NUFFIELD RESEARCH PLACEMENTS STUDY: COMPOSITION REPORT

Gillian Paull & Xiaowei Xu

April 2017

CONTENTS

1 INTRODUCTION

There is a severe shortage of skills in STEM (science, technology, engineering and mathematics) subjects in the UK. Over a third of employers report difficulties recruiting STEM skilled staff and the shortfall is expected to increase over the next decade.[1] One in four employers also report that applicants for STEM roles do not have adequate practical experience or laboratory skills for work.[2]

Students from disadvantaged backgrounds are particularly underrepresented in STEM. Students from lower socio-economic classes constitute only 27-28% of all entrants into mathematical and physical science subjects at university, compared to 35-37% of entrants into medicine, law and business administration.[3] In addition to reducing diversity and economic potential, low STEM participation among disadvantaged groups limits individual opportunities for social mobility: on average, employees in STEM occupations earn 20% more than those in other fields.[4]

Against this backdrop, the Nuffield Foundation provides talented post-16 students the opportunity to carry out STEM research placements through the Nuffield Research Placements (NRPs) programme. The programme aims to deepen students' understanding of STEM subjects and research and to encourage them to pursue further studies and careers in STEM. The programme is administered at a regional level by a network of 16 coordinators across the UK. All students who meet specific academic eligibility criteria are able to apply, although the Nuffield Foundation has focused on reaching students from more disadvantaged backgrounds in the last three years.

This report is part of the evaluation of the NRP programme being undertaken by Frontier Economies and CFE Research. The evaluation aims to provide robust evidence on the impact of the placements and to identify ways in which the programme might be improved. This report is the first output from the quantitative element, presenting a profile of the types of students who undertake placements and an analysis of the characteristics of pupils most likely to apply to the programme, with particular consideration given to whether those from disadvantaged backgrounds are more likely to take up the opportunity than other type of pupils. This analysis involved the linking of the application data from the programme to large-scale administrative data in order to obtain additional information on placement students and to allow the modelling of the selection process of placement students from the population of AS level pupils in England.[5]

Later work in the study will assess the impacts that participation has on choices about post-compulsory study and career paths in STEM using the linked large scale administrative data to track placement students through to their initial

---

[1] CBI (2015), *Inspiring growth: CBI/ Pearson Education and Skills Survey 2015*

[2] Ibid.

[3] Higher Education Statistics Agency (2015), *Statistical First Release 224*, https://www.hesa,ac,uk/free-statistics

[4] Institute of Education (2011), *The labour market value of STEM qualifications and occupations*

[5] It should be noted that the analysis could only be undertaken for NRP students in England as administrative data could not be readily matched for the devolved nations (Wales, Scotland and Northern Ireland) where placements are also offered.

career destinations at age 21. A qualitative element of the study, undertaken by CFE Research, will seek to understand how well the programme works for placement providers and participating schools and students and to provide deeper understanding on how the placements influence student choices. The study will complete in 2022.

The remainder of this report is structured in the following way:

- Chapter 2 describes the data sources and how the programme data was matched with the administrative data to identify applicants in the administrative data.

- Chapter 3 presents a description of the composition of NRP students and the types of placements undertaken.

- Chapter 4 analyses the selection of NRP students through application and selection for a placement.

- Chapter 5 further explores the targeting of pupils from disadvantaged backgrounds, examining regional variation in targeting and the targeting of pupils within schools.

- Chapter 6 concludes.

2 DATA SOURCES

This chapter describes the two data sources used to examine the composition and selection of pupils for the Nuffield Research Placements. The first section describes the data collected in the application process by the Nuffield Foundation, while the following section describes the data used from the National Pupil Database (NPD). The final section in the chapter describes the matching process between the two sources.

2.1 NRP application data

This report uses data for applications made to the NRP programme in three cohorts in years 2014, 2015 and 2016 provided by the Nuffield Foundation. This data contained 9,635 records but 20 duplicates were dropped, leaving an initial 9,615 applications across the three years[6]. Table 1 presents the numbers of applicants and number of successful applicants across the three cohorts. Just over one third (36 percent) of applicants were successful in being awarded a placement, with a slightly rising rate of success over the years (from 34 percent to 38 percent).

Table 1   **NRP applicants in the UK**

	All applicants	Successful applicants	Percentage successful
2014	3,185	1,081	34%
2015	3,264	1,150	35%
2016	3,166	1,193	38%
All cohorts	9,615	3,424	36%

Table 2   **NRP applicants in England**

	All applicants	Successful applicants	Percentage successful
2014	2,504	809	32%
2015	2,701	886	33%
2016	2,597	966	37%
All cohorts	7,802	2,661	34%

The NRP programme covers all nations in the UK, but, as explained in the following section, suitable matching administrative data was only available for pupils in England. Over the three cohorts, 81 percent of all applications were in

---

[6] These duplicates were identified on the basis of the applicant's first name, last name and date of birth. One record was kept from each pair, prioritising successful applications over unsuccessful ones (10 cases) the record which placed the applicant in the correct year for application according to the date of birth (5 cases), while the remaining 5 cases were dropped from identical records in the same application year. Overall, 7 records were dropped from the 2014 cohort, 11 from 2015 and 2 from 2016.

England, with 83 percent of unsuccessful candidates in England and 78 percent of successful candidates in England. Table 2 presents the numbers of applicants and successful applicants in England across the three cohorts. The success rate is slightly lower for applications in England (34 percent) than for the entire UK, but the pattern of a slightly rising trend over the cohorts can be observed within England as well as for the UK.

The application data contains a rich mixture of information on applicants' gender, age, ethnicity, residency/nationality, bursary, parental education, type of school, enrichment activities, qualifications and courses, and subject preferences. Data on the applicants' first name, last name, date of birth and school name was used to match applicants with administrative data, while data on gender, ethnicity and region was used to complete gaps in the administrative data and to compare the profiles of all applicants with those matched to the administrative data. In addition, data was also available from the Nuffield Foundation for successful applicants on the dates of their placement, the STEM subject involved, the type of organisation providing the placement and any placement expenses. This data was used to analyse the allocation of applicants across placement subjects and organisation types.

2.2 National Pupil Database (NPD)

The National Pupil Database (NPD) collects demographic data and academic attainment for all pupils in maintained schools in England. Equivalent datasets for the devolved nations are available[7], but there are substantial barriers to obtaining linked data and the analysis is primarily restricted to England, although comparative data from applicants across the UK is presented where possible. It should be noted that data for most pupils in independent schools is not available in the NPD, an issue which is considered in the following section.

NPD data was requested for all key stage 4 (KS4) students who would be of the correct age to apply to the NRP programme in the 2014 to 2016 cohorts. This generated a sample of 1,698,247 pupils, of whom 46 percent (785,459) were studying for AS level or the International Baccalaureate (IB) at key stage 5 (KS5).[8] This group of "AS Level pupils" is the pool from which applicants are drawn for the NRP programme and is used as the population base from which selection in to placements is analysed.

The eligibility criteria for NRPs requires students to be studying at least one STEM subject at AS level and to have at least five GCSEs at grade B or above, including maths, English and a science. Applying these criteria to the KS4 exam results and KS5 study subjects in the NPD data identified that 45 percent of AS Level pupils met the NRP eligibility criteria across the three cohorts, generating a sample of 348,511 eligible AS level pupils. Table 3 presents the numbers of AS level pupils and eligible AS level pupils across the three cohorts.

---

[7] The Welsh Pupil Level Annual Schools Census (PLASC) and Pupil Attainment dataset for Wales; Pupils in Scotland Census for Scotland; and the Northern Ireland Schools Census for Northern Ireland.

[8] This proportion was 48 percent for the 2014 cohort, 47 percent for the 2015 cohort and 43 percent for the 2016 cohort.

Table 3   **NPD sample**

	All AS Level pupils	All eligible AS Level pupils	Percentage eligible for NRP
2014	271,717	121,292	45%
2015	266,615	117,470	44%
2016	243,213	109,749	45%
All cohorts	781,545	348,511	45%

The NPD provides pupil level data on gender, ethnicity, eligibility for free school meals (FSM), the postcode-based "Income Deprivation Affecting Children Index" (IDACI) score and local authority (LA). The LA variable was used to match pupils into the regions used in the NRP programme. In addition, the NPD data provides a measure of pupil attainment in the GCSE points scored which is based on the grades in the pupil's best eight GCSEs.

School level data from the NPD was merged with the pupil level data using the key stage 5 unique school identifier (urn). This included measures of average disadvantage among pupils in the school, captured in the percentages for pupils ever having received free school meals (FSM) at key stage 4 and the percentage at key stage 5. Two measures of average school attainment were also added: the proportion of students who attainted five good (A* to C) GCSEs including English and maths at KS4 and the proportion of students attaining at least three A levels at key stage 5.

2.3   Combining data sources

NRP applicants must give permission for their application and placement data to be individually linked to administrative data including the NPD. No permission was requested for the 2014 cohort, permission was requested only from successful applicants in 2015 and permission was requested from all applicants in 2016. For cases where permission has not been given[9], it is possible to obtain NPD data which indicates individuals who were successful or unsuccessful applicants but without any identifier to link back to the individual in the NRP application data.[10] These are referred to as "non-identifiable" matches, while individuals who did give permission for linking are referred as "identifiable" matches. This means that there were potentially 4,319 identifiable matches (all 2014 applicants and 2015 unsuccessful applicants) and 3,483 identifiable matches (2015 successful applicants and all 2016 applicants) with the NRP data.

Matching with the NPD data was undertaken by the team at the Department for Education, using the pupils' first name, last name, date of birth, home postcode and, if needed, the current or previous school name.

---

[9] Permission was refused by 11 unsuccessful and 13 successful applicants in the 2016 cohort which meant that these 24 cases were "non-identifiable" in the 2016 cohort. Although 20 were matched, they could not be used in the analysis due to an error which meant they were all recorded as successful in the exchange of data with DfE.

[10] In practice, this meant that all those without permission had an identifier equal to 1 if they were unsuccessful and to 9999 if they were successful.

Table 4 presents a summary of the number of matches across successful and unsuccessful applicants in the three cohorts. Overall, 83 percent of unsuccessful applicants and 88 percent of successful applicants could be matched to NPD data, creating an NRP sample for the analysis consisting of 4,258 unsuccessful applicants and 2,339 successful applicants. The matching rate is broadly similar across successful and unsuccessful applicants and across cohorts (ranging from 81% to 91%).

Table 4   Matching NRP applicants with the NPD data

Number (% in column)	2014 cohort		2015 cohort		2016 cohort		All cohorts	
	Not succ.	Succ.	Not succ.	Succ.	Not succ.	Succ.	Not succ.	Succ.
Non-identifiable not matched	314 (19%)	66 (9%)	324 (18%)		11 (1%)	13 (1%)	649 (13%)	79 (3%)
Identifiable not matched				132 (15%)	234 (14%)	111 (11%)	234 (5%)	243 (9%)
Non-identifiable matched	1,381 (81%)	743 (91%)	1,491 (82%)				2,872 (56%)	743 (28%)
Identifiable matched				754 (85%)	1,386 (85%)	842 (87%)	1,386 (27%)	1,596 (60%)
Total	1,695	809	1,815	886	1,631	966	5,141	2,661

Notes: 24 cases in 2016 were non identifiable because the students refused permission for their data to be linked to administrative data.

The main reason for the cases where no match was found for NRP applicants in the NPD data is because pupils at independent schools and in Further Education institutions do not have data recorded in the NPD. Indeed, around 800 of the roughly 1,200 "unmatched" cases were matched by the NPD team to pupils with reference numbers issued to Further Education institutions and independent schools. In addition, the NRP application data indicates that 72 percent of the not matched identifiable applicants were at independent schools. Other possible reasons where matches have not been found could include cases where the pupil is not, by age, in the 2014 to 2016 cohorts (although the application data indicated that only 16 of the not matched identifiable cases had dates of birth lying outside of this timeframe), misspelling of names or change in home postcode.

