Heterogeneous effects of missing out on a place at a preferred secondary school in England

Gorman, Emma and Walker, Ian

Heterogeneous effects of missing out on a place at a preferred secondary school in England

Emma Gorman* and Ian Walker†

Abstract

Schools vary in quality, and high-performing schools tend to be oversubscribed: there are more applicants than places available. In this paper, we use nationally representative cohort data linked to administrative education records to study the consequences of failing to gain admission to one's first-choice secondary school in England. Our empirical strategy leverages features of the institutional setting and the literature on school choice to make a case for a selection-on-observables identifying assumption. Failing to gain a place at a preferred school had null to small impacts on short-run academic attainment, but was associated with reductions in mental health and increased smoking rates in early in adulthood. In areas which deployed a manipulable assignment mechanism to allocate school places, we detected detrimental effects on high-stakes examination outcomes. Our results show that schools are important in shaping more than just test scores, and that the workings of the school admission system play a fundamental role in ensuring access to good schools.

Keywords: Education; school choice; human capital; market design; risky behaviours

JEL codes: I21; I24; J24; H44; D47.

*We are grateful to the ESRC for funding this project via ES/R003629/1; OFSTED for providing (experimental) inspection outcomes data; David Church at the Centre for Longitudinal Studies UCL for his work preparing confidentialised distance-to-school variables for the project; John Coldron and colleagues for kindly providing data on variation in school admission policies; the UK Data Service for support and access to the data via their secure lab. This work was produced using statistical data from ONS. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates. The authors are grateful for helpful comments from participants at the 3rd IZA Economics of Education workshop and the VUW Applied Econometrics Workshop 2018, and to two anonymous reviewers whose comments have improved the paper. Declarations of interest: none.

*School of Organisations, Economy and Society, Westminster Business School, University of Westminster, United Kingdom (corresponding author). e-mail: e.gorman@westminster.ac.uk

†Department of Economics, Lancaster University Management School, Lancaster University, United Kingdom.

1 Introduction

Demand for places at popular schools exceeds supply, and not all children are able to attend their preferred school. Attending a preferred school may have important consequences for long term outcomes, if parents choose schools based on test score improvements or other school attributes important for child outcomes. In addition to academic attainment, parents may rationally value a wider range of outputs that schools might provide, for example non-cognitive skill development or a safe environment that supports pupil well-being—all of which can have important consequences for success in later life (Heckman et al., 2006; Mendolia et al., 2018).

The evidence to date on whether parents choose schools based on causal improvements in test scores, rather than peer mix, pedagogy or other school policies, is mixed (Gibbons and Silva, 2011; Burgess et al., 2014; Abdulkadiroglu et al., 2017b). The evidence on effects of attending a preferred school is also mixed, although the majority of studies document null, or at most modest, effects of attending a preferred school on short-run test scores (Cullen et al., 2006; Dobbie and Fryer, 2014; Deming et al., 2014; Abdulkadiroglu et al., 2014, 2017a; Beuermann and Jackson, 2020). These studies are often based on data from specific cities (for example, Boston or New York), and specific settings, such as academically "elite" schools. This paper provides new evidence by combining administrative data and detailed cohort data to study the short- and long-run outcomes associated with missing out on a place at a first-choice secondary school. Our setting is England, which, with a largely centralised school funding system and admissions policies, represents an informative laboratory for studying school choice.

The majority of school places in England are allocated using a centralised system, which endeavours to honour parental preferences subject to school capacity constraints. Our study focuses on secondary schools, which children attend from age 11 to at least 16 years. Secondary school test score averages are well-publicised and a number of empirical regularities have been established. We know that parents value secondary schools which perform well on accepted academic metrics (Gibbons and Silva, 2011; Burgess et al., 2014), and there is regional and demographic patterning in the share of families gaining entrance to their preferred school (Wel-

don, 2018). On average, parents value schools perceived as academically "good" schools, and are willing to travel to attend them (Weldon, 2018; Burgess et al., 2019). About 84% of children are offered a place at their first-choice secondary school. Yet, to date, there is little evidence on differences in outcomes for those attending (or not) their first-choice secondary school in the UK.

This paper contributes to the literature in at least two important ways. First, we consider longer-run outcomes, in contrast to many papers which focus on the effects of missing out on short-run test scores only. For instance, Ovidi (2020) studies the effects of missing out on preferred primary schools in London on short-run achievement outcomes. However, parents may choose schools based on other dimensions, the effects of which may not be revealed until later in life (Beuermann and Jackson, 2020). Hence in addition to short-run academic achievements, we consider longer run and broader outcomes, including mental health, fertility, income and smoking in early adulthood. Second, two mechanisms for allocating school places (given parental preferences, school priorities and capacity constraints) were employed at the time that the children in our cohort were making their school selections: "first-preference" (also called immediate acceptance, which is now illegal) and "equal preference" (also called deferred-acceptance, which remains in force today). Theoretically, first-preference is inefficient, because it incentivises parents, when they rank schools, to trade off their true preferences against acceptance probabilities (Ergin and Sönmez, 2006). Moreover it is, in practice, likely to be inequitable, because not all families have the knowledge to engage with this system to ensure a satisfactory school placement (Pathak and Sönmez, 2008). Geographical variation in the allocation system allows us to estimate effects separately for pupils exposed to these two mechanisms. While there is a body of theoretical work studying these mechanisms that suggest that their allocations might differ, there is little empirical evidence that tests whether there are differential consequences for child outcomes.

To apply for a place at a state secondary school in England, parents submit a ranking of schools to the local admissions authority. Parents can list between 3 and 6 schools depending on their local area. High-performing secondary schools tend to be oversubscribed: there are more

applicants than places available. In the case of oversubscribed schools, places are rationed based on a set of published criteria, including whether the applicant had special educational needs, has older siblings at the school, and the family's distance from the school. Faith will be considered for religious schools. If a pupil cannot be allocated to any of their nominated schools (often because the parents do not list enough schools), they are assigned to a school with spare capacity. We study the consequences of not attending a first-choice school, henceforth termed "preferred" school.

We employ data from the Longitudinal Study of Young People in England (LSYPE, also known as *Next Steps*), a nationally representative birth cohort study which tracks the lives of a cohort of young people in England through secondary secondary until age 25 years (Calderwood, 2018). This data is confidentially linked to the National Pupil Database (NPD), a database of administrative education records. From this data we can characterise the attributes of schools, and track the achievement of the LSYPE pupils in high-stakes examinations. An ideal study design would involve using administrative data on the parental rankings of schools, linked to complete data on admissions probabilities, in order to model the admissions process and identify the outcomes of pupils on the margins of acceptance to their preferred school. We do not have this ideal complete administrative parental preference and admissions data. Rather, we have reports of whether a pupil is attending their preferred school. This implies that strong assumptions must be imposed for our results to have a causal interpretation. We rely on a selection-on-observables strategy using a combination of regression-adjustment and geographical matching for estimation.

The validity of a selection-on-observed variables assumption relies, in part, on understanding the process governing assignment to treatment. In applications where little is explicitly known about this selection process, a selection-on-observed variables assumption has the potential to be less credible for causal inference compared with study designs involving quasi-experimental variation. In this application, however, we do have information on the rules governing selection into treatment (admission to a preferred school), which informs our choice of covariates and makes this assumption more tenable. Our choice of covariates is informed by

both the school admissions criteria and the literature on parental preferences. We have detailed information on the variables which determine school admission, proxies for parental preferences and engagement with the school choice system, socio-economic status, prior ability of the child, and variables characterising the density and quality of the local choice set of schools. One important advantage of our linked survey data is that we have information on important admissions-relevant variables—religion and presence of older siblings—which are absent from the available administrative data in England. To support our findings, we pair our analyses with tests of sensitivity to unobserved selection (Oster, 2019). However, we nonetheless recognise that strong assumptions are required for a causal interpretation of our estimates and that this is a key limitation of this paper—especially compared with alternative identification strategies exploited in this literature such as distance-based regression discontinuity design or school choice lotteries.

We use Ordinary Least Squares (OLS), and then a two-stage combination of geographical matching and regression adjustment, to estimate the average effects of failing to gain a place at a preferred school. For the matching procedure, we match each pupil who missed out on a place at their preferred school (the treatment group) to a set of control pupils who attend schools in close proximity to the treated pupil's home. Using this geographically matched sample, we use regression-adjustment to estimate the average treatment effect of missing out on a preferred school, where the contributions of the control observations are weighted in inverse proportion to their school's geographic distance from the treated pupil.

The total effect of missing out is informed by both between-school variation (e.g., school quality differences) and within-school variation (a pupil-specific "match effect"). We aim to explore these two channels by adding specifications which include school-level attributes, and separately, school fixed effects, to assess whether attending a non-preferred school generates poorer than average outcomes while holding school quality and other observed differences fixed. School choice is often promoted via the lens of improving child outcomes through competition on school effectiveness. However, choice in itself has the potential to improve outcomes, both individual and at the sector-level, if gains can be made from matching pupils to the right school

for them. Whether a given school is a good match for a child may be based on factors potentially unrelated to headline measures of "quality", such as anti-bullying policies or support for students with additional needs.

Our main findings show, at best, only weak evidence for detrimental effects of missing out a preferred school on high-stakes examination scores. We find no significant effects on gaining 5 or more 'good' passes in examinations taken at age 16 years (GCSEs), and a statistically significant, but small in magnitude, reduction of 3-5% in the total grade points achieved at this level. In terms of longer run outcomes, we fail to detect any effects on university attendance or staying on after the age of compulsory schooling. The results show small detrimental effects (1-2% reduction) on income at age 25 years, but sensitivity analyses indicate that this figure is not robust to unobserved selection. We do find robust evidence of negative effects on mental health in early adulthood, of about one-fifth of a standard deviation; a 5 percentage point increase in the probability of having a child by age 25 years; and an increase in the probability of being a current smoker at age 25 years of 7-9 percentage points. The findings for mental health and smoking persist after controlling for a comprehensive set of school attributes, and school fixed effects, respectively. These findings suggest that there are direct negative effects of attending a non-preferred school, over and above differences due to variations in school quality and other attributes. This is consistent with parents having information on what is a good school "match" for their child, and the inability to fulfil these preferences generating poorer outcomes—especially mental health and smoking rates in early adulthood.

Our subgroup analyses show little evidence for differences by ethnicity. The exception is mental health in early adulthood, where the consequences of missing out are larger for pupils of White ethnic origin, compared with Black, Asian and Minority Ethnic (BAME) pupils. The findings for differences by school allocation mechanism design are more clear, and show that the negative consequences of missing out are more pronounced in local areas which used a more manipulable mechanism ('first-preference first', FPF) for allocating school places. This mechanism is now illegal because it rewarded strategic behaviour by parents in ranking schools, giving rise to a concern that this system advantages families who could successfully engage

with this complexity. Conversely, under an 'equal-preference' (or 'deferred acceptance', DA) system, parents can (in theory[1]) do no better than reveal their true preferences over schools, and playing strategically yields no advantage. In FPF areas, we do find detrimental effects on short-run academic attainment and staying on in school after age 16 years, in addition to income at age 25 years. The findings are similar to Abdulkadiroğlu et al. (2017), who find test score and graduation rate improvements from switching to a less manipulable school choice mechanism in Boston.

Overall, our findings reveal only weak evidence for negative effects of missing out on a preferred school on short-run test scores, with null to small effect sizes—except in those areas which used a manipulable mechanism to assign school places, where we find more pronounced detrimental effects. These results suggest that the nature of the school choice process and assignment mechanism, and how parents understand and engage with it, are important for ensuring equal access to satisfactory schools. The results also demonstrate that the school a child attends shapes more than just test scores: we find that missing out on a preferred school is associated with poorer life outcomes in a broader sense, including mental ill-health and smoking rates.

