Leading education and social research
Institute of Education
University of London

Department of Quantitative Social Science

Using the Longitudinal Study of Young People in England for research into Higher Education access

Jake Anders

DoQSS Working Paper No. 12-13
December 2012

Disclaimer

Any opinions expressed here are those of the author(s) and not those of the Institute of Education. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

DoQSS Workings Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

Department of Quantitative Social Science, Institute of Education, University of London
20 Bedford Way, London WC1H 0AL, UK

Using the Longitudinal Study of Young People in England for research into Higher Education

Jake Anders[1]

Abstract

The Longitudinal Study of Young People in England (LSYPE) has the potential to be an important new resource for addressing research questions regarding access to Higher Education. This paper outlines the data available in the LSYPE and assesses its quality, particularly relative to other datasets that have been used to address similar questions in the past. The paper finds many positive features of the data. These include data collection from parents (including much information on family background characteristics) and good family income measurement compared with many previous studies. The LSYPE also measures a greater depth of HE-related outcomes than some previous datasets, including application, entry, subject studied and institution attended. However, comparison with official statistics suggests that this may be undermined by a large overestimation of the proportion of young people who enter Higher Education (as much as ten percentage points) than we would see in a truly nationally representative sample. There is also some evidence of underreporting of family income. Nevertheless, the paper concludes that analysis of the LSYPE has the potential to shed new light on university access in England.

JEL classification: I24, J62.

Keywords: Higher Education, Socioeconomic Gradient, Intergenerational Mobility, Longitudinal Research, Survey Data.

[1] Department of Quantitative Social Science, Institute of Education, University of London (jake@jakeanders.co.uk)

Acknowledgements

This research forms part of the Nuffield Foundation project ``Higher Education Funding and Access: exploring common beliefs''. The project is based at the Institute of Education and the Institute for Fiscal Studies. It is directed by Claire Crawford, Lorraine Dearden, John Micklewright and Anna Vignoles. Thanks to my supervisors Lorraine Dearden and John Micklewright for all their help.

1 Introduction

Interest in access to Higher Education (HE) in the UK has only intensified in recent years, particularly as the share of the population attending HE has risen and the share of the cost provided by students themselves has increased.

The Longitudinal Study of Young People in England (LSYPE) (Department for Education and National Centre for Social Research, 2012) has the potential to be an important new resource for research in this area with its rich set of questions, links to administrative data, and longitudinal design. The aim of this paper is to evaluate to what extent it lives up to this potential.

The LSYPE follows a sample of young people (YPs), born between September 1989 and August 1990, from the age of 13-14 to the age of 19-20. Its cohort members were first able to enter HE in September 2008. At this time, students faced maximum tuition fees of £3,145, and were eligible for a subsidised loan from the Student Loans Company to cover both this and living costs. This was after a major reform in September 2006, before which tuition fees were capped at around £1,000, but were payable upfront.

Given its timeframe, it is natural that the LSYPE includes questions directly related to the process of applying to university. Moreover, the LSYPE includes questions which previous research has indicated have a bearing on determining success (in terms of both application to and entering university). These include the prior attainment and socioeconomic background of YPs and the attitudes of YPs and their parents.

In drawing up and analysing the aspects of the survey I have particularly taken into account previous studies using the LSYPE for related purposes, such as Chowdry et al. (2009), and literature reviews on determinants of children's attainments, such as Haveman and Wolfe (1995). The paper takes a cross disciplinary approach: while it is mainly grounded in the economics of education it also draws, to a lesser degree, on relevant contributions from sociology and educational psychology. I also refer to the Department for Education's LSYPE User Guide (Department for Education, 2011a).

The paper proceeds as follows. In section 2 it considers the structure and features of the LSYPE, including timings of surveys and attrition rates. Section 3 turns to a discussion of previous research in the area, including key datasets that have been used in the past and

their strengths and weaknesses. Next, section 4 considers our main outcomes of interest, university application and attendance, considering how well the LSYPE captures the desired information. Section 5 consider related outcomes of interest.

Having considered outcomes, the paper turns to explanatory information included in the survey. Section 6 considers the role of family income; section 7 considers other socio-economic characteristics of the YPs and their families; section 8 deals with attitudes and expectations; section 9 assesses cognitive ability and prior attainment; and section 10 looks at school and neighbourhood effects. Finally, section 11 concludes by reflecting on whether these parts taken together amount to the building blocks of good models for research questions regarding HE. There are also appendices detailing A) questions referred to in the course of the paper and the relevant variables names in the Stata edition of the LSYPE dataset and B) technical details of the construction of income measures.

2 Structure and Features of the LSYPE

The LSYPE is made up of seven 'waves' conducted annually, beginning in Summer 2004 with children in Year 9 (aged 13-14) at the time. Interviews were conducted with YPs and their parents, covering information about the YPs themselves and their households. All individuals who took part in the LSYPE consented to having their responses linked with administrative data in the form of the National Pupil Database (NPD). This contains the results for statutory examinations beginning with Key Stage 2 SATS, taken in the last year of primary school (age 10-11)[1].

Wave 7, which is the final wave, covers YPs aged 19-20; it observes those who attend university straight after leaving school or after a single year's break. As a result, those who take what might be thought of as the 'traditional' route into HE will be included in analysis[2].

Although Wave 7 was the final wave of the LSYPE, the Department for Education is currently

[1]Taking and reporting SATS results is only mandatory for state schools. Hence much of the data for children in the independent sector at the relevant points in their educational careers is missing. This is discussed in Section 9.1.

[2]Just over 30% of individuals who receive an offer through UCAS do so after age 19. Complete figures are shown in Table 1 and discussed in Section 4.2.

in the process of commissioning a second cohort. This will follow many of the same features as the existing survey. Some changes, drawing on the experiences from the LSYPE have been suggested by Collingwood et al. (2010), an official DfE review of the first cohort.

2.1 Timing

Waves 1 to 4 were conducted through face to face interviews at the YP's home. The field-work for each wave ran from roughly April to October each year. This means the survey begins towards the end of an academic year, but some responses may not be made until after the start of the following one.

From Wave 5 onwards only YP interviews were conducted. These were carried out using a variety of methods: online, over the telephone, and face to face. The former two methods left a certain amount of autonomy over when the YP chooses to complete their survey, with responses normally beginning in May but some not coming in until October, presumably the deadline (see Figures 1, 2, 3 and 4). In common with the early waves, this means that while most are made towards the end of an academic year, some responses may not be made until the start of the following one.

Nevertheless, the vast majority of responses have been collected by the end of August, meaning the potential contamination of academic years this could cause should not have too great an effect. It is important to be mindful of this spread of response timings though, since it spans important discontinuities especially between Wave 5 and 6 in the shape of A Level results and potentially the start of university attendance (which could then show up a year earlier than intended).

This timing also means that most people who are going through university applications will have been through most stages of the process (see Section 4) and be holding offers by the time they complete the survey. The questionnaire will most likely (but not necessarily) be completed before critical A Level results and subsequently any university places found through 'Clearing'.

Figure 1: Wave 4 Young Person Interviews by Month

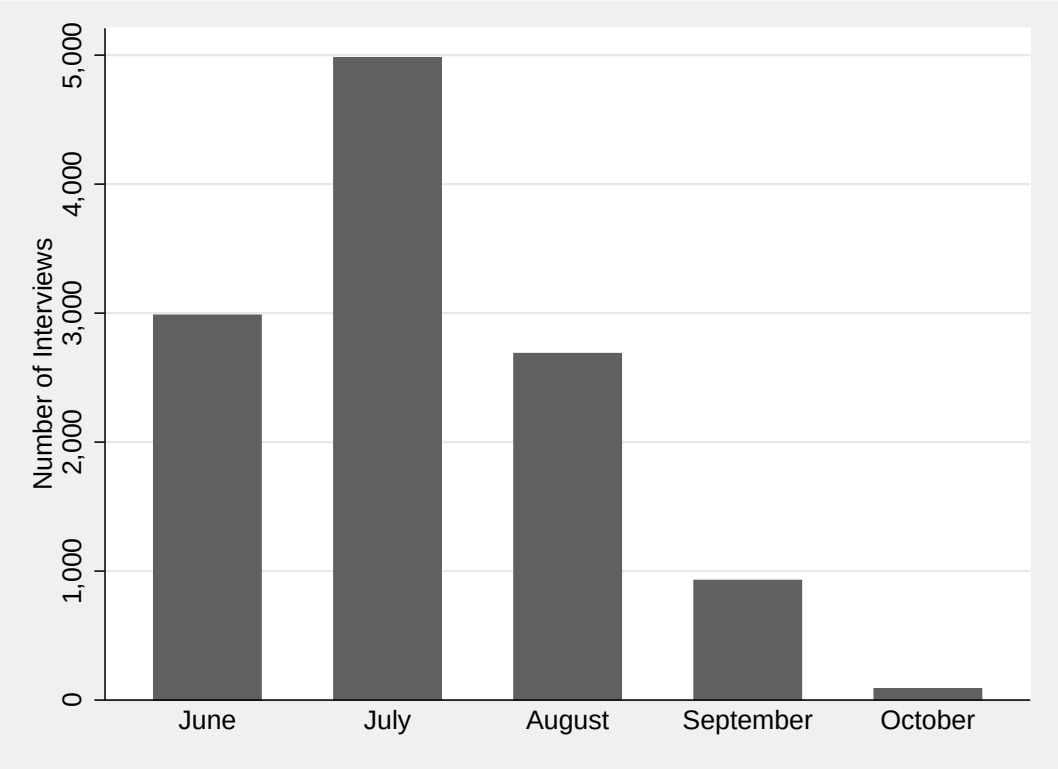


Figure 2: Wave 5 Young Person Interviews by Month

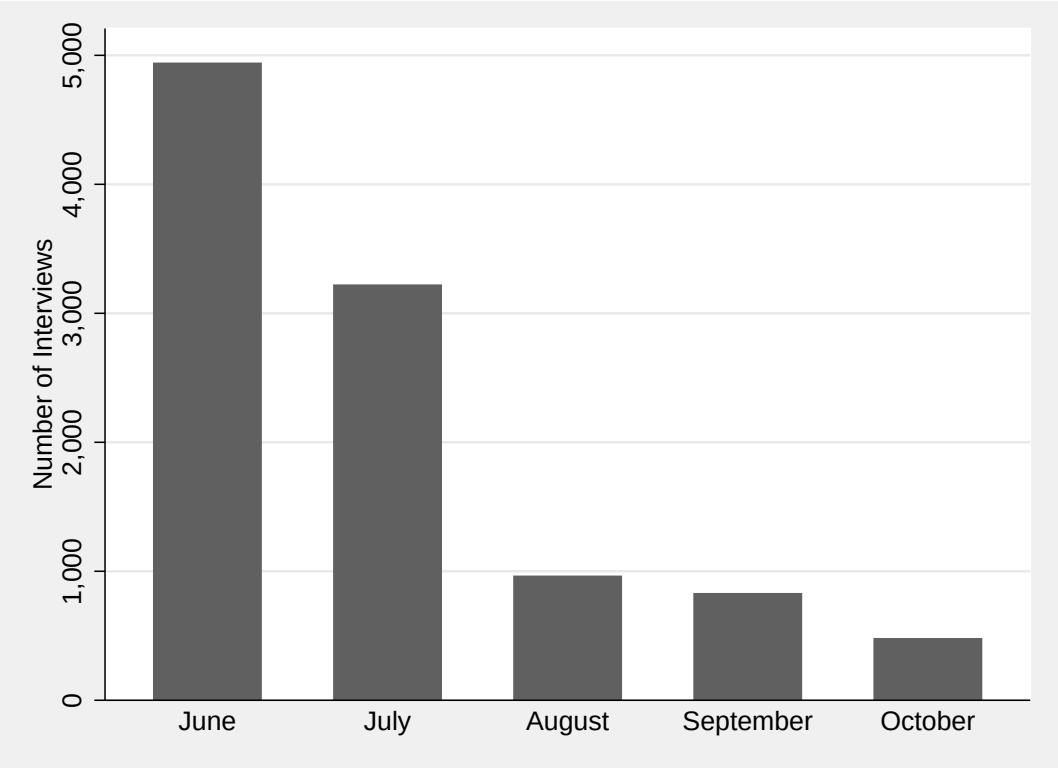


Figure 3: Wave 6 Young Person Interviews by Month

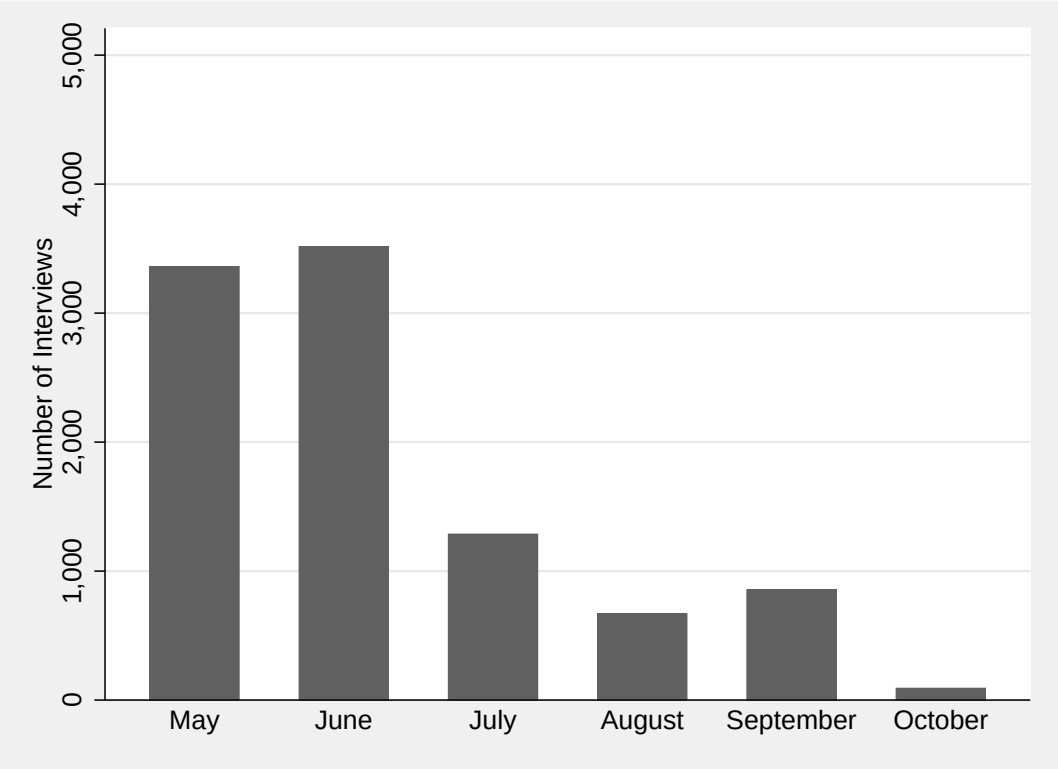
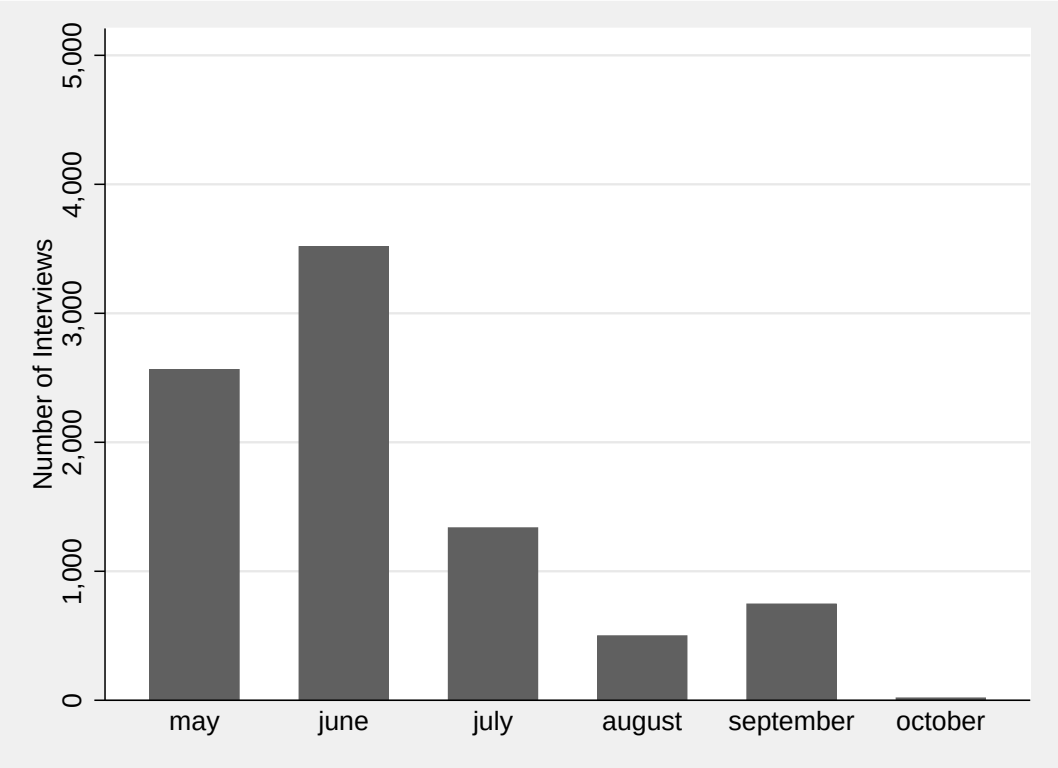


Figure 4: Wave 7 Young Person Interviews by Month



2.2 Sampling

The LSYPE was sampled using a probability proportional to size method, using schools as the primary sampling unit. It was additionally stratified on deprivation levels of those schools, oversampling more deprived schools and oversampling pupils from minority ethnic groups (Department for Education, 2011a, pp. 7-12). After stratification, "the school selection probabilities and the pupil selection probabilities ensured that within a [school] deprivation stratum, all pupils within an ethnic group had an equal chance of selection" (Department for Education, 2011a, p. 7).

This sampling method means that the sample size from each school varies depending upon the ethnic composition of the school. In initial sampling there were a mean number of YPs per school of 33.3, however this was significantly reduced by unit non-response. As such, in Wave 1 the mean number of responding young people per school is 24. By Wave 7 this has fallen further to 13.2.

As a result, it is important to take this survey design into account in any analysis. The deliberate skew introduced to achieve reasonable minimum sample sizes across deprivation strata and ethnic groups could result in highly misleading point estimates without weighting, while the standard errors would be too small without adjusting for the clustering of individuals in schools.

This may be done using variables provided in the data indicating the primary sampling unit (i.e. the school) and the stratum the YP is drawn from. Design weights are provided in Wave 1 to return the panel to representative proportions of individuals from each ethnic group and deprivation stratum (Department for Education, 2011a, pp. 55-76).

At Wave 4, an ethnic minority boost sample of six hundred Black African and Black Caribbean young people was added. This sample was selected from schools who did not co-operate in initial sampling. This aims to increase the sample size which research focussing on these groups has to work with.

2.3 Attrition and Non-Response

As with almost any longitudinal survey, the LSYPE experiences attrition and non-response across its waves. Initial unit non-response is 26%, with 21,000 individuals sampled but only 15,570 partial responses achieved (of which 13,914 are full responses, including both parents and YPs) (Department for Education, 2011a, p.13).

There is also a high level of non-cooperation by schools. Collingwood et al. (2010, p.52) calculates that if we include notional individuals from uncooperative schools in our figure for initial unit non-response the response rate falls to 53%. That is, only just over half of those individuals who should have been, or ultimately were, invited to participate ended up responding to a Wave 1 questionnaire.