One way to consider whether the matched sample of applicants differs from the original complete sample is to compare the variables which appear in both data sources.[11] Comparing gender, ethnicity and region does not indicate any major

---

[11] Variables which appear in only one data source cannot be usefully compared because of the non-identified matches which means that there would be missing values in one of the samples.

bias in the matching process: with a few minor exceptions[12], the proportions in each category are within a percentage point in the two samples.

It should be noted that matching the NPD data with the application data on an individual (identifiable) basis is only required in this report in order to analyse the placement subject and organisation type across the NPD variables (including disadvantage).[13] This means that the lack of identifiability for some of the matching is not a major issue as it only means that the 2014 cohort cannot be included in the analysis of placement type

2.4 Summary

This chapter has described how the data sample for the analysis was derived by matching application data from the Nuffield Foundation with applicants' records in the National Pupil Database and extracting a large sample of comparable AS level pupils. This served two purposes: it provided additional information on the applicants in terms of measures of disadvantage, educational attainment and school background and also provided two alternative large samples of AS level pupils and AS level pupils meeting the eligibility criteria to apply for the NRP programme to model which types of pupils are most likely to apply for placements. The final sample contains 7,802 applicants for the years 2014 to 2016 and around 800,000 AS level students (of which around 350,000 would be eligible to make an NRP application).

There are two main caveats on this data:

■ Because the NPD only covers England, the sample excludes NRP applications in the devolved nations (around 20 percent of applications) and this report does not capture how application and selection processes may differ in those other nations. However, the ongoing qualitative research for this evaluation does cover these nations and can provide some insight on whether the programme operates very differently outside of England.

■ Most independent schools do not record data with the NPD and it is likely that a substantial proportion of the 15 percent of NRP applicants in England who were not matched in the NPD data attend independent schools. Although the profile of applicants in terms of gender, ethnicity and region was broadly similar between the matched and unmatched samples, it should be noted that the analysis is for pupils primarily in state maintained schools.

---

[12] For ethnicity, the proportion white is 2 percentage points higher and the proportion Asian 2 percentage points lower in the matched sample. For region, the proportion in the London region is 2 percentage points higher in the matched sample and the proportion in the West Midlands region is 2 percentage points lower.

[13] In a small number of cases (30 for ethnicity and 166 for region), data from the NRP application was used to complete missing data in the NPD.

3 CHARACTERISTICS OF NRP STUDENTS

This chapter describes the characteristics of NRP students (those who undertook a placement) in terms of their demographic profile, educational attainment, school background and types of NRP placements they undertake. Some of these statistics were already available from the NRP data sources, but the NPD provides some additional information and patterns are tracked over the three cohorts. In addition, this chapter offers supporting evidence for the subsequent analysis in the choice of variables considered and the combining of the analysis across all three cohorts.

3.1 Demographic characteristics

Table 5 presents the proportions of NRP students who are male and female across the three cohorts. On average, 54 percent of students are female with some variation across the cohorts but no marked trend.

Table 5   Distribution of NRP students by gender

	2014	**2015**	**2016**	**All cohorts**
Male	45%	48%	44%	46%
Female	55%	52%	56%	54%
Number of students	743	754	842	2,339

Source: Matched NPD data

For all 3,424 NRP students, including those in the devolved nations and those in England and not matched to the NPD data, the gender proportions reported in the application data are almost identical: the proportion female was 53 percent, 52 percent and 56 percent for the three cohorts and an average 54 percent across all three cohorts.

A substantial proportion of NRP students are in a non-white ethnic minority group (table 6): across all three cohorts, 58 percent are white, 7 percent black, 26 percent Asian and 8 percent of mixed or other ethnic group.[14] In 2016, there was a slight drop in the proportion who are white and corresponding increase in the Asian proportion, but this was driven by a rise in the absolute numbers of students being accounted for almost entirely by an increase in the number of Asian students (64 of the 78 increase in numbers of students were in the Asian group) rather than a decrease in the number of white students (which declined only by 3 between the two years).

---

[14] It should be noted that the ethnic composition of placement students (and the other characteristics considered in this chapter) is compared to the wider eligible population in the next chapter.

Table 6   **Distribution of NRP students by ethnic group**

	2014	2015	2016	All cohorts
White	60%	61%	54%	58%
Black	8%	7%	8%	7%
Asian	25%	24%	29%	26%
Mixed and other	6%	9%	9%	8%
Number of students	734	754	842	2,330

Source: Matched NPD data

The ethnic mix reported in the NRP application data is slightly different among all NRP students, including those in the devolved nations and those in England not matched to the NPD data: the proportions white, black, Asian and mixed / other are 63 percent, 6 percent, 25 percent and 7 percent across all three cohorts. The slightly higher proportion of students who are white is consistent with the ethnic mix across the four nations in the UK. The pattern of slight change in the ethnic composition across cohorts for all NRP students is similar to that for the matched NPD sample, which would be expected given that the NPD sample constitutes a substantial proportion of all students.

Table 7   **Distribution of NRP students by region**

	2014	2015	2016	All cohorts
North East	3%	4%	4%	4%
North West	11%	8%	9%	10%
Yorkshire region	5%	8%	9%	8%
Merseyside	14%	10%	8%	11%
West Midlands	7%	10%	12%	10%
East Midlands	7%	7%	6%	7%
East Anglia	6%	6%	5%	6%
Home counties	12%	14%	11%	12%
London	19%	17%	15%	17%
Kent	3%	4%	4%	4%
Hampshire + Sussex	5%	5%	6%	5%
Wessex	3%	5%	5%	4%
South West	4%	4%	5%	4%
Number of students	670	754	842	2,266

Source: NRP application data

Notes: The region covering Manchester, Cumbria and Lancashire is abbreviated to North West. The region covering Yorkshire, Lincolnshire and Humberside is abbreviated to the Yorkshire region. The region covering Herts, Beds, Berks, Bucks and Oxon is termed the Home Counties. The Greater London and Surrey region is abbreviated to London. The region covering Devon, Dorsett and Cornwall is abbreviated to South West.

The distribution of NRP students across the 13 programme regions in England are presented in table 7. A comparison of the region reported in the NRP application data and in the NPD data highlighted a small number of cases where pupils were in different but bordering regions. In order to ensure that applicants are placed in the region according to their application, region from the application data is used and table 7 therefore presents the distribution for the NPD sample using the region reported in the application. The table notes list how some of the region names have been abbreviated from the full Nuffield Foundation name.

There have been some minor shifts in the regional distribution across the three cohorts: the Yorkshire region and West Midlands have increased their share, while the proportions of students in Merseyside and London have declined (although the absolute number of students in London has remained virtually unchanged).[15]

Considering the programme across all the countries in the UK using the NRP application data shows that the proportion of students in Wales has remained fairly constant across the three cohorts with a small rise from 6 percent to 7 percent (and the absolute numbers have risen), while the share in Scotland has declined from 13 percent to 8 percent over the three years and the proportion in Northern Ireland declined from 6 percent to 4 percent (both with declines in absolute numbers of students).

3.2 Disadvantage

Free school meals (FSM) are an indicator of background disadvantage: children are eligible for free school meals if their family is in receipt of income related benefits or tax credits. Just over one fifth (22 percent) of NRP students have been eligible for free school meals at some time in the previous six years (table 8). This proportion has changed little over the three cohorts.[16]

Table 8   Proportion of NRP students by eligibility for Free School Meals

	2014	2015	2016	All cohorts
Never eligible	79%	80%	77%	79%
Ever eligible	21%	20%	23%	22%
Number of students	743	754	842	2,339

Source: Matched NPD data

Notes: Percentages do not sum to 100 in the final column due to rounding.

An alternative measure of disadvantage is the IDACI (Income Deprivation Affecting Children Index) score, which reflects the level of disadvantage in the neighbourhood where the child lives. NRP students appear to be approximately

---

[15] This shift in proportions of students across regions may be due to change in the programme which switched from using school population to post-16 populations in each region to set target numbers for regional coordinators.

[16] Comparative data for free school meals, IDACI, GCSE score and the school measures are not available from the NRP application data.

representative of the distribution of pupils across this measure of neighbourhood deprivation: around one fifth of students are in each of the quintiles of the IDACI score, although the proportion from the most deprived quintile is very slightly higher over all three cohorts (table 9). The proportions change little across the three cohorts.

Table 9   Proportion of NRP students by IDACI quintile

	2014	2015	2016	All cohorts
Lowest deprivation	20%	21%	20%	20%
2	18%	18%	16%	17%
3	19%	20%	19%	19%
4	19%	22%	22%	21%
Highest deprivation	23%	20%	23%	22%
Number of students	741	754	840	2,335

Source: Matched NPD data

Notes: Percentages do not sum to 100 in the final column due to rounding. The IDACI quintiles are derived for each cohort from the distribution of the IDACI score in the Census data (KS4) for all pupils in the NPD data.

In the following analysis, the FSM measure rather than the IDACI measure is used to consider the degree to which the NRP programme selects students from more disadvantaged backgrounds. There are several reasons for this:

- The FSM measure and IDACI score are closely related: the correlation coefficient between FSM and the IDACI score is 0.39 and between FSM and the IDACI quintile is 0.37 (both are statistically significantly positive at the 99 percent level) within the entire NPD sample of all AS level students. Preliminary analysis indicated that both measures drew qualitatively similar conclusions regarding the selection into the NRP programme.

- The FSM measure is a more direct measure of student disadvantage: higher deprivation in the IDACI measure can include less disadvantaged students living in disadvantaged areas and may exclude more disadvantaged students living in less disadvantaged areas.

- The FSM measure, being a dichotomous measure of disadvantage, is more straightforward to analyse and interpret than multiple discrete categories (such as the quintiles) for the IDACI score.

- Eligibility for FSM is more commonly used as a measure of disadvantage in the existing literature and is more widely understood.

3.3 Educational attainment

Table 10 presents the distribution of the NRP students across GCSE points quintiles, capturing their level of educational attainment at key stage 4, the summer prior to application for the NRP programme. Given the GCSE and A level requirements for application, it is not surprising that the vast majority of

students (83 percent) are in the top quintile with almost none in the lowest three quintiles.[17] Although the middle cohort has slightly different proportions in the top two quintiles from the other cohorts, there is no strong pattern in the distribution over the three years.

Table 10   Distribution of NRP students by GCSE points quintile

	2014	2015	2016	All cohorts
Lowest	0%	0%	<1%	<1%
2	1%	<1%	2%	1%
3	3%	2%	2%	2%
4	17%	12%	16%	14%
Highest	79%	86%	81%	82%
Number of students	743	754	842	2,339

Source: Matched NPD data

Notes "<1%" indicates a percentage between 0 and 0.5. The GCSE points quintile bands are 0 - 164, 165 - 274, 275 - 335, 335.5 - 386 and 387 - 549 in 2014; 0 - 186, 187 - 284, 284,5 - 338, 338.5 - 386 and 386.5 - 502 in 2015; and 0 - 214, 215 - 296, 296.5 - 344, 344.5 - 386 and 386.5 - 530.5 in2016.

3.4 School characteristics

Table 11 presents the average proportion of students who have been eligible for free school meals in the previous 6 years in the schools attended by NRP students at the time of application. The proportions are lower at key stage 5 than at key stage 4: this may be because students eligible for free school meals are less likely to remain in these schools at key stage 5 (that is, they may be more likely to go on to study at other types of educational institutions) or, possibly, because the NRP programme is targeting schools with high proportions at key stage 4 and natural churning means that the proportions are closer to average levels at key stage 5. There is almost no change in the proportions across the three cohorts.

Table 11   Average FSM proportion in NRP students' schools

	2014	2015	2016	All cohorts
Proportion in school eligible for FSM in previous 6 years in Year 11 (KS4)	28%	28%	28%	28%
Proportion in school eligible for FSM in previous 6 years in Year 12 (KS5)	21%	21%	20%	21%

Source: Matched NPD data

Notes: The number of NRP students with matched school level proportions was 2,339 for KS4 and 2,313 for KS5.