2 Related literature

There are a number of studies that combine market design theory with modern methods for policy evaluation to assess how missing out on a preferred school impacts outcomes (Pop-Eleches and Urquiola, 2013; Cullen et al., 2006; Dobbie and Fryer, 2014, 2015; Abdulkadiroglu et al., 2017a; Ovidi, 2020). Several of these papers focus on peer effects as a possible channel (Abdulkadiroglu et al., 2014; Hoekstra et al., 2018). Nonetheless, the majority of these studies do not find any causal effect of attending a preferred secondary school in academic outcomes. This puzzle could have a number of possible explanations. First, many of these studies use a compelling empirical design—comparing the outcomes of pupils who "just" got into a school compared with those who just missed out, either side of a threshold based on a continuous test

[1]Constrained list lengths induce some degree of strategic behaviour in practice.

score. On this margin, there are at least two possible offsetting effects. First, pupils who just get in will be at the bottom of the ability rank in a top tier school, whereas those who miss out will be at the top of the ability rank. Rank has been documented to have a direct effect on outcomes (Elsner and Isphording, 2017). This issue is less relevant for our results, as for most schools in England selection on ability is explicitly prohibited—meaning we escape the confounding effect of ability rank faced by other studies, and our estimates are more generalisable across the ability distribution. Second, parents and families may increase their effort invested in their child if they miss out, again exerting an offsetting effect (Pop-Eleches and Urquiola, 2013).

More generally, parents may not choose schools based on school effectiveness; either because they are not aware of the effectiveness of different schools, or they mistake the quality of the intake mix for school effectiveness (Abdulkadiroglu et al., 2017b), or they value something else. Parents may value other school attributes, such as "elite" status of the school, diversity in the school, or expected skill improvements on other margins, such as non-cognitive skills. Indeed, choosing a school based on considerations other than test scores can be a rational choice, which might show up in a wider set of outcomes, such as well-being and mental health, and longer-term outcomes. For instance, Beuermann and Jackson (2020) provide compelling evidence that while attending one's preferred school in Barbados does not improve short-run test scores, it does reduce early pregnancy and enhance employment rates among women.

There are number of papers looking at the effects of attending a "good school" in the UK. However, most are focused on attendance at academically selective "grammar schools", which have largely been phased out, and currently comprise only about 3% of pupils (see Online Appendix A for a description of school types). Studies looking at the effects of gaining a place at an academically selective "grammar" schools typically find positive effects on secondary school educational tests scores and years in school. The effects on other outcomes are more mixed: Clark and Del Bono (2016) find large positive effects on income and wages, and reduced fertility among women, but no similar effects among men, in a sample from one district in England. Pastore and Jones (2019) study the effects of attending academically selective grammar schools in a nationally representative sample, and do not detect any effects on adult labour market out-

comes or measures of health and well-being. There is little evidence in the English literature studying the effects of attending a preferred secondary school. One exception is Ovidi (2020), who estimates the causal effects of missing out on a place at a preferred state primary school in London, focusing on school-pupil match effects. Exploiting quasi-experimental variation from the centralised admission system, the findings show positive effects on student achievement from attending a preferred school while holding school quality fixed. These results suggest that attending a preferred school can generate better-than-average outcomes, over and above the effects of school quality.

3 Setting

"School choice" is defined as policies which allow parental preferences to influence which school their child will attend (Cantillon, 2017). The central problem in designing a school choice system is to allocate pupils to schools, while accounting for parental preferences, school capacities and wider policy objectives (e.g. diversity in schools). One motivation for school choice focuses on the importance of respecting parental preferences, and allowing them to choose from a diversity of school types. A second motivation, not necessarily implied by the first, is to induce competition between schools with the aim of improving school standards. Both of these motivations have been referenced by different UK governments in relation to school choice policy. In England, features of parental "choice" and a "quasi-market" in compulsory schooling started to emerge from 1988, formalising the role of parental preferences and tying funding to enrollment numbers.

The allocation of pupils across schools is relevant for several reasons. While many changes in the operation of the school choice system have occurred over recent decades (West et al., 2011), the share of children being offered a place at their first choice school has remained static at about 84%. Taking the stock of "good schools" as fixed, this statistic has both equity and allocative efficiency concerns at its core. First, if particular segments of society are more likely to miss out on a place at a preferred school, this raises an equity concern about equal access

to good schools. In England, there are large differences in attending a good school by ethnic group, which are unlikely to be explained by differing preferences (Weldon, 2018). A second issue is potential school-pupil match effects. If attending a preferred school generates positive outcomes, over and above the effects of observed school attributes, this would suggest the existence of pupil-specific match effects. If gains can be made by matching the right pupils to the right school, the allocation of pupils to schools has potential to improve outcomes.

Generally speaking, the process of allocating school places in England works as follows: parents submit a ranking of their preferred schools to the local government. They can list between 3 and 6 places depending on their local area. For oversubscribed schools, the allocation of places is prioritised based on a set of criteria ("school priorities"). In 2000/01, when the LSYPE cohort were applying for admissions, these criteria were similar to today's setting, including: whether the child has exceptional "medical or social needs"; whether the child has an older sibling at the school; by distance from the school as the tie-breaker [2]. For religious schools, faith is often considered. Once the families' preference rankings have been submitted, a procedure is required to allocate school places, given school capacities and priorities. There are a number of approaches for doing this. The two used in 2000 were: the *first-preference* (FPF), algorithm (also termed the "immediate-acceptance", or Boston, mechanism); and the *equal-preference* algorithm, also termed "deferred-acceptance" (DA).

The FPF approach proceeds in rounds. First, each school starts by assigning places to pupils *who put that school as their first choice*, based on the ranking determined by the school's admission criteria—until either there are no places left, or all students who put that school as first choice are assigned. The next round conducts the same process for pupils who put that school as their second-choice, and so on. This approach maximises the share of families getting their first-choice school. The drawback of this approach is that families who are not offered a place at their first preference school can also be rejected by their second preference school in favour of a child of parents who placed that school first, despite living further away. If families put

[2] In 2000, it was not required to give "looked-after" children first priority. From 2006, it was a statutory requirement that such children in care should be given top priority in the event of a school being oversubscribed, and in 2008, almost all schools (99%) had an admissions criterion relating to children in care compared with 2% in 2001 West et al. (2011).

their true preferred school as their first choice, but where they have a very low chance of acceptance, they may not only fail to secure a place at their preferred school, but also fail to be placed at any school that they deem acceptable. This means that families have an incentive to trade-off their true preferences against their probability of acceptance, in order to increase the chance of gaining a place at a satisfactory, but not most preferred, school. 41% of schools used a 'first-preference' approach to offer places for the September 2001 intake (West et al., 2004).

This practice of using preference order as a priority was prohibited in the School Admissions Code 2007: [schools must not] *"give priority to children according to the order of schools named as preferences by their parents, including 'first preference first' arrangements;"* (DfES (2007), 2.16 (b)); *"the 'first preference first' criterion made the system unnecessarily complex to parents"* and *"forces many parents to play an 'admissions game' with their children's future."* (DfES (2007), Foreward, p.7.). In the nationally representative survey conducted in Flatley et al. (2001), about 25% of parents indicated that they took into account the nature of the oversubscription criteria when considering which schools to apply to. Parents who were familiar with the oversubscription criteria in their preferences tended to be more highly educated, owner occupier and of white ethnic origin.

In contrast, the DA approach does not use the preference ranking as a way for schools to rank applicants. In this method, parents submit their preference ranking over schools. The families' application is considered at each school that they listed as a preference without reference to where the school was ranked by parents. Places are offered by ordering applicants by the oversubscription criteria, such that priority goes to families who meet these criteria to the greatest extent. This may lead to a parent whose child does not meet the criteria of their first preference school being offered a place at a nearby school for which they have nominated as a second preference. Because proximity is almost always the tie-breaker, one effect is to confer advantage to families who live near to several popular schools.

The LSYPE cohort were applying to secondary schools to start in September 2001. Admissions were similar in principle to today, but rather less regulated and subject to less stringent reporting standards. The legal guidance on admissions at this time was 2001 Schools Admis-

sions Code (DfEE, 1999), enacted on 01 April 1999, and in force for admissions from September 2000 onward. The Admissions Code is a document providing guidance on best practice in admissions. The 2001 and subsequent editions of the School Admissions Code contain descriptions of the primary legislation from which they stem, extensive guidance, and reference to statutory responsibilities which must be met, but are not themselves a legal document. Schools and admission authorities were required to "have regard to" the indications in the Code. The 2001 Code encourage increased uptake of common admission processes: *"LEAs should consider, with other admission authorities, having coordinated admission arrangements - including standard application forms and common timetables - for all schools"* (DfEE (1999), para. 3.9). West et al. (2004) collected survey evidence to explore the use of various admission criteria at this time, and identified older siblings at the school (96%), distance from home to the school (86%), medical/social need (73%), catchment area (61%) and 'first preference' (41%) as the most common admission criteria practices in the presence of oversubscription. We do not have data on the exact admission arrangements used in each school or admission authority at this time, therefore it is not possible to precisely model the admissions process or to rely on this type of modeling for causal inference.

There are a range of sources of uncertainty about whether one will be admitted to a particular school, such that while parents can increase their chances of admission for their child, for instance by moving closer to a school, the probability of admission remains uncertain and difficult to predict in advance. The changing interaction of sibling, distance and the other criteria makes it difficult for parents to predict their chances of gaining a place at a particular school. Distance to school is itself an unpredictable criterion, because families do not know in advance whether or not they live close enough to a school in any given year. Although some individual schools and Local Authorities publish information over several years about the 'cut-off point' for distance, typically it is not clear how near 'near enough' will be the next year (Williams et al., 2001).

4 Data

Longitudinal Study of Young People in England

We employ data from the Longitudinal Study of Young People in England (LSYPE, or *Next Steps*). This is a nationally representative birth cohort study which tracks the lives of a cohort of around 15,000 young people in England who were born in school year 1989/90 (Calderwood, 2018; University College London et al., 2017, 2018). The study begins when the children are in Year 9, the third year of secondary school, and follows them until they are aged 25 years (annually until age 19/20, at which point there is a break until the age 25 wave). The LSYPE contains detailed information on school choice, including whether the child attends their preferred school, as well as family background, experiences in school, and crucially labour market and university outcomes.

The sampling frame comprised all pupils attending publicly-funded (maintained) schools, private (independent) schools and pupil referral units (for children with behavioural difficulties) in England in February 2004. LSYPE used a two-stage sampling design (first sampling schools, then pupils within those schools), with oversampling of ethnic minority groups to ensure sufficient sample sized for subgroup analyses. The first seven waves of the study were funded by the Department for Education (DfE), commissioned to explore the factors shaping educational attainment and transitions out of compulsory schooling. The final wave, at age 25 years, was funded by the Economics and Social Research Council, and management was transferred to the Center for Longitudinal Studies, now with a broader remit to explore wider aspects of the transition to the labour market.

The LSYPE records have been confidentially linked to the National Pupil Database (NPD), a database of administrative records on schools and pupils. The NPD contains individual test scores, individual characteristics, and school-level attributes (e.g., socio-economic and ethnic mix). From this data we can track the achievement of the LSYPE pupils in national standardised examinations from age 11 years through ages 17/18 years.

Treatment variable

In the LSYPE survey, each parent is asked whether their child's current school was their first-choice school. From this, we construct the variable *Missed out* which is equal to one if the current school is not the first-choice school, and zero otherwise. The proportion of families attending their first-choice school is 85%, very similar to administrative figures of 84%.

Outcomes

We examine a number of outcomes, both "short-run" academic outcomes, as well as broader long-run outcomes. We first consider variables summarising attainment at secondary school. The English school system implements national standardised testing at regular intervals ("Key Stages (KS)"). In this paper we use results from Key Stage 2 examinations at age 11 years, Key Stage 4 results assessed at ages 15/16 years (General Certificate of Secondary Education and equivalents), and Key Stage 5 (academic and vocational qualifications assessed through ages 17/18 years). In Online Appendix A, we provide further detail on the nature of these qualifications and the English school system. The outcomes variable comprise: a binary variable indicating whether the pupil passed at least five Key Stage 4 GCSE subjects, or GNVQ equivalents, at grade C or above ("5+ A*-C"); the total grade points attained in Key Stage 4 examinations ("KS4 points"); a binary variable indicating whether the pupil stayed on after age 16 years to study for any Key Stage 5 qualification, which could be vocational, academic, or a mix of the two ("Stayed on"). We then consider whether the cohort members go on to attend university, at any time until age 25 years ("Attends Uni"). In terms of longer run outcomes, we look at mental ill-health measured using the GHQ 12 point scale ("Mental ill-health"), where a higher score indicates poorer mental health. We also look at fertility: Does the young person have a child by age 25 years ("Fertility")? We consider the log of income measured at age 25 years ("ln(income)"). Finally, we examine whether the young person currently smokes at age 25 years ("Smoker").

To classify Local Education Authorities (LEAs) by use of deferred-acceptance or first-preference mechanism, we use data collected by Coldron et al. (2008) and Pathak and Sönmez (2013).