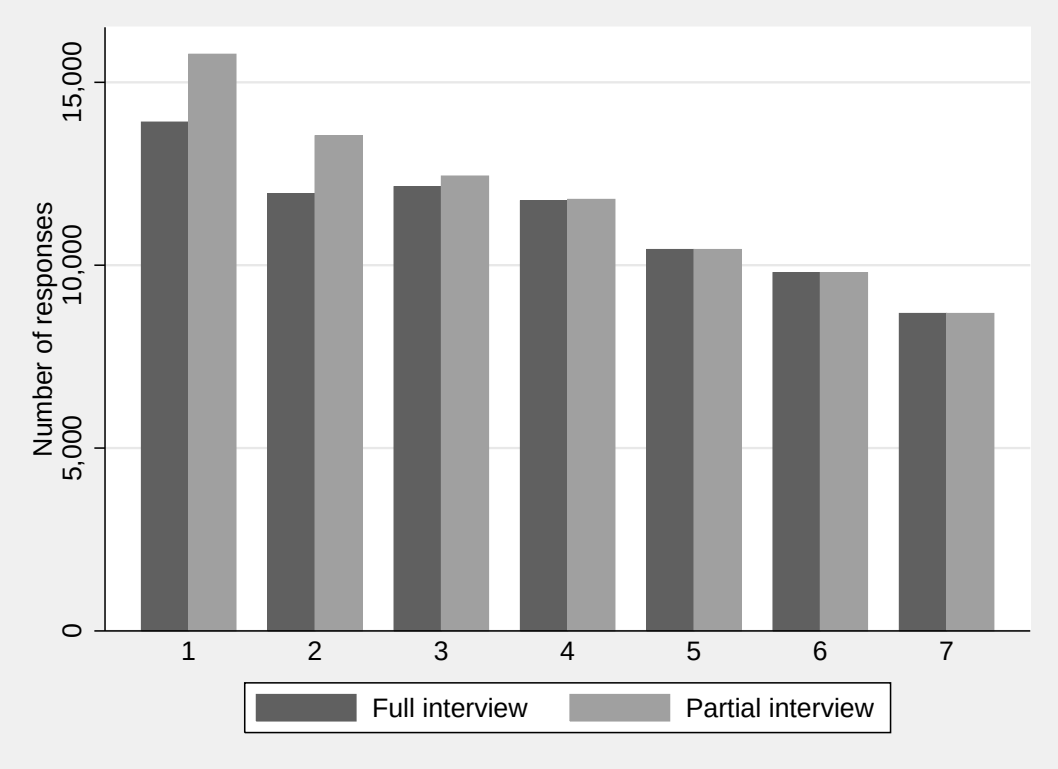
This initial non-response of individuals and non-cooperation of schools may be more difficult to deal with than later attrition. This is because there are no survey responses with which non-response can be modelled. The weights for this provided in the LSYPE rely on cell weighting using the school's deprivation status (a binary variable based on the proportion of pupils entitled to free school meals) and its region (specifically whether it is in London or not) (Department for Education, 2011a, p.56-57). Other characteristics were assessed, but found not to be significant predictors, including the proportion of pupils from non-White ethnic groups and the proportion with 5 or more GCSEs at grades A* to C.

In the case of individual non-response, almost all YPs are linked to the National Pupil Database, allowing for some analysis using administrative data from this source. The model providing the LSYPE's supplied Wave 1 weights includes region in which the individual lived, their ethnicity and a three category coding of their GCSE performance (Department for Education, 2011a, p.57). Those for whom no match in the NPD was found (around 3%) were given an average weighting.

From this point, the total number of full responses falling from 13,914 in Wave 1 down to 8,682 by Wave 7 (see Figure 5), a fall of just under 40%. This should not be too much of a problem if those dropping out of the survey are doing so randomly, although it may have some impact on the asymptotic efficiency of estimates of population parameters (Griliches et al., 1978).

More concerning is if such drop out is correlated with factors relevant to HE participation

Figure 5: Full and Partial Response to the LSYPE, by Wave



Notes: Partial response is considered to be any wave where at least one respondent (e.g. main parent, YP) completed a valid survey. Since there is only one interview in Waves 5, 6 and 7 full and partial response are the same by definition.

questions and explanatory variables included in analysis, potentially biasing estimates of variables of interest. This being true is distinctly plausible for at least some research questions in this area. Much of this bias may be reduced by using the non-response weights included with each wave of the dataset (Department for Education, 2011a, pp. 55-76).

However, weights can only be modelled on the basis of observable characteristics, so can only reliably control for attrition explained by such factors. Of course, relevant unobservables may be correlated with these, but we cannot be sure of this. As such, it may also be necessary to consider the use of selection methods such as those suggested in Heckman (1979) to model the probability of attrition and hence control for this in calculating estimates. Miranda and Rabe-Hesketh (2010, pp. 19-20) suggest that survey company engaged to provide the LSYPE fieldwork may provide an instrument for such selection models.

Further details on missing data in the LSYPE may be found in Piesse and Kalton (2009).

3 Previous Research, Datasets and Literature

Repeated government policy changes and underlying concern in goals such as social mobility have motivated a great deal of research into the area of HE participation (The Sutton Trust, 2008; Vignoles et al., 2008; Gayle et al., 2002). Concern that those from more disadvantaged backgrounds are underrepresented in HE, particularly among high-status institutions, has led to analysis that attempts to quantify and uncover the causes of this.

3.1 Administrative Data

One source of data used to examine these effects in the past has been individual-level linked administrative data. For example, Chowdry et al. (2010a) build a model of HE participation from administrative data by linking the Department for Education's National Pupil Database (containing information on prior attainment, limited socioeconomic background and other potential controls) with Higher Education Statistics Agency (HESA) records on university attendance. This has the advantage of a huge sample size, but the drawback of very limited information on an individual's family background (typical in administrative data), hence making it potentially difficult to control for this.

The authors were able to state with confidence their core finding of large gaps in HE participation being substantially reduced once prior attainment had been controlled for, but were less able to examine the processes leading to this earlier in secondary school.

Furthermore, the data source used gave the study a purely zero-one indicator of HE participation, not allowing us to explore at what stage in the application process gaps do emerge. Linkage to Universities and Colleges Applications Service (UCAS) data would solve this particular problem but is not possible at present.

While the LSYPE does not have the sample sizes that administrative data have, it is able to cover some of these data shortfalls, while maintaining the "extremely detailed information" (Chowdry et al., 2010a, p. 3) on prior attainment used by this study, through links with the National Pupil Database.

3.2 Youth Cohort Studies

In other papers, the Youth Cohort Studies (YCS) (Department for Children Schools & Families, 2008, for example) have been used for analysis of HE access. The YCS are a set of longitudinal studies, beginning in 1985 following young people, initially from England and Wales but now only from England, from the minimum school leaving age of sixteen for roughly three years hence. Each cohort consists of three postal surveys undertaken across three sweeps, sometimes annually or sometimes bi-annually.

Marcenaro-Gutierrez et al. (2007) use cohorts 6, 7, 8 and 9 (corresponding to university entry years of 1994, 1996, 1998 and 2000) of the YCS to describe HE participants during this period. However, similarly to studies in section 3.1, the authors note the absence of key explanatory variables, such as family income. There is a distinct possibility that this leads to omitted variable bias in the estimates produced by their models. While it is conceptually difficult ever to claim to have all explanatory variables needed, the LSYPE covers much broader topic areas and over a longer timescale, meaning this barrier to estimating causal impacts should be overcome to some extent.

Unlike for administrative data studies, there is also a large problem of attrition in the YCS, potentially biasing results. While all longitudinal surveys suffer from attrition, it appears particularly large in the YCS with, for example, only 28% of the initial target sample for

ages 16-17 still participating at the 18 year old sweep in cohort 9 (Marcenaro-Gutierrez et al., 2007, p. 338).

The YCS also only consists of interviews with the young person; there are no questionnaires to or interviews with parents. This likely gives poor quality and/or missing information on important household characteristics, such as parental occupation. Using PISA data, Kreuter et al. (2010) analyses the problems with children's reports of parental characteristics and the measurement error. In particular, it notes "a negative relationship between the presence of reporting errors and test scores" (Kreuter et al., 2010, p. 131). This suggests that YCS estimates may well be biased by such reporting errors.

While the YCS is often able to provide very timely analysis of HE access, given its regularity, it falls short of the LSYPE in several respects. Its shorter timeframe, lack of parental interviews and high levels of attrition all give clear reasons to prefer LSYPE data.

3.3 National Birth Cohorts

Earlier research into university participation used the national birth cohorts, the National Child Development Survey 1958 (NCDS58) and the Birth Cohort Study 1970 (BCS70). These followed, and continue to follow, individuals all born around the same time through early childhood, into education and thence into the labour force. These studies have some informational advantages compared to the LSYPE. They follow participants from birth (rather than the age of 13-14) and contain information gathered contemporaneously during early childhood, including testing aimed at measuring ability rather than attainment. However, they are now rather out of date for the purpose of considering university participation.

The institutional features of HE were significantly different at the time their surveyed individuals would have entered (1975-6 and 1988-9), with a much smaller share of the population remaining into this phase of education: at age 19 for each of these cohorts the UK Age Participation Index (API) was around 13% and 18% respectively (Elias and Purcell, 2004, p. 4). Unfortunately, for comparison purposes, the Department for Education and Skills replaced the API with the Higher Education Initial Participation Rate (HEIPR) in 2001, ceasing to collect the different data required for the API and hence no longer calculating it. As a re-

sult, directly comparable figures for the LSYPE university participants' predominant entry years are not available.

The API is calculated by dividing the number of initial entrants to HE aged under 21 by the population of Great Britain aged 18-19 divided by two (Elias and Purcell, 2004, p. 4). The HEIPR, however, is calculated at each age between 17 and 30 by dividing the number of initial entrants to HE courses[3] by the total population of Great Britain at that age; the "headline" HEIPR figure is the sum of these individual age HEIPRs.

To construct a more comparable figure (henceforth referred to as an API-HEIPR) the number of Initial Entrants aged between 18 and 20 is added together and divided by the population of 18 and 19 year olds, and again divided by two. For 2008-09 this yields an API-HEIPR of 35.9% and for 2009-10 an API-HEIPR of 36.9% (Department for Business Innovation & Skills, 2012). These figures are still not comparable due to other potential differences in eligible population between the two measures, however we can be confident that a hypothetical API for these years would be significantly higher than it was for the NCDS and BCS years.

Individuals in the next national birth cohort study, the Millennium Cohort Study, will not reach university age until 2018 at the earliest.

In summary, the national birth cohorts have a great many strengths. However, none of them (yet) cover the current period of mass HE participation. Thus, none of the birth cohort studies are able to provide data for analysis of the current institutional setup, as the LSYPE can.

3.4 ALSPAC

Another potential dataset is the Avon Longitudinal Study of Parents and Children (ALSPAC). Described as a study of children's health and development, it invited all prospective mothers due to give birth between 1st of April 1991 and 31st of December 1992. 85% of these consented, resulting in an initial sample of almost 15,000. Data are collected (through both questionnaires and biological sampling) on those parents and their children from pre-

[3]Technically the HE Courses must also be expected to last longer than six months and the individual enrolling on them must be enrolled on the course for more than six months.

childbirth to the present day.

There are some potential problems with ALSPAC usage in this research area. Previous studies have noted weaknesses in income measurement (Burgess et al., 2004, p. 9), with individuals' responses recorded only in five large bands. While data collection is frequent (compared to the national birth cohorts), income data is not collected at every wave, reducing this advantage.

In addition, ALSPAC is, as its name suggests, not a national survey; it only surveys parents and children from Local Authorities which make up the former county of Avon. While cross-checking of observable characteristics against those of census data suggests is it relatively representative of the national picture, some differences are found suggesting a slightly lower proportion of low income individuals (Burgess et al., 2004, p. 6). There may also be some other unobservable differences.

Given ALSPAC's many strengths it seems likely that it will become a very useful resource for research questions into access to HE within the next few years. However, questions over income data suggest, at the very least, a role for the LSYPE in examining research questions relating to this. Furthermore, we might still want data from a national sample, to ensure it is representative.

4 Admissions Process

In any statistical analysis it is important that outcomes are comprehensively and accurately measured. In this case, we want to be sure that the LSYPE is able to provide accurate information on what actually happens to the young people as they proceed through the university admissions process.

The English university admissions process, almost exclusively organised through the Universities and Colleges Admissions Service (UCAS), consists of several stages. The more comprehensive the LSYPE's questions and linked data describing YP's progress through the application process the wider the set of possible research questions that can be meaningfully answered.

4.1 The English HE Admissions System

The mainstream route into HE, through UCAS Route A, allows individuals to make an application to up to five universities. Those universities are provided with information on the individuals including GCSE results, AS Level results (or A Level results if this is a post-Year 13 application) or equivalents and a personal statement. Such applications must be made by a deadline of mid-January, although for some courses (medicine, dentistry, all courses at Oxford and Cambridge etc.) this deadline is brought forward to mid-October. For some other courses (mostly art and design courses) the deadline is extended to late March. Universities can consider applications made after these dates, but only if there are still vacancies on the relevant course.

Universities may also undertake additional selection processes, such as selection tests and interviews. Once these are complete, the universities each decide whether or not to offer that individual a place, either conditional on certain grades or unconditionally. After all replies have been received, the applicant may choose one of these offers as a "firm" offer and, optionally, another as an "insurance" offer (normally with lower conditional requirements).

If the applicant did not receive any offers, or chose not to accept any of the offers they received, they may apply for additional courses that still have vacancies, one at a time, through UCAS Extra. An individual can only choose to enter UCAS Extra once all applications have been unsuccessful or they have declined any offers they did receive. This option is open between the end of February and a date in early July.

For most applicants holding offers conditional on their A Level results, news of whether they have been deemed to have fulfilled either their firm or insurance offer will come on their A Level results day in mid-August.

At this point, if an individual has met the conditions of their Firm offer this place will be confirmed. Likewise, if they have met the conditions of their Insurance offer (but not the conditions of the Firm offer) this place will be confirmed. In both cases the YP is committed to take up the place, with the only alternative being complete withdrawal from UCAS for that application cycle.

A new feature called Adjustment, slightly complicating the above, was introduced in the

application cycle for 2009 entry. As such, this affects members of the LSYPE cohort who applied to university during the academic year following their A Level results. It does not affect those who applied while still at school for either an immediate start or deferred entry place. Under Adjustment, if YPs exceed the conditions of their Firm offer, they can apply for other courses over the space of five days without risking their confirmed place. If they apply to and are accepted onto another course they are then committed to this choice, otherwise they remain committed to their Firm offer.

If they did not fulfil either of their offers, or for those still not holding any offers at this point, applicants can choose to enter Clearing. In Clearing individuals apply (with their grades in hand) to any courses still with vacancies. In principle, Clearing is open between July and September, however it is only widely used after A level results are published.

Finally, virtually all YPs with a fulfilled offer, an Adjustment offer or who gained an offer through Clearing begin their courses.

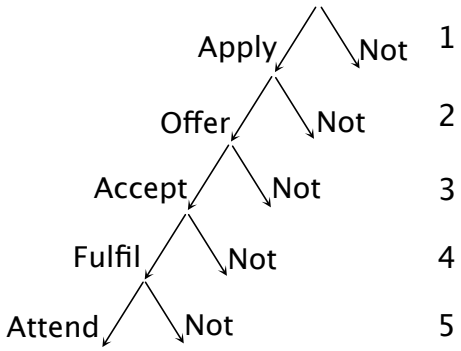
4.2 Measurement of Admissions

This is a necessarily complex system, every last nuance of which the LSYPE does not, and was never likely to, cover. Instead, what we really want to assess is whether it can provide the information needed for acceptably abstracted models. For example, it would seem sensible to model the admissions process as a simplified sequential structure, removing complexities such as UCAS Extra and Clearing, and indeed the complexities inherent in applying to multiple universities and multiple courses at once.

A simplified model is illustrated in Figure 6. This includes firstly whether individuals apply or not; if they have applied, whether they have received an offer or not; if they have received an offer, whether they accept at least one of these offers; if they accept an offer, whether they fulfil the terms of at least one of their offers; and if they fulfil, whether they do indeed attend university in the end.

The LSYPE provides the basic details of the university admissions process for its young people well, but not perfectly. We should know from all young people if they did attend university in the two years following school or college, with any missing data here only reflecting attrition and non-response more generally (see section 2.3 for a discussion of

Figure 6: Simplified Admissions Model



this).

The question of whether a person is at university is asked in both Wave 6 and again in Wave 7, allowing us to know at what point they entered HE. Data on the name of the university attended (where applicable), is available but only under special licence.

It may well be objected, with good reason, that only knowing about those who attend university in the two years after they leave school (i.e. either going straight from school to university, or taking a single gap year) excludes a potentially large and interesting population. However, over the past few years it has always been the case that over 60% of accepted UCAS applicants are aged under 20, corresponding to the two years after leaving school (UCAS, 2011a) (see Table 1), meaning that this population of later entrants is certainly the minority, albeit quite a sizeable one.

Secondly, given that the LSYPE is no longer planned to extend beyond Wave 7, in the absence of linkage with administrative data (either UCAS data, which could give the complete admissions process, or HESA data, which could only give final attendance) there is no gain to waiting for further data collection. So, while any analysis must be clear about only including this partial picture, most would agree that this is better than nothing.

The questions also ascertain what point in the application process YPs who do not end up going to university got up to, covering the major milestones of putting in an application, receiving any offers, and accepting an offer. These questions are asked at Wave 5, for those who continue straight from A Levels to University, again at Wave 6, for those who apply the following year and finally in Wave 7. The proportion of applicants making it to these milestones is detailed in Table 2.

Table 1: UCAS - Percentage breakdown of the age of applicants who received offers, for the years 2005 to 2010

Age	2005	2006	2007	2008	2009	2010
17 and under	2.3	2.3	2.2	2.1	2.0	1.9
18	46.4	47.1	44.9	43.6	42.7	42.3
19	21.1	20.4	22.5	21.9	22.3	23.0
20	8.3	8.2	8.6	9.2	9.1	9.3
21	4.6	4.7	4.7	5.0	5.2	5.1
22	1.6	3.0	3.2	3.2	3.2	3.3
23	3.1	2.1	2.1	2.2	2.3	2.3
24	2.2	1.6	1.6	1.6	1.7	1.7
25-39	4.5	4.5	4.4	4.8	4.9	4.8
30-39	4.0	3.9	3.7	4.1	4.2	4.2
40 and over	2.0	2.1	2.1	2.4	2.5	2.3
Total	100	100	100	100	100	100

Notes: Columns may not sum exactly, due to rounding. Source: Universities and Colleges Admissions Service (2011a)

Table 2: Percentages of Young People Achieving Key Application Milestones

	Overall	Female	Male
University apply	50.5 (0.94)	54.3 (1.21)	46.5 (1.27)
Sample size	*8677*	*4476*	*4199*
Receive offer	47.6 (0.94)	51.2 (1.23)	43.9 (1.28)
Sample size	*8677*	*4476*	*4199*
Accept offer	45.8 (0.95)	49.3 (1.26)	42.1 (1.27)
Sample size	*8677*	*4476*	*4199*
University attend	38.6 (0.90)	42.1 (1.21)	35.0 (1.22)
Sample size	*8677*	*4476*	*4199*
Uni. attend, conditional on applying	76.5 (0.78)	77.5 (1.02)	75.3 (1.21)
Sample size	*5323*	*2905*	*2416*
HE attend	43.3 (0.92)	46.7 (1.21)	39.7 (1.24)
Sample size	*8677*	*4476*	*4199*
Russell Group attend	9.5 (0.52)	10.4 (0.75)	8.6 (0.72)
Sample size	*8665*	*4470*	*4193*
Russell Group attend, conditional on uni.	24.6 (1.02)	24.7 (1.42)	24.4 (1.51)
Sample size	*4194*	*2315*	*1878*

Notes: Standard errors in parentheses. Weighted using Wave 7 LSYPE Weights, which attempt to adjust for oversampling and attrition. Application, Offers, Acceptances and Attendance calculated across Wave 5, 6 and 7. Sample: All Wave 7 respondents.