---

[17] A number of successful NRP applicants (around 9 percent) did not meet the eligibility requirements according to the GCSE results and AS level study recorded in the NPD data. Of the matched 2,339 students, 13 did not meet the requirements at either level, 10 did not meet the AS level requirements and 177 did not meet the GCSE requirements. The proportion not meeting both requirements among the unsuccessful candidates was slightly higher at 12 percent, with a similar pattern of 43 not meeting the requirements at either level, 25 not meeting the A level requirements and 454 not meeting the GCSE requirements (of 4,258 unsuccessful applicants).

Table 12 presents the average educational attainment in the schools attended by NRP students at the time of application. Again, there is very little change over the three cohorts.

Table 12   Average educational attainment in NRP students' schools

	2014	**2015**	**2016**	**All cohorts**
Proportion in school achieving five good (A*-C) GCSEs including English and Maths (KS4)	64%	65%	61%	63%
Proportion in school achieving at least 3 A Levels (KS5)	77%	75%	75%	76%

Source: Matched NPD data

Notes: The number of NRP students with matched school level proportions was 2,065 for KS4 and 2,096 for KS5.

In the following analysis, the key stage 5 measures are used for school background but not the key stage 4 measures for the following reasons:

- The key stage 4 and key stage 5 measures are closely related: the correlation coefficient between the FSM proportion at key stage 4 and at key stage 5 is 0.75 and between educational attainment at key stage 4 and key stage 5 is 0.38 (both are statistically significantly positive at the 99 percent level) within the entire NPD sample of all AS level students.

- The key stage 5 measures are the more relevant ones for time of application during key stage 5 study.

3.5 Placement type

Table 13 presents the combinations of subjects and organisation types for NRP students in the 2015 and 2016 cohorts. Almost one third of placements are in the subject area of biology, biomedical and environmental sciences undertaken in a university. Indeed, almost three quarters of all placements are undertaken in universities with the private sector providing most of the remaining placements and research institutes and other types of organisations offering only 4 percent and 7 percent respectively. Some 40 percent of all placements are in the biology subject group with a more even spread across the four remaining subject areas, although the gap in the proportions in the highest of these subject group (23 percent in computing, engineering and manufacturing) and the lowest (9 percent in maths, statistics and data science) is notable.

The picture presented in table 13 is only for NRP students in England in the 2015 and 2016 cohorts. However, the picture for all NRP students including those in the 2014 cohort, in the devolved nations and others not matched to the NPD data is very similar: the main differences are that the proportion undertaking placements in the biology subject group in universities is 3 percentage points higher and the proportion undertaking placements in the computing subject group

in universities is 2 percentage points lower (all other cell differences are no more than 1 percentage point different).

Table 13 **Distribution of NRP students by placement subject and organisation type**

	Private Sector	Research Institute	University	Other	All
Astronomy and physics	1%	>1%	11%	1%	13%
Biology, biomedical and environmental sciences	2%	2%	31%	5%	40%
Chemistry, biochemistry and forensics	5%	<1%	10%	<1%	15%
Computing, engineering and manufacturing	7%	1%	15%	1%	23%
Maths, statistics and data science	1%	1%	7%	<1%	9%
All	16%	4%	73%	7%	100%

Notes: The number of students is 1,589 (those in cohorts 2015 and 2016). "<1%" indicates a percentage between 0 and 0.5. Other organisation type includes environmental, museum/education centre, zoo and medical/hospital Row and column percentages may not sum to the all column and row due to rounding.

A key aspect to note from table 13 is that the proportions across subjects for each organisation type are broadly similar and that the proportions across subject for each organisation type are broadly similar. The one exception is that the private sector tends to have relatively more placements in the computing subject group and relatively less in the biology subject group then other organisation types. However, overall, the allocations of placement subject and organisation type are broadly independent of each other[18] meaning that it is reasonable to analyse them separately. Separate analysis is also convenient as the low numbers in some cells would make it unlikely that differences in placement types across the explanatory factors will be identified.

Tables 14 and 15 present the distributions of NRP students across placement subject and placement type for the two cohorts to be used in the subsequent analysis. The main difference across the two cohorts is the lower proportion of students in placements in the chemistry subject group in 2016 and corresponding lower proportions in placements in the astronomy and physics group and the maths group, but it is difficult to draw any conclusions about trends from two years.

[18] The correlation coefficient is statistically significantly different from zero, although there is a relatively weak correlation of 0.20.

Table 14   Distributions of NRP students by placement subject

	2015	2016	All cohorts
Astronomy and physics	12%	15%	13%
Biology, biomedical and environmental sciences	41%	40%	40%
Chemistry, biochemistry and forensics	18%	12%	15%
Computing, engineering and manufacturing	22%	23%	23%
Maths, statistics and data science	7%	10%	9%
Number of students	749	840	1,589

Table 15   Placement of NRP students by placement organisation type

	2015	2016	All cohorts
Private Sector	15%	16%	16%
Research Institute	4%	4%	4%
University	73%	73%	73%
Other	7%	7%	7%
Number of students	749	840	1,589

3.6 Summary

This chapter has described the characteristics of NRP students from the 2014 to 2016 cohorts in terms of their demographic profile, educational attainment and school background for the NPD matched sample for England. The key statistics are:

- Just over half (54 percent) of students are female.

- A substantial proportion of students are from ethnic minority groups (7 percent are black, 26 percent Asian and 8 percent of other or mixed ethnicity).

- Just under a quarter (22 percent) of students are currently eligible for Free School Meals or have been eligible in the previous six years.

- Students are drawn fairly evenly from across the distribution of neighbourhood deprivation as measured in the IDACI score.

- Most students (82 percent) are from the top 20 percent of academic achievers at GCSE which is unsurprising given the GCSE requirement for NRP application.

- The proportions of students in the Yorkshire and West Midlands regions have increased slightly over the three years, while the proportions from Merseyside and London have declined.

- In schools attended by NRP students, the average proportion of pupils who are currently eligible or have been eligible for Free School Meals in the previous 6 years is 28 percent at Key Stage 4 and 21 percent at Key Stage 5.

- In schools attended by NRP students, the average proportion of pupils attaining five good GCSEs including maths and English at Key Stage 4 is 63 percent and the average proportion attaining at least 3 A levels at Key Stage 5 is 76 percent.

An examination of the types of placements undertaken showed that:

- The most common placement subject area is biology, biomedical and environmental sciences, with 40 percent of placements in this field.

- The vast majority of placements (73 percent) are undertaken at universities.

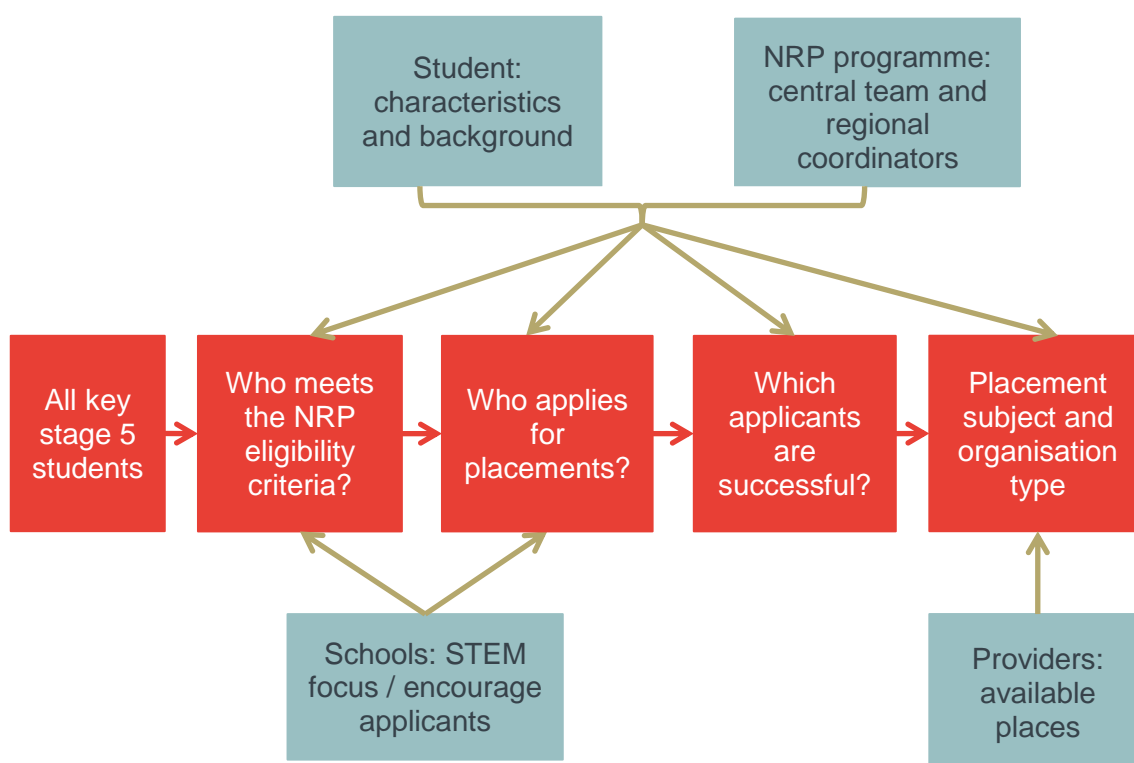- Placement subject and organisation type are not closely related.

4 SELECTION FOR PLACEMENTS

This chapter analyses the selection of NRP students through several stages from the population of all AS level students to those who are eligible to apply, to those who do actually apply, and to those offered a placement. It also considers the selection into the types of placement undertaken. Regression models are estimated to identify the factors which are related to application, offer and type of placement from a range of potential explanatory factors covering individual demographic characteristics and educational attainment and school background.

4.1 Methodology

Figure 1 presents a simple overview of the steps involved in the selection of NRP students. The red box show the steps of the selection process, while the teal coloured boxes highlight the factors influencing which students may progress through each step.
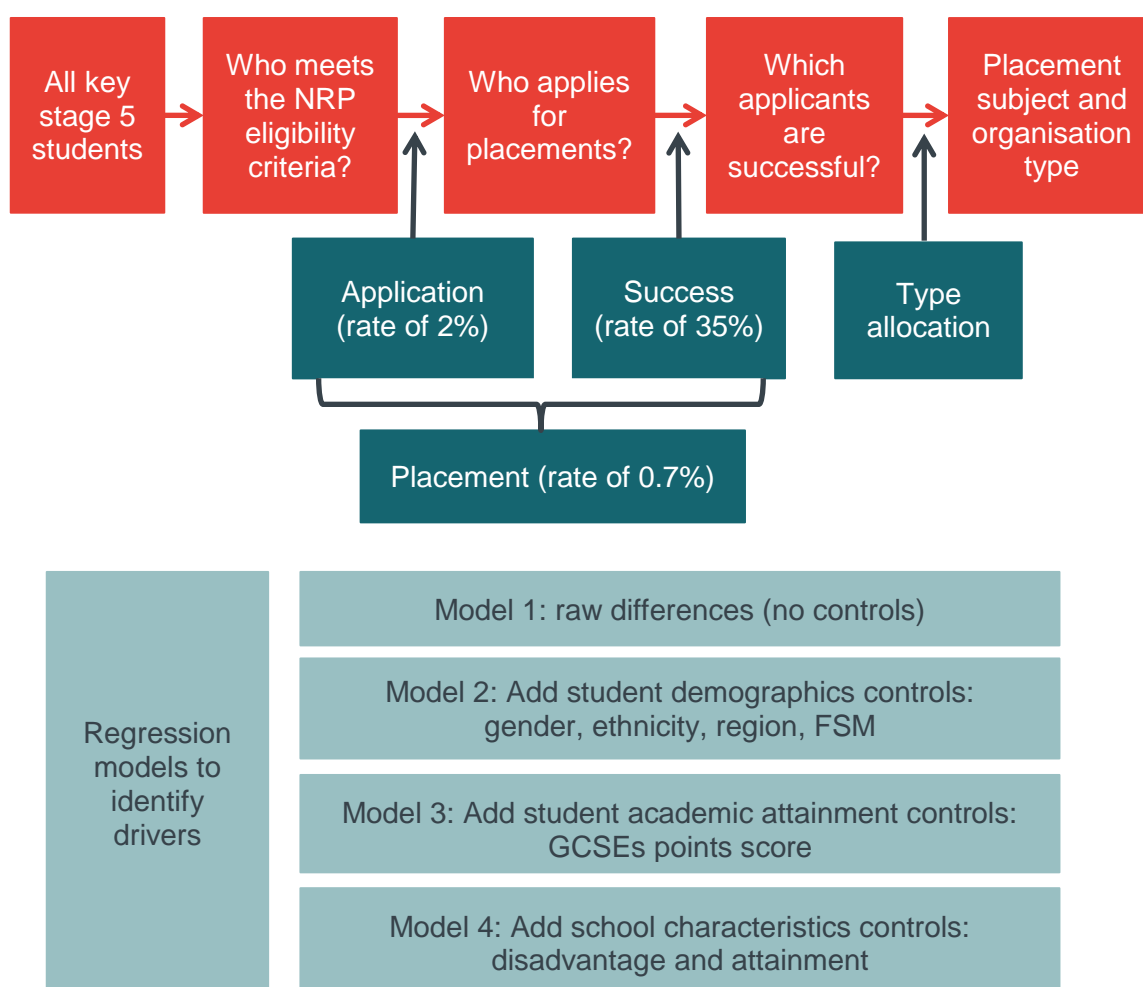
Figure 1: Overview of selection steps