Coldron et al. (2008) surveyed the admissions policies of English Local Authorities for pupils entering the 2006 academic year, indicating whether schools in an area used a 'first-preference' approach (30% of schools used FPF in 2006, compared with 41% in 2001). To identify the few areas which had switched from FPF to DA before 2006, and hence while coded as DA in 2006, would have used FPF in 2001, we use the data in Pathak and Sönmez (2013), who compile a list of the year in which areas switched from FPF to DA. An important caveat is that, in 2000, when our cohort were applying to schools, there was variation both between and within LEAs in use of first-preference as a mechanism, and it is not possible to completely characterise an area by type of admissions policy. While for the majority of schools (i.e., government-controlled maintained schools), the Local Authority coordinated and implemented admissions, a minority of schools, especially church schools, were their own admissions authority and may not necessarily have coordinated to have the same mechanism as the rest of that Local Authority. Therefore, this area-based classification is inevitably fuzzy, but still gives an indication of areas where at least some places, if not all, were allocated by FPF.

4.1 Sample descriptives

Figure 1 characterises the geographical distribution of the proportion missing out on their preferred school, and the geographical distribution of the final sample. Evidently, the probability of missing out is higher in London, where the school market and population is more dense. The (survey-weighted) final sample is relatively evenly distributed across regions, with a slightly higher concentration in the South East. Appendix Figure A.1 maps the geographical distribution of the 80th percentile of distance from home to school; the mean distance from home to school; and the proportion of sample members in urban areas. The statistics and sample sizes underlying these maps are reported in Online Appendix B, Table WA.1.

[Figure 1 about here]

892 schools were selected for the study, with 658 schools responding. Among responding schools, the LSYPE achieved pupil-level cross-sectional responses rates ranging from a mini-

mum of 51% (in wave 8), to a maximum of 92% (in Waves 3 and 4). As reported in Appendix Table A.1, there were n=15,770 productive responses at Wave 1, of which n=7,707 partook at Wave 8; and n=5,426 partook in all 8 Waves. The analytical sample is restricted to study members who remain until Wave 8 and have non-missing data on all covariates and outcomes used in the analyses, which yields a final sample of 4,931 study members attending 531 schools. In this sample there are an average of 9.28 pupils per school, with a standard deviation of 3.71.

Hence, between Wave 1 and our final sample there is substantial attrition. This type of attrition could pose a problem for our study if, conditional on observed covariates, survey drop-out is related to the true relationship between missing out and later outcomes, which is our parameter of interest. While it is not possible to completely rule out this possibility, because it relates to unobserved selection, we do investigate the issue in our analyses by: (a) comparing the final sample to the full sample observed in Wave 1; (b) comparing the final sample to national averages where possible; (c) performing all analyses with the Wave 8 survey weights, which account for differential (school and pupil) non-response and attrition. The Wave 8 weights incorporate a comprehensive set of variables from previous Waves in adjusting for drop-out, including: basic family background and demographic characteristics, as well as the young person's characteristics at later Waves, such as housing tenure, labour market and education status, risky behaviours, interview month and interview mode.

Table 1 shows mean differences in key covariates by whether a pupil got into (control group) or missed out on (treated group) their preferred school. Two samples are shown, one is the "full sample" of all cohort members who responded at Wave 1 (n=15,770). The second is the estimation sample ("final sample") which comprises those who remain to Wave 8 and have non-missing data on all the covariates and outcomes (n=4,931). Overall, the distributions of key covariates are fairly similar in the full and final sample. Comparing by treatment status, the key pupil characteristics which over-represented among those gaining a place at their preferred school include having resident older siblings, being of White ethnic origin, being an owner occupier, and being of a Christian faith. There is little difference by prior attainment (Key Stage 2), and only small differences by parental education. This is consistent with most studies based

on more recent administrative preferences data, which tend to find the strongest patterning in admissions by ethnicity, rather than by socio-economic status (Weldon, 2018; Burgess et al., 2019).

[Table 1 about here]

Online Appendix B, Table WA.2 reports a comparison of national averages with the final sample. The population data is from an extract of the 'National Pupil Database' Schools Census, linked to attainment data, for the entire population of England for the years corresponding to the LSYPE cohort who were Year 9 in 2003/04. The final LSYPE sample is identical to the population on gender; and similar for English as First Language, gaining 5 or more A*-C grades and Key Stage 4 grade points. However, the final LSYPE sample does have a slightly higher share of White British pupils (88% compared to 81% in the population data), and a slightly lower share of pupils eligibility for Free School Meals (12% compared to 16% in the population data). The analyses are weighted throughout using the survey weights. We have also reproduced the results without the weights, which does not alter the conclusions. While it is not possible to completely rule out this possibility of non-ignorable attrition, because it relates to unobserved selection and data, these investigations provide some reassurance that our findings are not driven by differential survey drop-out.

5 Empirical strategy

Which school a child attends is shaped by parental preferences and admission constraints. Our aim is to estimate the average outcomes associated with missing out on a place at a first-choice school. There is some conditionally random variation in whether families miss out a preferred school, because the distance cut-offs for admission to schools are unpredictable to parents in advance, and they change year-on-year. The children of families who look observationally very similar could get in or miss out on a school due to living slightly different distances from a school, in a way which is unrelated to child outcomes. However, because we do not have com-

plete data on parental preferences and admissions probabilities, we cannot employ a regression discontinuity, or similar design, to explicitly compare pupils on the margin of acceptance to a preferred school. Rather we must impose more stringent assumptions for a causal interpretation, in particular a selection-on-observed variables assumption—that the child's potential outcomes are conditionally independent of missing out on a preferred school. We condition on variables determining school admission probability, proxies for parental preferences, detailed information on socio-economic status, prior ability of the child and the nature of the feasible choice set of schools, and we explore the results of robustness checks to probe the sensitivity of our estimates to this assumption.

We estimate the parameters of a parametric linear model using Ordinary Least Squares (OLS) with a specification as detailed in Equation 1. Y_{ijk} denotes the outcome under consideration, for pupil i, in school j in Local Authority k. c_o denotes a constant. D_{ijk} is the treatment indicator: equal to 1 for pupils are not attending their first-choice school, and 0 for pupils who are attending their first-choice school. X_{ijk} is a vector of covariates, LA_k are Local Authority fixed effects and ϵ_{ijk} is an idiosyncratic error term.

$$Y_{ijk} = c_o + \beta D_{ijk} + X'_{ijk}\gamma + \text{LA}_k + \epsilon_{ijk} \tag{1}$$

Second, we incorporate detailed geographical information in a two-stage procedure.[3] Nearby pupils who gained a place at their preferred school are informative about the counterfactual outcome of those who missed out on their preferred school, controlling for the socio-economic and other variables. We match observations based on geographical distance, and use regression adjustment on this matched sample, to estimate the average effect of missing out on a preferred school among those who missed out, the Average Treatment effect on the Treated (ATT).

Specifically, we match each pupil who missed out on a place at their preferred school (treated) to a set of pupils who gained a place at their preferred school (controls)[4]. Potential control pupils

[3]Implemented using the Stata user-written software *kmatch* (Jann, 2017).

[4]Using those who miss out as the treatment group, rather than the other way around, ensures a larger pool control observations, as those who miss out are the minority group.

are identified based on the straight-line distance between the population-weighted centroid of the treated pupil's home Lower Super Output Area (LSOA), and the post-code centroid of the school attended by a control pupil. The idea is that these control pupils attending a preferred school within a close proximity to the pupil who missed out are informative about the counterfactual outcome of the treated pupil—had they got into a potential preferred school.[5] The bandwidth within which control pupils are considered candidate donors is set at 4.2km, the 80th percentile distance from home-to-school in the sample.[6]

Using this geographically matched sample, the second-stage uses regression-adjustment to estimate the Average Treatment effect on the Treated of missing out on a preferred school. The contributions of the control pupils are weighted in inverse proportion to their straight-line distance from their school to the treated pupil's home, using an biweight kernel. The purpose of this two-stage approach (henceforth termed 'distance-matching') is to ensure the sample is balanced on geography, such that less reliance is put on capturing geography correctly in the functional form of the regression. Residential location is a choice which may also act as a sufficient statistic for unobserved variables—unmeasured wealth or preferences for example.

There are a range of factors shaping parental preferences over schools, and the probability of admission to a school. The variables we include in X'_{ijk} can be loosely grouped into three (overlapping) categories: admissions-relevant variables; variables proxying for preferences over schools; and socio-economic variables. To capture factors which shape admissions probabilities, we adjust for the following variables. First, the presence of resident older siblings in the household (interacted with region), and an indicator for whether the child has a certificate of Special Educational Needs (a small group of high need pupils who may qualify for admission under the "medical and social needs" category).[7] The density and quality of schools available

[5]The Centre for Longitudinal Studies has kindly derived the relevant straight-line distances from cohort member home Lower Super Output Area population-weighted centroid to the other schools in the area attended by LSYPE cohort members in a secure environment for this project. A LSOA contains between 400 and 1200 households.

[6]We have also confirmed the robustness of the results to bandwidths selected using data-driven bandwidth selection procedures, such as cross-validation and other "optimal" selection procedures. These bandwidths are typically larger than 4.2km. In Online Appendix D, Table WA.6., we report results using the default option in Jann (2017), which is based on the bandwidth selection methods developed in Huber et al. (2015) which select the bandwidth as 1.5 times the 90% quantile of the (non-zero) distances in pair matching with replacement.

[7]At the time when our cohort were applying for school, it wasn't required by law to admit pupils with a Special Educational Needs (SEN) statement as it is now. It was also not required to admit pupils in state care, as this was

to families varies by area, which correlates with application strategy, the probability of getting into a preferred school and subsequent outcomes. Using data on the population of schools (from Edubase), we construct a variable characterising the local market for schools; a distance-weighted share of good-, very good-, or excellent-rated schools by OFSTED (the Office for Standards in Education is the regulatory monitor of school quality based on school inspections), within the 80th percentile of home-to-school distances travelled in the pupil's Local Authority. We have data on whether the child has been in state-care; however, the proportion is so small that it is not reliable to include this variable.

Socio-economic and demographic variables are included to capture indirectly aspects of both admission probability and preferences, as well as being associated with child outcomes. For example, family income shapes a family's ability to move closer to their preferred school. In this category we control for (i) the child's prior academic test score—their average points score from their KS2 tests completed at the end of primary school; average KS2 points score from child's primary school; the child's month-of-birth; family income, housing tenure, parental education and occupational social class (NS-SEC).

We adjust for variables which proxy for the families' preferences (the type of schools they prefer). In this category we have: religious denomination of parents; ethnicity of the child; whether English is the main language spoken at home and the information used to choose a school. Religion and ethnicity capture parental preference for attending religious schools, and a school with children of the same ethnicity. The LSYPE asks about school preferences. One question asks what information the parents used in their school application decision, listing options such as "looked at league tables on the internet", used "Local Authority brochure", and so on. We use Principal Components Analysis to condense these responses into one summary variable (the first principal component), intended to capture the extent to which parents seek out information about schools.

Another important issue is the interaction between the state-school non-selective system with

introduced as explicit advice (but not a statutory requirement) in the Education Act 2002 and associated School Admissions Code 2003. However these children could be given priority by schools based on the commonly used "medical and/or social needs" priority.

the 'outside options': publicly-funded selective grammar schools, and private (Independent) schools. Families' choices will depend on the availability of these alternative options, generating variation in the association between missing out and outcome by area. In the LSYPE, around 4% of pupils attend private schools (the weighted estimate is 7%), and 3% of the sample attend grammar schools (the weighted figure is about 3%). The main specifications control for whether the Local Authority has any grammar schools. Regarding the private school sector, we control for the distance from home to the closest private school using external data on the full population of schools.

To summarise, the covariates included in the main specifications are: child's gender, month-of-birth, ethnicity (White British; Indian, Pakistani or Bangladeshi; Black Caribbean or Black African; mixed-ethnicity or other minority ethnicity), an indicator for any resident older siblings, whether English is the main language spoken at home, Free School Meal eligibility (an indicator of low income); whether the child has special educational needs, housing tenure, family income, parental occupational social class, parental education, parental religion, Principal Components Analysis summary of information used to choose schools, presence of local grammar school, distance to closest private school, distance-weighted share of good, very good or excellent rated schools within the 80th percentile of home-to-school distances travelled in that pupil's Local Authority, and Local Authority fixed effects.