It is important to note that, because of the longitudinal nature of the data, the selection of which waves' measurement of application and attendance should be used in modelling varies according to the particular question of interest. To illustrate this, let us consider a few likely examples.

If our research question only considers whether individuals apply to university we can use all the application data in Waves 5, 6 and 7. It doesn't matter that we will not observe whether Wave 7 applicants are successful in getting a place.

If we are only interested in whether university is attended we can use attendance data in Waves 6 and 7 (it is not measured in Wave 5, and very few YPs are likely to attend a year early in any case).

If, however, we are interested in a sequential model of attendance conditional on application, it is only possible to use application data from Waves 5 and 6 with attendance data from Waves 6 and 7. Those who are first seen to apply in Wave 7 are treated as not applying. The reason for this is that otherwise those who indicate that they are applying for a place during Wave 7 (who then may get a place in a notional Wave 8) will be counted as having failed to get a place when, in reality, this is not observed. This is the approach taken for the purposes of any figures in this paper.

Finally, let us turn to the penultimate decision point in our basic model, that of offer fulfilment. Unfortunately, although the question of whether the individual fulfilled his/her offers is asked indirectly (and somewhat imperfectly) at Waves 6 and 7, it is only asked to those who are in HE. This leaves us unable to tell apart those who fail to fulfil their offers and those who simply decide not to attend university at the last minute.

Looking at proportion of YPs who reach each stages of this admissions model (see Table 2) reveals that very few applicants receive no offers and very few individuals who receive an offer do not accept any of them. As such, analysis on these separate sections of the would be virtually impossible due to small sample sizes.

Although the LSYPE does cover most elements of this basic model, we may wish to introduce some nuances to such a model, accounting for factors such as the quality of university applied to. The LSYPE does not go into as much detail as would be ideal in this respect, nor have linkages, such as with UCAS data, been achieved that would allow such information

to be obtained from these sources.

For example, it would be useful to know details of which universities young people had applied to; which (if any) they had received offers from; and which (if any) of these offers the YP had held on to as their firm and insurance offers. Instead, in Wave 5 the young people were asked how many universities they applied to and the name of only the firm offer they held on to[4].

These questions were then not repeated for those going through the application process during Wave 6, meaning even this limited data is missing for those who chose to make their application to university in the year after their A Level results. Inclusion of such additional information would, for example, allow a model to include considerations of the rate of success for individual applications and consider success rates for particular groups of institutions of interest, such as the Russell Group universities. These omissions could be rectified were data linkages with UCAS data achieved, however it does not seem likely this will happen in the immediate future.

Thus, the LSYPE seems to provide the information for a basic model of university admissions, with the exception of offer fulfilment, but not for the ability to explore potentially more complex models. While this may prove to cause some limitations, overall there seems a good case that this still represents strong measurement of the picture of admissions for our population of interest.

4.3 Comparison

The Higher Education Initial Participation Rate (HEIPR)[5] for English domiciled young people ages 17-19 for 2008/09 is 32.9% and for 2009/10 is 34.1% (Department for Business Innovation & Skills, 2012). Since the LSYPE observes entry in both of these academic years we might expect its estimate to fall somewhere between these two. Instead the LSYPE HE attendance measure is notably larger at 43.3%.

Official figures from the LSYPE and YCS released by the Department for Education (2011b, p.16) give the HE participation rate at age 19 as 40%. This is slightly lower than my HE

[4] As with university name above, access to this variable is only available under special licence.
[5] See Section 3.3 for details on the calculation of the HEIPR.

attendance calculation, but still much higher than the administrative figures suggest. This figure only looks at Wave 7 (and its equivalent in the YCS), whereas my figure includes anyone who attended even if only for a more limited period. As such we would expect it to be slightly higher than this paper's figure because of drop out from HE between Waves 6 and 7. This suggests errors in analysis are unlikely to account for the differences.

Such drop out from HE is not negligible, particularly within the first year of study. Statistics from the HESA give non-continuation rates by young entrants after the first year as 6.5%. In addition, the rate of drop-out varies by socioeconomic status: for individuals from low participation neighbourhoods (using the POLAR2 measure of area-based participation rates, something of a proxy for socioeconomic status) it is markedly higher at 8.7%, while for all other neighbourhoods it is 6.2% (HESA, 2010). Therefore, differential drop-out could induce bias rather than just under-measurement; however, the direction is uncertain.

The HEIPR only includes individuals who undertake at least six months of HE. Exactly how long individuals have been studying for an HE qualification in the LSYPE is slightly less clear. Its responses are generally collected towards the end of an academic year, or into the summer break between academic years. However, the question is worded as follows: "Have you been studying for any [Higher Education] qualifications since September [last year]?"[6]. This leaves some ambiguity over whether the individual is still required to be studying for the qualification at this time (in which case we only capture those who last roughly an academic year on the course) or just has to have studied it at some point since that date (in which case we capture those who have dropped out, including after less than six months study). Either way, it seems unlikely the definition will correspond exactly with the HEIPR definition.

Turning to applications, UCAS state that for 2008 and 2009 [7] the number of applicants was 588,689 and 639,860, respectively, while the number of applicants finally accepted onto a course was 456,627 and 481,854. This implies that for 2008 77.6% and for 2009 75.3% of all applicants were finally accepted onto a course (UCAS, 2011b). This is definitionally different both from our definitions of both accepting an offer and actually attending, however there should be very little difference between those finally accepted and those who attend.

[6]The question as delivered includes the possible Higher Education qualifications by name, and includes the relevant year.

[7]A UCAS Year refers to the year in which applicants in that cycle begin HE. Hence UCAS Year 2008 should correspond to HEIPR Year 2008-09 and so on.

According to Table 2, the percentage of those applying is 50.5% and the percentage of those attending university is 43.3%. Hence the proportion of those who have applied that attend is $\frac{\text{Proportion Attending}}{\text{Proportion Applying}} = \frac{0.433}{0.505} = 0.857$. Again, this estimate is around 10 percentage points higher than administrative data suggests it should be.

UCAS also provide a measure of the proportion of the 18 year old population apply to university[8] (UCAS, 2012). For UCAS year 2008 this measure is given as 29.2%. To see whether this is similar to the LSYPE measure I cannot use the same university application rate as in the rest of this paper, as that includes Wave 6 appliers (the vast majority of whom will be aged 19). The university application rate from the LSYPE, based solely on Wave 5 (age 17-18), is 35.2%[9].

In summary, this suggests the proportion of YPs in the LSYPE applying to university, attending university, and the proportion of applicants successfully getting a place are all too high.

This leaves a major unexplained over-report of HE participation in the LSYPE. Possible reasons for this include the definitional differences between the LSYPE and the HEIPR (meaning, for example, that drop-out is captured differently); attrition on characteristics that the weighting scheme does not control for; and bias caused by initial non-response between sampling and Wave 1 (both caused by school non-cooperation and individual unit non-response, and again not accounted for by weights). Finally, it is never completely possible to rule out errors in analysis, however comparison with official DfE figures suggests these are unlikely.

5 Other outcomes

There are other outcomes beyond the admissions process itself that are of interest in relation to HE access.

[8] This measure only includes those who have applied by the January deadline, described in Section 4.1. As such, some additional applications could be made after the deadline.

[9] This figure is derived using the `w5heapplyyp` variable, provided in the public release edition of the LSYPE. It is weighted using the Wave 5 cross-sectional weight.

5.1 University attended

It seems likely that there will be differing rates of return to qualifications from differing universities (Chevalier and Conlon, 2003). As such, socioeconomic gaps in which universities individuals attend are of interest in considering university access. The LSYPE asks participants which university they are attending (if any) in Waves 6 and 7. There is a relatively high level of non-response for these variables, with just over 30% of Wave 7 participants who attend university not having their HE institution recorded.

The individual institution variable is not included in the general release of the LSYPE, but is available under special licence. Instead, flags indicating attendance at the universities of Oxford and Cambridge, and at a Russell Group institution are provided. While the small sample sizes involved for attendees at Oxford and Cambridge[10] make this variable not especially useful, the Russell Group indicator may be of more use.

The Russell Group is a group of research intensive HE institutions. It is made up of the following twenty universities[11]: University of Birmingham, University of Bristol, University of Cambridge, Cardiff University, University of Edinburgh, University of Glasgow, Imperial College London, King's College London, University of Leeds, University of Liverpool, London School of Economics and Political Science, University of Manchester, Newcastle University, University of Nottingham, University of Oxford, Queen's University Belfast, University of Sheffield, University of Southampton, University College London and University of Warwick. Although far from a perfect measure, Russell Group membership might be seen as a proxy of quality and hence associated with higher returns to HE.

9.5% of the LSYPE Wave 7 sample (after weighting) attend Russell Group institutions, representing just short of a quarter of those who go to university. Unfortunately, as noted in Section 4.2, we do not observe the institutions individuals apply to. This prevents us from considering issues such as different admissions process experiences by status of university applied to.

[10] Unweighted, 94 LSYPE cohort members are observed to attend either Oxford or Cambridge.

[11] As of March 2012 four additional institutions have joined the Russell Group. However, given the timeframe of the data collection, for our purposes the Russell Group is made up of its original twenty members.

5.2 Subject choice

In the same way as with university prestige, previous research has suggested differential rates of return to different subjects (Bratti et al., 2008; Conlon and Patrignani, 2011). As such, there is potential interest in considering HE subject choice. The LSYPE asks relevant YPs which HE subject they are studying in Waves 6 and 7. There is a very low level of item non-response with over 97% of Wave 7 participants in HE giving details of their subject.

These variables are not included in the general release of the LSYPE, but are available under special licence. Instead, a variable on broad subject groups, based on the Joint Academic Coding System (JACS) 2.0, is provided. However, given the wide array of subjects and the small sample sizes that would result for some of these, this is less of a loss than it might seem.

A group of subjects we might particularly want to consider are those making up "STEM". STEM stands for Science, Technology, Engineering and Mathematics, and these subjects are seen as critical to the UK's economic wellbeing.

Although there is some controversy, a reasonable definition of STEM subjects can be drawn from the JACS codings provided in the LSYPE. The resulting measure would allow consideration of research questions on the importance of family background to choosing a STEM subject for HE.

6 Family income

An estimate of family income in LSYPE is obtained through several questions, dependent on the type of employment (employed, self-employed, unemployed etc.) and hence sources of income (wages, benefits etc.) of the parent(s) in the household. This process was repeated, in slightly differing ways, between Waves 1 and 4 (see Table 3), with Wave 2 designed to be the most comprehensive attempt to assess family income.

In Waves 1 and 4 the main estimate of income comes from a single question to the main parent for themselves and their partner (where applicable). Levels of measurement error in cases such as this, where details of income for the whole household has been sought

from a single member, are thought to be particularly large (Micklewright and Schnepf, 2010) because of the large amount of information they are required to know. For example, it is particularly difficult where the main parent must attempt to work out how much their partner earns, when in some cases finances may in fact be kept quite separate.

Technical details regarding the construction of income measures used in this paper may be found in Appendix A.

Table 3: Income measurement at each wave

Wave	Income Measurement
1	Main parent only's estimate of gross income for themselves and partner and individual parent estimates of gross earnings and benefits
2	Main parent and second parent estimates of gross and net earnings and benefits
3	Joint interview (where possible and applicable) estimating gross income of main and second parent
4	Main parent only's estimate of gross income for themselves and partner
5	None
6	None
7	None

Table 4: Gross family income statistics at each wave

Variable	Mean	Median	Standard Deviation	Skewness	N
Wave 1	31,911	22,089	35,797	5.40	6343
Wave 2	35,342	29,161	31,045	7.31	6472
Wave 3	33,175	26,862	23,695	1.19	7092
Wave 4	33,515	25,753	25,708	1.24	7084
Permanent	32,283	25,726	24,491	2.35	8374

Notes: Incomes held constant at Wave 1 (2004) prices using Annual RPI. Weighted using LSYPE Wave 7 Respondent weights. Sample: Wave 7 respondents with valid income data from at least one of Waves 1-4.

6.1 Wave 1

In Wave 1, total family income was assessed in two ways. Firstly, a two part show card question, directed to the main parent[12]. It asked for an estimate for total gross income including "earnings from work, benefits and anything else". The first show card included amounts up

[12]The main parent is defined as the parent most involved in the YP's education.

to £36,400 per year[13] and if the top category was chosen, the second show card with bands up to £400,000 a year was presented to gather the appropriate information.

Gross earnings, calculated through hourly earnings rates and hours worked, where known, or alternatively income through self-employment (before and after relevant self-employment related deductions), are sought separately and from questions directed to each parent.

6.2 Wave 2

Wave 2 questions were more directed at specifically assessing earnings of each parent, through questions directly to each. No question on estimated overall gross family income was asked, instead there were questions to individual parents on gross pay, take home pay (including after application of tax credits) and benefits. If they are on a fixed hourly rate, this rate was also asked. If no response was gained for these more specific questions, questions on banded gross pay were asked. These banded questions were used in pre-release data cleaning to fill in approximate values for those who did not wish to respond to the more detailed questions.

As in Wave 1, self-employment income, if appropriate, was assessed through questions tailored to this employment status. Again, a banded estimate was requested for cases where no response was gained to the more specific questions.

6.3 Waves 3 and 4

In Wave 3, the documentation states that there is no second person interview, with individual parent questions being answered by the main parent on behalf of the second parent, where appropriate. However, the questionnaire states that if the second parent was present the individual parent questionnaire could be conducted with both parents jointly. As such, income is assessed with a single question to the main parent or to the parents collectively. They select a banded estimate of total gross ("before tax and similar deductions") family income from any sources. The top band includes all individuals with gross income above £52,000. Joint interviews only seem to have been conducted in roughly 5% of the Wave 3

[13]These options were given in weekly, monthly and annual amounts to assist the choice.

sample.

For Wave 4 the process was very similar to Wave 3. One exception is that the Wave 4 parental questionnaire does not discuss the potential for joint completion. However, the documentation does note that questions relating to the second parent (such as employment status) were to be addressed to that individual where this was possible. Since the section on family income followed this, it is plausible that joint estimation would occur in some cases. Unlike in Wave 3 no record is kept of whether the second parent was present.

6.4 'Permanent' Income

The way income is estimated varies from wave to wave, making each wave's variable not directly comparable. This has advantages and disadvantages. It is an advantage that there is more likely to be a wave with income measurement which suits differing research questions. However, incomparable data across waves precludes using the measures in a number of ways. For example, it is not possible simply to replace a missing income measure in one wave with the, appropriately inflation adjusted, income measure from another wave. Similarly, one cannot use differences in income between waves as an explanatory factor in any analysis.

However, if one assumes that families were treated similarly within waves and that there is no particular reason to believe the probability of response to income questions at each wave does not vary by family structure characteristics, this doesn't rule out averaging income across the waves. By this is meant calculating the mean of an individual family's responses across waves, excluding those where no response was gained.

While this may not be a wholly justified assumption there are several potential benefits to making it. Firstly, it reduces our missing data problem where different families' incomes were assessed in different waves. Secondly, it potentially reduces measurement error where families had their income assessed multiple times, by reducing the importance of any outliers in an individual's stream of responses. This benefit would not hold if responses are systematically biased in one direction or another. Thirdly, and importantly given the structure of the LSYPE, this allows us to use information from both Wave 2, when hopefully income was surveyed most thoroughly for initial respondents, and Wave 4, hence including

ethnic minority boost respondents.

Finally, averaging reduces the effect of transient changes in income that happened to be captured in a particular survey, getting closer to a measure of permanent income. Although far from a true measure of permanent income, getting closer to one may help to identify a greater proportion of the effect of income, since it has been suggested that "'[p]ermanent' income has a greater effect on outcomes than 'current income'" (Jenkins and Schluter, 2002, p. 2). A histogram showing the distribution of 'permanent income' calculated in this way may be found in Figure 8, while the difference in the cumulative distribution of the income measures across waves is shown in Figure 7.

Table 5: Proportion of Permanent Income values made up of each number of waves of data

Number of Waves	Percentage
1	6.3
2	11.9
3	28.3
4	53.4
Total	100

Notes: Adjusted using LSYPE Wave 7 respondent weights. Items may not quite sum to 100% due to rounding.

Table 5 shows the proportion of 'permanent' income observations that are made up from differing numbers of waves of income data. A majority of those still in the survey at Wave 7 have data from all four waves, and only just over 6% have data from only a single wave. It is reassuring that so few are simply relying on one observation.

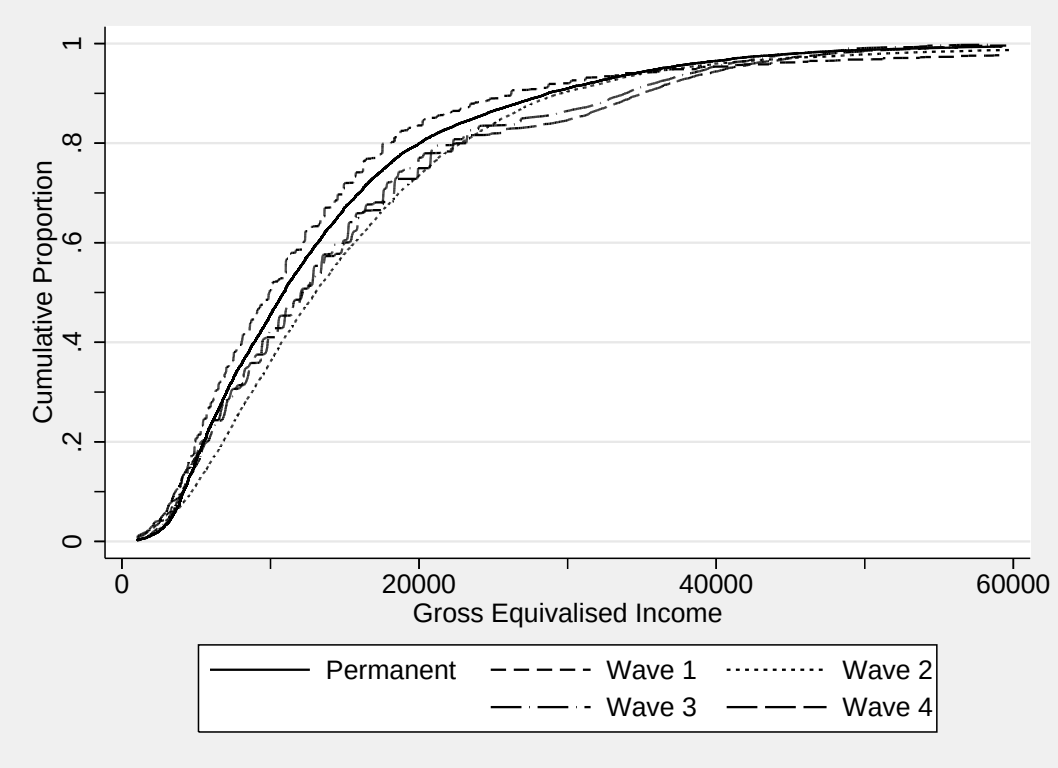
Table 6 reports the same income statistics as Table 4 but includes only those who answer income questions at all four Waves where it is asked about. Unsurprisingly, this subsample have higher mean and median incomes for each of the waves and the 'permanent' measure.

