changes

56. Alongside these substantive methodological changes, we have made some minor changes in the definitions of the characteristics being used.

57. First, the way we treat unknown data for POLAR4, income deprivation affecting children index (IDACI) and index of multiple deprivation (IMD)[26] has changed. Where data is not available, such as non-English postcodes for IMD, we continue to group these students as not applicable. However, where data is missing without reason, we assign a quintile to these individuals. This is because we did not want to lose all the information about these individuals by removing them from the model, but the number of unknowns was too small for them to be included in the model as a group of their own. Previously, we categorised these individuals in the middle quintile of each of the geographic measures (IDACI, IMD and POLAR4). For these updated measures, we have examined the access and continuation rates of the unknown groups and put them in the quintiles with the most similar access or continuation rate.

58. Second, for ABCS continuation, we have updated the ethnicity grouping to better reflect the way in which the data is collected. This has resulted in us collapsing the ‘white – Irish’, ‘white – other’ and ‘white – English, Welsh, Scottish, Northern Irish, British’ into a single ‘white’ group. Most data collected by HESA only allows this single ‘white’ option[27], and so this is a better representation of the data. The ethnicity groupings for ABCS access will reflect the way in which the data is collected on the NPD[28], with the exception of the Gypsy/Traveller and Irish – Traveller groups, which have been collapsed into a single group to give a large enough group

[26] See www.gov.uk/government/statistics/english-indices-of-deprivation-2015

[27] See www.hesa.ac.uk/collection/c19051/a/ethnic

[28] See find-npd-data.education.gov.uk/en/concepts/7fce3af0-c88d-467c-a5cc-f31565f7c75a

to use in our modelling. As a result, the ethnicity groupings will differ between ABCS access and ABCS continuation.

59. The final change is in the local or distance learner characteristic. In the original ABCS continuation measure, this was split into ‘local or distance learner’ or neither. For this update, we have been able to split the local and distance learners into two separate groups. Local students are those whose home address is in the same travel to work area as their provider. Distance learners as those who are not in attendance at the provider for most of their course. That is, they are studying at a distance from their provider.

60. Definitions of all the variables used in the three ABCS measure can be found in Annex B.

We are keen to receive any feedback regarding:

- **how the ABCS continuation measure might be used**
- **the methodology for the statistical modelling**
- **the methodology for creating the continuation groups.**

Please email Annalise Ruck at official.statistics@officeforstudents.org.uk.