Robustness

While we have carefully chosen set of control variables based on institutional rules determining selection, we cannot rule out the possibility of bias from unobserved variables. To explore the sensitivity of our estimates to such selection, we report a diagnostic statistic developed in Oster (2019), alongside the OLS estimates. The premise is that the degree of selection on observed variables can be informative about the potential degree of selection on unobserved variables. In what follows, we outline the features of this general framework (with subscripts omitted) as far as they are relevant for this analysis, with the full details available in Oster (2019). In Equation 2, Y denotes the outcome of interest; D denotes the treatment variable; β is the coefficient of interest; X_1 contains all observed control variables multiplied by their coefficients; X_2 contains

unobserved control variables multiplied by their coefficients; ϵ captures an idiosyncratic error, for instance measurement error in Y, which is unrelated to the treatment D. Hence, X_2 contains only the omitted variables which are related to D.

$$Y = \beta D + X_1 + X_2 + \epsilon \tag{2}$$

The idea that selection on observed variables may act as a guide to selection unobservables is formalised by defining the parameter δ—the coefficient of proportionality—to be such that the following equality holds:

$$\frac{cov(D, X_1)}{var(X_1)} = \delta \frac{cov(D, X_2)}{var(X_2)} \tag{3}$$

δ denotes the degree to which selection on unobserved variables relates to selection on observed variables. For example, if $\delta = 1$, observed and unobserved variables are equally important in explaining selection; if $\delta = -1$, observed and unobserved variables are equally important in explaining selection, however the selection effects have opposite signs.

Using the Stata program psacalc (Oster, 2016), one can calculate the size of δ required to drive the treatment effect β to zero. In other words, how large would the selection on unobserved variables, as a proportion of selection on observed variables, need to be to overturn the result? A higher value of δ would suggest it is less plausible that unobserved selection explains the result. A rule of thumb suggested in Oster (2019) is to assess whether δ exceeds 1, with values above 1 suggesting conclusions may be robust to selection on unobservables. To derive this threshold, Oster (2019) reviewed a set of articles in highly ranked economic journals from 2008-2010. Calculating a bias-adjusted treatment effect, assuming equal proportional selection ($\delta = 1$), lead to 90% of the randomized, and 50% of the non-randomised, results to continue being statistically significant. This motivates using $\delta = 1$ as a rule-of-thumb cut-off consistent with a degree of 'acceptable' selection.[8]

[8]The assumptions underlying the calculation of δ can be varied. In particular, the researcher can vary the

Channels

Attending a preferred school is a bundle of treatments: the preferred school many have smaller class sizes, better teachers, better peers, or some idiosyncratic pupil-level factors which mean a school is a good fit for a particular child. Hence, the coefficient β in Equation 1 is informed by comparisons between pupils, who get in or miss out, both between schools and within schools.

Missing out on preferred school could impact outcomes through two channels. First, the indirect effect of being exposed to different school attributes at a non-preferred school compared with a preferred school. Non-preferred schools on average have lower headline attainment and OFSTED inspection ratings. Second, a direct "match effect". Pupils may do worse than average, holding fixed the given school characteristics, solely due to the fact that it is not their preferred school. In particular, parents choose schools based on headline measures of "school quality", but also on factors specific to their child: parents are likely to have information about their child's suitability for a given school which is unobserved to the researcher.

To disentangle these two channels, we explore two further specifications which control for between-school variation. First, we add a set of school-level controls to Equation 1 to adjust for between-school quality and peer attributes. The control variables are: % gaining 5+ A*-C grades in 2004; KS2-KS4 value-added; OFSTED rating of good or above; % with English as first language; % of pupils White British; Free School Meal band of the school; % with SEN; school type. Second, instead of the school-level controls, we add school fixed effects to account for unobserved school-specific attributes, such that β is now informed only by within-school variation (this specification omits Local Authority fixed effects). Looking at the reduction in the size of the coefficient β in these specifications, compared with that in Equation 1, provides some indication what share of the total effect is mediated via school attributes, as opposed to the remaining variation which could be construed as a direct "match effect".

assumed value of *R-max*, the R-squared from a hypothetical regression of the outcome on the treatment and both observed and unobserved controls. The rule of thumb proposed in Oster (2019) is to set *R-max* equal to 1.3 times the R-squared from a regression of the outcome on the treatment and observed control variables (denoted $\tilde{R}$). Therefore, we report δ based on a *R-max* of $1.3\tilde{R}$.

6 Findings

Descriptive evidence

Table 2, Panel A, shows the raw differences in pupil outcomes by whether they attend their preferred school. There are relatively small, insignificant, differences in key academic outcomes—for instance, pupils who miss out have fewer points at GCSE, but a similar probability of having 5 or more 'good' GCSE passes. In terms of longer run outcomes (mental ill-health, income, fertility and smoking) there are large, and statistically significant, differences. Table 2, Panel B, reports the attributes of schools. Preferred schools tend to have a higher share of pupils whose first language is English, higher OFSTED inspection outcomes and higher value-added. They also have a lower share of FSM-eligible pupils (as measured by FSM Bands, a grouping variable which categorises schools by the share of pupils eligible for free school meals from Band 1, less than 8%, to Band 5, more than 50%).

[Table 2 about here]

Estimation results

Turning to the regression findings, Table 3 reports two sets of estimates characterising the association between missing out on a preferred school on short run academic outcomes, and then longer run outcomes measured at age 25 years. Panel A reports estimates from OLS. Panel B reports estimates from the distance-matching procedure: the Average Treatment Effect on the Treated and an estimate of Y_o, the model-adjusted outcome mean in the control group. Because we do not have exogenous variation in attending a preferred school, we use the method that has been developed in Oster (2019) to explore the sensitivity of estimates to unobserved selection. This method generates estimates of the parameter δ, which measures the degree of selection on unobserved variables proportional to selection on observed variables. δ is reported alongside the OLS estimates, with a higher value indicating increased robustness to selection-on-unobservables. Following Oster (2019), we assess whether δ exceeds 1, as a threshold indicating an "acceptable" level of selection.

Considering the academic outcomes first, the results suggest only a small average effect of missing out on KS4 achievement. There is no significant effect on gaining 5 or more 'good' GCSE passes across either estimation method. While the effect on KS4 points is statistically significant, at at least the 10% level across both specifications, the effect size is small, at a reduction in KS4 points of 3.4% from the control mean in the distance-matched specification (denoted Y_o), and a reduction of 5.1% in the OLS specification. The figures for the probability of staying on in academic or vocational education post-16 years and attending university are not statistically significant.

Turning to the corresponding figures for the longer run outcomes (mental ill-health, income, fertility and smoking), pupils who miss out on attending a preferred school have significantly poorer mental health at age 25 years by about one-fifth of a standard-deviation (17.5% of a standard deviation in the OLS result, and 22.1% in the distance-matched results). Income is reduced by 1-2%, although this estimate has a small value of δ, less than 1, suggesting that this figure is likely to be sensitive to selection on unobservables. In addition to the low value of δ, the estimate of the effect on income is not statistically significant across both the regression-adjustment and distance-matched sample, which together suggests this figure is not robust. The probability of having a child by age 25 years ("Fertility") is increased by 5 percentage points in both specification, and is statistically significant in the OLS specification. The probability of smoking increases by 7 and 9 percentage points in the OLS and distance-matched results, respectively.

[Table 3 about here]

Heterogeneous effects

Table 4 now splits the sample by whether the Local Authority was, at that time, using a deferred-acceptance (DA) or first-preference (FPF) mechanism to allocates places. There is a theoretical suggestion that missing out under FPF may be worse than under DA since in the latter missing out on one's first preference may increase the chance of missing out on your second too. There is evidence for group differences, with statistically significantly larger effect sizes

in the FPF areas, particularly for short-run academic outcomes. For example, considering the outcomes which have statistically significant differences across groups in the OLS results, the effects of missing out in the FPF sub-sample show that missing out is associated with a 9 percentage point reduction in the probability of gaining 5 or more 'good' GCSE passes, compared with a figure of -0.01 for the DA group; a 13 percentage point reduction in the probability staying on in education, compared with a 0.00 percentage point reduction in the DA group; and a 4% reduction in income at 25 years, compared with a figure of 0.00% in the DA group. The difference in the mental health estimates between groups is large, but the difference is not statistically significant at conventional levels. These patterns are also reflected in the distance-matched results.

[Table 4 about here]

These differences could, however, be confounded by other characteristics that differ between areas. Table 5 reports the means of various attributes by treatment status and FPF vs DA area-type. FPF areas tend to be fairly similar to DA areas in terms of school value-added, KS2 attainment and share of pupils with Special Educational Needs (SEN). However, schools in FPF areas a higher proportion of White pupils, a higher proportion with English as first language, and slightly lower OFSTED inspection outcomes. Given that there does not appear to be a clear differential pattern of observed variables by area-type, there is not a clear case for confounding based on observed variables. However, there may remain selection based on unobserved factors, such as area-level variation in school choice strategies. This potential for further omitted variable bias make it difficult to assign a causal effect to this mechanism effect based on this cross-sectional comparison.

In terms of mechanisms, the difference in the distance-to-school between those who get in and miss out is larger in FPF areas. Those who miss out in an FPF area travel 0.13km further than those who get in, compared DA areas where those who get into their preferred school travel 0.08km further. This suggests that, under FPF, pupils who miss out are being allocated a school further from their home, perhaps further than they would prefer to travel and further from local friends. This may be an important factor behind our results, especially if it indicates a poorer

match on other, unobserved, dimensions.

[Table 5 about here]

We have also considered other effect modifiers, informed by the literature. Previous research has found little difference in gaining a place at a first-choice school by socio-economic status, but large differences by ethnicity. For instance, 90% of White British families were offered a place at their first-choice school, compared with 72% Asian and 62% of Black families applying for schools in England in the 2014/15 academic year (Burgess et al., 2019). This motivates splitting the sample by ethnic origin of the child: White British, or Black, Asian or other minority ethnicity (BAME). The only outcome for which the difference in effect size between subgroups is statistically significant across both specifications is mental ill-health, where the effects appear larger among White pupils. This pattern could have several explanations, including differential school preferences by ethnicity, variation in residential sorting, or the fact the White families tend to list fewer schools than ethnic minority families (Burgess et al., 2019). If White parents are not listing enough schools, when they do miss out they may be allocated a school which they would not have chosen at all. However, without further data is difficult to be confident about the specific explanatory mechanism(s). There is less evidence for across the board ethnic differences; we report these results and associated discussion in Online Appendix C, Tables WA.3 and WA.4.

Channels

Table 6 adds school-level controls, and, separately, school fixed effects, to the previous specifications. Our intention here is to assess the extent to which the findings are explained by between or within-school variation. That is, missing out on a preferred school could impact outcomes via an indirect effect of attending a poorer quality school. Alternatively, part of the missing out effects could arise solely from 'attending a school that they do not want to be attending', over and above any between-school quality differences.

The estimates including either school-level attributes, and alternatively school fixed effects, in Table 6 follow a similar pattern of magnitude to those in the initial specification, however

the estimates for KS4 points and fertility become smaller and statistically insignificant.[9] The effects for mental health and smoking remain, which suggests that these results are not entirely explained by between-school differences and the within-school effect has relevance. The size of the coefficients for mental health and smoking increase when adding the school fixed effects compared with OLS. However, this increase is partly explained by compositional effects, as schools which do not have any within variation in the treatment are included in the OLS estimation but do not contribute information to the fixed effects estimates. Re-estimating the OLS figures using only schools with within-school variation in the treatment again produces large statistically significant estimates, but reduces the difference in magnitude between the OLS and fixed effects results (reported in Online Appendix D, Table WA.5). We also see the subgroup differences persist after controlling for school attributes as reported in Appendix Table A.2 and Online Appendix C, Table WA.4 (we do not include school fixed effects in the subgroup analyses to avoid small number of pupils per school within subgroup).

One interpretation of these findings is that the effect of 'being at a school you do not prefer' is especially detrimental to mental health and risky behaviours. These findings may suggest that parents and children know something unobserved to the researcher about their idiosyncratic suitability for a given school.

[Table 6 about here]

7 Conclusions

This paper examines the association between missing out on a place at a first-choice ("preferred") secondary school in England on short- and long-run outcomes. Employing nationally representative cohort data, confidentially linked to administrative records on education outcomes, we compare a range of outcomes between pupils who get into their preferred school and those who miss out. The data follows pupils from the academic year 2003/04, in their third

[9] Estimates using the distance-matching approach with controls for school attributes are reported in Online Appendix D, Table WA.7

year of secondary school, until they are aged 25 years.