Table 6: Gross family income statistics at each wave for a fixed sample

Variable	Mean	Median	Standard Deviation	Skewness	N
Wave 1	35,132	27,258	35,989	5.18	4281
Wave 2	37,042	30,975	32,029	7.01	4281
Wave 3	35,960	31,534	23,694	1.10	4281
Wave 4	36,662	30,396	26,033	1.17	4281
Permanent	36,199	29,453	24,316	2.18	4281

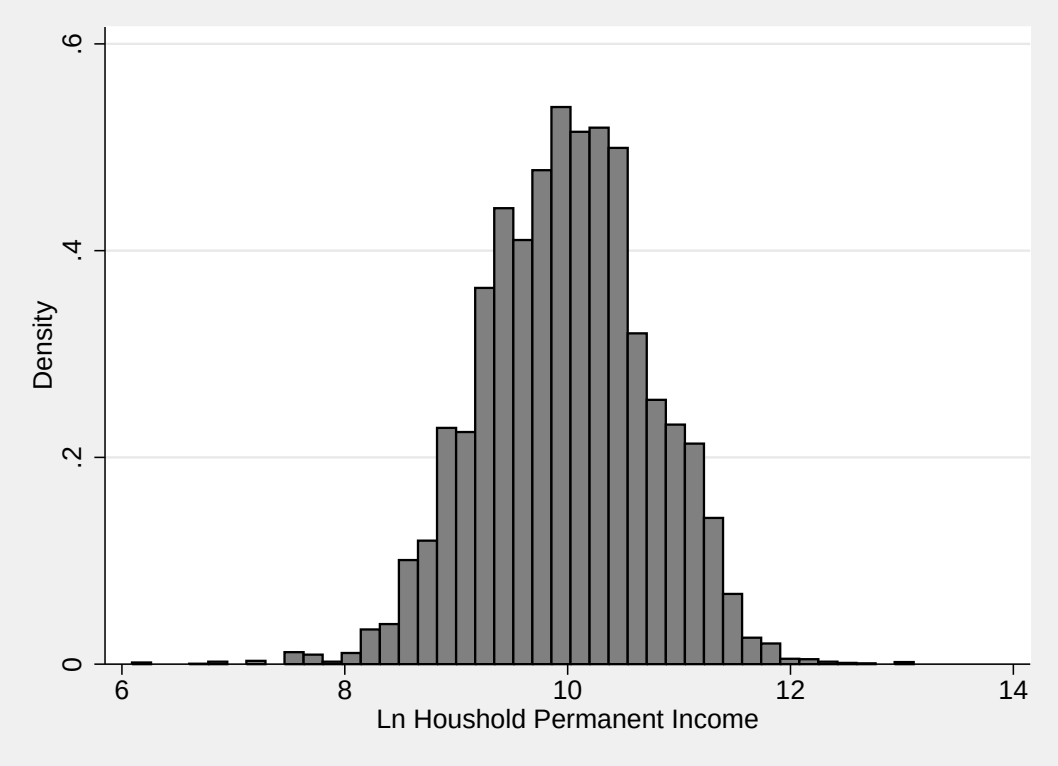
Notes: Incomes held constant at Wave 1 (2004) prices using Annual RPI. Weighted using LSYPE Wave 7 Respondent weights. Sample: Wave 7 respondents with valid income data in all of Waves 1-4.

Figure 7: Cumulative distribution of family gross income



Notes: Incomes were held constant at Wave 1 prices using the ONS Annual Retail Prices Index. Weighted using LSYPE Wave 7 Participant Weights. Sample: Wave 7 participants, however sample size varies by wave income measure.

Figure 8: Distribution of logarithmic 'permanent' family gross income



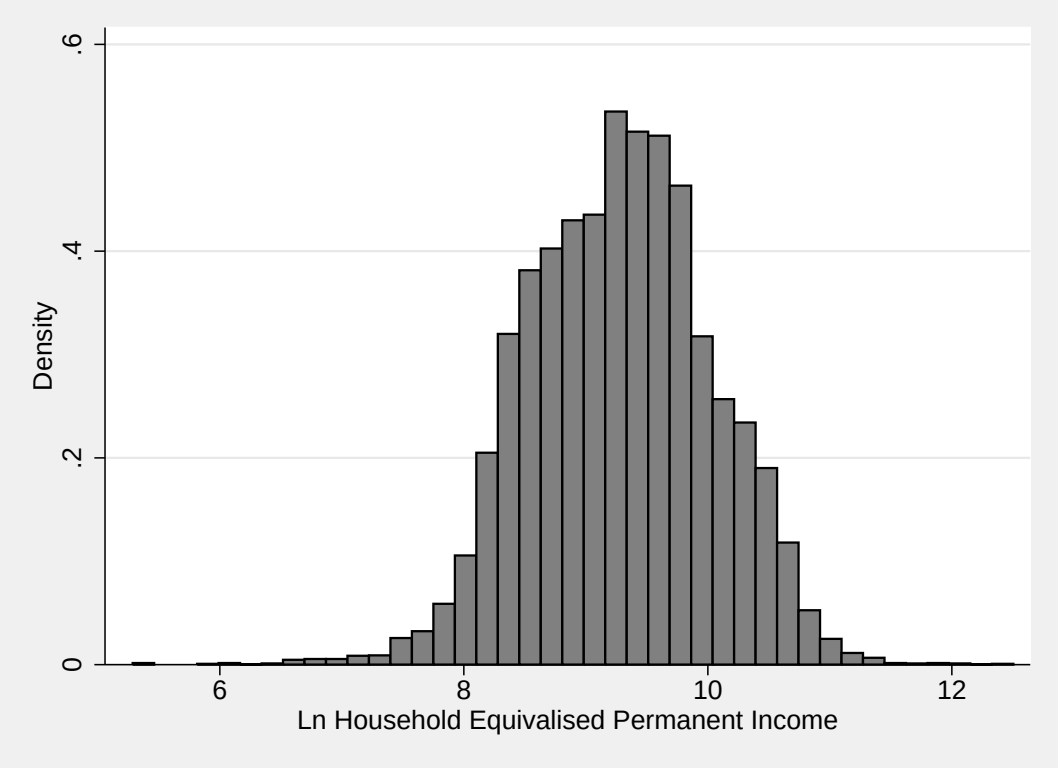
Notes: Calculated using Waves 1 – 4. Incomes were held constant at Wave 1 prices using the ONS Annual Retail Prices Index. Sample: Participants with valid income data from at least one of Waves 1-4.

6.5 Equivalised Income

Since we are, at least to some extent, concerned with the amount of financial resources available to help build human capital, how far that income has to be spread will also be important. To this end, family structure, including number of parents and numbers of siblings in the household, and their ages, is important to calculating an equivalised income. See Figure 9 for a histogram showing the distribution of permanent income equivalised by dividing by the square root of the number of resident family members.

These data on family structure are collected by the LSYPE, and may of course also be considered important factors in their own right; further discussion of the effects of family size may be found in section 7.2.

Figure 9: Distribution of logarithmic equivalised 'permanent' family gross income



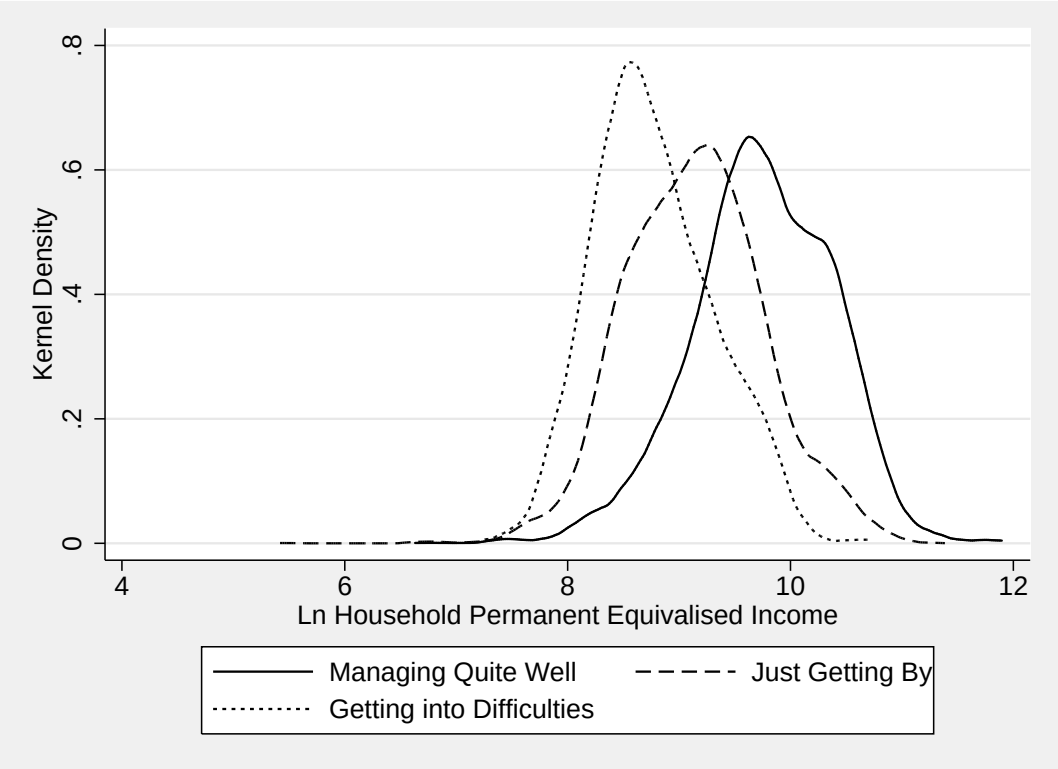
Notes: Calculated using Waves 1 - 4 and equivalised by dividing by the square root of family size at relevant wave. Incomes were held constant at Wave 1 prices using the ONS Annual Retail Prices Index. Sample: Participants with valid income data from at least one of Waves 1-4, valid data on number of resident siblings and valid data on number of parents in the household.

6.6 Other

How well the family feels they are managing on their income, whatever it is, is also asked in waves where a parent is interviewed. Seeing as family income does not necessarily reflect disposable income (although it will of course be correlated with it) this additional factor might be of some importance.

The question is posed as a choice between the three options "Managing quite well, able to save or spend on leisure", "Just getting by, unable to save if wanted to" and "Getting into difficulties". Figure 10 demonstrates the difference in income densities these correspond to.

Figure 10: Equivalised income density by how well household is managing on income



Notes: Equivalised permanent family gross income calculated using Waves 1 - 4. Incomes were held constant at Wave 1 prices using the ONS Annual Retail Prices Index. Question on how well household is managing on its income asked at Wave 1 (`w1managhhmp`). Sample: Wave 7 Participants with valid income data from at least one of Waves 1-4 and a valid response to question on how well household is managing on its income.

There is very significant overlap between all three groups. For example, around a quarter of those who report they are 'getting into difficulties' have larger equivalised incomes than the median family that self-identify as 'just getting by'. In the same way, roughly 15% of those saying they are 'just getting by' have family incomes greater than the median of those

who are 'managing quite well'. Finally, at the most extreme, around 5% of those who say they are 'getting into difficulties' have a greater equivalised family income than the median family reporting that they are 'managing quite well'.

One interpretation of this is showing the importance of taking into account disposable income. Alternatively, it may just demonstrate the subjectivity of the income that individuals perceive they need. Nevertheless, a clear trend towards "Managing quite well" is seen as income increases.

In addition to the income estimates, the LSYPE also includes information on whether the parent has ever purchased extra tuition for the YP. This could be seen to some extent as an indicator of the propensity of the family to use its financial resources for building young people's human capital.

6.7 Comparison with Family Resources Survey

In order to check that the income distribution generated through the above process, a simple comparison measure from the Family Resources Survey (FRS) has been derived for the same years. This generates what should be a comparable distribution. Summary statistics for each year of this variable may be found in Table 7.

Table 7: Income distribution statistics from Family Resources Survey for families in England with at least one child between the ages of 11 and 15

Year	Mean	Median	Standard Deviation	Skewness	N
2004	36,774	28,808	39,208	9.46	2675
2005	38,128	28,732	43,178	8.35	2587
2006	37,904	29,552	43,537	11.70	2392
2007	37,739	28,571	32,193	3.38	2157

Notes: Income is Total Gross Benefit Unit Income. All incomes in 2004 prices, adjusted using the ONS Annual Retail Prices Index. Weighted using gross3 grossing factor.

The comparative variables were constructed using the FRS derived Total Gross Benefit Unit Income variable (`buinc`), grossed up to be representative of the UK population using the "`gross3`" grossing factors provided, and adjusted using the same annual RPI correction factor to convert into Wave 1 (2004) prices as for the LSYPE measures above. A benefit unit is "an adult plus their spouse (if applicable) plus any dependent children they are living with"

(Palmer, 2012), hence this should correspond well with the LSYPE family income measures. For convenience I use the term family to mean benefit unit hereafter.

Only families with dependent children between the ages of 13 and 15, living in England (the FRS covers the whole of the UK) were included in the calculations to make the sample more comparable. The variables for this comparison were not equivalised.

To assess the measurement in the LSYPE with that in the FRS the (inflation adjusted) individual observations from all years are pooled and compared with the LSYPE permanent income measure. Various characteristics of the distributions are shown in Table 8. They are also shown graphically in Figure 11.

Table 8: LSYPE vs. FRS gross family income summary statistics

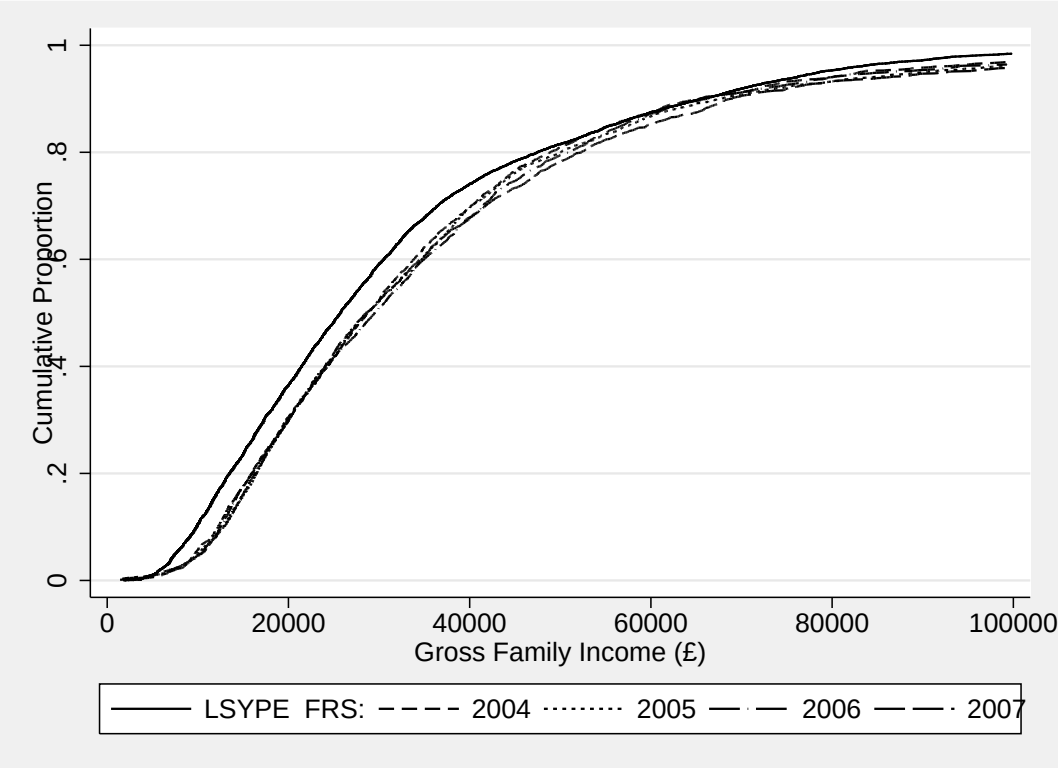
Characteristic	**LSYPE**	**FRS**
Mean	32,283	37,633
Standard Deviation	24,491	39,837
Minimum	451	141
Maximum	341,241	1,144,523
1st Percentile	4,917	5,463
10th Percentile	9,834	12,532
25th Percentile	15,434	17,998
Median	25,726	28,964
75th Percentile	40,961	44,772
90th Percentile	65,581	67,432
99th Percentile	110,354	170,484
N	8,374	9,811

Notes: All incomes in 2004 prices, adjusted using ONS Annual Retail Prices Index. **LSYPE**: Income is Permanent Family Gross Income, weighted using Wave 7 weights. **FRS**: Income is Total Gross Benefit Unit Income. Families with no children between the ages of 11 and 15 or outside England have been excluded. Weighted using gross3 grossing factor.

On Figure 11, the line representing the LSYPE is consistently to the left of those for the FRS until above the 75th percentile, or between £40-50,000. This suggests underreporting across most of the income distribution. Table 8 confirms this picture. The median of the LSYPE is just over 10% below that of the FRS. The LSYPE's mean is a slightly larger 15% below that of the FRS. Above the middle of the distribution the proportional underreport declines: it remains steadier in absolute terms with between £2-4,000 less income reported in the LSYPE than in the FRS.

The exception to this comes once we reach the very top, where the LSYPE's 99th percentile

Figure 11: Cumulative distribution of gross family income in LSYPE and FRS



Notes: Incomes were held constant at Wave 1 (2004) prices using the ONS Annual Retail Prices Index. **LSYPE:** Weighted using LSYPE Wave 7 Participant Weights. Sample: Wave 7 Participants with valid income data from at least one of Waves 1-4 and a valid response to question on how well household is manageing on its income. **FRS:** Income is Total Gross Benefit Unit Income. Weighted using gross3 grossing factor. Sample: FRS Households in England with at least one child between the ages of 13 and 15.

is a much larger amount below the FRS. Given that much of the measurement of income in LSYPE is based on banded questions, underestimation of the incomes of those at the very top[14] is to be expected.

There are also other differences in the income distribution, for example the LSYPE income data have a smaller standard deviation than that of the LSYPE. This will partly be driven by the fact that the FRS better accommodates top end outliers, as witnessed by its significantly higher maximum value. The FRS can also be used to compare another feature of the LSYPE's data that will be relevant to equivalised income, family size. In the FRS the mean family size is 3.8, whereas in the LSYPE it is significantly larger at 4.2. This seems to be driven by a wider spread distribution, although it can't just be explained by the LSYPE's longer top tail.

The LSYPE's apparent underreporting of income is perhaps not surprising since it does not put the same effort into measuring income as surveys where it is a particular focus, such as the FRS. As discussed above, in more than one wave income is collected using, to all intents and purposes, a single question. Finally, it should be remembered, of course, that the FRS is not 'the truth' against which we can compare. It has its own problems in terms of non-response and mis-reporting (Maher, 2006).

Nevertheless, the LSYPE seems to do a reasonable job of assessing family income, particularly compared to many of its predecessor cohort studies.

7 Socioeconomic Characteristics

Previous research shows a strong association between the socioeconomic and family background characteristics of YPs and their educational attainment in general (Björklund and Salvanes) and likelihood of attending university in particular (Galindo-Rueda et al., 2004). Household income and parental education play a particular role in determining young people's education in human capital theory, hence good measurement of each will be important to most economic models of university participation.

The LSYPE contains a wide range of questions on such characteristics, surveyed in particu-

[14]For those in the top income band, we only observe a lower bound, but no upper bound. This makes it harder for interval regression to place these individuals correctly.

lar depth at Wave 1 with a view to ascertaining family history, but followed up until surveys were concentrated purely on YP interviews after Wave 4.

7.1 Parental Education

The correlation of parents' education with that of their children can be down to a number of factors. Firstly, given heredity in genetic endowments, parents with greater innate ability are more likely to have children with greater innate ability, with the correlation between their education levels down to the correlation between ability and education for each party (Haveman and Wolfe, 1995, p. 1833). As such, interpreting the simple correlation between parental education and children's educational outcomes as being causal, and hence giving potential policy implications for raising educational levels of parents, would be incorrect.

However, there exist other potential causal routes through which an increase in parental education could have an impact on the education of their children. The first of these is so-called 'cultural transmission' in which increasing a parent's education makes them more likely to provide an environment conducive to the education of their children.

This could still be a potentially difficult causal route to isolate, since it also requires identifying whether it is the case that an increase in education will spur parents to improve the home environment with respect to education, or alternatively if it is the same underlying character of the parent that drives them both to seek more education and to provide a home environment conducive to education.