To be eligible for a placement, pupils must be in the first year of A level study (first red box in the figure) and be studying at least one STEM subject at A level and to have at least five GCSEs at grade B or above, including maths, English and a science (second red box). The next step is application by the student, followed by selection for a placement and allocation to a particular type of placement. All four steps in the selection may be related to students' individual characteristics and background as well by the NRP programme in terms of how it

sets the eligibility requirements, actions taken to solicit applications and the selection process and placement allocation mechanism. In addition, schools may play a role in influencing how many and which students meet the eligibility criteria through their approach to supporting the study of STEM subjects and in their encouragement of students to apply for a placement. Placement providers will determine the availability of different types of placements and critically shape the types of placements undertaken by students.

This report focuses on a statistical analysis of the factors potentially driving the second and third steps in this process (application and success) although some consideration is also given to the type of placement that students are allocated. Other elements of the evaluation will consider the processes underlying these relationships using qualitative approaches involving case study research with programme co-ordinators, schools and colleges, providers and students.

Figure 2: Modelling selection

Note: The overall application rate in England is 2.2 percent but the placement rate in the NPD sample is 1.9 percent due to the absence of unmatched applicants in the NPD data.

Figure 2 summarises the focus of this report and outlines the regression models employed to indicate the factors which may drive the steps of applying to the programme and being successful in that application. It also considers the effect of the factors on the two steps combined (termed "placement") to test how mixed influences on application and success may have an overall influence on an eligible student securing a placement. Some consideration is also given to the "selection" of students from all AS level students to those that are eligible for application, but this is primarily for background information on the pool from which applicants are selected.

Four different regression models were estimated for most factors within the stacking of models[19]. The exception to this was for region, for which only models 1 and 4 were estimated due to the large number of discrete categories for this variable. An initial model testing the raw differences was estimated to allow for the fact that the NPD sample is a subsample of the entire NRP student population.[20] This model simply considers whether the probabilities of application and success are different across the characteristic in the entire population of NRP students, within a reasonable degree of confidence (set at 95 percent throughout[21]). The subsequent models add different combinations of control variables, testing whether any raw differences are explained by other related factors. Although the stacking of these models indicates some prioritising of the potential explanatory factors, each potential influence is also considering individually as well as in the final model with all factors included. All models were estimated as logit regressions as the dependent variables are all probabilities.[22] Versions of model 4 were also estimated with controls for the different cohorts but the inclusion of these controls had no qualitative effect on the findings and the models are not presented.

Each section below begins with the presentation of a bar group showing the distributions of each factor for all AS level students, for eligible AS level students, for applicants and for successful applicants. Increases (decreases) in the proportions across the four bars highlight greater (lower) probabilities that particular types of pupils will be eligible to apply for the programme; will apply; and will be successful in their application.

4.2 Selection by demographic characteristics

Figure 3 presents the proportions of pupils who are female across the four groups of all AS level students; eligible students; applicants and successful applicants. Just over half of all AS level students are female and very similar

---

[19] That is, only models 1, 3 and 4 are estimated for variables added in models 3 and 4 and only models 1 and 4 are estimated for those added in model 4.

[20] If the sample were complete, there would be no need for any statistical tests of differences between different types of students because the proportions would be population means without any sampling variance.

[21] A single level was considered to reduce the burden of analysing the large number of tests including variables with multiple discrete categories and to simplify the presentation of the findings.

[22] In the cases of placement subject and placement type, the models were initially estimated as multinomial logit models, estimating the inter-related probabilities across the five and four outcome categories respectively. Presentation of these findings was challenging (comparing multiple alternative outcomes across multiple category explanatory variables), so a simpler approach was taken to estimate a logit model for each outcome category separately. This generated no dissimilar results (the two are not exactly equivalent) and the key findings could be more easily presented.

proportions are eligible to apply for NRP placements, do apply and are successful in securing a placement.

These simple proportions suggest that gender is not an important influence in the selection process for placements. Indeed, regression models 1 and 2 indicate that there is no gender difference in the probability of application or selection (table 16). However, when academic attainment is allowed for (model 3), female pupils are less likely to apply than males ones. In other words, given their level of academic attainment, female pupils would be expected to be more likely to apply than male pupils. There is some limited evidence (at a lower confidence level of 90 percent rather than 95 percent) that being female is associated with a lower probability that an eligible pupil will secure a placement than an equivalent male counterpart.
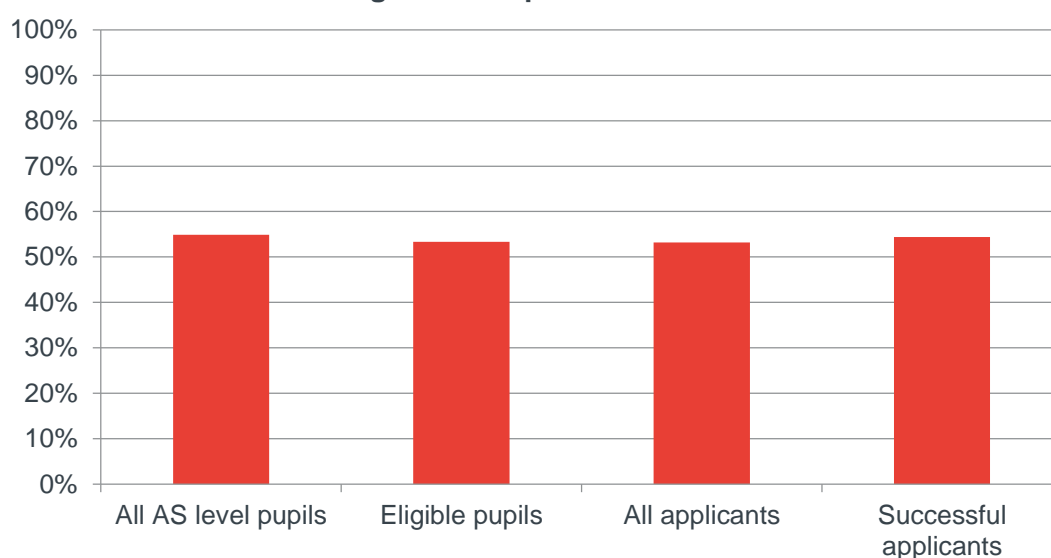
Figure 3: Proportion female

Table 16   Regression results for gender differences in selection

	Probability apply given eligible	Probability successful given apply	Probability placement given eligible
Model 1: Raw differences	No difference	No difference	No difference
Model 2: Addition of demographic controls	No difference	No difference	No difference
Model 3: Addition of academic attainment	Male > female	No difference	No difference
Model 4: Addition of school characteristics	Male > female	No difference	Male > female (at 10% sig.)

Notes: The symbol ">" indicates where one category has a statistically significantly higher probability than another category at the 95 percent confidence level (except in the case of only 90 percent in the final cell of the table).

The ethnic mix of eligible pupils is very similar to that for all AS level students, but those in non-white groups, particularly those in the Asian group, are much more

likely to apply than those in the white group (figure 4). Conditional on applying, the proportion of white pupils who are successful is slightly higher than those for the other ethnic groups, although the proportions of the ethnic minority groups undertaking placements are still notably higher than in the eligible student population.
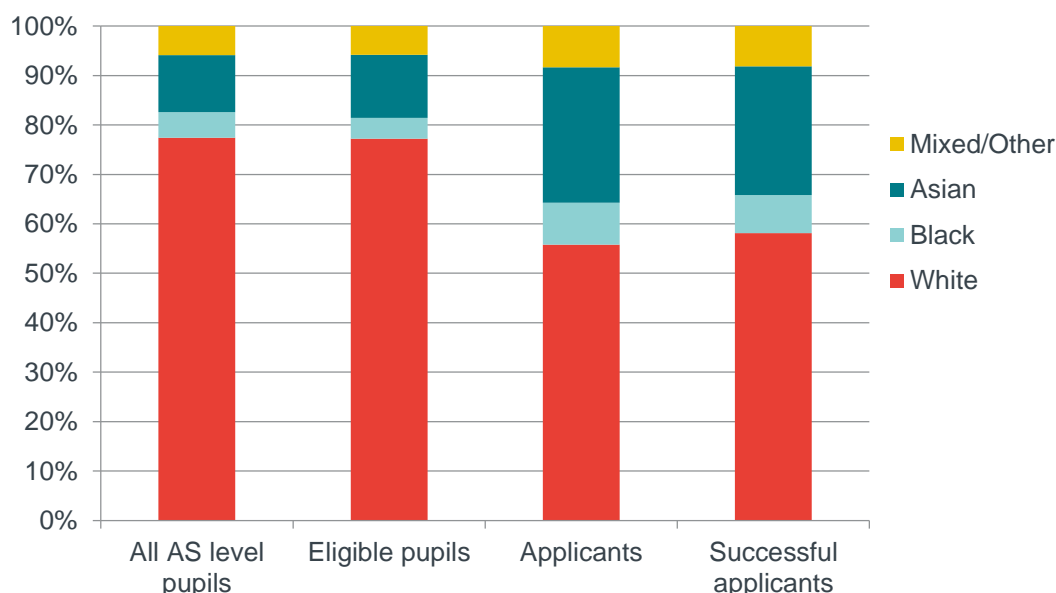
Figure 4: Proportions by ethnic group