We contribute to a growing literature tracing the consequences of attending schools which are preferred by parents, and we provide new evidence on this topic for secondary schools in England. Our empirical strategy leverages features of the institutional setting and the literature on school choice to make a case for a selection-on-observed variables assumption, which we pair with sensitivity tests to selection-on-unobservables. We use regression-adjustment and geographical matching to estimate average treatment effects of missing out on a place at a preferred school. Our findings show at best small negative effects of missing out on a preferred school on short-run test scores taken at age 15/16 years. We do not find evidence for effects on rates of leaving school at age 16 years, the age of mandated compulsory schooling during our sample, or university attendance. In terms of longer run outcomes, we find a detrimental effect on mental health, increased fertility by age 25 years, and increasing smoking rates. The results show small negative effects on income at age 25 years, but this finding was not robust to unobserved selection. The research highlights the importance of considering broader outcomes when assessing the effects of schooling: the mental health effect has with a relatively large effect size of about one-fifth of a standard deviation.

We find that the link with poorer mental health and smoking behaviours at age 25 years persists after controlling for a comprehensive set of school attributes, and separately, school fixed effects. These patterns suggest that there are direct negative effects of being allocated a non-preferred school. That is, in addition to difference in attributes between schools being relevant (i.e., teaching quality), simply being at a school that one does not want to be attending yields negative consequences. One interpretation of these findings is that parents have information unobserved to the researcher on what is a good school "match" for their child, and the inability to fulfil these preferences is associated with poorer outcomes—especially mental health and smoking rates in early adulthood.

During our time period of interest, two mechanisms for allocating school places were used, given preferences, school priorities and capacities. 'First-preference-first' rewarded strategic behaviour by parents in ranking school, giving rise to a concern that this system advantages

families who could successfully play this admissions "game". Conversely, under a 'deferred acceptance' system, parents can (at least, in theory) do no better than reveal their true preferences over schools, and playing strategically yields no advantage. Geographical variation in the allocation system allows us to estimate treatment effects separately for pupils exposed to these two systems. In these areas we find negative effects on short-run academic attainment and rates of staying on in education or training after the age of compulsory schooling, in addition to reduced income in adulthood. These findings suggest that legislation outlawing FPF, enacted in 2007 for 2008 secondary school admissions, may have lead to improved outcomes across multiple dimensions for young people in England.

Overall, our findings have a number of suggestive policy implications. The results show that which school a child attends is important for more than just test scores: missing out on a preferred school is associated with poorer life outcomes in a broader sense, including mental health and smoking rates. These broader measures could fruitfully be considered alongside traditional measures of school quality. Second, the nature of the assignment mechanism, and how parents understand and engage with it, is important for ensuring equal access to schools. The school choice process, and oversubscription criteria which act as the gatekeeper to popular schools, are areas ripe for innovation to ensure fair access to schools.

References

Abdulkadiroglu, A., Angrist, J. D., Narita, Y., and Pathak, P. A. (2017a). Research design meets market design: Using centralized assignment for impact evaluation. *Econometrica*, 85(5):1373–1432.

Abdulkadiroglu, A., Angrist, J. D., and Pathak, P. A. (2014). The elite illusion: Achievement effects at Boston and New York exam schools. *Econometrica*, 82(1):137–196.

Abdulkadiroglu, A., Pathak, P., Schellenberg, J., and Walters, C. (2017b). Do parents value school effectiveness? NBER Working Paper No. 23912.

Abdulkadiroğlu, A., Agarwal, N., and Pathak, P. A. (2017). The welfare effects of coordinated assignment: Evidence from the New York City high school match. *American Economic Review*, 107(12):3635–3689.

Beuermann, D. W. and Jackson, C. K. (2020). The short and long-run effects of attending the schools that parents prefer. *Journal of Human Resources [published online ahead of print]*.

Burgess, S., Greaves, E., and Vignoles, A. (2019). School choice in england: evidence from national administrative data. *Oxford Review of Education*, 45(5):690–710.

Burgess, S., Greaves, E., Vignoles, A., and Wilson, D. (2014). What Parents Want: School Preferences and School Choice. *Economic Journal*, 125(August 2014):1262–1289.

Calderwood, L. (2018). Next steps user guide (second edition). sweep 8 – age 25 survey. Centre for Longitudinal Studies, University College London.

Cantillon, E. (2017). Broadening the market design approach to school choice. *Oxford Review of Economic Policy*, 33(4):613–634.

Clark, D. and Del Bono, E. (2016). The long-run effects of attending an elite school: Evidence from the United Kingdom. *American Economic Journal: Applied Economics*, 8(1):150–176.

Coldron, J., Tanner, E., Finch, S., Shipton, L., Wolstenholme, C., Willis, B., Demack, S., and Stiell, B. (2008). Secondary School Admissions. Department for Children, Schools and Families Research Report RR020, London.

Cullen, J. B., Jacob, B. A., and Levitt, S. (2006). The effect of school choice on participants: Evidence from randomized lotteries. *Econometrica*, 74(5):1191–1230.

Deming, B. D. J., Hastings, J. S., Kane, T. J., and Staiger, D. O. (2014). School Choice, School Quality and Post-Secondary Attainment. *American Economic Review*, 104(3):991–1013.

DfEE (1999). *Code of Practice on School Admissions*. Department for Education and Employment (DfEE), London.

DfES (2007). *School Admissions Code*. Department for Education and Skills (DfES), London.

Dobbie, W. and Fryer, R. G. (2014). The impact of attending a school with high-achieving peers: Evidence from the New York City exam schools. *American Economic Journal: Applied Economics*, 6(3):58–75.

Dobbie, W. and Fryer, R. G. (2015). The medium-term impacts of high-achieving charter schools. *Journal of Political Economy*, 123(5):985–1037.

Elsner, B. and Isphording, I. E. (2017). A big fish in a small pond: Ability rank and human capital investment. *Journal of Labor Economics*, 35(3):787–828.

Ergin, H. and Sönmez, T. (2006). Games of school choice under the Boston mechanism. *Journal of Public Economics*, 90(1-2):215–237.

Flatley, J., Connolly, H., Higgins, V., Williams, J., Coldron, J., Stephenson, K., and Logie, A. (2001). Parents' experiences of the process of choosing a secondary school. Department for Education and Skills Research Report RR 278.

Gibbons, S. and Silva, O. (2011). School quality, child wellbeing and parents' satisfaction. *Economics of Education Review*, 30(2):312–331.

Heckman, J. J., Stixrud, J., and Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3):411–482.

Hoekstra, M., Mouganie, P., and Wang, Y. (2018). Peer Quality and the Academic Benefits to Attending Better Schools. *Journal of Labor Economics*, 36(4):881–884.

Huber, M., Lechner, M., and Steinmayr, A. (2015). Radius matching on the propensity score with bias adjustment: tuning parameters and finite sample behaviour. *Empirical Economics*, 49(1):1–31.

Jann, B. (2017). KMATCH: Stata module module for multivariate-distance and propensity-score matching, including entropy balancing, inverse probability weighting, (coarsened) exact matching, and regression adjustment. Statistical Software Components S458346, Boston College Department of Economics, revised 30 Jul 2019.

Mendolia, S., Paloyo, A. R., Walker, I., and Mendolia, S. (2018). Heterogeneous effects of high school peers on educational outcomes. *Oxford Economic Papers*, 70(3):613–634.

Oster, E. (2016). Psacalc: Stata module to calculate treatment effects and relative degree of selection under proportional selection of observables and unobservables. Statistical Software Components S457677, Boston College Department of Economics, revised 18 Dec 2016.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37(2):187–204.

Ovidi, M. (2020). Do parents know better? Parental choice and students achievement in London primary schools. Unpublished working paper.

Pastore, C. and Jones, A. M. (2019). Human capital consequences of missing out on a grammar school education. Health Econometrics and Data Group WP 19/08, University of York.

Pathak, P. A. and Sönmez, T. (2008). Leveling the playing field: Sincere and sophisticated players in the Boston mechanism. *American Economic Review*, 98(4):1636–1652.

Pathak, P. A. and Sönmez, T. (2013). School admissions reform in Chicago and England: Comparing mechanisms by their vulnerability to manipulation. *American Economic Review*, 103(1):80–106.

Pop-Eleches, C. and Urquiola, M. (2013). Going to a better school: Effects and behavioral responses. *American Economic Review*, 103(4):1289–1324.

University College London, UCL Institute of Education, and Centre for Longitudinal Studies (2017). Next Steps: Sweeps 1 and 8, 2001 and 2016: Geographical Identifiers, 2001 Census Boundaries: Secure Access. [data collection]. 4th Edition. UK Data Service. SN: 8189.

University College London, UCL Institute of Education, and Centre for Longitudinal Studies (2018). Next Steps: Sweeps 1-8, 2004-2016: Secure Access. [data collection]. 4th Edition. UK Data Service. SN: 7104.

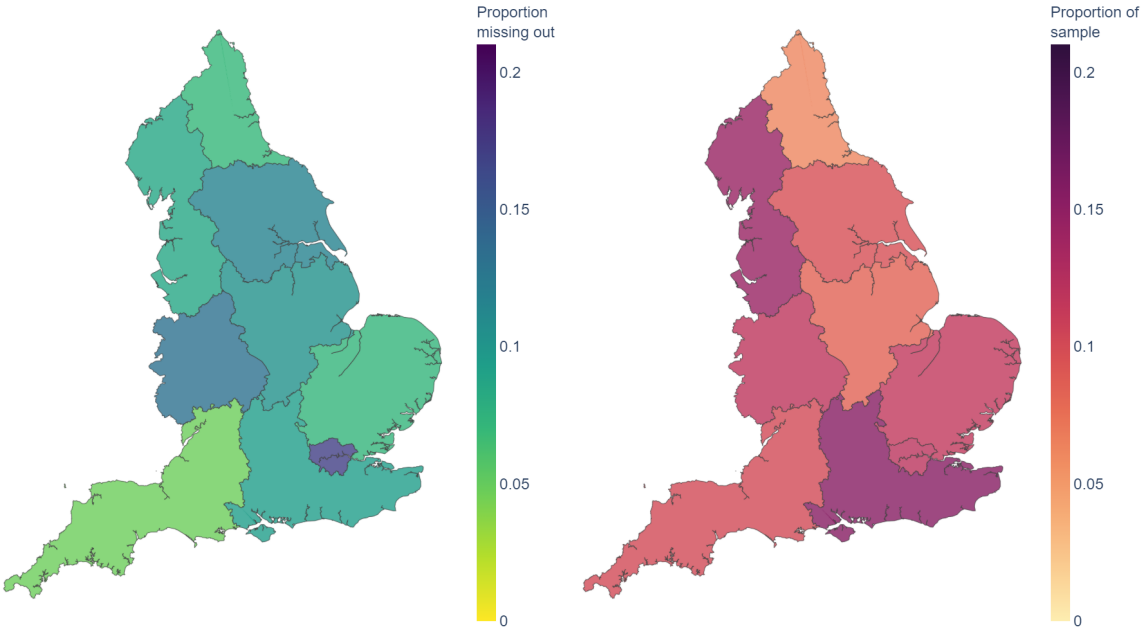
Weldon, M. (2018). Secondary school choice and selection: insights from new national preferences data. Department for Education Research Report, August 2018.

West, A., Barham, E., and Hind, A. (2011). Secondary school admissions in England 2001 to 2008: Changing legislation, policy and practice. *Oxford Review of Education*, 37(1):1–20.

West, A., Hind, A., and Pennell, H. (2004). School admissions and 'selection' in comprehensive schools: Policy and practice. *Oxford Review of Education*, 30(3):347–369.

Williams, J., Coldron, J., Fearon, J., Stephenson, K., Logie, A., and Smith, N. (2001). An analysis of the policies and practices of admission authorities in England. BERA Conference proceedings, September 2001.

Figure 1: Distribution of sample and treatment



(a) Proportion missing out

(b) Proportion of sample

Notes: These figures report the survey-weighted proportion of the final sample missing out on a place at a preferred school, in sub-figure (a), and the proportion of the weighted sample by Government Office Region, in sub-figure (b).