Similarly, parental education could have a causal effect on children's education if increased parental education affects the opportunity cost of educating their children (Ermisch and Francesconi, 2001, p. 139). The effect could be argued to be in either direction. If greater parental education leads to higher labour market rewards, the substitution effect of the higher wages they earn could lead to a reduction in the time a utility maximising parent would spend on educating their child.

On the other hand, if increased education of their own makes it easier and more effective for parents with greater education to pass on knowledge and skills to their children, this would reduce the opportunity cost of imparting the same level of education. Given the

literature and empirical findings relating to household income (see section 6) the latter of these two effects would seem to be more plausible, or at least more important. These routes are subject to the same caveats as with 'cultural transmission' above.

Basic information is gathered from each parent on their education, such as school leaving age, leaving age of additional full time education and qualifications gained. A small snapshot of information on earlier generations' education is also gained, through questions on whether the parents' mothers and fathers have degrees.

7.2 Family Size and Birth Order

Studies relating the impact of family size and birth order (Iacovou, 2008) to children's educational attainment, whether through the 'confluence model' or the 'resource dilution model', suggests the importance of considering such factors in models of university attendance.

Some theoretical predictions have been given empirical backing, such as that "middle and youngest-born children perform significantly worse throughout their lives than first-born children. The magnitude of these effects varies according to sibship size" (Iacovou, 2008, p. 19). Similarly, "[although] family size affects children from different socio-economic groups in different ways, there appears to be no heterogeneity between socio-economic groups in the effects of birth order. However, there is evidence that the effects of birth order are heterogeneous between the sexes" (Iacovou, 2008, p. 19).

The LSYPE allows us to build a complete picture of both of these factors, with data collected on the number of older and younger siblings at Wave 1, and hence either to control for or investigate these effects. The LSYPE also collects data on whether siblings are resident or non-resident at this stage, potentially allowing for investigation of whether this affects the operation of the theoretical models.

7.3 Other Socioeconomic Status

From a more sociological point of view, the importance of mother and father's occupational class may be considered a potential factor in access to HE. In the LSYPE this is measured

through the National Statistics Socioeconomic Classification (NS-SEC) of each parent's current job (measured at each wave from 1 to 4, though with the same caveats regarding interviews as discussed in section 6), allowing for not only the individual parent's classification, but also the highest classification in the household. Similarly, more general employment status, housing tenure and lone parent status are measured repeatedly in the first few waves and to some degree in the historical questions, along with mother's age.

In addition, summaries of employment history, for purposes such as ascertaining periods of unemployment in the household, are sought. Both of these may form explanatory factors of interest. Response rates to these questions at some stages of the survey are over 90% for the main parent and 60% for the second parent.

The LSYPE's measurement of many socioeconomic characteristics is likely to be rather more accurate than asking the YP about such characteristics. This compares favourably with previous survey data, such as the Youth Cohort Study, in which questions were only asked of young people themselves. Reasons for this were discussed in section 3.2.

8 Attitudes, Aspirations and Expectations

The association between, and potential causal impact of, the attitudes, aspirations and expectations of young people and their parents and educational outcomes has increasingly become a source of research interest. Although such mental states will only affect such outcomes through how they change behavioural decisions and are likely to be partially determined by other background characteristics, they may nevertheless continue to have impacts if those characteristics are also controlled for.

8.1 Attitudes towards Education

There are also questions in Wave 4 on more abstract views of HE. These include beliefs regarding whether the YP needs a university education to pursue their preferred career path (should they have one yet), whether in general a degree leads to better paying jobs later, whether the "best jobs" go to those who have been to university, whether their friends are planning to go to university and whether "people like me" go to university.

Other potentially interesting attitudes questions ask about the YP's attitude towards education in general and their schooling in particular. These may be seen as tracking, to some extent, their valuations of the intrinsic and extrinsic rewards of education.

However, it should be noted that using these answers in models of HE participation leads to a strong risk of endogeneity bias: an individual's beliefs about their likelihood of applying to university are themselves likely to affect their attitudes towards education and their schooling, resulting in a correlation between the explanatory variable and the error term in any regression model including them. This violates one of the fundamental assumptions of regression techniques and leads to biased estimates. To put this another way, the problem stems from the likelihood that as those intending to undertake HE get nearer to this stage of their education, they are likely to adjust their attitudes in order to fit in with the realities they face. Such models must, therefore, be treated with great caution.

8.2 Attitudes and Risky Behaviours

Previous research using the LSYPE has considered the associations of attitudes and expectations with GCSE results noting a seemingly positive impact on GCSE results if:

> "parents think it likely that they will go on to HE; spend time sharing family meals and outings; quarrel with their child relatively infrequently; devote material resources towards education including private tuition, computer and internet access; the child: has a greater belief in his or her own ability at school; believes that his or her own actions make a difference and that he or she can control events that affect him or her (captured in this study by having an 'external economic locus of control'); finds school worthwhile; thinks it is likely that he or she will apply to, and get into, HE; avoids risky behaviours such as frequent smoking, cannabis use, antisocial behaviour, truancy, suspension and exclusion; does not experience bullying" (Chowdry et al., 2010b, p. 34).

While the factors Chowdry et al. have been able to examine depend upon the fact that they were using the LSYPE, they give no indication they feel that this imposed unacceptable constraints upon their research. Given the plausible assumption that such factors continue to affect academic success, including application to and attendance at university, these

variables continue to provide a potential question of interest.

A useful model of how such attitudes, aspirations and expectations can act as transmissions mechanisms, stemming both from observable family background / socioeconomic factors and other unobservable characteristics is described in Chowdry et al. (2009, p. 12) and reproduced in this document (see Figure 12), as adapted by Jerrim (2011, p. 39).

8.3 HE Preferences and Expectations

Questions specifically related to university application and preferences begin all the way back in Wave 1[15]. YPs are asked how likely they think it is they will apply for a place at university and, unless they say there is no chance of this happening, what they think the likelihood of such an application being successful is. These same questions are repeated verbatim until Wave 4, when the second one is dropped in favour of questions on planned timing of such applications and reasons for wanting to go to university.

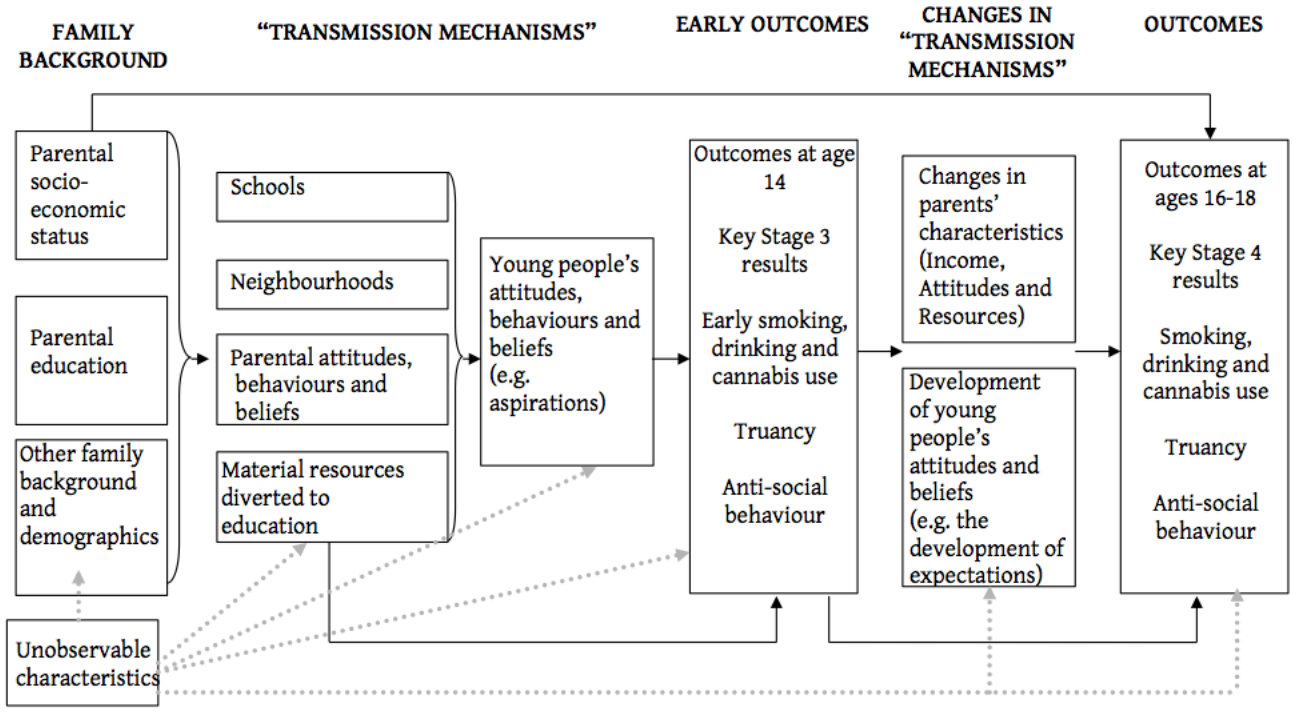
By Wave 5, with some already having applied to university, the question on likelihood of application is modified to be only for those who have not yet applied, along with the same question on timing. Similarly, in Wave 6, it is adapted to be only for those not already in university or who have applied that year. Also in Waves 4 and 5, a question is included on whether financial aspects of university have made them consider not going.

Young people are also asked if they are doing any of their present courses in order that they can apply to university. This is followed up with a question on whether they have decided what course they would like to study at university, with a coding for different groups of subjects and some variables corresponding to reasons for wanting to do this course. It would be potentially interesting to look at the courses this question refers to, in order to consider if they are appropriate courses for such a university course.

Once YPs reach Wave 5, corresponding to the end of Year 13[16], those who are planning to attend university straight after they have finished school should, on the whole, have completed their applications, received offers and most likely be awaiting results on which their offers depend. As well as questions on whether they have made an application, whether

[15]Corresponding to the end of Year 9.
[16]The school year in which A Levels are normally finished

Figure 12: Model of links and transmission mechanisms between family background, aspirations, expectations and outcome, from Chowdry et al. (2009), as adapted by Jerrim (2011, p. 39)



they have received offers and whether they have accepted any of those offers, as noted above, questions on subject choice are repeated. Financial concerns for those who feel they are likely to go to university are probed at this point too, asking if such considerations have ever made the young people seriously consider not going to university, what financial concerns they have about university and which of these they consider the most important. Linked to this, they were asked how well they feel they understand the financial support systems available to students. As well as the financial concerns about university, questions regarding other potential concerns or worries about going to university are included. They are also asked what they would do instead if they had decided not to go to university as a result of these concerns, financial concerns or other reasons.

Information about how they plan to study at university is also sought, in particular whether if they are successful in getting into their first choice university they will study full or part time and whether they think they will live with parents or away from home. In addition, they are asked if they are planning to take a Gap Year before beginning university and if so what they plan to do on this[17].

8.4 Parental Aspirations and Expectations

Both aspirations and expectations for the YP post-16 are sought from their main parent. They are asked both what they expect the YP to do at this age, and what they would like them to do. However, when it comes to HE, the only relevant question asked is whether they think the YP is likely to go to university. This is still, of course, useful information, since the parent's perception of this may itself affect the likelihood of the YP going to university through affecting family resources diverted to the YP's education and through shaping the YP's own aspirations and expectations in turn.

[17] A particularly interesting point might be differences between those wanting to take a Gap Year to save up money for university and those more interesting in pursuing other, perhaps more enjoyable, activities. This line of enquiry is currently being pursued by Claire Crawford at the Institute for Fiscal Studies.

8.5 Attitudes to debt and discount rates

One such attitude is young people's views on debt. Young people in the LSYPE cohort faced maximum tuition fees of £3,145 per year[18], with the option to pay none of this upfront, instead funding the studies with a loan from the Student Loans Company with repayments dependent upon income earned above a £15,000 per year threshold. Given that most HE participants fund their studies with loans from the Student Loan Company (Callender and Jackson, 2005; Pennell and West, 2005) it may be useful that the LSYPE includes questions on attitudes towards such debt from Wave 4 onwards.

These questions may be used both to take into account an individual's view of debt when they make key decisions and to be included in the model and to see how changes in such attitudes over time could also be associated with other relevant events. Interestingly these include questions on both debt in an abstract sense, personal attitudes and perceptions of others' attitudes towards student debt. Linked in with this are questions about the YP's expectations (and then in later waves the reality) of how they will fund their time at university, since clearly different funding will potentially mitigate the effects of attitudes towards debt and may themselves affect choices relating to university. Also covered here is whether they think they will get a grant or bursary to go to university and if not, why not.

For another important aspect of the effects of attitudes, it helps to return to our underlying human capital model. In choosing whether to attend university young people are choosing between two potential income streams, one which starts immediately if they do not choose to choose to apply for HE and one which begins only once they have completed their university degree. As such, other things being equal, how they discount future income should affect their decision, with those with a higher discount rate seeing income delayed for three or more years as worth less than they would with a lower discount rate. Participation in risky behaviours may act as a proxy (albeit a weak one) for higher discount rates and hence, ceteris paribus, a lower probability of applying for university. The LSYPE regularly asks questions regarding indulgence in risky behaviours, such as drug taking and truancy, which if they are not otherwise biased, could allow such a relationship to be tested.

[18]This was the tuition fee limit for the academic year 2008-09, it would then have been uprated in line with inflation for following years of study and begin at a consequently higher rate of £3,225 for those who took one gap year.

9 Prior Attainment and Cognitive Ability

Because of the correlation between young people's attainment and their other characteristics, for some research questions it may be necessary to attempt to control for prior attainment. To illustrate this point, let us take as an example a research question on the contribution of school type to HE admission. Given that young people are not randomly assigned to different types of school, but rather there is at least some degree of choice potentially taking into account the child's attainment, failure to control for attainment could result in upward biased estimates for school types which attracted high ability children in the first place.

However, this does have drawbacks. Since attainment itself is likely affected by socio-economic background, any quantitative models including prior attainment will only identify the effects of such characteristics which take impact after attainment was measured. For some research questions this will not be a problem: taking our above school type example further, since it seems unlikely that the kind of secondary school attended will affect attainment measured prior to starting secondary school, our statistical model should still be capable of isolating the appropriate relationship of interest. On the other hand, for a research question considering the effects of household income on HE admission it seems likely that household income does impact upon intermediate attainment outcomes. As such only the additional impacts of income, after the prior attainment measure but before applying to university, will then be identified by a model including this attainment measure.

9.1 NPD Attainment Data

To provide this data on prior attainment, the LSYPE comes linked to a subset of the National Pupil Database (NPD), providing useful administrative data on pupils' academic achievement in key measures. These data are available for all young people for whom a match was successfully made against the data in the NPD. Data are missing for just 79 YPs who deliberately weren't matched for some reason and for 362 others for whom a match does not appear to have been successfully found on the National Pupil Database. Agreement to have data linked to the NPD was a pre-requisite for inclusion in the LSYPE.

These correspond to roughly 0.5% and 2.7%, respectively, of the 16,122 entries[19] in the LSYPE linked NPD. In addition, attainment data are not available for examinations at Key Stage 2[20] for approximately a further 813 (around 5%) and at Key Stage 3[21] for approximately a further 506 (around 3%), most likely because these young people attended independent schools at these relevant points, which have no obligation to require these tests to be sat, or to publish the results if they are.

The linked NPD that comes with the LSYPE data includes KS5/AS and A Level results. Performance in these exams is a key part of the university admissions process. However, there are challenges with its use.

Firstly, A Levels present us with a sample selection problem: since these examinations are taken after the end of compulsory education, they are not observed for our whole population of interest, only a subsample. As such, models including A Level results could only possibly provide valid inference for this subpopulation at best.

Secondly, there is a high risk of endogeneity bias driven by A Level results' prominent place in the university application process. When A Level examinations are taken, individuals are likely to have already decided whether they will apply to HE, indeed it this must be the case for those who intend to go straight on to university without taking a Gap Year. The line of causality cannot be running against time.

Even for university attendance, offers have likely been given conditional on achieving certain A Level grades, inducing perhaps extra work to ensure these offers are met. Likewise for post-results applications, an individual planning to apply is likely to behave differently when it comes to their A Levels compared to an individual who is not. For example, someone not planning to apply to university might lack the same motivation to work hard specifically to achieve higher A Level grades.

In any case, earlier examinations, especially when coupled with other indicators, are likely to be able to predict A Level results rather well. This implies that the additional information that is being bought at this large cost in terms of potential bias is not actually as great as might be first thought.

[19]This is higher than the 15,570 initial partial responses at Wave 1 on account of the ethnic minority boost sample, mentioned in Section 2.2.

[20]Taken during Year 6, at age 10-11

[21]Taken during Year 9, at age 13-14

The LSYPE hence includes details on attainment at Key Stage 2 SATS[22]; Key Stage 3 SATS[23]; Key Stage 4 GCSE and equivalents[24]; and Key Stage 5 A levels and equivalents[25]. See Table 9 for this placed into the context of the LSYPE waves.

Table 9: Structure of the LSYPE

Time / Source	History	Young Person	Parental Attitudes	Family Background	National Pupil Database
Pre-Wave 1					Key Stage 2
Wave 1: Summer 2004	Yes	Yes	Yes	Yes (Both Parents)	
Year 9					Key Stage 3
Wave 2: Summer 2005	Yes	Yes	Yes	Yes (Both Parents)	
Wave 3: Summer 2006		Yes	Yes	Yes (Main Parent/Joint)	
Year 11					GCSE
Wave 4: Summer 2007		Yes	Yes	Yes (Main Parent)	
Year 12					AS Levels
Wave 5: Summer 2008		Yes			
Year 13					A Levels
Wave 6: Summer 2009		Yes			
Wave 7: Summer 2010		Yes			

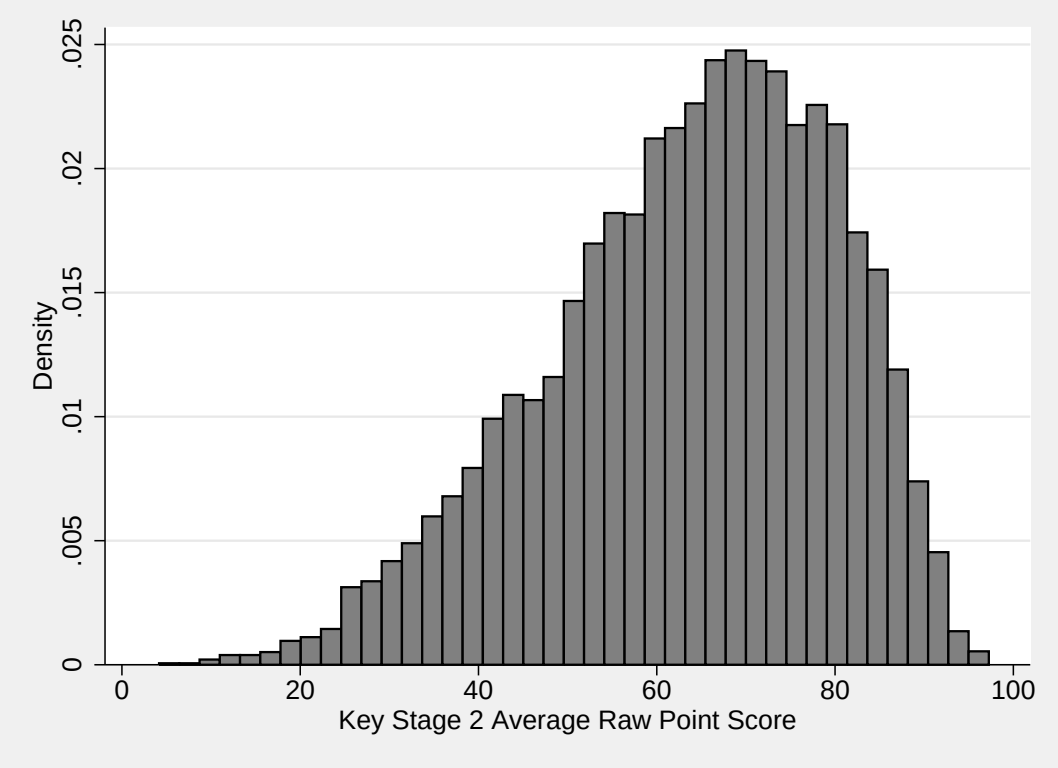
Information on prior attainment is available from as early as Key Stage 2. This is hopefully sufficiently long enough before university application not to suffer from the potential problems of endogeneity that we might see in results from exams taken around the time of university application (A Levels and equivalents) and to some extent ones from only a couple of years earlier (GCSEs and equivalents).