Table 17   Regression results for ethnic differences in selection

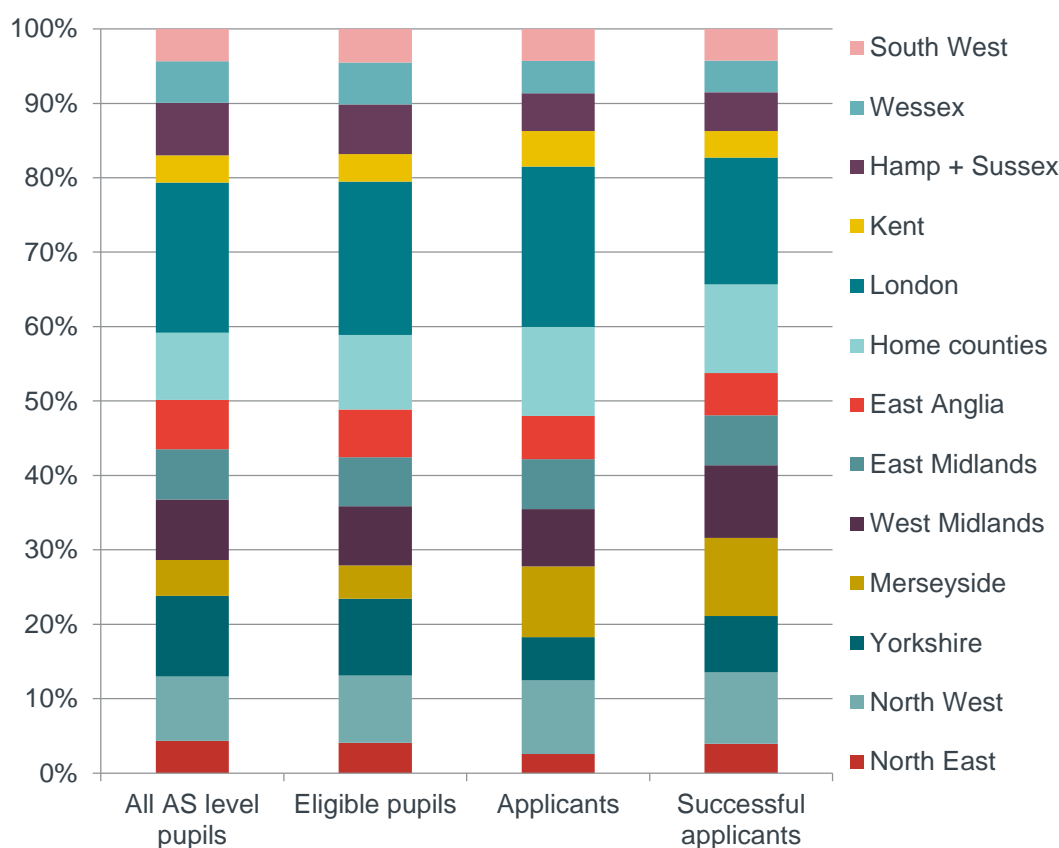|	Probability apply given eligible	Probability successful given apply	Probability placement given eligible
Model 1: Raw differences	Black & Asian > mixed/other > white	Black & Asian > white	Black & Asian > mixed/other > white
Model 2: Addition of demographic controls	Black & Asian > mixed/other > white	No differences	Black & Asian > mixed/other > white
Model 3: Addition of academic attainment	Black & Asian > mixed/other > white	No differences	Black & Asian > mixed/other > white
Model 4: Addition of school characteristics	Black & Asian > mixed/other > white	White > Asian	Black & Asian > mixed/other > white

Notes: The symbol ">" indicates where one category has a statistically significantly higher probability than another category at the 95 percent confidence level.

Even allowing for all other control factors (including the FSM measure of disadvantage), the regression results show that black and Asian pupils are more likely to apply than those of mixed or other ethnicity, who in turn are more likely to apply than white students (table 17). The raw differences in the success rates indicate that black and Asian pupils are more likely to be successful than white pupils, but controlling for school background indicates that white pupils are more

likely to be successful than Asian ones (indicating that the reason for the Asian and black higher success rate was due to the types of schools they attend). Overall, the much greater difference in the application rates dominates the final outcome that eligible pupils from ethnic minority groups are more likely to secure a placement than their white counterparts.

The analysis of application and success rates across regions considers the 13 English regions as presented in figure 5. It should be noted that some of these regions do not correspond directly to the official Government regions of the same name. For example, the south west covers only Devon, Dorset and Cornwall, London includes Greater London and Surrey and the "South East" is divided into the "Home counties" (consisting of Hertfordshire, Bedfordshire, Berkshire, Buckinghamshire and Oxfordshire), Hampshire & Sussex and Kent.

Figure 5: Proportions by region

Notes: The region covering Manchester, Cumbria and Lancashire is abbreviated to North West. The region covering Yorkshire, Lincolnshire and Humberside is abbreviated to the Yorkshire region. The region covering Herts, Beds, Berks, Bucks and Oxon is termed the Home Counties. The Greater London and Surrey region is abbreviated to London. The region covering Devon, Dorsett and Cornwall is abbreviated to South West.

Analysing patterns across 13 categories is challenging, yet figure 13 shows some notable patterns. First, the proportions in the first two columns are quite similar, indicating that AS level pupils in all the regions are equally likely to be eligible for application to the NRP programme. Second, comparing the middle two columns indicates that the rate of application is notably higher in Merseyside and lower in Yorkshire, with slightly higher rates in the Home Counties, London, Kent and Hampshire & Sussex. Third, the final two columns indicate that the rate of

success is higher in the North East, Yorkshire, Merseyside and the West Midlands and distinctly lower in London. Overall, the combined "placement rate" from eligibility to securing a place (comparing the second and fourth columns), appears higher in Merseyside and the West Midlands and lower in Kent and Yorkshire.

The large number of categories is not a barrier to statistically testing some of these differences in regression models because the NPD sample is so large. In particular, the application and placement rates are calculated using the large numbers of all eligible pupils. Statistical tests for the success rate are more challenging because the base sample is only applicants and fewer statistically significant results would be expected for the probability of success for this reason. Nevertheless, the regression models have two major advantages. First, they can identify patterns which may be difficult to see in the figure, particularly among the regions with smaller proportions of pupils. Second, model 4 with all controls for other potential influences identifies relationships which may be pure region effects and not due to some particular characteristic of the region.

Table 18   Regression results for regional differences in selection

Probability apply given eligible	Probability successful given apply	Probability placement given eligible
Model 1: Raw differences		
Almost all in: Merseyside > Kent > Home counties > North West > London > East Midlands, West Midlands, East Anglia, South West > Wessex, Hamp + Sussex > North East, Yorkshire	North East > all except Yorkshire Yorkshire > all except North East, West Midlands All > London, Kent	Merseyside > all Almost all in: North West, West Midlands, East Midlands, Home counties > Yorkshire, East Anglia, London, South West, Hamp + Sussex, Wessex, South West North East, Kent, South West > Yorkshire
Model 4: Addition of all controls		
Almost all in: Merseyside > Kent > South West > East Anglia, Home counties > North West, East Midlands > West Midlands, Wessex > Hamp + Sussex > North East > Yorkshire, London	North East, Yorkshire, West Midlands > North West, East Midlands, Wessex, South West All except North West, London, Wessex > Kent All > London	Merseyside > all East Anglia, Home Counties > North East All except London > Yorkshire All > London

Notes: The symbol ">" indicates where one category has a statistically significantly higher probability than another category at the 95 percent confidence level.

Table 18 presents a summary of the regression results. In some cases, the term "almost all in" has been used to draw out and summarise key patterns rather than listing every nuanced variation. The first column shows that almost all differences between the regions are statistically significant. Model 1 (raw differences), Merseyside, Kent and the Home Counties have the highest application rates, while North East and Yorkshire lie at the lower end. In contrast, the second column shows that the success rate is significantly higher in the North East and Yorkshire and significantly lower in London and Kent, suggesting some complementarity in application and success rates: areas with lower relative numbers of applicants are more likely for their applicants to be successful. Indeed, the greater bunching of regions in the final column (there are two main groups in the middle) suggests that success may rates may, to some degree, be compensating for differences in application rates. However, the final column also shows that the rate of application tends to dominate the overall placement rate for the more outlying regions, with Merseyside having a higher placement rate than all other regions and Yorkshire and the North East have some of the lowest placement rates in spite of their high success rate.

Controlling for the characteristics of different regions has relatively little impact on these findings (model 4 in the bottom row of table 18) with a couple of exceptions for the rate of application. First, London drops from being in the top half of the regions in the raw application rate to one of the lowest. The reason for this is that the high proportion of non-white pupils and FSM pupils in London means that the raw application rate in this region should be higher than in other areas once allowance is made for these factors.[23] Combined with the low success rate in London, allowing for the ethnic and FSM composition means that London has the overall lowest placement rate. Second, the South West (and to a lesser extent East Anglia) have relatively higher application rates and overall placement rates once allowance is made for other factors because the low proportions of non-white and FSM pupils in these areas explain some part of their rates being lower than in other areas.

Overall, there is a considerable degree of variation in application rates across regions, suggesting that there may be some diversity in approach to recruiting applicants (an issue to be explored in the qualitative strand of this evaluation). The compensating effect of the converse variation in success rates may be not be unexpected if the programme selection process aims towards a reasonably even spread of placements across regions relative to the eligible pupil population (again, something which can be explored in the qualitative research).

4.3 Selection by disadvantage

A key objective of the NRP programme is to provide opportunities for experience in STEM to pupils from more disadvantaged backgrounds. This is explored in this section using eligibility for Free School Meals in the previous six years as the measure of disadvantage.

---

[23] The previous section presented the ethnic differences, while the following section presents higher application and success rates among FSM pupils than non-FSM pupils.

Pupils studying at AS level who are or have been eligible for FSM are less likely to be eligible to apply to the NRP programme than non-FSM pupils (figure 6). However, among those eligible, FSM pupils are more likely to apply and, among applicants, are more likely to be successful. While FSM pupils constitute 11 percent of all eligible pupils, they account for 19 percent of applicants and 22 percent of those offered placements, providing strong evidence that recruitment for placements is successfully targeting those from more disadvantaged backgrounds.

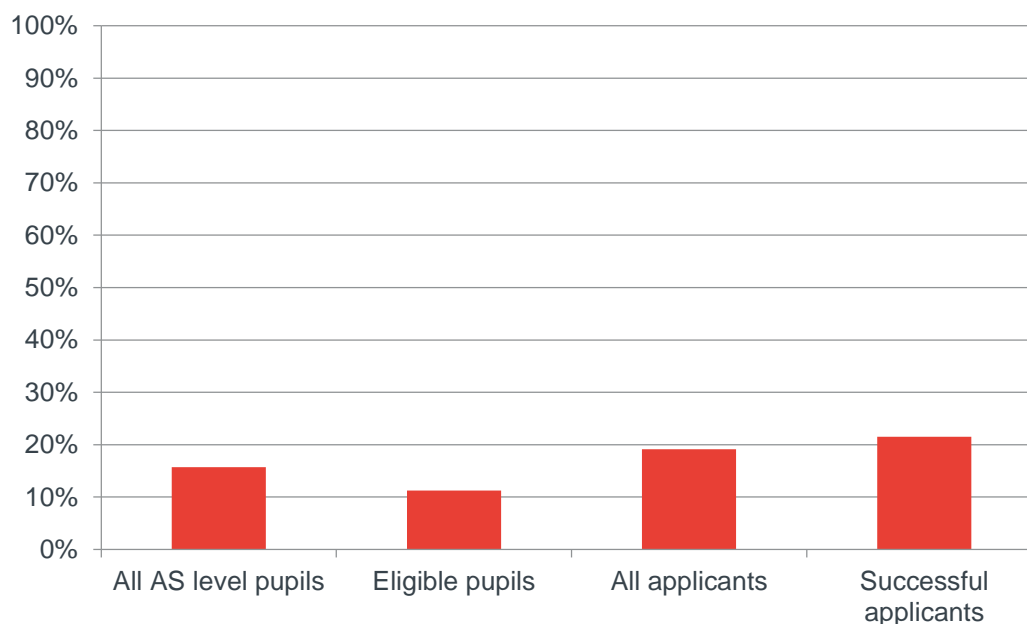**Figure 6: Proportion Free School Meals**

Table 19   Regression results for differences in selection by FSM status

	Probability apply given eligible	Probability successful given apply	Probability placement given eligible
Model 1: Raw differences	FSM > non-FSM	FSM > non-FSM	FSM > non-FSM
Model 2: Addition of demographic controls	FSM > non-FSM	FSM > non-FSM	FSM > non-FSM
Model 3: Addition of academic attainment	FSM > non-FSM	FSM > non-FSM	FSM > non-FSM
Model 4: Addition of school characteristics	FSM > non-FSM	No difference	FSM > non-FSM

Notes: The symbol ">" indicated where one category has a statistically significantly higher probability than another category at the 95 percent confidence level.