Table 1: Summary statistics for pupils

	Got in	Missed out	Got in	Missed out
	n=15,770		n=4,931	
	Full sample		*Final sample*	
Female	0.49	0.50	0.49	0.53
Any resident older siblings?	0.86	0.81	0.83	0.80
Child's ethnic group				
White	0.88	0.76	0.89	0.74
Indian, Pakistani, Bangladeshi	0.05	0.08	0.05	0.11
Black African, Black Caribbean	0.03	0.08	0.02	0.07
Other, Mixed	0.04	0.08	0.04	0.08
MP highest qual				
Higher ed.	0.24	0.20	0.22	0.21
A-levels, A-C GCSE	0.14	0.13	0.15	0.12
Lower GCSE	0.41	0.39	0.41	0.44
Other qualification	0.02	0.02	0.03	0.03
No qualification	0.20	0.26	0.19	0.20
Housing tenure				
Owner occupier	0.71	0.60	0.71	0.66
Renting from council/LA	0.23	0.31	0.24	0.30
Private renter/Other	0.07	0.09	0.05	0.03
MP religion				
None	0.23	0.25	0.23	0.25
Christian	0.68	0.61	0.68	0.60
Buddhist, Hindu, Jewish	0.04	0.06	0.03	0.07
Muslim	0.05	0.08	0.05	0.09
KS2 average points score	27.04	26.56	27.24	26.94
Distance to school (km)	3.17	3.33	2.94	2.93

Notes: This table reports survey-weighted proportions (or means, for continuous variables) of selected covariates, by treatment status, for the full sample of respondents to Wave 1, and then final sample of those who respond to Wave 8 and have non-missing data on all relevant covariates.

Table 2: Summary statistics (pupil outcomes and school attributes))

	Got in	Missed out	Difference	p-value
	A: Measured at pupil-level			
5+ A*-C GCSE	0.57	0.55	0.02	0.43
KS4 points	363.98	350.37	13.62	0.16
Stayed on	0.49	0.47	0.03	0.33
Attended university	0.35	0.39	-0.04	0.09
GHQ mental ill-health	2.28	2.85	-0.57	0.00
Ln(income)	9.66	9.59	0.06	0.00
Fertility	0.25	0.31	-0.05	0.06
Smoker	0.20	0.26	-0.06	0.02
	B: Measured at school-level			
Value-added (2004)	986.65	977.72	8.93	0.00
OFSTED rating good or above	0.63	0.54	0.09	0.01
% first language English (2004)	91.90	85.56	6.34	0.00
School FSM band (2004)	3.24	4.01	-0.77	0.00
% SEN (with or without statement)	15.45	19.89	-4.44	0.00

Notes: This table reports the survey-weighted means of outcomes reported by pupils attending their first-choice school schools of pupils reporting not to be at their first-choice school. The difference in means and p-values from a test of equality of means are also reported. p-values are computed from robust standard errors clustered by school (the Primary Sampling Unit).

Table 3: OLS and distance-matched effects of missing out on a first-choice school

	5+ A*-C	KS4 points	Stayed on	Ever uni	Mental ill-health	ln(income)	Fertility	Smoker
				A: OLS				
β	-0.03	-15.60**	-0.04	0.03	0.55***	-0.01	0.05**	0.07***
(s.e.)	(0.02)	(6.49)	(0.02)	(0.02)	(0.20)	(0.01)	(0.02)	(0.03)
δ	-22.13	-74.40	-11.78	4.126	22.03	0.821	37.10	-54.16
N	4,931	4,931	4,931	4,931	4,931	4,931	4,931	4,931
				B: Distance-matching				
ATT	-0.02	-14.75*	-0.03	0.03	0.70***	-0.02**	0.05	0.09***
(s.e.)	(0.03)	(8.07)	(0.02)	(0.03)	(0.24)	(0.01)	(0.03)	(0.03)
Y_0	0.58	365.08	0.50	0.36	2.18	9.59	0.26	0.18
N	3039	3039	3039	3039	3039	3039	3039	3039

Notes: β: OLS coefficient on treatment variable; ATT: average treatment effect on the treated; Y_0: model-adjusted outcome mean for control group; N: observations; δ: size of proportional selection on unobservables to drive ATT to zero; estimates in brackets are robust standard errors clustered by school. *** p<0.01, ** p<0.05, * p<0.1. All analyses are weighted using the survey weights. Each specification includes the following covariates: gender; month-of-birth; child ethnicity; older siblings; whether English is the main language spoken at home; Free School Meal eligibility; whether the child has special educational needs; Key Stage 2 scores; average Key stage 2 scores at the child's primary school; housing tenure; family income; parental occupational social class; parental education; parental religion; Principal Components Analysis summary of information used to choose schools; presence of local Grammar school; distance to closest private school; distance-weighted share of good-, very good- or excellent-rated schools within the 80th percentile of home-to-school distances travelled in that child's Local Authority; Local Authority fixed effects. *** p<0.01, ** p<0.05, * p<0.1. All analyses are weighted using the survey weights.

Table 4: Estimation results by area-type subgroup

	5+ A*-C	KS4 points	Stayed on	Attended uni	Mental ill-health	ln(income)	Fertility	Smoker
				OLS				
DA								
β	-0.01	-14.81*	-0.00	0.02	0.36*	-0.00	0.04	0.08***
(s.e.)	(0.03)	(7.93)	(0.03)	(0.03)	(0.22)	(0.01)	(0.03)	(0.03)
δ	-2.026	-26.14	-0.343	1.790	17.19	0.114	-20.55	44.28
N	3,460	3,460	3,460	3,460	3,460	3,460	3,460	3,460
FPF								
β	-0.09**	-14.05	-0.13***	0.05	1.03**	-0.04**	0.08*	0.02
(s.e.)	(0.04)	(10.55)	(0.04)	(0.04)	(0.43)	(0.02)	(0.04)	(0.06)
δ	13.70	5.271	11.89	-5.946	19.91	2.872	7.233	8.084
N	1,471	1,471	1,471	1,471	1,471	1,471	1,471	1,471
p-value	0.09	0.99	0.00	0.62	0.17	0.05	0.41	0.50
				Distance-matching				
DA								
ATT	0.02	-7.46	0.03	0.00	0.57**	-0.01	0.04	0.09**
(s.e.)	(0.03)	(9.26)	(0.03)	(0.03)	(0.27)	(0.01)	(0.04)	(0.04)
N	2,156	2,156	2,156	2,156	2,156	2,156	2,156	2,156
FPF								
ATT	-0.15***	-24.63*	-0.18***	0.08	0.91*	-0.05**	0.01	-0.02
(s.e.)	(0.05)	(13.60)	(0.05)	(0.05)	(0.51)	(0.02)	(0.07)	(0.08)
N	828	828	828	828	828	828	828	828
p-value	0.01	0.14	0.00	0.92	0.22	0.12	0.31	0.25

Notes: β: OLS coefficient on treatment variable; ATT: average treatment effect on the treated; N: observations; δ: size of proportional selection on unobservables to drive ATT to zero; estimates in brackets are robust standard errors clustered by school. *** p<0.01, ** p<0.05, * p<0.1. All analyses are weighted using the survey weights. p-value: p-value from a test of no significant difference in the ATT between subgroups. Each specification includes the following covariates: gender; month-of-birth; child ethnicity; older siblings; whether English is the main language spoken at home; Free School Meal eligibility; whether the child has special educational needs; Key Stage 2 scores; average Key stage 2 scores at the child's primary school; housing tenure; family income; parental occupational social class; parental education; parental religion; Principal Components Analysis summary of information used to choose schools; presence of local Grammar school; distance to closest private school; distance-weighted share of good-, very good- or excellent-rated schools within the 80th percentile of home-to-school distances travelled in that child's Local Authority; Local Authority fixed effects. DA: sub-group of pupils in areas using Deferred Acceptance mechanism; FPF: sub-group of pupils in areas using first-preference mechanism.

Table 5: Means of selected variables by area-type and treatment status

	DA		**FPF**	
	Got in	Missed out	Got in	Missed out
Value-added (2004)	987.31	977.94	985.27	977.12
OFSTED good or above	0.64	0.55	0.62	0.51
% first language English (2004)	90.46	84.57	94.93	88.39
% White (2004)	82.71	74.37	88.99	81.92
School FSM band (2004)	3.32	4.02	3.08	3.98
% SEN (with or without statement)	15.34	19.70	15.68	20.42
KS2 average points score	27.29	27.06	27.15	26.60
Urban	0.81	0.91	0.79	0.82
Distance to school (km)	3.04	2.96	2.72	2.85

Notes: Means are weighted using the survey weights. DA: Sub-sample of pupils attending schools in areas using a deferred-acceptance mechanism to allocate school places; FPF: Sub-sample of pupils attending schools in areas using a first-preference mechanism to allocate school places. Got in: attends preferred school; Missed out: does not attend preferred school.

Table 6: OLS effects of missing out with school fixed effects

	5 A*-C	KS4 points	Stayed on	Attended uni	Mental ill-health	ln(income)	Fertility	Smoker
				A: OLS excluding school fixed effects				
β	-0.03	-15.60**	-0.04	0.03	0.55***	-0.01	0.05**	0.07***
(s.e.)	(0.02)	(6.49)	(0.02)	(0.02)	(0.20)	(0.01)	(0.02)	(0.03)
δ	-22.13	-74.40	-11.78	4.13	22.03	0.82	37.10	-54.16
				B: OLS including school-level controls				
$\beta_{\text{sch\_X}}$	-0.01	-7.25	-0.02	0.05*	0.57***	-0.01	0.03	0.06**
(s.e.)	(0.02)	(6.35)	(0.02)	(0.02)	(0.20)	(0.01)	(0.03)	(0.03)
δ	2.28	2.98	12.00	-56.32	22.45	0.61	3.76	31.15
				C: OLS including school fixed effects				
$\beta_{\text{sch\_FE}}$	-0.02	-9.75	-0.05*	0.04	0.72***	-0.01	0.02	0.08**
(s.e.)	(0.03)	(6.41)	(0.03)	(0.03)	(0.22)	(0.01)	(0.03)	(0.03)
δ	6.43	4.17	-11.77	10.39	139.50	0.75	1.79	1.52
N	4,931	4,931	4,931	4,931	4,931	4,931	4,931	4,931

Notes: β: OLS coefficient on treatment variable; ATT: average treatment effect on the treated; N: observations; δ: size of proportional selection on unobservables to drive ATT to zero; estimates in brackets are robust standard errors clustered by school. *** p<0.01, ** p<0.05, * p<0.1. All analyses are weighted using the survey weights. Each specification includes the following covariates: gender; month-of-birth; child ethnicity; older siblings; whether English is the main language spoken at home; Free School Meal (FSM) eligibility; whether the child has special educational needs (SEN); Key Stage 2 scores; average Key stage 2 scores at the child's primary school; housing tenure; family income; parental occupational social class; parental education; parental religion; Principal Components Analysis summary of information used to choose schools; presence of local Grammar school; distance to closest private school; distance-weighted share of good-, very good- or excellent-rated schools within the 80th percentile of home-to-school distances travelled in that child's Local Authority. The specifications in panel A also include Local Authority fixed effects; those in panel B also include the following school-level controls: % gaining 5+ A*-C grades in 2004; KS2-KS4 value-added; OFSTED rating of good and above?; % with English as first language; % of pupils White British; FSM band of the school; % with SEN; school type. Panel C includes school fixed effects.