For simplicity, I assess a single measure of attainment at each stage. For KS2, I take the average of individuals' point scores for English, Maths and Science in their SATS. The distribution of this Average Raw Point Score is shown in Figure 13. At KS4, I use the capped GCSE measure. In this individuals are awarded a certain number of points for each grade they achieve in a GCSE or equivalent qualification (58 for an A*, 52 for an A, and so on). The

[22]Taken at age 10-11, in the final year of primary school.
[23]Taken at age 13-14, in school year 9.
[24]Taken at age 15-16, Year 11, the final year of compulsory education.
[25]Taken at age 16-18, Years 12-13, during post-compulsory education

scores from their eight best qualifications are then summed together to produce the measure. This is intended not to favour individuals who take lots of GCSEs, skewing it towards quality rather than quantity. The distribution of this score is shown in Figure 14.

Figure 13: Distribution of Key Stage 2 SATS Average Raw Point Score



Notes: Sample: Participants with valid KS2 attainment data, linked from NPD.

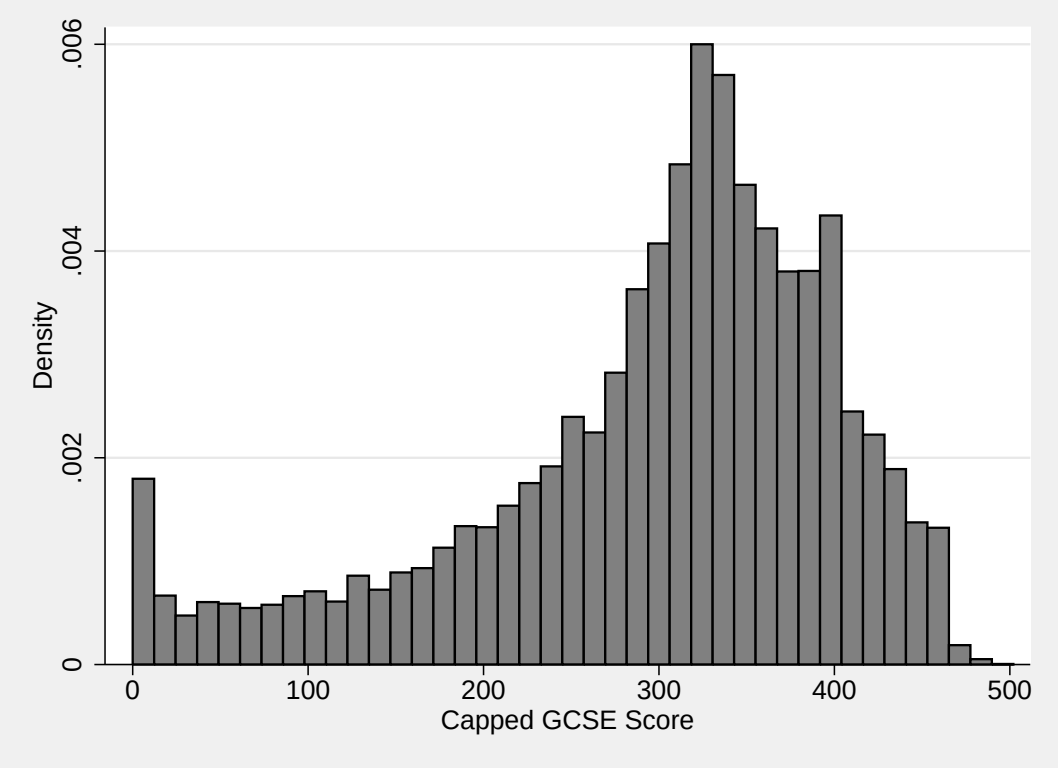
Since these results are drawn from the National Pupil Database (NPD), they can easily be compared with the results for the entire English cohort. This allows assessment of how representative the LSYPE sample is by the time it reaches Wave 7, treating the full cohort's NPD results as a population against which we can compare[26].

Tables 10 and 11 show summary statistics for KS2 and KS4 results for the weighted LSYPE sample at Wave 7, and for the comparable population in the complete NPD sample. In the case of KS2 scores the differences between the distributions are very small with even extreme points of the distribution (1st and 99th percentiles) matching up well.

On the other hand, comparison of KS4 results presents a different picture. Here the differences seem more significant. For example, the mean of the LSYPE distribution is 15 points above that of the NPD, corresponding to around 13% of a standard deviation (in the NPD

[26]I am grateful to Rebecca Allen, at the IoE, for assistance with these comparisons.

Figure 14: Distribution of capped GCSE (and equivalents) new style point score



Notes: Sample: Participants with valid KS4 attainment data, linked from NPD.

Table 10: KS2 SATS Average Point Score' - LSYPE vs. NPD Attainment Summary Statistics

Characteristic	**LSYPE**	**NPD**
Mean	66	64
Standard Deviation	16	16
Minimum	5	0
Maximum	97	99
1st Percentile	26	24
10th Percentile	43	42
Median	67	66
90th Percentile	84	84
99th Percentile	92	91
N	8,166	616,201

Notes: Weighted using LSYPE Wave 7 Respondent weights. Sample: Wave 7 respondents with relevant linked NPD data.

Table 11: Capped GCSE Score' - LSYPE vs. NPD Attainment Summary Statistics

Characteristic	LSYPE	NPD
Mean	306	291
Standard Deviation	104	114
Minimum	0	0
Maximum	483	540
1st Percentile	0	0
10th Percentile	150	106
Median	326	320
90th Percentile	422	416
99th Percentile	464	464
N	8,624	664,654

Notes: Weighted using LSYPE Wave 7 Respondent weights. Sample: Wave 7 respondents with relevant linked NPD data.

data). This is certainly not negligible. A closer look at the distributional statistics also suggests that the differences are concentrated in the lower half. For example, the 10th percentile of the LSYPE is 46 points above that of the NPD.

Why the LSYPE should match up so much better with the general population at KS2 is not clear. One hypothesis could be that those who attrit from the LSYPE (and whose attrition is not dealt with by the weighting scheme) are less likely to make expected progress between ages 11 and 16.

The relatively favourable comparisons between the LSYPE and the NPD are also in contrast to the apparent overestimation of the proportion applying to and attending HE, as discussed in Section 4.3. We might have expected correspondingly higher exam performance among the LSYPE cohort, however there is little evidence of this beyond the modest differences seen in the lower half of the KS4 distribution.

10 School and Neighbourhood Effects

Although an individual's characteristics and behaviours are of course a very important part of the story regarding access to HE, the decisions and actions of those around them will also have an impact. This holds for institutional actions and decisions, as in the case of school characteristics and quality, and also individual actions and decisions, as in the case of peer effects (be those peers in the local neighbourhood or also within school).

10.1 School Effects

Assuming that schooling has some impact on educational attainment, the ability to be able to allow for school effects to be held constant in models of access to HE will be important. Whether individual models will wish to do so will, of course, depend upon their particular research questions. Since school choice is not exogenously determined, but rather affected by parental characteristics, inclusion of school characteristics prevents us from identifying the full effect of those parental characteristics. School choice may be an important way in which intergenerational transmission of advantage occurs.

Previous studies, using different approaches, have attempted to isolate the impact of school characteristics and quality on educational attainment in general (Rivkin et al., 2005) or in some cases specifically on access to HE (Curtis et al., 2008). There would seem to be two broad ways in which this could be achieved within the context of the LSYPE.

Firstly, because the LSYPE was sampled at school level[27] it should be possible to use statistical models that take such clustered data into account. Both random effects and fixed effects modelling would seem to be potential candidates for such analysis. However, random effects models make the assumption that the school level errors have no correlation with our explanatory variables. Given the many socioeconomic factors that go into school choice, including indirect factors such as choice of neighbourhood, this would seem to be an untenable assumption.

Instead, we could turn to a fixed effects model. This method, by effectively including a dummy variable for each school, attempts to soak up the variation due to between school variation. This still presents drawbacks, for example it allows no wider inference regarding school effects to be drawn, it simply attempts to prevent them from biasing other estimates, hence ruling out research questions about school impacts on access to HE. In addition, for effective estimation of each group's effect a fixed effects model requires relatively large group sizes. By Wave 7 remaining samples from each school in the LSYPE range from just two up to one hundred and sixty eight[28]. The schools with small remaining sample sizes could potential lead to some poorly estimated group effects.

[27] An anonymous school identifier may be found in the variable `samppsu`.

[28] Only a couple of schools have more remaining YPs than the initial maximum per school of 34: these are from the ethnic boost sample added at Wave 4.

The alternative to such methods would be instead to use questions from the survey itself or linked data from the school level census that is available, on request, as part of the LSYPE. Questions on school quality in the survey itself are largely restricted to "how much do you agree" questions posed to the YP from Waves 1 to 3. Such questions are likely to have major problems of subjectivity and lack of knowledge school quality on the part of the YP. For example, it seems likely that YPs would be less inclined to agree with positive statements about their school if they are not performing well in it. Nevertheless, should the researcher believe these potential problems can be circumvented, this possible route may exist.

On the other hand, these linked data include a range of information on schools, including admissions policy, institution type, specialist school type, total number of pupils, whether it has a sixth form, average attainment at Key Stage 3 and 4 and average Key Stage 2 attainment of its intake. The LSYPE does not include school characteristics for an individual's primary school, however, since our earliest prior attainment measures are for the end of primary school/Key Stage 2 we would expect to see any impacts of school quality at this age to affect prior attainment indicators at this point as an intermediate variable.

In contrast to the fixed effects method, using these data would allow statistical inference to be drawn regarding the associations of those particular characteristics that have been measured with access to university. It could also estimate the overall magnitude of school effects on access to HE, if the linked data describes all variation due to school type and quality. Taking this approach, as with any simple regression model, then makes the implicit assumption that the resulting estimates do not suffer from omitted variable bias.

Choice of model will of course depend upon the particular research question. For example, fixed effects would seem to be more appropriate where schools are not the particular characteristic of interest but simply need to be prevented from biasing other estimates.

10.2 Neighbourhood Effects

An individual's peer group and their beliefs and behaviour may also have an impact on their behaviour. Arguably, the family (especially parents, as discussed above) are just a special case of this. However, should we wish to control for any potential peer group effects, to consider fully these effects we need to cast the net wider, considering the effects of

young people's neighbourhood and school peers on their educational aspirations and expectations. Of course, because a family has at least some control over their neighbourhood and the school the child is sent to, such peer effects are potentially inseparable from family background characteristics, leading us to overestimate peer effects if family background is not fully described (Evans et al., 1992; Ginther et al., 2000).

Nevertheless, the LSYPE allows us to model such potential effects, should we wish to, largely through linked data. On application, data that allow us to get a good idea of the YP's neighbourhood can be obtained, including variables covering the area's urbanity; Income Deprivation Affecting Children Index (IDACI) score and ranking; and the Index of Multiple Deprivation (IMD) score and ranking. Likewise, an impression of the other members of their school can be gained from data on its gender of entry, ethnic composition, Special Educational Needs (SEN) composition, percentage eligible for Free School Meals (FSM) and percentage whose first language is not English.

It also conducts more direct measurement of potential peer group effects through questions to the YP on the educational intentions of their friends. For example, young people are asked whether they think most of their friends will stay on in education after statutory leaving age or not. One could also potentially consider the extent to which people are likely to be influenced into educational decisions through a question on why they chose the GCSE option subjects that they did, including the non-exclusive option that their friends were doing the same subjects.

11 Conclusions

By considering previous research and relevant literature, this paper attempted to assess the utility of the LSYPE to researchers concerned with HE access. While there are a few areas where it does not provide all the information and details that might be ideal, on the whole it seems to provide a strong dataset with which research to this end can be conducted.

The LSYPE has significant advantages over previous datasets used for similar analysis. It has deeper background information than administrative data, albeit with much smaller sample sizes. More in depth questioning, over a longer time period, with lower attrition give it significant advantages of the Youth Cohort Studies. All these factors taken together

potentially allow statistical inference that is more likely to be causal in nature.

Furthermore, we can undertake much more up to date analysis using LSYPE data. Previous analysis that used the national birth cohort studies referred to institutional arrangements that have now changed drastically. Work using the LSYPE is able to consider HE access during a period of mass HE participation. However, additional changes to the student funding system in 2012, notably including the near tripling of tuition fees and the introduction of the National Scholarship Scheme for students from less advantaged backgrounds, means the target has shifted yet again. Hopefully future datasets, including the LSYPE's own successor, will allow us to keep up.

The LSYPE also allows for consideration of a more in depth set of HE-related outcomes. These can consist of not just entry or otherwise to HE, but various milestones in the admissions process, subject studied and university attended. This increased detail could add significantly to our understanding of access to HE, relative to previous research on similar questions.

However, a major note of caution is that validation of the attendance measures against administrative data suggests that there is large overestimation of the proportion of young people who enter Higher Education. There is roughly a 10 percentage point discrepancy between the LSYPE figure, and official statistics.

There are various possible reasons for this data problem. These include definitional differences between the LSYPE and the HEIPR (such as treatment of individuals who drop out); attrition on characteristics that the weighting scheme does not control for; bias caused by non-cooperation of schools who were initially chosen to be sampling units (unaccounted for by weights); and individual non-response at Wave 1. In future work I plan to investigate which of these seems most plausible and what, if anything, can be done about it.

On the explanatory side, the LSYPE covers, in detail, factors that previous research and theory suggest should be included in any model of the proposed type. This includes a range of explanatory variables on attitudes, aspirations, expectations, socioeconomic characteristics (notably family income) and prior attainment. These are covered in detail and longitudinally, allowing us all the benefits this affords.

Validation suggests some problems with underreporting of family income across most of

the income distribution. For example, the LSYPE's median family income is reported as just over 10% below that of the FRS. Nevertheless, it still represents a large improvement on many previous datasets.

On balance, we can be confident both that the LSYPE contains the necessary information and structure to further our understanding of HE access and admissions in England and that such work will bring new insights to academic and policy debate in this field. It is important, however, to keep the caveats discussed in mind where they could confound our analysis.

Bibliography

Björklund, A. and Salvanes, K. G. *Handbook of the Economics of Education*, chapter Education and Family Background: Mechanisms and Policies. Elsevier, London.

Bratti, M., Naylor, R., and Smith, J. (2008). Heterogeneities in the returns to degrees: evidence from the British Cohort Study 1970. Departmental Working Papers 2008-40, Department of Economics, Management and Quantitative Methods at Università degli Studi di Milano.

Burgess, S., Propper, C., and Rigg, J. (2004). The Impact of Low Income on Child Health: Evidence from a Birth Cohort Study. CASEpaper 85, London School of Economics, London.

Callender, C. and Jackson, J. (2005). Does the Fear of Debt Deter Students from Higher Education? *Journal of Social Policy*, 34(04):509–540.

Chevalier, A. and Conlon, G. (2003). Does It Pay to Attend a Prestigious University? CEE DP 33, Centre for the Economics of Education, London School of Economics, London.

Chowdry, H., Crawford, C., Dearden, L., Goodman, A., and Vignoles, A. (2010a). Widening participation in higher education: analysis using linked administrative data. IFS Working Paper W10/04, Institute for Fiscal Studies, London.

Chowdry, H., Crawford, C., and Goodman, A. (2009). Drivers and Barriers to Educational Success. Evidence from the Longitudinal Study of Young People in England. Research Report RR102, Department for Children, Schools and Families, London.

Chowdry, H., Crawford, C., and Goodman, A. (2010b). Outcomes in the secondary school years: evidence from the Longitudinal Study of Young People in England. Report, Joseph Rowntree Foundation, York.

Collingwood, A., Cheshire, H., Nicolaas, G., D'Souza, J., Ross, A., Hall, J., Armstrong, C., Prosser, A., Green, R., Collins, D., Gray, M., and McNaughton Nicholls, C. (2010). A review of the Longitudinal Study of Young People in England (LSYPE): recommendations for a second cohort. DfE Research Report DFE-RR048, Department for Education/NatCen.

Conlon, G. and Patrignani, P. (2011). The Returns to Higher Education Qualifications. Research Paper 45, Department for Business Innovation and Skills, London.

Curtis, A., Power, S., Whitty, G., Exley, S., and Sasia, A. (2008). Primed for success? The characteristics and practices of state schools with good track records of entry into prestigious UK universities. Research-reports monograph, The Sutton Trust, London.

Department for Business Innovation & Skills (2012). Participation Rates in Higher Education: Academic Years 2006/2007–2010/2011 (Provisional). Statistical first release, Department for Business, Innovation & Skills, London.

Department for Children Schools & Families (2008). Youth Cohort Study of England and Wales, 2004-2007; Cohort 12, Sweep One to Four [computer file].

Department for Education (2011a). LSYPE User guide to the datasets: Wave One to Wave Seven. User guide, Department for Education, London.

Department for Education (2011b). Youth Cohort Study and Longitudinal Study of Young People in England: The Activities and Experiences of 19 year olds: England 2010. Statistical Bulletin B01/2011, Department for Education, London.

Department for Education and National Centre for Social Research (2012). Longitudinal Study of Young People in England: Waves One to Seven, 2004-2000 [computer file].

Elias, P. and Purcell, K. (2004). The Earnings of Graduates in their Early Careers. Researching Graduate Careers Seven Years On, Research Paper 5, Warwick Institute for Employment Research, Warwick.

Ermisch, J. and Francesconi, M. (2001). Family matters: Impacts of family background on educational attainments. *Economica*, 68(270):137–156.

Evans, W. N., Oates, W. E., and Schwab, R. M. (1992). Measuring Peer Group Effects: A Study of Teenage Behavior. *The Journal of Political Economy*, 100(5):966–991.

Galindo-Rueda, F., Marcenaro, O., and Vignoles, A. (2004). The Widening Socio-Economic Gap in UK Higher Education. CEE Discussion Papers 44, Centre for the Economics of Education, London.

Gayle, V., Berridge, D., and Davies, R. (2002). Young people's entry into higher education: quantifying influential factors. *Oxford Review of Education*, 28(1):5–20.

Ginther, D., Haveman, R., and Wolfe, B. (2000). Neighborhood Attributes as Determinants of

Children's Outcomes: How Robust Are the Relationships? *The Journal of Human Resources*, 35(4):603–642.

Griliches, Z., Hall, B. H., and Hausman, J. A. (1978). Missing Data and Self-Selection in Large Panels. *Annales de l'inséé*, 30/31:137–176.

Haveman, R. and Wolfe, B. (1995). The Determinants of Children's Attainments: A Review of Methods and Findings. *Journal of Economic Literature*, 33(4):1829–1878.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47(1):153–161.

HESA (2010). PIs: Non-continuation rates (including projected outcomes). Tables T3a and T3b (2008/9 versions) from `http://www.hesa.ac.uk/index.php?option=com_content&task=view&id=2064&Itemid=141`.

Iacovou, M. (2008). Family Size, Birth Order, and Educational Attainment. *Marriage & Family Review*, 42(3):35–57.

Jenkins, S. P. and Schluter, C. (2002). The Effect of Family Income During Childhood on Later-Life Attainment: Evidence from Germany. IZA Discussion Paper 604, Institute for the Study of Labor (IZA), Bonn.