Table 19 shows that even allowing for other related factors (including ethnicity, academic achievement and school background), FSM pupils are more likely to apply and to be offered a placement. Only the addition of school characteristics in model 4 removes the statistically significant difference between FSM and non-FSM students. This suggests that, in determining which pupils are successful in their application, school background (covering the proportion of pupils who are or
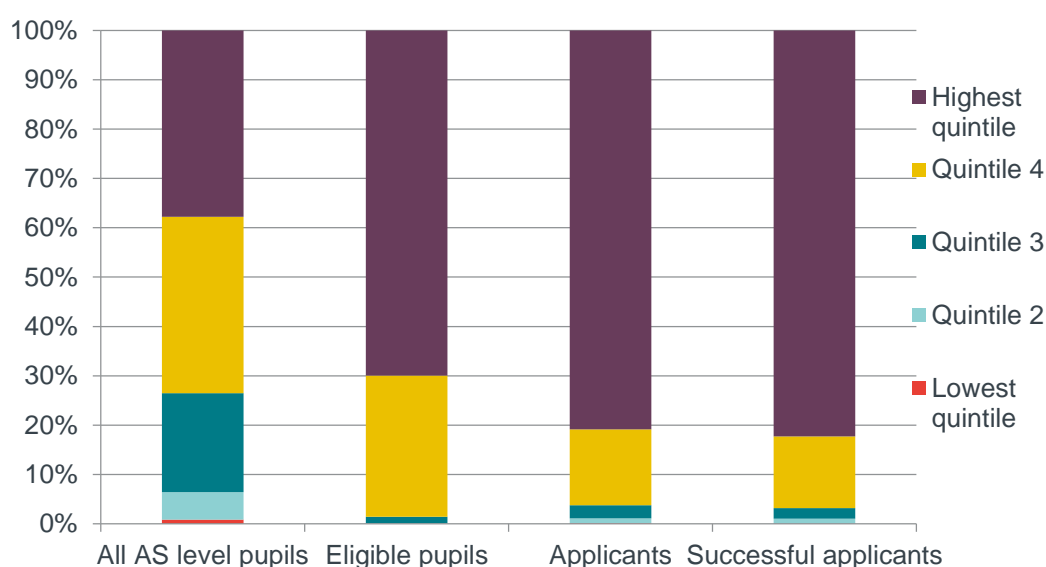
have been eligible for FSM and average educational attainment within the school) is the driving factor rather than individual FSM status[24].

4.4 Selection by pupil attainment

Pupil attainment is a type of "intermediate" factor in that it will be influenced by the demographic factors considered above but may also have an independent influence on application and success rates, reflecting that it may capture individual characteristics not covered in the other variables such as motivation. It is therefore useful to include pupil attainment as a control in the regression models and of interest to consider whether it has an effect over and above the other factors.

Among pupils studying at AS level, those eligible for application to the NRP programme have considerably higher academic attainment: almost all are in the top two quintiles of the GCSE point score, while the distribution is less skewed towards the top among all AS level pupils (first two columns in figure 7). This is to be expected given the eligibility requirements for NRP application and provides support for the analysis of application rates based on the eligible groups of AS level students rather than all AS level students. The application rate is higher among the highest and lower quintiles of achievement, with pupils in the fourth quintile least likely to apply. This suggests that there may be a dichotomy in the drivers of application with one set of factors influencing those at the top of educational attainment and another affecting those at the lower end. Interestingly, success rates are very similar across academic attainment, suggesting that this is not an important factor in obtaining a place once a pupil has applied.

Figure 7: Proportions by GCSE point quintile

Because of the very small numbers of pupils in the bottom two quintiles of the GCSE point score among eligible pupils, these two categories were combined in

---

the regression analysis (table 20). The patterns observed in figure 7 broadly reflect statistically significant differences across the quintiles, but the regression findings also show that pupils in the highest quintile are more likely to be successful than the other quintiles, particularly when controls for school background are included (reflecting the finding presented below that applicants from schools with higher average educational attainment are less likely to be successful). There is also a potentially spurious finding that those in the very lowest two quintiles are most likely to apply: this is likely to reflect the fact that this very small group of pupils will contain the small number of NRP applicants who did not meet the eligibility criteria but who were nevertheless included in the sample which will push up the application rate.

Table 20   Regression results for differences in selection by academic achievement

	Probability apply given eligible	Probability successful given apply	Probability placement given eligible
Model 1: Raw differences	Q1/2 > Q3 > Q5 > Q4	Q5 > Q3	Q1/2 > Q3 > Q5 > Q4
Model 3: Addition of demographic controls	Q1/2 > Q3 > Q5 > Q4	Q5 > Q3, Q4	Q1/2 > Q3, Q5 > Q4
Model 4: Addition of school characteristics	Q1/2 > Q3, Q5 > Q4	Q5 > Q1/2, Q3, Q4	Q1/2 > Q3 > Q5 > Q4

Notes: The symbol ">" indicates where one category has a statistically significantly higher probability than another category at the 95 percent confidence level.

4.5 Selection by school characteristics

The final set of factors considered reflect school background in terms of the proportion of pupils in the school who are or have been eligible for FSM and the average educational attainment at AS level.

Interestingly, the average proportion of FSM pupils within a school has a fairly weak relationship with the probability that a pupil will be eligible for NRP application (first two columns in figure 8), but the school FSM proportion tended to be higher for pupils who apply and higher again for those who are successful in their application. Among all eligible pupils, the average school FSM proportion is 14 percent compared to 18 percent for pupils who apply and 21 percent for those who are successful.

The regression models show that the positive relationships between application and success (and overall placement rate) and the school FSM proportion is statistically significant and remains even when allowance is made for all other factors, including the pupil's individual FSM status (table 21). This indicates that being from a school with a higher proportion of FSM pupils increases the likelihood of application and success regardless of whether the student is or has been eligible for FSM themselves.
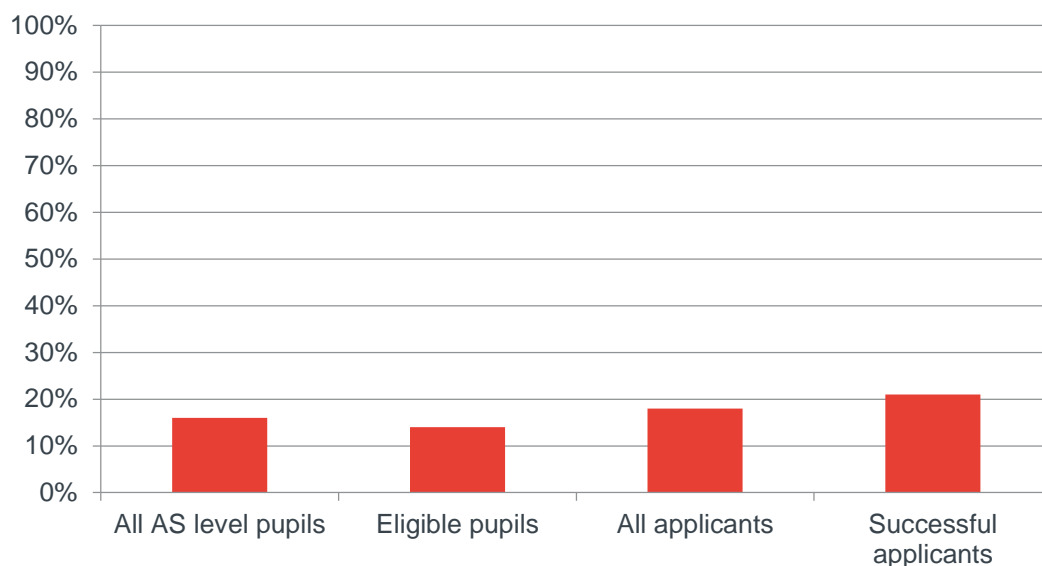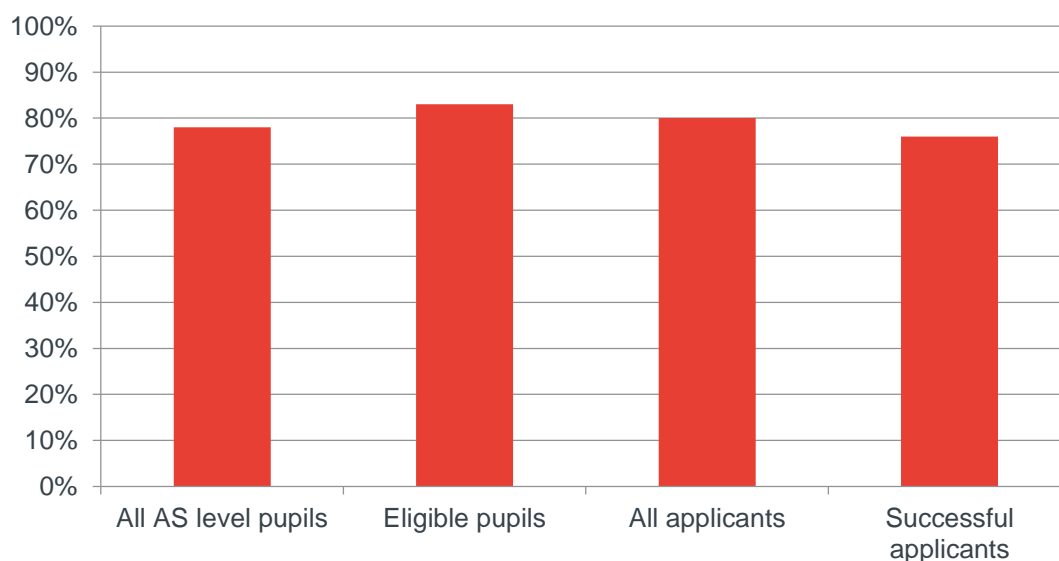
Figure 8: Average school proportion FSM

Table 21   Regression results for differences in selection by school FSM

	Probability apply given eligible	Probability successful given apply	Probability placement given eligible
Model 1: Raw differences	Rises with FSM proportion	Rises with FSM proportion	Rises with FSM proportion
Model 4: Addition of all controls	Rises with FSM proportion	Rises with FSM proportion	Rises with FSM proportion

The relationships with average school academic attainment measured as the average proportion of pupils who achieve three or more AS levels is the inverse of the picture for school FSM (which is not surprising is schools with lower proportions of FSM pupils tend to have higher attainment).

Figure 9: Average school proportion with 3+ A levels

As shown in figure 9, while eligible pupils are more likely to be in schools with higher attainment, pupils who apply and succeed tend to be in schools with slightly lower attainment. However, the regression results (table 22) show that that the probability of application and success (and the combined probability of placement) declines with average school academic attainment even controlling for the school FSM proportion (model 4). This suggests that application recruitment may be targeting schools not only with higher proportions of FSM but also schools who have lower academic achievement.

Table 22   Regression results for differences in selection by school academic achievement

	Probability apply given eligible	Probability successful given apply	Probability placement given eligible
Model 1: Raw differences	Declines with 3 A levels proportion	Declines with 3 A levels proportion	Declines with 3 A levels proportion
Model 4: Addition of all controls	Declines with 3 A levels proportion	Declines with 3 A levels proportion	Declines with 3 A levels proportion

4.6 Allocation across placement types

This final section of this chapter briefly considers the allocation of types of places across gender, ethnicity and pupil's FSM status. This analysis is more limited than that presented for application and application success rates for several reasons. First, the sample is substantially smaller for placement types because it is limited to successful candidates in the matched NPD data in the 2015 and 2016 cohorts. Second there is considerable additional complexity of examining placement type across several discrete categories which makes the presentation of findings complex and the likelihood of statistically significant quantitative findings lower. Finally, the allocation of successful applicants to placement types is a more complex mix of student preferences and availability of places and may be better suited to consideration in the qualitative strand where greater use of NRP data can be made.

As described in chapter three, the allocation by placement subject and by placement organisation type are relatively independent and the analysis presented considers the two aspects separately.

Table 23 presents the proportions of male and female students across the five categories of placement subjects and four organisation types. Female students are substantially more likely to be in biology, biomedical and environmental science placements than male students and are less likely to be in placements in the astronomy and physics subject area and the computing, engineering and manufacturing subject area. Regression models for the subject choice confirm that these differences are statistically significant even in model 4 which includes controls for a broad range of other factors, and also shows that female students are also more likely to undertake placements in the chemistry, biochemistry and forensics subject area than male students.