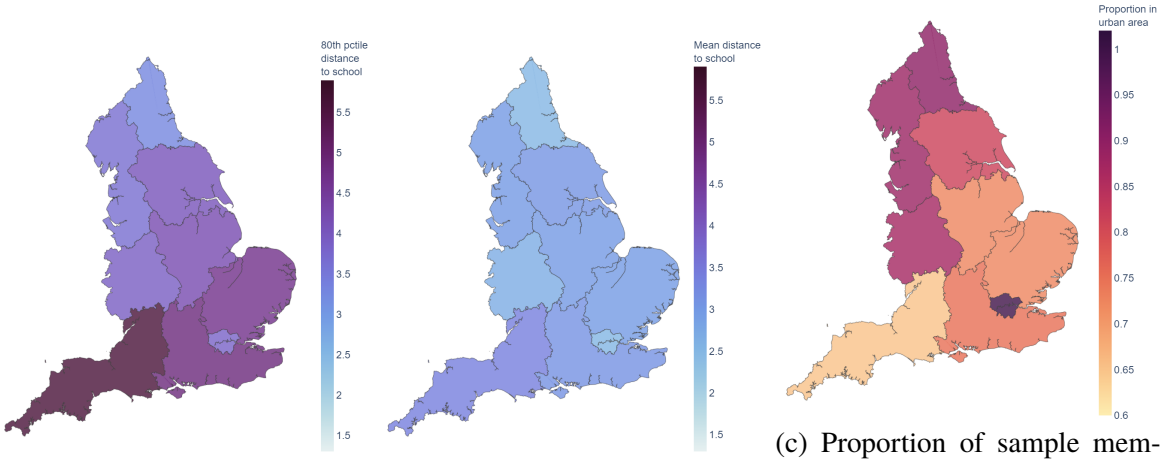
Appendix

Table A.1: Data structure and sample accounting

	Unweighted sample size
LSYPE panel structure	
Wave 1	15,770
Wave 2	13,539
Wave 3	12,439
Wave 4	11,801
Wave 5	10,430
Wave 6	9,799
Wave 7	7,707
All waves	5,426
Number of non-missing observations for:	
All pupil-level covariates	5,789
All school-level covariates	7,062
All outcomes	6,436
Final sample: *all covariates & outcomes*	4,931
Number of schools and pupils per school:	
Number of schools in Wave 1 sample	658 (n=15,770)
Number of schools in final sample	531 (n=4,931)
Mean [Pupils/school] in final sample	9.28 (n=4,931)
SD [Pupils/school] in final sample	3.71 (n=4,931)
Number of schools in final sample with within-school variation in the treatment	303 (n=2,607)

Notes: This table reports the structure of the LSYPE panel (how many study members appear at each Wave); the number of study members who have non-missing data on key covariates and outcomes; the numbers of schools, and pupils per school in the initial and final estimation sample.

Figure A.1: Geography of distance to school and urbanicity



(a) Within-area 80th percentile home-school distance

(b) Within-area mean home-school distance

(c) Proportion of sample members in urban small-areas by region

Notes: This figure reports the survey-weighted 80th percentile distance from home-to-school by Government Office Region, in sub-figure (a); the survey-weighted mean distance from home-to-school by Government Office Region, in sub-figure (b); the proportion of sample members living in a small-area geographical unit ("Output Area") classified as urban rather than rural using the ONS Urban-Rural classification, by Government Office Region, in sub-figure (c).

Outcome variable definitions

- "5+ A*-C": This is a binary variable indicating whether the pupil gained at least 5 GCSEs (or the vocational GNVQ equivalents) at grade C or above. GCSE is the examination taken at the end of Key Stage 4, when pupils are aged 15 or 16 years. This threshold was considered a "good" passing grade required progression to many routes of further study.

- "KS4 points": The grade received for each GCSE subject is assigned points, and we use the variable "GCSE points" which is the sum of points over GCSE subjects (and other KS4 equivalents such as GNVQ).

- "Stayed on": This is binary variable indicating whether a pupil stayed on after the age of compulsory schooling (16 years) to enter for any qualification at Key Stage 5. The options at Key Stage 5 include A-levels, but also vocational qualifications (such as BTECS, qualifications provided by the Business and Technology and Education Council).

- "Attends Uni": This is a binary variable indicating whether a cohort member reports attending university by age 25 years.

- "Mental ill-health": This variable is the General Health Questionnaire 12 (GHQ12), a screening tool for identifying minor psychiatric disorders in the general population, and suitable for use in adolescents and upward. For each of 12 questions, participants are asked whether symptoms of mental ill-health are 'not at all present', present 'no more than usual', present 'rather more than usual', or present 'much more than usual'. The responses are coded following a conventional approach, where the first two responses receive a score of zero, the latter two receive a score of one. These scores are summed to give a continuous variable ranging from 0 to 12 where a higher value indicate poorer Mental ill-health.

- "Fertility": This is a binary variable indicating whether a cohort member has any children by age 25 years (the final wave of the study).

- "ln(income)": gross weekly income measured at age 25 years.

- “Smoker”: Binary variable indicating the respondent is a current smoker at age 25 years.

Table A.2: Estimation results by area-type subgroup - with school attributes as controls

	5 A*-C	KS4 points	Stayed on	Attended uni	Mental ill-health	ln(income)	Fertility	Smoker
				OLS				
DA								
β	0.01	-6.48	0.01	0.04	0.40*	-0.00	0.02	0.07**
(s.e.)	(0.03)	(7.91)	(0.03)	(0.03)	(0.22)	(0.01)	(0.03)	(0.03)
δ	-5.63	3.14	4.04	15.60	32.44	0.094	4.74	17.55
N	3,460	3,460	3,460	3,460	3,460	3,460	3,460	3,460
FPF								
β	-0.08*	-8.57	-0.11***	0.07*	1.03**	-0.04**	0.06	0.02
(s.e.)	(0.04)	(10.34)	(0.04)	(0.04)	(0.42)	(0.02)	(0.05)	(0.06)
δ	5.98	1.88	5.91	-4.79	14.83	2.25	3.21	2.99
N	1,471	1,471	1,471	1,471	1,471	1,471	1,471	1,471
p-value	0.10	0.97	0.00	0.59	0.16	0.06	0.47	0.47
				Distance-matching				
DA								
ATT	0.03	-3.84	0.04	0.02	0.58**	-0.01	0.02	0.10***
(s.e.)	(0.03)	(9.16)	(0.03)	(0.03)	(0.26)	(0.01)	(0.04)	(0.03)
N	2,156	2,156	2,156	2,156	2,156	2,156	2,156	2,156
FPF								
ATT	-0.15***	-20.90*	-0.17***	0.09*	0.99*	-0.05**	0.08	-0.04
(s.e.)	(0.06)	(12.54)	(0.05)	(0.05)	(0.53)	(0.02)	(0.07)	(0.08)
N	828	828	828	828	828	828	828	828
p-value	0.01	0.27	0.00	0.22	0.48	0.13	0.45	0.12

Notes: β: OLS coefficient on treatment variable; ATT: average treatment effect on the treated; N: observations; δ: size of proportional selection on unobservables to drive ATT to zero; estimates in brackets are robust standard errors clustered by school. *** p<0.01, ** p<0.05, * p<0.1. All analyses are weighted using the survey weights. p-value from a test of no significant difference between subgroups. Each specification includes the following covariates: gender; month-of-birth; older siblings; child ethnicity; whether English is the main language spoken at home; Free School Meal (FSM) eligibility; whether the child has special educational needs (SEN); Key Stage 2 scores; average Key stage 2 scores at the child's primary school; housing tenure; family income; parental occupational social class; parental education; parental religion; Principal Components Analysis summary of information used to choose schools; presence of local Grammar school; distance to closest private school; distance-weighted share of good-, very good- or excellent-rated schools within the 80th percentile of home-to-school distances travelled in that child's Local Authority; Local Authority fixed effects. School-level controls: % gaining 5+ A*-C grades in 2004; KS2-KS4 value-added; OFSTED rating of good and above?; % with English as first language; % of pupils White British; FSM band of the school; % with SEN; school type. DA: sub-group of pupils in areas using Deferred Acceptance Mechanism; FPF: sub-group of pupils in areas using first-preference mechanism.

A Online Appendix: English education system

Ages and stages:

The English school curriculum is divided into blocks of years called "Key Stages" (KS). Children are assessed based on national standards at the end of each Key Stage. Key Stage 1 runs through ages 5 to 7 years, and Key Stage 2 through ages 7 to 11 years. Key Stage 2 is assessed at the end primary school at age 11 years. Children then transition to secondary school, where they must attend until age 16. Key Stage 3 runs from ages 11 to 14 years. During Key Stage 4, ages 14 to 16 years, children study a set of subjects in which they will sit the high-stakes General Certificate of Secondary Education examinations (GCSEs), or equivalents, at age 15/16 years. Pupils typically take between 5 and 10 subjects, and usually a passing grade in at least five of them—including Mathematics and English—is required to progress to many academic or other qualifications. A common school performance metric is the proportion of pupils achieving at least five GCSEs at grades A*-C (including English and Mathematics).

With satisfactory performance in GCSEs, pupils can study for A-levels, from age 16–18. A-Levels are the most common route to university entrance, and are taken either at the same secondary school, or some pupils move to another school or to a specialised "sixth form college" to take their A levels. Further Education (FE) colleges are an alternative to staying in an secondary school, and they provide vocational training or a mix of academic and vocational. Pupils typically enter Higher Education (University) from age 18.

School types:

The types of secondary schools in England in 2001 are described as follows. The first category are "state schools" which are controlled and funded by central government, and on which this study is based. All children in England between the ages of 5 and 16 are entitled to a free place at a state school. These schools follow the National Curriculum and are inspected by Ofsted (the government's Office for Standards in Education, Children's Services and Skills). About 93% of pupils attended such schools when our data was collected (97% unweighted). The sec-

ond category are "Independent Schools" (private schools) which about 7% of pupils attended (3% unweighted).

Within the category of state schools, there are community schools (67% of pupils in our data), voluntary-controlled (3%), voluntary-aided (11%), foundation schools (15%). *Community schools* are entirely run by the local council (Local Authority). *Foundation or Trust schools* are run by a local governing body. These schools were formerly called "Grant-maintained" schools. This was an initiative to allow more flexibility in provision of education where, by majority parental vote, schools could opt out of Local Authority control and be run by a governing body with more control over admissions and staffing. *Voluntary-aided schools* are typically religious or faith schools, which can admit pupils on religious affiliation grounds. *Voluntary-controlled schools* are almost all faith schools. They are a mix between community and voluntary-aided schools: similar to a community school, the local authority employs the staff and sets the entrance criteria, but the school land and buildings are owned by a charity, often a church, which also can appoint some members of the governing body.

Regardless of governance arrangement, all non-selective state schools have to comply with the school admissions code of practice which sets guidelines for fair admissions. The exception to this are grammar schools: about 3% of state schools are *Grammar schools*, a minority of academically selective state-funded schools. They select all or most of their pupils based on academic ability, assessed by the so-called "11 plus" exam. The 11-plus exam is distinct from the national Key Stage assessments that take place at the age of 11. The former are not nationally co-ordinated. Sometimes they are set and graded by individual grammar schools, and sometimes (usually in local authorities with several grammar schools) they are co-ordinated by the local authority.

B Online Appendix: Sample characteristics

Table WA.1: Regional distribution of sample

	N	Missed out	Distance to school (km)			Weighted N	Pr(Urban)
Region		Mean	Mean	Median	80th percentile		Proportion
North East	254	0.071	2.527	1.526	3.285	302.46	0.903
North West	703	0.086	2.986	1.877	3.670	702.85	0.885
Yorks & Humber	453	0.117	3.079	2.040	4.077	475.37	0.816
East Midlands	383	0.104	2.984	1.801	4.221	408.29	0.720
West Midlands	601	0.132	2.696	1.463	3.940	578.56	0.873
East of England	578	0.072	2.971	1.806	4.679	566.75	0.726
Greater London	796	0.170	2.534	1.796	3.852	589.01	1.000
South East	748	0.094	3.114	1.894	4.859	756.69	0.752
South West	415	0.048	3.401	1.837	5.615	496.23	0.651
Total	4931	0.101	2.938	1.788	4.200	4876.21	0.815

Notes: This table shows the proportion (mean) missing out on their first choice school ("missed out"); the mean, median and 80th percentile of distance from home to school in kilometres; the weighted sample size; the share of sample members living in an urban area, Pr(Urban); by region. The unweighted sample size underlying all of these statistics is denoted as N.

Table WA.2: Comparison of key variables with population counterparts

	Male	White British	English First Language	KS4 points	5+ A*-C	FSM eligible
Final sample	0.51	0.88	0.94	361.61	0.57	0.12
Population	0.51	0.81	0.91	347.61	0.56	0.16

Notes: This Table reports the means (or proportions for binary variables), and samples sizes, of a set of demographic, socio-economic and attainment variables. The variables are selected based on the availability of variables in the population data available for this project. *Final LSYPE sample* reports these statistics for the final analytical sample used in the paper comprised of records with non-missing data on all outcomes and covariates (n=4,931). *Population* reports these statistics for the population counterparts: all 'on roll' pupils who were in Year 9 in the academic year 2003/04. The data source is the 2003/04 Pupil-Level Annual School Census (PLASC) linked to Key Stage attainment data. These population statistics were produced using statistical data from ONS via the Secure Research Service. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

C Online Appendix: Further results

Ethnicity sub-group analyses

Previous research has found negligible differences in the probability of gaining a place at a first-choice school by socio-economic status, but large differences by ethnicity. For instance, 90% of White British families were offered a place at their first-choice school, compared with 72% Asian and 62% of Black families applying for schools in England in the 2014/15 academic year (Burgess et al., 2019). As shown in Weldon (2018), this pattern is observed across the regions of the England—it is not, for instance, explained by clustering of black and ethnic minority families in the denser London school market. The reasons for this variation remain unclear, but one possible explanation is differences by ethnicity in engagement with the school choice process. White British families are more likely to list only one school: 41% of White British households list one school, compared to 17% of Asian households and 12% of Black households (Burgess et al., 2019). Ethnic minority families are more likely to list many schools, and select, on average, more ambitious (further away and better quality) schools as their first choice. This may lead to a higher chance of missing out based on the oversubscription criteria.