Jerrim, J. (2011). Disadvantaged children's "low" educational expectations: Are the US and UK really so different to other industrialized nations? DoQSS Working Paper 11-01, Institute of Education, London.

Kreuter, F., Eckman, S., Maaz, K., and Watermann, R. (2010). Children's Reports of Parents' Education Level: Does it Matter Whom You Ask and What You Ask About? *Survey Research Methods*, 4(3):127–138.

Maher, J. (2006). A description of non-respondents to the Family Resources Survey 2004-2005. Report, National Centre for Social Research, London.

Marcenaro-Gutierrez, O., Galindo-Rueda, F., and Vignoles, A. (2007). Who actually goes to university? *Empirical Economics*, 32(2):333–357.

Micklewright, J. and Schnepf, S. V. (2010). How reliable are income data collected with a sin-

gle question? *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 173(2):409–429.

Miranda, A. and Rabe-Hesketh, S. (2010). Missing ordinal covariates with informative selection. DoQSS Working Paper 10-16, Institute of Education, London.

Palmer, G. (2012). Households, families and benefit units. Retrived from `http://www.poverty.org.uk/summary/households.shtml` on 5/11/2012.

Pennell, H. and West, A. (2005). The Impact of Increased Fees on Participation in Higher Education in England. *Higher Education Quarterly*, 59(2):127–137.

Piesse, A. and Kalton, G. (2009). A Strategy for Handling Missing Data in the Longitudinal Study of Young People in England (LSYPE). DCSF Research Report DCSF-RW086, Department for Children, Schools and Families/WESTAT, London.

Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, 73(2):417–458.

The Sutton Trust (2008). *Wasted talent: attrition rates of high-achieving pupils between school and university*. The Sutton Trust, London.

UCAS (2011a). UCAS - Age. `http://www.ucas.com/about_us/stat_services/stats_online/data_tables/age`.

UCAS (2011b). UCAS - Data summary. `http://www.ucas.com/about_us/stat_services/stats_online/data_tables/datasummary`.

UCAS (2012). UCAS - Analysis of UCAS January deadline application rates by country. `http://www.ucas.ac.uk/about_us/media_enquiries/media_releases/2012/300112b`.

Vignoles, A., Goodman, A., Machin, S., and McNally, S. (2008). Widening participation in higher education: a quantitative analysis. Research report, Economic and Social Research Council, Swindon.

A Appendix: Construction of Income Measures

This appendix describes the steps taken to tidy and clean income variables in the LSYPE raw data into those used by tables and figures in this paper. All data manipulation was performed using Stata 12 SE.

Summary statistics for the various measures may be found in Table 4.

A.1 Interval Regression

As in much survey data, most income measurement in the LSYPE was collected using banding. This has the advantage of helping to prompt the interviewee, making response easier. However, for the data user the use of banding presents a significant problem. Even making the assumption that true income lies between these two bands[29] this still leaves a significant range of potential values. A simplistic method of dealing with this is to assign the midpoint of the bands to any individuals falling into that band. However, there are still problems with data treated in this way: it is not as continuous as the underlying data would be and unless income is uniformly distributed across the band summary statistics for the data will be distorted.

Thankfully, we can do better relatively easily. Interval regression fits a regression model with both lower and upper bounds as dependent variables. Independent variables included are then other variables expected to affect the true dependent variable in one direction or the other. The interval regression assumes that the underlying dependent variable will be distributed normally allowing us to achieve a more realistic distribution of any predictions of the underlying dependent variable that we then make. Furthermore, it is possible to obtain the predictions as being the expected value of the underlying dependent variable conditional on being between the two bands, thus taking into account as much information as possible.

[29] Of course, this could not be the case, however we can't really proceed without making such an assumption. The general problems of income measurement will not be discussed here.

A.2 Wave 1

The income measure for Wave 1 is constructed using interval regression (`intreg`) using the upper and lower bounds placed on family income by the main parent's banded estimate of family income (`w1inc1est` and `w1inc2est`) and socioeconomic background variables which seemed likely to affect or reflect family income.

Those used are family gross earnings (`w1grssyrhh`) and benefits (`w1benammp`) (estimated through questions to each parent on hourly earnings rate and hours worked or earnings through self employment plus questions on amounts of benefits received), indicators of mother and father's full time and part time work status (`w1wrk1amp` and `w1wrk1asp`), if the family received means tested benefits (`w1ben2mp`), if the family received tax credits (`w1ben3mp`), if the family received any other benefits (`w1ben2mp`), mother and father's highest level of education (`w1hiqualgmum` and `w1hiqualgdad`), mother and father's NS-SEC classification (`w1nsseccatmum` and `w1nsseccatdad`) and the family's housing tenure (`w1hous12`).

Interval regression assumes a normal distribution in its output. In the case of an income distribution, where the distribution is not generally considered to be normal, this presents a problem. However, if we assume instead that the income distribution is log normal, a much better founded assumption, the interval regression may be run using log normal bounds and logged gross earnings and benefits to replicate this distribution in its output. Because of this assumption of log normality, mean and median incomes from the interval regression predicted variable are lower than from the unadjusted banded income variables.

Predicted values for each family's income are then obtained using the interval regression results. As a check on the interval regression not leading to predictions outside the known range, the predictions were constrained to fall within the bands provided by the main parent's estimate of total family income. This was done using the Stata command: `predict predictedincome, e(lowerbound, upperbound)`

A.3 Wave 2

In Wave 2 no overall estimate of family income is recorded, instead an estimate was obtained in the following way.

Firstly main parent and second parent's (where appropriate) estimation of their gross personal income are summed together (w2grssyrmp and w2grssyrsp)[30]. This is treated as valid even if one of the parent's data are missing. This does give a risk of underestimation for such families, but it is only for a small number of families and I judge the bias from excluding these families from the sample would be greater.

To this summed variable is added the variable for total annual benefits for the family (w2bentotanam)[31]. Where both parents' earnings report is missing, but they state that they are not employed, I set the earnings to be zero, so long as there is some benefit income[32] reported.

Due to these raw data, our variable will not include income from non-earned sources e.g. investment income. As such, Wave 2 income could be underestimated, particularly higher up the income distribution where investment income is likely to be a more important part of total income. However, this could be offset by the increased detail of the questions in this wave.

The indicator was converted into Wave 1 (2004) prices using the annual Retail Prices Index (RPI).

A.4 Wave 3

Wave 3 returns to banded estimates of family income. Again an interval regression was estimated, using the same method as for Wave 1. Unfortunately fewer background characteristics were available to refine the interval regression.

As such, background characteristics used are indicators of mother and father's full time and part time work status (w3wrkfullmum and w3wrkfulldad), if the family received means tested benefits w3incsourmp0e), if the family received tax credits w3incsourmp0g), if the family received any other benefits w3incsourmp0f), mother and father's highest level of ed-

[30]Unlike in the other Waves, these variables are continuous. Banded questions for the same information are stored in variables w2inc1estmp, w2inc2estmp, w2inc1estsp and w2inc2estsp and the midpoints of these bands are included in the pre-release derivations of w2grssyrmp and w2grssyrsp. Less than 2% of the sample do not answer the continuous income questions but do answer the banded income questions.

[31]This variable is not raw data, but rather has been derived from the raw data as part of the pre-release data processing. The LSYPE documentation warns of the potential for double counting in this total annual benefits variable.

[32]Benefit income for this purpose must include more than child benefit, since many individuals seem to report this while refusing to answer any other benefit receipt questions. I am grateful to Jean-Baptiste Gilbert (Institute of Education) for bringing this particular feature to my attention.

ucation (from Wave 2: w2hiqualgmum and w2hiqualgdad), mother and father's NS-SEC classification (w3cnsseccatmum and w3cnsseccatdad) and the family's housing tenure (w3hous12).

The indicator was converted into Wave 1 (2004) prices using the annual RPI.

A.5 Wave 4

Similarly to Wave 3, an interval regression approach is adopted, again with different socioeconomic characteristics used.

Background characteristics used are indicators of mother and father's full time and part time work status (w4wrkfullmum and w4wrkfulldad), mother and father's highest level of education (w4hiqualgmum and w4hiqualgdad), mother and father's NS-SEC classification (w4cnsseccatmum and w4cnsseccatdad) and the family's housing tenure (w4hous12).

The indicator was converted into Wave 1 (2004) prices using the annual RPI.

A.6 Equivalised Income

In order to allow for the effects of family size on the spending power of different incomes to be taken into account, income was equivalised. This was achieved by dividing by the square root of the number of resident family members, measured at the relevant wave.

A.7 'Permanent' Income

'Permanent' income was constructed by taking the mean of non-missing income variables constructed in Waves 1 to 4 above (after they had been adjusted for inflation). This was achieved using Stata's rowmean function. In the case of 'permanent' equivalised income, the mean was calculated from Wave 1 to Wave 4 figures after equivalisation.

As such, so long as income questions were answered at one wave a value is obtained for the 'permanent' income measure. Where income questions were not answered at all waves this variable is set to missing.

B Appendix: Listing of referenced variables

List of variables the properties of which have been discussed in the main review, along with others that may be of interest, their names in LSYPE Stata release and summarised description of content. For further information refer to LSYPE documentation.

B.1 Admissions Process

B.1.1 Wave 2

Variable name	Description
w2alevifyp	are you staying on in full time education do you can get A Levels, AVCEs or something else?
w2alevuniyp	will you be doing A levels, AVCEs or something else so you can apply to university later on?

B.1.2 Wave 3

Variable name	Description
w3dec16ayp	when decided to stay in full time education after year 11
w3reas16ayp1a–w3reas17ayp1s	reason decided to stay in full time education after year 11
w3alevif	are you staying on in full time education do you can get A Levels, AVCEs or something else?
w3alevuniyp	will you be doing A levels, AVCEs or something else so you can apply to university later on?
w3unisubayp	have you decided what course or subject you would like to study at university?
w3unisubyp1–w3unisubyp23	what subject?

B.1.3 Wave 4

Variable name	Description
w4alevuniyp	are you doing any of these courses so you can apply to uni later on?
w4unisubayp	if so, have you decided what course you want to study at uni yet?
w4hesubXyp	coded for what kind of course want to study where: X = 1: medicine and dentistry 2: subjects allied to medicine 3: biological sciences 4: veterinary science 5: agriculture and related 6: physical sciences 7: mathematical sciences 8: computer science 9: engineering and technology 10: architecture, building and planning 11: social studies 12: law 13: business and administrative 14: mass communications and docum 15: languages 16: historical and philosophical 17: creative arts and design 18: education 19: combined or general courses 20: other
w4ifuniyp0a– w4ifuniyp0i	why do you want to do this course?
w4unisubotyp	have you decided what course to study at uni?
w4subreas1yp	chose subject to get a specific job or career (1: very important - 4: not at all important)

Variable name	Description
w4subreas2yp	chose subject to make it more likely can get a well paid career (1: very important - 4: not at all important)
w4subreas3yp	chose subject as interested in it (1: very important - 4: not at all important)
w4subreas4yp	chose subject as good at it (1: very important - 4: not at all important)
w4subreas5yp	chose subject because parents want you to do it (1: very important - 4: not at all important)
w4subreas6yp	chose subject because you think it will be easier to get in with this subject than others (1: very important - 4: not at all important)

B.1.4 Wave 5

Variable name	Description
w5heapplyyp	have you applied to start on a course September 2008 or September 2009?
w5routesyp	route a (1), route b (2) or both (3)?
w5extrasyp	did you use UCAS Extra at all?
w5offersyp	have you received any offers yet?
w5acceptyp	have you accepted any of these offers yet?
w5heposs9yp	if haven't applied yet, how likely is it you think you will apply? (1 very likely - 4 not at all likely)
w5whenapplyyp	if likely when? 1: next two years 2: beyond that 3: never
w5unisubayp	have you decided what subject you'd like to do at uni?
w5unisubbyp	if so, yes? if you've received offers already what course for (preferred if more than one)? if you've accepted an offer what was firm accept for?
w5ifuniyp	do you want to do this course at uni to qualify for a specific job or profession? (details in w5ModAp6b2yp)

Variable name	Description
w5subreas1yp	chose subject to get a specific job or career (1: very important - 4: not at all important)
w5subreas2yp	chose subject to make it more likely can get a well paid career (1: very important - 4: not at all important)
w5subreas3yp	chose subject as interested in it (1: very important - 4: not at all important)
w5subreas4yp	chose subject as good at it (1: very important - 4: not at all important)
w5subreas5yp	chose subject because parents want you to do it (1: very important - 4: not at all important)
w5subreas6yp	chose subject because you think it will be easier to get in with this subject than others (1: very important - 4: not at all important)
w5nchoicesyp	how many universities did you apply to? (1-5)
w5heinstyp	what is the name of the university you have applied to that is your firm choice?
w5whychosyp	why did you choose to apply to that uni?
w5hepref1	if you get into 1st choice uni will you 1: live at home 2: live away from home 3: undecided
w5hepref2	if you get into 1st choice uni will you do the course 1: full time 2: part time 3: undecided
w5subconfyp	how well informed do you feel about financial support available to uni students? (1 very well informed - 4 not at all informed)
w5gapyearyp	will you go on a gap year?
w5gapdoyp	what main thing do you expect to do on gap year? 1: travelling 2: working abroad 3: working in britain 4: training 5: something else

Variable name	Description
w5gapwhyyp	why do you want to do a gap year? 1: break from studying 2: become more independent or gain more experience 3: earn money 4: get work experience 5: apply for a course after you have got your exam results 6: friends are doing the same 7: something else

B.1.5 Wave 6

Variable name	Description
w6univyp	Are you at university?
w6heflag	Are you currently doing an HE qualification?
	Asking people who are actually in HE:
w6heinstyp	What your university is called? (May also be found in w6scollegeyp)
w6hequalyp–w6hequal2yp	What HE qualification they're studying for? 1: Degree 2: Foundation Degree 3: Teacher Training and others (if totally other then will be found in w6hequaloyp)
w6hesubyp	What subjects your qualification involved?
w6hefullyp	1 for full time, 2 for part time
w6hehomeyp	1 for live at home, 2 for live away from home
w6hehomewyp	why they live at home? 1: save money 2: near friends 3: live with family. others: in w6hehomwoyp
w6instprefyp	1 if first choice institution, 2 if not first choice
w6instprefnoyp	1 if didn't get an offer from 1st choice, 2 if failed to meet offer requirements, 3 other (not sure where this is stored)
w6subprefyp	1 if first choice subject, 2 if not first choice - says what first choice was in w6subpref2yp
w6subprefnoyp	1: didn't get offer, 2: didn't meet offer conditions 3: changed mind 4: other (found in w6subprefnooyp)

Variable name	Description
w6gradebetyp	1 is got the grades they needed for university offer exactly, 2 is actual grades were better than offer ones, 3 is worse than
	Asking people not yet in HE (but with some chance of going there) about aspects of application
w6heapplyyp	have you applied to start on a course September 2009 or September 2010?
w6heoffersyp	if so, have you received any offers of any kind yet?
w6heacceptyp	if so, have you accepted any of these offers yet?
w6gapyryp	if so, are you on a gap year? "a year between exam results and actually starting uni course"
w6heapevyp	(asked if said haven't applied this year and stated didn't apply in wave 5 interview (or no wave 5 interview) Have you ever applied for a university course?
w6heoff2yp	(asked if said have ever applied or stated in wave 5 that have applied but not received any offers yet) Did you receive any offers?
w6heacpt2yp	(asked if said ever received offers above, or if said received offers in wave 5 but said hadn't accepted any of them yet) Did you accept any of those offers?
w6heacptnoyp	(asked if didn't accept any of those offers) Why did you not accept any of these offers - coded into w6heacptnoyp0a-o
w6nowgap	what have you been doing on your gap year (1: travelling 2: working abroad 3: working in britain 4: volunteering in Britain 5: volunteering abroad 6: doing training 5: retakes)
w7gapdec	when did you decide to take a gap year (1: before results 2: after results)

Variable name	Description
w7heapplyyp	have you applied to start on a course September 2010 or September 2011?
w7heoffersyp	if so, have you received any offers of any kind yet?
w7heacceptyp	if so, have you accepted any of these offers yet?
w7gapyryp	if so, are you on a gap year? "a year between exam results and actually starting uni course"

B.1.6 Wave 7

Variable name	Description
w7tcurrentactyp	what is the main thing you are currently doing? The raw question 'are you currently at university' (as in w6univyp) is not included in the dataset, but this derivation should give the same thing.
w7heflag	Are you currently doing an HE qualification?
	Asking people who are actually in HE:
w7heinstyp	What your university is called? (May also be found in w7scollegeyp)
w7nowgap	what have you been doing on your gap year (1: travelling 2: working abroad 3: working in britain 4: volunteering in Britain 5: volunteering abroad 6: doing training 5: re-takes)
w7gapyryp	if so, are you on a gap year? "a year between exam results and actually starting uni course"

B.2 Socioeconomic Characteristics

B.2.1 Wave 1

Variable name	Description
w1hous12HH	Housing tenure

Variable name	Description
w1ben1MP0a– w1ben1MP0i	State benefits currently received by main parent or partner (Child Benefit, Guardian's allowance, Invalid Care Allowance, Severe Disablement Allowance, Disability Living Allowance, Attendance Allowance)
w1ben2MP0a– w1ben2MP0o	State benefits currently received by main parent or partner (Job Seekers Allowance, Income Support, Incapacity Benefit, Statutory Sick Pay, Industrial injury disablement benefit, Maternity allowance, Statutory maternity pay from a current or former employer, Widows benefit, Widows pension, Widows allowance, Housing benefit, Carer's allowance, Council tax benefit)
w1ben3MP0a– w1ben3MP0e	State benefits currently received by main parent or partner (Working Tax Credit, Child Tax Credit)
w1benchMP	Does main parent receive child benefit for the young person participating in the LSYPE
w1benamMP	Total received by family for working tax credit and/or child tax credit
w1benchildMP	Does main parent or partner currently receive any benefits for families with children?
w1benunempMP	Does main parent or partner currently receive any benefits for unemployed people?
w1benlowincMP	Does main parent or partner currently receive any benefits for people on low income?
w1bendisMP	Does main parent or partner currently receive any benefits for people with disabilities?
w1benberMP	Does main parent or partner currently receive any benefits for bereaved people?
w1benothMP	Does main parent or partner currently receive any other benefits?

Variable name	Description
w1inc1estMP	Banded gross income estimate, including work, benefits and anything else, for main parent and partner (lower band)
w1inc2estMP	Banded gross income estimate, including work, benefits and anything else, for main parent and partner (higher band)
w1managhhMP	How well is the family managing on its current income?
w1ed1mp	At what age did main parent first leave school?
w1ed1amp	Did main parent return to full time education after leaving school?
w1ed1bmp	At what age did main parent leave full time education altogether?
w1apprentmp	Has main parent completed an apprenticeship?
w1oversmp	Does main parent have any other overseas qualifications?
w1otypemp	Rough equivalence of highest overseas qualification
w1ed3mp	Did main parent's father get a degree?
w1ed4mp	Did main parent's mother get a degree?
w1wrk1amp	What is main parent's current employment status?
w1wrkymp	Year they began being in this employment status
w1wrk2mp	Has main parent ever had a paid job or been self-employed?
w1wrk3mp	In what year did last job end?
w1wrkstatmp	Is main parent's current job, or was previous job, employed or self-employed?
w1wrk10mp	Does main parent have formal supervisory responsibility in their job?
w1wrk11mp	How many employees at main parent's workplace?
w1wrkfixhrmp	Is main parent paid on hourly rate?
w1fixramp	Main parent's hourly rate (pounds)
w1fixra2mp	Main parent's hour rate (pence)

Variable name	Description
w1salar1mp	Main parent's gross pay on last occasion they were paid
w1salar2mp	The period this gross pay covered
w1salar3mp	Main parent's take home pay on last occasion they were paid
w1jjbhrsmp	How many hours does main parent work per week?
w1hrsovermp	How many hours of overtime does main parent work per week?
w1paidovrmp	How many hours of paid overtime does main parent work per week?
w1wrk12amp	For self-employees, does main parent work alone or have employees?
w1wrk12bmp	If so, how many employees?
w1seiinc1mp	What were main parent's earnings while self employed?
w1seiinc2mp	What was main parent's estimated income from self-employment in first full years, before tax, after expenses, in bands?
w1seiinc3mp	What was main parent's income from self-employment in last tax year, before tax, after expenses?
w1hea1mp	What was main parent's general health in the last twelve months?
w1hea2mp	Does main parent have a long-standing illness, disability of infirmity?
w1noldsibhs	How many older siblings does the young person have?
w1noldbrohs	How many younger siblings does the young person have?
w1nressibhs	How many older, non-resident siblings does the young person have?
w1sibs	Number of siblings to young person in the family?
w1sibs2	Number of siblings to young person (including non-resident)?
w1ed1sp	At what age did second parent first leave school?