Table 23   Distribution of placement subjects by gender

	Male	Female
Astronomy and physics	19%	8%
Biology, biomedical and environmental sciences	26%	52%
Chemistry, biochemistry and forensics	13%	17%
Computing, engineering and manufacturing	33%	14%
Maths, statistics and data science	9%	9%
Total	100%	100%

Female students are also more likely to undertake placements in universities and less likely to undertaken them in the private sector than male students (table 24). Again, these differences are statistically significant even in models with the broad range of controls, indicating that gender in itself is a key driver of the type of placement.

Table 24   Distribution of placement organisation types by gender

	Male	Female
Private Sector	20%	12%
Research Institute	4%	4%
University	69%	77%
Other	7%	8%
Total	100%	100%

The main distinctions in placement type across ethnic groups are between white students and the other ethnic groups. White students are more likely than other students to undertake placements in the astronomy and physics subject area and in the chemistry, biochemistry and forensics subject area and are less likely to undertaken a placement in the biology, biomedical and environmental sciences area (table 25). This distinction is statistically significant even in the models with all controls for other factors (table 26), although the difference with the Asian group appears to be explained by school background for two of the subject areas, while only holds with the Asian group for the chemistry subject group. In addition, there is some tendency for Asian students to be more likely to undertake placements in the maths, statistics and data science subject area and the raw difference with white students is statistically significant, but is explained by other related factors (that is, the difference is not significant in models 2 to 4).

There is also a higher propensity for white students to undertake placements in the private sector rather than universities (table 27). The raw differences for white and Asian students are statistically significant, but are explained by other related factors and are not significant in the regression models with controls.

Table 25   Distribution of placement subjects by ethnicity

	White	Black	Asian	Mixed / other
Astronomy and physics	16%	5%	10%	11%
Biology, biomedical and environmental sciences	36%	55%	43%	46%
Chemistry, biochemistry and forensics	18%	11%	11%	11%
Computing, engineering and manufacturing	22%	21%	23%	23%
Maths, statistics and data science	7%	8%	12%	8%
Total	100%	100%	100%	100%

Table 26   Regression results for ethnic differences in placement subject

Probability allocated to subject	Model 1: Raw differences	Model 2: Addition of demographic controls	Model 3: Addition of academic attainment	Model 4: Addition of school characteristics
Astronomy and physics	White > black, Asian	White > black, Asian	White > black, Asian	White > black
Biology, biomedical and environmental sciences	All > white Black > Asian	All > white	Black, mixed/other > white	Black, mixed/other > white
Chemistry, biochemistry and forensics	White > Asian	White > Asian	White > Asian	White > Asian
Computing, engineering and manufacturing	No differences	No differences	No differences	No differences
Maths, statistics and data science	Asian > white	No differences	No differences	Asian > mixed/other

Notes: The symbols "<" (">") indicate where one category has a statistically significantly lower (higher) probability than another category at the 95 percent confidence level.

Table 27   Distribution of placement organisation types by ethnicity

	White	Black	Asian	Mixed / other
Private Sector	18%	13%	12%	11%
Research Institute	4%	5%	3%	4%
University	70%	76%	77%	74%
Other	7%	5%	8%	11%
Total	100%	100%	100%	100%

Table 28 presents the distributions of placement subject by students' FSM status. FSM eligible students have a slightly higher tendency to undertake placements in the biology, biomedical and environmental sciences area and in the maths, statistics and data science subject area. These differences are statistically significant in the regression models for raw differences, but are not so when other

control variables are included indicating that the differences are driven by other related factors (including ethnicity as most FSM students are not white).

Table 28   Distribution of placement subjects by Free School Meals status

	Not eligible	Eligible
Astronomy and physics	14%	11%
Biology, biomedical and environmental sciences	39%	45%
Chemistry, biochemistry and forensics	15%	14%
Computing, engineering and manufacturing	24%	19%
Maths, statistics and data science	8%	11%
Total	100%	100%

Finally, table 29 shows that FSM eligible students are slightly more likely to undertake placements in universities rather than in the private sector, but, again while the raw differences are statistically significant, the differences appear to be explained by other related factors as the differences are not significant in models with controls for other factors.

Table 29   Distribution of placement organisation types by Free School Meals status

	Not eligible	Eligible
Private Sector	17%	12%
Research Institute	4%	3%
University	72%	78%
Other	7%	6%
Total	100%	100%

Overall, there is no evidence that FSM status in itself influences the type of placement undertaken. However, some caution should be exercised in concluding that FSM status is not important as the relatively small sample of placement students and the high number of control variables could mean that any effects are too small to detect in this analysis.

4.7 Summary

This chapter has analysed the selection of NRP students by examining the variation in application rates and success rates across the range of individual demographic characteristics and educational attainment and school background variables, using regression models to isolate potential drivers of differences in the selection process. The key findings are:

■ Female eligible pupils are less likely to apply (given their other characteristics) but have the same probability of success if they apply as their male

counterparts. Overall, there is limited evidence that eligible female pupils are less likely to undertake placements than males.

- Eligible ethnic minority pupils are more likely to apply than white pupils, but Asian applicants (given their other characteristics) are less likely to succeed than white applicants. Overall, eligible ethnic minority pupils are more likely to undertake placements than white pupils.

- Eligible pupils in the Merseyside, Kent and Home counties regions are more likely to apply than those in other regions, but those in Kent and London are less likely to be successful. Pupils in the Yorkshire and North East regions are less likely to apply than those in other regions but are more likely to succeed. While there is some balancing in these opposing differences across regions, the application rate appears to dominate with eligible pupils in Merseyside being most likely to undertake a placement and those in the Yorkshire and North East regions least likely.

- Eligible pupils from disadvantaged backgrounds (measured by Free School Meals eligibility) are more likely to apply and are more likely to be successful in their application than other pupils, although the higher success rate is explained by related school characteristics. Overall, disadvantaged eligible students are more likely to undertake a placement than other pupils.

- Eligible pupils with lower academic attainment or the highest level (in the lower or highest GSCE score quintile) are more likely to apply, but applicants in the highest quintiles have the greatest probability of success. Overall, pupils in the lower or highest quintiles of achievement are more likely to undertake placements than pupils in the middle of the range.

- Eligible pupils in schools with higher proportions of Free School Meals pupils and lower average educational attainment are more likely to apply and more likely to succeed if they apply (with a higher rate of placements overall).

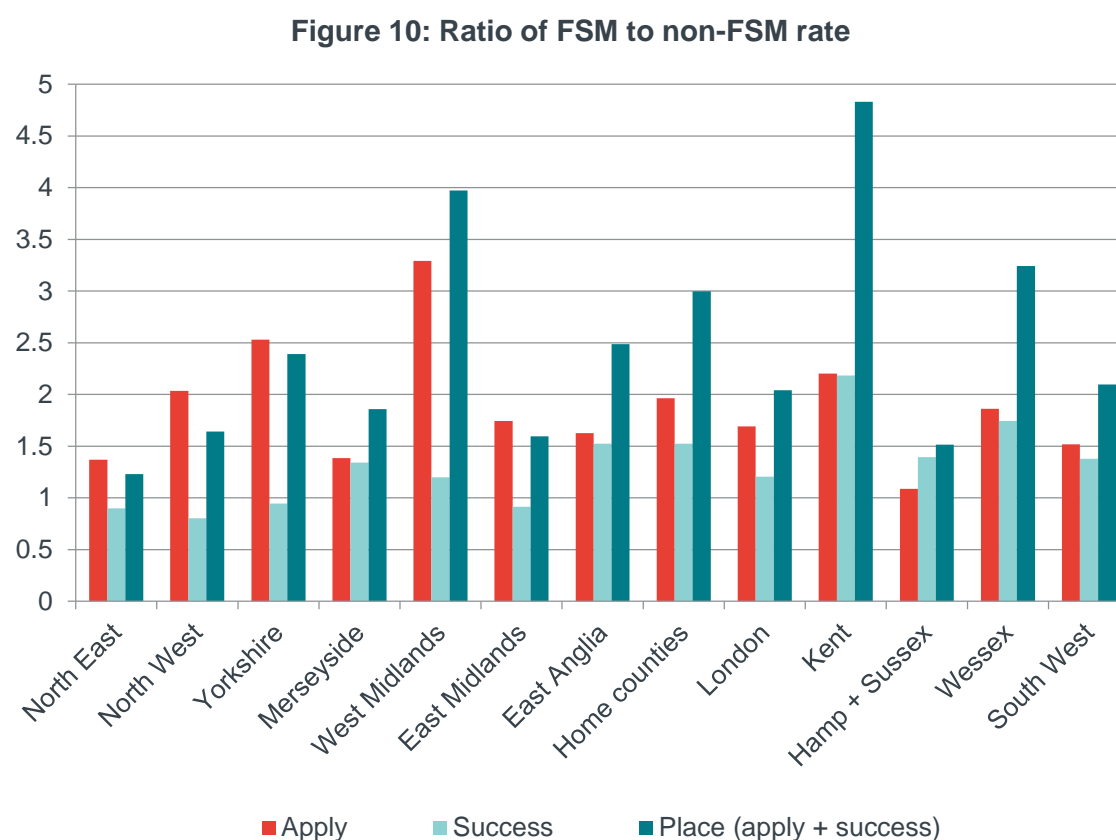A brief review of the distribution of placement types also showed:

- Female NRP students are more likely than males to have placements in biology and chemistry related subjects than in physics, computing or engineering related subjects. They are also more likely to have placements in universities than in the private sector.

- White NRP students are more likely than ethnic minority students to have placements in physics and chemistry related subjects than in biology related subjects.

5 TARGETING DISADVANTAGED PUPILS

This chapter focuses on two particular aspects of the aim of the NRP programme to target pupils from disadvantaged backgrounds: a comparison of the targeting of disadvantaged students across the NRP regions and an examination of whether disadvantaged students are targeted within schools. However, it should be noted that, as explained in chapter 2, the NPD data contains very few students from independent schools and this analysis does not include most NRP students from independent schools.

5.1 Targeting across regions

Figure 10 presents the ratio of the application, success and placement rates for FSM students to non-FSM students across the 13 NRP English regions. Bars which are greater than one indicate where a rate is greater for FSM students than non-FSM students and bars less than one indicate where a rate is greater for the non-FSM students.

Figure 10: Ratio of FSM to non-FSM rate

The figure highlights several important points. First, all of the application (red) bars are greater than one: the application rate for FSM students is greater than non-FSM students in every region. Second, in almost all areas, the success bar is also greater than one, indicating the FSM applicants are more likely to be successful than non-FSM applicants in almost all areas. This is not the case in only in four areas (North East, North West, Yorkshire and East Midlands). Taken

together, these two points mean that probability of an eligible pupil obtaining a placement is higher for FSM students than non-FSM students in all regions.

The third point to note is that a few regions have considerably higher placement ratios. For example, in Kent, an eligible pupil is 4.5 times more likely to obtain a placement if they are or have been eligible for FSM than a student who has not, while the ratio is almost 4 in the West Midlands and over 3 in Wessex. Moreover, the regions with the highest placement ratios achieve this in one of two ways: either by having very high recruitment ratios (for example, Yorkshire and West Midlands) or by having more moderate but relatively higher ratios for both recruitment and success (for example, Kent and Wessex). This suggests a potential for learning across the regions in how they recruit and support pupils from disadvantaged backgrounds to apply for placements.