This motivates splitting the sample by ethnic origin of the child. Table WA.3 reports OLS and distance-matched results separately by ethnic group: White British, or Black, Asian or other minority ethnicity (BAME). The distance-matched results are less precise than OLS, as they are based on smaller sample sizes due to the restricted bandwidth, especially within subgroups. The only outcome for which the difference in effect size between subgroups is statistically significant across both specifications is mental ill-health, where the effects appear larger among White pupils. The effects on staying on in school after age 16 years, the minimum age of compulsory schooling at this time, are also consistently larger across OLS and the distance-matching results among White pupils, however the difference between subgroups is only statistically significant in the OLS results. This pattern could have several explanations, including differential school preferences by ethnicity, variation in residential sorting, or differential list lengths as discussed earlier. If White parents are not listing enough schools, when they do miss out they may be

allocated a school which they would not have chosen at all; While if BAME parents list more schools, in the way the system is designed to be used, and they miss out then the child has a higher chance of ending up in a fallback local school that is at least acceptable. However, without further data is difficult to be confident about the specific explanatory mechanism(s).

Table WA.3: Estimation results by ethnicity subgroup

	5+ A*-C	KS4 points	Stayed on	Attended uni	Mental ill-health	ln(income)	Fertility	Smoker
				OLS				
BAME								
β	-0.02	-18.36	0.03	0.05	-0.15	-0.01	-0.00	0.02
(s.e.)	(0.04)	(12.82)	(0.04)	(0.04)	(0.26)	(0.01)	(0.03)	(0.04)
δ	-0.94	-3.73	1.88	2.90	-1.45	-2.86	-0.71	5.34
N	1,337	1,337	1,337	1,337	1,337	1,337	1,337	1,337
White								
β	-0.04	-16.09**	-0.06**	0.01	0.74***	-0.01	0.06**	0.09***
(s.e.)	(0.03)	(7.07)	(0.03)	(0.02)	(0.24)	(0.01)	(0.03)	(0.03)
δ	8.98	8.57	10.81	-1.77	70.13	3.93	6.33	50.44
N	3,594	3,594	3,594	3,594	3,594	3,594	3,594	3,594
p-value	0.27	0.47	0.01	0.29	0.03	0.88	0.16	0.20
				Distance-matching				
BAME								
ATT	-0.06	-19.80	-0.00	0.06	-0.13	-0.03**	0.06	0.06
(s.e.)	(0.05)	(14.79)	(0.04)	(0.06)	(0.32)	(0.02)	(0.04)	(0.04)
N	977	977	977	977	977	977	977	977
White								
ATT	-0.05	-12.56	-0.06*	0.02	1.04***	-0.02	0.05	0.10**
(s.e.)	(0.03)	(9.27)	(0.03)	(0.03)	(0.32)	(0.01)	(0.04)	(0.04)
N	1,625	1,625	1,625	1,625	1,625	1,625	1,625	1,625
p-value	0.75	0.68	0.31	0.56	0.01	0.43	0.82	0.53

Notes: β: OLS coefficient on treatment variable; ATT: average treatment effect on the treated; N: observations; δ: size of proportional selection on unobservables to drive ATT to zero; estimates in brackets are robust standard errors clustered by school. *** p<0.01, ** p<0.05, * p<0.1. All analyses are weighted using the survey weights. p-value is the p-value from a test of equality of coefficients between groups. BAME; child's reported ethnicity is Black, Asian and Minority Ethnicity. White; child's reported ethnicity is White British. Each specification includes the following covariates: gender; month-of-birth; older siblings; whether English is the main language spoken at home; Free School Meal eligibility; whether the child has special educational needs; Key Stage 2 scores; average Key stage 2 scores at the child's primary school; housing tenure; family income; parental occupational social class; parental education; parental religion; Principal Components Analysis summary of information used to choose schools; presence of local Grammar school; distance to closest private school; distance-weighted share of good-, very good- or excellent-rated schools within the 80th percentile of home-to-school distances travelled in that child's Local Authority; Local Authority fixed effects.

Table WA.4: Estimation results by ethnicity subgroup - school attributes in controls

	5+ A*-C	KS4 points	Stayed on	Attended uni	Mental ill-health	ln(income)	Fertility	Smoker
				OLS				
BAME								
β	-0.01	-17.23	0.02	0.06	-0.18	-0.02	-0.01	0.03
s.e.	(0.04)	(12.89)	(0.04)	(0.04)	(0.27)	(0.01)	(0.03)	(0.04)
δ	-0.49	-3.80	1.48	4.15	-1.56	-2.90	-1.49	12.52
N	1,337	1,337	1,337	1,337	1,337	1,337	1,337	1,337
White								
β	-0.02	-5.97	-0.04*	0.04*	0.75***	-0.01	0.04	0.08**
s.e.	(0.03)	(6.88)	(0.03)	(0.02)	(0.24)	(0.01)	(0.03)	(0.03)
δ	1.61	1.17	3.57	-2.51	52.57	2.02	2.02	12.57
N	3,594	3,594	3,594	3,594	3,594	3,594	3,594	3,594
p-value	0.35	0.60	0.02	0.42	0.03	0.95	0.23	0.23
				Distance-matching				
BAME								
ATT	-0.06	-23.58	-0.04	0.08	-0.38	-0.04**	0.03	0.07*
s.e.	(0.05)	(15.74)	(0.05)	(0.06)	(0.35)	(0.02)	(0.04)	(0.04)
N	977	977	977	977	977	977	977	977
White								
ATT	-0.04	-8.96	-0.05	0.04	1.03***	-0.01	0.04	0.09**
s.e.	(0.03)	(8.59)	(0.03)	(0.03)	(0.32)	(0.01)	(0.04)	(0.04)
N	1,625	1,625	1,625	1,625	1,625	1,625	1,625	1,625
p-value	0.72	0.41	0.81	0.49	0.00	0.22	0.98	0.74

Notes: β: OLS coefficient on treatment variable; ATT: average treatment effect on the treated; N: observations; δ: size of proportional selection on unobservables to drive ATT to zero; estimates in brackets are robust standard errors clustered by school. *** p<0.01, ** p<0.05, * p<0.1. All analyses are weighted using the survey weights. p-value is the p-value from a test of equality of coefficients between groups. BAME; child's reported ethnicity is Black, Asian and Minority Ethnicity. White; child's reported ethnicity is White British. Each specification includes the following covariates: gender; month-of-birth; older siblings; whether English is the main language spoken at home; Free School Meal (FSM) eligibility; whether the child has special educational needs (SEN); Key Stage 2 scores; average Key stage 2 scores at the child's primary school; housing tenure; family income; parental occupational social class; parental education; parental religion; Principal Components Analysis summary of information used to choose schools; presence of local Grammar school; distance to closest private school; distance-weighted share of good-, very good- or excellent-rated schools within the 80th percentile of home-to-school distances travelled in that child's Local Authority; Local Authority fixed effects. School-level controls: % gaining 5+ A*-C grades in 2004; KS2-KS4 value-added; OFSTED rating of good and above?; % with English as first language; % of pupils White British; FSM band of the school; % with SEN; school type.

D Online Appendix: Robustness checks

Table WA.5: OLS results restricted to the sample with within-school variation

	5+ A*-C	KS4 points	Stayed on	Attended uni	Mental ill-health	ln(income)	Fertility	Smoker
				A: OLS				
β	-0.04	-14.10**	-0.05**	0.04	0.67***	-0.01	0.03	0.09***
(s.e.)	(0.02)	(6.63)	(0.02)	(0.02)	(0.21)	(0.01)	(0.03)	(0.03)
δ	-3.08	-4.96	-4.57	4.37	56.05	1.27	158.4	-17.00
N	2,607	2,607	2,607	2,607	2,607	2,607	2,607	2,607
				B: OLS with school attributes in controls				
β	-0.02	-8.24	-0.04*	0.05**	0.70***	-0.01	0.02	0.08***
(s.e.)	(0.02)	(6.50)	(0.02)	(0.02)	(0.21)	(0.01)	(0.03)	(0.03)
δ	-2.93	-7.92	-5.26	8.08	134.8	0.99	6.61	-95.8
N	2,607	2,607	2,607	2,607	2,607	2,607	2,607	2,607

Notes: The analyses in this Table are restricted to pupils in schools which have within-school variation in the treatment. β: OLS coefficient on treatment variable; N: observations; δ: size of proportional selection on unobservables to drive ATT to zero; estimates in brackets are robust standard errors clustered by school. *** p<0.01, ** p<0.05, * p<0.1. All analyses are weighted using the survey weights. Each specification includes the following covariates: gender; month-of-birth; older siblings; whether English is the main language spoken at home; Free School Meal (FSM) eligibility; whether the child has special educational needs (SEN); Key Stage 2 scores; average Key stage 2 scores at the child's primary school; housing tenure; family income; parental occupational social class; parental education; parental religion; Principal Components Analysis summary of information used to choose schools; presence of local Grammar school; distance to closest private school; distance-weighted share of good-, very good- or excellent-rated schools within the 80th percentile of home-to-school distances travelled in that child's Local Authority; Local Authority fixed effects. Panel B additionally includes the following school-level controls: % gaining 5+ A*-C grades in 2004; KS2-KS4 value-added; OFSTED rating of good and above?; % with English as first language; % of pupils White British; FSM band of the school; % with SEN; school type.

Table WA.6: Distance-matched effects of missing out on first-choice school - optimal bandwidth

VARIABLES	5+ A*-C	KS4 points	Stayed on	Attended uni	Mental ill-health	ln(income)	Fertility	Smoker
ATT	-0.03	-13.27*	-0.04	0.03	0.66***	-0.02**	0.04	0.07***
(s.e.)	(0.02)	(7.52)	(0.02)	(0.03)	(0.22)	(0.01)	(0.03)	(0.03)
Y_0	0.58	363.44	0.50	0.36	2.20	9.61	0.27	0.18
N	3389	3389	3389	3389	3389	3389	3389	3389
Bandwidth	6.918	6.918	6.918	6.918	6.918	6.918	6.918	6.918

Notes: ATT: average treatment effect on the treated; Y_0: model-adjusted outcome mean for control group; Bandwidth is the radius in kilometres from which control units are selected, selected using a data-driven bandwidth selection procedure. N: unweighted cell count for all statistics in this table; Estimates in brackets are the standard errors.

Table WA.7: Distance-matched effects of missing out on first-choice school with school-attributes in controls

VARIABLES	5+ A*-C	KS4 points	Stayed on	Attended uni	Mental ill-health	ln(income)	Fertility	Smoker
ATT	-0.01	-8.52	-0.00	0.04	0.69***	-0.02	0.05	0.07**
(s.e.)	(0.03)	(11.63)	(0.03)	(0.03)	(0.23)	(0.02)	(0.03)	(0.03)
Y_0	0.56	358.85	0.47	0.35	2.20	9.60	0.26	0.19
N	3039	3039	3039	3039	3039	3039	3039	3039

Notes: ATT: average treatment effect on the treated; Y_0: model-adjusted outcome mean for control group; N: unweighted cell count for all statistics in this table; Estimates in brackets are the standard errors. Each specification includes the following covariates: gender; month-of-birth; older siblings; whether English is the main language spoken at home; Free School Meal (FSM) eligibility; whether the child has special educational needs (SEN); Key Stage 2 scores; average Key stage 2 scores at the child's primary school; housing tenure; family income; parental occupational social class; parental education; parental religion; Principal Components Analysis summary of information used to choose schools; presence of local Grammar school; distance to closest private school; distance-weighted share of good-, very good- or excellent-rated schools within the 80th percentile of home-to-school distances travelled in that child's Local Authority; Local Authority fixed effects. School-level controls: % gaining 5+ A*-C grades in 2004; KS2-KS4 value-added; OFSTED rating of good and above?; % with English as first language; % of pupils White British; FSM band of the school; % with SEN; school type.