Table 30   Regression results for regional differences in selection in targeting disadvantaged pupils

Ratio of FSM to non-FSM		
Probability apply given eligible	**Probability successful given apply**	**Probability placement given eligible**
Model 1: Raw differences		
West Midlands > all except Yorkshire Yorkshire > North East, Merseyside, East Anglia, London, Hamp + Sussex, South West North West > North East, Merseyside, Hamp + Sussex Home counties, Kent > Merseyside, Hamp + Sussex	Kent > North East, North West, Yorkshire, West Midlands, East Midlands, London Wessex > North East, North West, Yorkshire, East Midlands Home counties > North West, Yorkshire, East Midlands Merseyside, West Midlands, East Anglia, London, Hamp + Sussex > North West	West Midlands, Kent > all except East Anglia, Home counties, Wessex Home counties > North East, North West, Merseyside, East Midlands, London Wessex > North East, East Midlands Yorkshire > North East
Model 3: Addition of all demographic and educational attainment controls		
West Midlands > all except Yorkshire, Wessex Kent > Merseyside, London, Hamp + Sussex Yorkshire > London, Hamp + Sussex	Kent > North East, North West, Yorkshire, West Midlands, East Midlands, London Wessex > North East, North West, Yorkshire, East Midlands Home counties > North West, Yorkshire, East Midlands, London East Anglia > East Midlands Merseyside, West Midlands, East Anglia, London > North West	Kent > all except West Midlands, East Anglia, Wessex West Midlands > North East, North West, Yorkshire, Merseyside, East Midlands, London, Hamp + Sussex Wessex > North East, North West, East Midlands, London Home counties > North East, North West, East Midlands

Notes: The symbol ">" indicates where one category has a statistically significantly higher probability than another category at the 95 percent confidence level.

Table 30 presents the results from regression models testing the statistical significance of differences in the ratios across the regions, both for the raw differences in models without any controls and in a model with demographic and educational attainment controls.[25][26] As with the cross-regional analysis above, the set of interactions are complicated and the main objective is to identify regions which have generally higher and generally lower ratios than elsewhere once allowance has been made for variation in other potentially related factors.

In general, the inclusion of control variables only alters the relative performance of regions in a few cases in the FSM/non-FSM ratio for the rate of application. Allowance for other related factors means that the London ranks lower, while the North West and Home Counties regions do not rank so highly. Overall, the regression models confirm the patterns from figure 10:

■ West Midlands, Kent and Yorkshire have the highest FSM to non-FSM ratios for the application rate, while London and Hampshire & Sussex have the lowest.

■ Kent, Wessex and the Home counties have the highest FSM to non-FSM ratios for the success rate, while the North West and East Midlands have the lowest.

■ Kent, West Midlands, Wessex and the Home counties have the highest FSM to non-FSM ratio for the combined placement rate, while the North East, North West and East Midlands have the lowest.

5.2 Targeting within schools

As seen above, higher application rates and success rates are associated both with a pupil being or having been eligible for FSM and with a pupil attending a school with a higher proportion of FSM pupils. This suggests that the targeting of disadvantaged pupils occurs both directly at the individual level and through the targeting of schools with more FSM pupils. This section considers more directly whether FSM pupils are targeted within schools.

By way of background, table 31 presents the distributions of applicants across different schools, derived using the school identifier in the NPD data. Within each cohort, around 1,100 schools were identified as having applicants to the programme, while around 2,000 different schools had an applicant at some time during the three years.[27] Just over half of schools had a single applicant in each year, with almost a quarter having two applicants in each year. Summing the number of applicants across all three years leads to moderately higher proportions at higher numbers of applicants, indicating that there is some churning in the specific schools which offer applicants.

---

[25] These models included a dummy variable for each region and a dummy variable interacted with the FSM status for each region. Testing the differences for the interaction terms indicated the differences in the FSM to non-FSM ratio.

[26] The models were not estimated with school background controls because this would have removed any targeting of FSM pupils via targeting schools with high proportions of FSM pupils.

[27] It should be noted that this number only covers schools in England and those with applicants or placements matched to the NPD data.

Table 31   Numbers of applicants per school

	2014	2015	2016	All cohorts
Proportion of schools with:				
1 applicant	55%	55%	58%	36%
2 applicants	21%	22%	21%	22%
3 applicants	11%	12%	9%	12%
4 applicants	6%	4%	6%	9%
5-9 applicants	6%	6%	5%	16%
10-20 applicants	1%	<1%	1%	5%
20+ applicants	0%	0%	0%	<1%
Total	100%	100%	100%	100%
Mean number of applicants per school	1.98	1.94	1.90	3.19
Number of schools	1,074	1,160	1,174	2,067

Source: Matched NPD data

Note: School numbers only include schools in England with NRP applicants pupils matched to the NPD data

Table 32   Numbers of placements per school

	2014	2015	2016	All cohorts
Proportion of schools with:				
1 placement	76%	76%	74%	57%
2 placements	19%	19%	16%	23%
3 placements	3%	3%	6%	11%
4 placements	1%	1%	2%	3%
5-9 placements	1%	1%	1%	5%
10-20 placements	0%	0%	0%	<1%
Total	100%	100%	100%	100%
Mean number of placements per school	1.33	1.34	1.40	1.87
Number of schools	560	563	601	1,250

Source: Matched NPD data

Note: School numbers only include schools in England with NRP applicants pupils matched to the NPD data

Table 32 presents the analogous picture for the distribution of placements across schools. Around 600 schools have pupils undertaking placements in each cohort, while around double that number (1,250) had a placement pupil at some point during the three years. Around three quarters of the schools had a single placement in each cohort, while over half only had a single placement across the three years. Again, summing the placements across all three years shows

moderately higher proportions at higher number of pupils, indicating that there are not a large number of schools repeatedly having pupils with placements.

Table 33 presents a summary of the number of years with an applicant and the number of years with a placement for each school with at least one applicant over the three years. These schools constitute 58 percent of all schools identified in the NPD data with AS level students: in other words, over half of the schools in England who have AS level students who record data for the NPD had at least one applicant to the NRP programme during these three years.

Some 40 percent of schools with an applicant at some time during the three years did not have any pupils who succeeded in obtaining a place over the three years.[28] At the other extreme, 35 percent of schools (the sum of the leading diagonal) achieved at least placement in every year that they had an least one applicant. The remaining 25 percent had applicants in multiple years and achieved places in some years and not in others. Interestingly, most schools (52 percent) had an applicant in only one of three years, while a small proportion (17 percent) had an applicant in all three years.

Table 33   Numbers of years schools have applicants and placements

Number of schools (proportion)	No placements	Placements in 1 year	Placements in 2 years	Placements in 3 years	Total
Applications in 1 year	596 (29%)	479 (23%)	n/a	n/a	1,075 (52%)
Applications in 2 years	182 (9%)	288 (14%)	173 (8%)	n/a	643 (31%)
Applications in 3 years	39 (2%)	95 (5%)	129 (6%)	86 (4%)	349 (17%)
Total	817 (40%)	862 (42%)	302 (15%)	86 (4%)	2,067 (100%)

Source: Matched NPD data

Note: School numbers only include schools in England with NRP applicants pupils matched to the NPD data

Overall, this background review indicates that it is challenging to identify whether particular pupils are targeted for application within schools because of the small number of applicants and placements relative to the number of schools involved. The regression analysis used a model with fixed effects for each individual school and several approaches were tested to identify one which could provide robust estimates. The only models which could be robustly estimated were for the

---

[28] The schools with no successful applicants were broadly similar to those with a successful applicant in terms of the proportion of FSM pupils in the school (means of 25 percent and 27 percent at key stage four and means of 17 percent and 19 percent at key stage five respectively) and in terms of school attainment (means of 62 percent and 63 percent achieving five good GCSEs including English and maths at key stage four and means of 77 percent and 76 percent achieving at least three A levels at key stage five respectively). The proportion of schools without any successful candidate varied to a greater extent across regions: Merseyside and the North East had the lowest proportions (23 percent and 29 percent respectively), while another five (North West, Yorkshire, West Midlands, East Midlands and Home Counties) had lower than average proportions between 34 percent and 39 percent and the remaining regions had proportions in the range of 44 percent to 47 percent.

probability of application for each cohort separately with a single explanatory variable for the pupil's FSM status.[29] These models had the following results:

- FSM eligibility had a marginal effect of 0.996, 0.898 and 1.072 on the probability of application for the 2014, 2015 and 2016 cohorts. For all three cohorts, the effect was statistically significant at the 1 percent level.

This means that current or recent eligibility for FSM is associated with an average higher probability of application of around 1 percentage point over other students in the same school. Given the average application rate of 1.9 percent in the NPD matched sample considered here, this is a sizable difference.

5.3 Summary

This chapter has focused on two particular aspects of the targeting of pupils from disadvantaged backgrounds. First, a comparison of the targeting of disadvantaged students across the NRP regions has shown that:

- The application rate for FSM students is greater than non-FSM students in every region and the success rate is higher for FSM students in all bar four regions. Overall, the probability of an eligible pupil obtaining a placement is higher for FSM students than non-FSM students in all regions.

- A few regions have considerably higher placement rates for FSM pupils over non-FSM pupils, suggesting some potential for learning across the regions in how they recruit and support pupils from disadvantaged backgrounds to apply for placements.

- Kent, West Midlands, Wessex and the Home counties have the highest FSM to non-FSM ratio for the overall placement rate, while the North East, North West and East Midlands have the lowest.

Second, an examination of whether disadvantaged students are targeted within schools has shown that:

- Over half of the schools in England who have AS level students who record data for the NPD had at least one applicant to the NRP programme during the three years 2014 to 2016.

- Placement students are drawn from a broad range of schools and the evidence suggest that very few schools have multiple placement students year on year.

- Current or recent eligibility for FSM is associated with an average higher probability of application of around 1 percentage point over other students in the same school. This is a substantial effect given the average application rate of 1.9 percent.

---

[29] Other approaches tested included a model combining all cohorts, models which also contained other control variables and models for placement rather than application, but these models were unable to converge to a unique solution.

6 CONCLUSIONS

A key aim of the NRP programme is to offer pupils from disadvantaged background the opportunity to obtain research experience in STEM subjects. There is substantial evidence that the programme is successfully targeting disadvantage pupils in several ways:

- Pupils from disadvantaged backgrounds (measured by FSM eligibility) are more likely to apply for and undertake a placement than other pupils.

- Pupils in schools with higher proportions of FSM pupils and lower average academic attainment are more likely to apply and to be successful in their application.

- Pupils eligible for FSM are more likely to apply than non-FSM pupils in all 13 of the English NRP regions and are more likely to undertake a placement in all regions.

- Pupils eligible for FSM are considerably more likely to apply than other pupils in their school.

The substantially higher degree of targeting of disadvantaged pupils in some regions suggests that there could be potential learning across the regions in how they recruit and support pupils from disadvantaged backgrounds to apply for placements.

In addition, pupils from ethnic minorities also appear to benefit disproportionately from the programme and are more likely to apply for and undertake placements than other pupils. Indeed, ethnicity appears to have an influence over and beyond any association with disadvantage. On the other hand, there is evidence that female pupils should be applying at a higher rate than they are given their background characteristics. However, the fact that more than half of placements are undertaken by female students may be deemed a success given the historically low representation of women in STEM.

There are two considerations to be drawn from this report for the subsequent research to be undertaken in the evaluation. First, it can inform on the best comparison group for the impact analysis to be undertaken on the subsequent study and career choices for placement students. This report has confirmed a prior expectation that NRP applicants closely resemble placement students in a number of characteristics, but it has also shown that AS level students who meet the NRP eligibility requirements in terms of GCSE qualifications and A level study choices resemble placement students much more closely than the entire population of AS level pupils. It has also shown that the sample size of these NRP eligible pupils is substantial. Hence, using AS level pupils selected on the basis of meeting the eligibility requirements as a comparison group is a better option than using all AS level students because of the superior comparability with relative little loss in sample size. However, using unsuccessful applicants as an additional comparison group has the advantage that this be able to include pupils from independent schools and in the devolved nations which cannot be matched with the NPD data.

The second consideration involves the linking of this analysis with other elements of the evaluation, particularly the ongoing qualitative research and further analysis of the application data and student survey. Several of the findings presented here will benefit from the qualitative analysis of the recruitment and selection process, for example, the regional variations in the process and the dichotomy in application across the highest and lower levels of academic achievement. Further examination of the determinants of placement type may also help explain the gender and ethnic variations observed in this report. On the other hand, the linking with the NPD data has provided an exceptionally rich data source on NRP applicants and placement students and holds the potential for further analysis to help address questions raised by the other work strands. For example, the pattern in applications across schools or other determinants of placement types might be explored further.