Variable name	Description
w1ed1asp	Did second parent return to full time education after leaving school?
w1ed1bsp	At what age did second parent leave full time education altogether?
w1apprentsp	Has second parent completed an apprenticeship?
w1overssp	Does second parent have any other overseas qualifications?
w1otypesp	Rough equivalence of highest overseas qualification
w1ed3sp	Did second parent's father get a degree?
w1ed4sp	Did second parent's mother get a degree?
w1wrk1asp	What is second parent's current employment status?
w1wrkysp	Year they began being in this employment status
w1wrk2sp	Has second parent ever had a paid job or been self-employed?
w1wrk3sp	In what year did last job end?
w1wrkstatsp	Is second parent's current job, or was previous job, employed or self-employed?
w1wrk10sp	Does second parent have formal supervisory responsibility in their job?
w1wrk11sp	How many employees at second parent's workplace?
w1wrkfixhrsp	Is second parent paid on hourly rate?
w1fixrasp	Second parent's hourly rate (pounds)
w1fixra2sp	Second parent's hour rate (pence)
w1salar1sp	Second parent's gross pay on last occasion they were paid
w1salar2sp	The period this gross pay covered
w1salar3sp	Second parent's take home pay on last occasion they were paid
w1jjbhrssp	How many hours does second parent work per week?
w1hrsoversp	How many hours of overtime does second parent work per week?

Variable name	Description
w1paidovrsp	How many hours of paid overtime does second parent work per week?
w1wrk12asp	For self-employees, does second parent work alone or have employees?
w1wrk12bsp	If so, how many employees?
w1seiinc1sp	What were second parent's earnings while self employed?
w1seiinc2sp	What was second parent's estimated income from self-employment in first full years, before tax, after expenses, in bands?
w1seiinc3sp	What was second parent's income from self-employment in last tax year, before tax, after expenses?
w1agemp	Age of main parent
w1agesp	Age of second parent
w1marstatmp	Marital status of main parent
w1marstatsp	Marital status of second parent
w1wrkagemp	Whether main parent is of working age
w1wrkagesp	Whether second parent is of working age
w1empsmp	Employment status of main parent
w1empssp	Employment status of second parent
w1workfullmp	Whether main parent works full or part time
w1workfullsp	Whether second parent work full or part time
w1empmonmp	Duration of continuous employment of main parent in months
w1empmonsp	Duration of continuous employment of second parent in months
w1depkids	Number of dependent children in the family
w1ch0_2hh	Number of children aged 0-2 in the family
w1ch3_11hh	Number of children aged 3-11 in the family
w1ch12_15hh	Number of children aged 12-15 in the family

Variable name	Description
w1ch16_17hh	Number of children aged 16-17 in the family
w1stepfam	Whether family is a step family
w1famtyp	Family composition
w1lounemp	Whether main parent is long-term unemployed (6 months or longer)
w1lounesp	Whether second parent is long-term unemployed (6 months or longer)
w1nsseccatmp	Main parent's National Statistics Socioeconomic Classification (NS-SEC) operational category
w1nsseccatsp	Second parent's National Statistics Socioeconomic Classification (NS-SEC) operational category
w1ethgrpyp	Ethnic group of young person
w1grssyphh	Gross annual salary of family
w1grssyrhhbands	Banded family income
w1inc1est	Estimate of gross family income

B.2.2 Wave 2

Variable name	Description
w2hous12hh	Housing tenure
w2wrkstatusmp	Main parent's current employment status
w2wrkymp	Year main parent started current activity
w2wrk2mp	Whether main parent has ever had a paid job or been self employed
w2wrk3mp	Year in which main parent's last job or period of employment ended
w2wrkstatmp	Whether current or last job or main parent was employed or self-employed
w2fixhrmp	Whether main parent paid a fixed hourly wage
w2fixramp	Main parent's hourly rate (pounds)

Variable name	Description
w2fixra2mp	Main parent's hourly rate (pence)
w2salar1mp	Main parent's gross pay on last occasion
w2salar2mp	Time period this gross pay related to
w2salar3mp	Main parent's take home pay on last occasion
w2salar3bmp0a–w2salar3bmp0f	Whether main parent's take home pay includes any of Statutory Sick Pay, Statutory Maternity Pay, Tax credits, Income tax refund
w2salar4mp	Time period this take home pay related to
w2inc1estmp	Ambiguous. Only asked to those with no more specific income data.
	UKDA Dictionary: What is total gross income from work, benefits and anything else for Main parent (and partner) (lower bands) Questionnaire: What is gross pay for Main parent (lower bands)
w2inc2estmp	Ambiguous. Only asked to those with no more specific income data.
	UKDA Dictionary: What is total gross income from work, benefits and anything else for Main parent (and partner) (higher bands) Questionnaire: What is gross pay for Main parent (higher bands)
w2jjbhrsmp	How many hours does main parent work per week?
w2hrsovermp	How many hours overtime does main parent work per week?
w2paidovrmp	How many hours paid overtime does main parent work per week?
w2wrk12amp	For self-employees, does main parent work alone or have employees?
w2wrk12bmp	If so, how many employees?

Variable name	Description
w2seiinc2mp	What was main parent's estimated income from self-employment in first full years, before tax, after expenses, in bands?
w2seiinc3mp	What was main parent's income from self-employment in last tax year, before tax, after expenses?
w2inc5estmp	What was main parent's income from self-employment in last tax year, before tax, after expenses, in bands? (lower bands)
w2inc6estmp	What was main parent's income from self-employment in last tax year, before tax, after expenses, in bands? (higher bands)
w2numalexmp	how many A Levels does main parent have?
w2numGCSEmp	how many GCSEs does main parent have?
w2childbmp	Whether the main parent is currently the named recipient of Child Benefit

Variable name	Description
w2ben1qmp0a–w2ben5qmp0l	Whether the main parent is currently the named recipient of Guardian's Allowance, Carer's Allowance, State Retirement Pension, Widows Benefit or Pension, Bereavement Allowance, Widowed Parent's Allowance, War Pensions, Severe Disablement Allowance, Disability Living Allowance, Attendance Allowance, Job Seeker's Allowance, Income Support, MIG, Pension Credit, Incapacity Benefit, Statutory Sick Pay, Industrial Injury Disablement Benefit, Maternity Allowance, Statutory Maternity Pay, Working Tax Credit, Child Tax Credit, Social Fund grant for funeral expenses, Social Fund grant for maternnity expenses, Sure Start Maternity Grant, Social Fund loan or Community Care grant, Back to Work bonus, 'Extended Payment' of housing Benefit, rent rebate, Council Tax Benefit, Bereavement Payment, Child Maintenance Bonus, Lone Parent's Benefit Run-On, Any National Insurance or State Benefit not mentioned earlier.
w2rechbenmp	Does main parent receive Housing Benefit or rent rebate, renting from Council or New Town?
w2rechben2mp	Does main parent receive Housing Benefit or rent rebate, renting from Housing Association or shared ownership?
w2mhelp1mp0a–mhelp1mp0i	Does anyone in the family receive anything towards mortgage or house loan regularly from DWP, their employer, other organisations, friends and relatives, mortgage protection or insurance policy, or other?
w2copaymp	Does main parent or any other resident pay Council Tax?
w2nopaymp	Why isn't council tax paid?
w2rebatmp	Does main parent get Council Tax Benefit or a rebate?

Variable name	Description
w2othsourcmp0a–w2othsourcmp0h	Does main parent regularly receive Occupational pensions from former employer(s), Occupational pensions from spouse's former employer(s), Private pensions or annuities, Redundancy payments from former employer(s), or Government training schemes?
w2reglrpmmp0a–w2reglrpmmp0f	Does main parent regularly receive Educational grant; Payments from friends and relatives outside family; or Maintenance, alimony or separation allowance?
w2rentpaymp	Does main parent currently receive rent from any property or subletting?
w2grssyrmp	Gross annual salary of main parent
w2netyrmp	Net annual salary of main parent
w2wrkstatussp	Second parent's current employment status
w2wrkysp	Year second parent started current activity
w2wrk2sp	Whether second parent has ever had a paid job or been self employed
w2wrk3sp	Year in which second parent's last job or period of employment ended
w2wrkstatsp	Whether current or last job or second parent was employed or self-employed
w2fixhrsp	Whether second parent paid a fixed hourly wage
w2fixrasp	Second parent's hourly rate (pounds)
w2fixra2sp	Second parent's hourly rate (pence)
w2salar1sp	Second parent's gross pay on last occasion
w2salar2sp	Time period this gross pay related to
w2salar3sp	Second parent's take home pay on last occasion
w2salar3bsp0a - w2salar3bsp0f	Whether second parent's take home pay includes any of Statutory Sick Pay, Statutory Maternity Pay, Tax credits, Income tax refund
w2salar4sp	Time period this take home pay related to

Variable name	Description
w2inc1estsp	Ambiguous: UKDA Dictionary: What is total income from work, benefits and anything else for Second parent (and partner) (lower bands) Questionnaire: What is gross earnings for Second parent (lower bands)
w2inc2estsp	Ambiguous: UKDA Dictionary: What is total income from work, benefits and anything else for Second parent (and partner) (higher bands) Questionnaire: What is gross earnings for Second parent (higher bands)
w2jjbhrssp	How many hours does second parent work per week?
w2hrsoversp	How many hours overtime does second parent work per week?
w2paidovrsp	How many hours paid overtime does second parent work per week?
w2wrk12asp	For self-employees, does second parent work alone or have employees?
w2wrk12bsp	If so, how many employees?
w2seiinc2sp	What was second parent's estimated income from self-employment in first full years, before tax, after expenses, in bands?
w2seiinc3sp	What was second parent's income from self-employment in last tax year, before tax, after expenses?
w2inc5estsp	What was second parent's income from self-employment in last tax year, before tax, after expenses, in bands? (lower bands)

Variable name	Description
w2inc6estsp	What was second parent's income from self-employment in last tax year, before tax, after expenses, in bands? (higher bands)
w2numalexsp	how many A Levels does second parent have?
w2numGCSEsp	how many GCSEs does second parent have?
w2childbsp	Whether the second parent is currently the named recipient of Child Benefit
w2ben1qsp0a–w2ben5qsp0l	Whether the second parent is currently the named recipient of Guardian's Allowance, Carer's Allowance, State Retirement Pension, Widows Benefit or Pension, Bereavement Allowance, Widowed Parent's Allowance, War Pensions, Severe Disablement Allowance, Disability Living Allowance, Attendance Allowance, Job Seeker's Allowance, Income Support, MIG, Pension Credit, Incapacity Benefit, Statutory Sick Pay, Industrial Injury Disablement Benefit, Maternity Allowance, Statutory Maternity Pay, Working Tax Credit, Child Tax Credit, Social Fund grant for funeral expenses, Social Fund grant for maternity expenses, Sure Start Maternity Grant, Social Fund loan or Community Care grant, Back to Work bonus, 'Extended Payment' of housing Benefit, rent rebate, Council Tax Benefit, Bereavement Payment, Child Maintenance Bonus, Lone Parent's Benefit Run-On, Any National Insurance or State Benefit not mentioned earlier.
w2rechbensp	Does second parent receive Housing Benefit or rent rebate, renting from Council or New Town?
w2rechben2sp	Does second parent receive Housing Benefit or rent rebate, renting from Housing Association or shared ownership?

Variable name	Description
w2mhelp1sp0a– w2mhelp1sp0i	Does anyone in the family receive anything towards mortgage or house loan regularly from DWP, their employer, other organisations, friends and relatives, mortgage protection or insurance policy, or other?
w2copaysp	Does second parent or any other resident pay Council Tax?
w2nopaysp	Why isn't council tax paid?
w2rebatsp	Does main parent get Council Tax Benefit or a rebate?
w2othsourcsp0a– w2othsourcsp0h	Does main parent regularly receive Occupational pensions from former employer(s), Occupational pensions from spouse's former employer(s), Private pensions or annuities, Redundancy payments from former employer(s), or Government training schemes?
w2reglrpmsp0a– w2reglrpmsp0f	Does main parent regularly receive Educational grant; Payments from friends and relatives outside family; or Maintenance, alimony or separation allowance?
w2rentpaysp	Does main parent currently receive rent from any property or subletting?
w2grssyrsp	Gross annual salary of second parent
w2netyrsp	Net annual salary of main parent
w2bentotanam	Total annual amount of benefit received

B.2.3 Wave 3

Variable name	Description
w3parpreshh	Is there a parent or guardian living in the household?
w3hous12hh	What is household's housing tenure?
w3usevchh	Does anyone in the household have use of a motor vehicle?
w3vehnohh	How many vehicles are there in the houshold?

Variable name	Description
w3condur4mp	Main parent: Is there a telephone in the household?
w3condur5mp	Main parent: Is there a computer in the household?
w3condur6mp	Main parent: Is there internet access in the household?
w3condur4yp	Young person: Is there a telephone in the household?
w3condur5yp	Young person: Is there a computer in the household?
w3condur6yp	Young person: Is there internet access in the household?
w3wrk1amp	Main parent's economic activity
w3wrk10mp	Does Main parent have formal supervisory responsibility at work?
w3wrk11mp	How many employees are there at main parent's workplace?
w3wrk12amp	Does main parent work alone or have employees?
w3wrk12bmp	If has employees, how many?
w3wrk1asp	Second parent's economic activity
w3wrk10sp	Does second parent have formal supervisory responsibility at work?
w3wrk11sp	How many employees are there at second parent's workplace?
w3wrk12asp	Does second parent work alone or have employees?
w3wrk12bsp	If has employees, how many?
w3incsourmp0a–w3incsourmp0m	Which of the following are sources of income for main parent and partner? Earnings from employment or self-employment, pensions from a former employer, state pension, Child Benefit, Income Support, other State Benefits, Tax Credits, Interest from savings etc., other kinds of regular allowance from outside the household, other.
w3incestmp	Estimate of total gross family income (yearly). This variable is referred to as w3inc1estmp in the documentation, but by this name is the data files.
w3empsmp	Employment status of main parent

Variable name	Description
w3empssp	Employment status of second parent
w3cempszmp	Employment status and organisation size for main parent
w3cempszsp	Employment status and organisation size for second parent
w3cnsseccatmp	Main parent's current NS-SEC operational category
w3cnsseccatsp	Second parent's current NS-SEC operational category
w3cnsseccatfam	Family's current NS-SEC operational category

B.2.4 Wave 4

Variable name	Description
w4parpreshh	Is there a parent or guardian living in the household?
w4hous12hh	What is household's housing tenure?
w4usevchh	Does anyone in the household have use of a motor vehicle?
w4vehnohh	How many vehicles are there in the household?
w4condur4mp	Main parent: Is there a telephone in the household?
w4condur5mp	Main parent: Is there a computer in the household?
w4condur6mp	Main parent: Is there internet access in the household?
w4wrk1amp	Main parent's economic activity
w4wrk10mp	Does Main parent have formal supervisory responsibility at work?
w4wrk11mp	How many employees are there at main parent's workplace?
w4wrk12amp	Does main parent work alone or have employees?
w4wrk12bmp	If has employees, how many?
w4wrk1asp	Second parent's economic activity
w4wrk10sp	Does second parent have formal supervisory responsibility at work?

Variable name	Description
w4wrk11sp	How many employees are there at second parent's workplace?
w4wrk12asp	Does second parent work alone or have employees?
w4wrk12bsp	If has employees, how many?
w4inc1estmp	Estimate of total gross family income

B.3 Attitudes, Aspirations and Expectations

B.3.1 Wave 1

Variable name	Description
w1heposs9yp	how likely do you think it is you'll apply to university?
w1hlikeyp	how likely do you think it is you'll get in to university, if you apply?
w1truantyp	played truant in last 12 months?
w1truant1yp	longest period of time truanted
w1truant2yp	main reason for doing it (bullying, bored, didn't like school, didn't like particular teachers, didn't like particular subject, something else)
w1truant3yp	parent assisted truanting in last 12 months?
w1truant4yp	how often did this happen?
w1namesyp	been called names?
w1youbulnyp0a	how often?
w1excpalyp	excluded from a group of friends or activities in last 12 months?
w1youbulnyp0b	how often?
w1moneyyp	forced to hand over money in last 12 months?
w1youbulnyp0c	how often?
w1thhityp	threatened with violence by other students in last 12 months?

Variable name	Description
w1youbulnyp0d	how often?
w1achityp	experienced violence from other students in last 12 months?
w1youbulnyp0e	how often?
w1racmotyp	do you think these were racially motivated?
w1cignowyp	do you ever smoke cigarettes at all?
w1cigfreqyp	how often/how much?
w1alceveryp	have you ever had a proper alcoholic drink?
w1alcfreqyp	how often over the last twelve months?
w1canntryyp	have you ever tried cannabis?
w1sprayyp	written on walls with spray cans in last 12 months?
w1smashyp	smashed public property in last 12 months?
w1shopyp	shoplifted since in last 12 months?
w1fightyp	fight in public in last 12 months?
w1kidskolmp	Main parent's view of overall quality of young person's school
w1extrtu1mp	Whether main parent has, in last twelve months, paid for private lessons also taught to young person in school
w1extrtu4mp	Whether main parent has, in last twelve months, paid for private lessons in other subjects
w1patt1mp	Main parent agrees that nowadays you need qualifications in order to get a job with having
w1patt2mp	Main parent agrees that leaving school at 16 limits young people's career opportunities later in life
w1patt3mp	Main parent wants young person to have a better education than they did
w1parasp2mp	What main parent would like young person to do when they reach school leaving age?
w1parasp1mp	What main parent thinks young person will do when they reach school leaving age?

Variable name	Description
w1hepossmp	What is main parent's view of likelihood of young person going into higher education?
	If not, reasons why not:
w1henomp0a	Won't get necessary grades
w1henomp0b	Can't afford it
w1henomp0c	No interest in in going
w1henomp0d	Has job in mind and no need for this qualification in career
w1henomp0e	SEN/learning problems
w1henomp0f	Wants to start working/earning
w1henomp0g	Wants to take education somewhere other than university e.g. college

B.3.2 Wave 2

Variable name	Description
w2heposs9yp	likelihood of yp applying for university
w2hlikeyp	likelihood of yp getting into university, if apply
w2truantyp	played truant in last 12 months?
w2truant1yp	longest period of time truanted
w2truant2yp	main reason for doing it (bullying, bored, didn't like school, didn't like particular teachers, didn't like particular subject, something else)
w2truant3yp	parent assisted truanting in last 12 months?
w2truant4yp	how often did this happen?
w2namesyp	been called names?
w2youbulnyp0a	how often?
w2excpalyp	excluded from a group of friends or activities in last 12 months?
w2youbulnyp0b	how often?

Variable name	Description
w2moneyyp	forced to hand over money in last 12 months?
w2youbulnyp0c	how often?
w2thhityp	threatened with violence by other students in last 12 months?
w2youbulnyp0d	how often?
w2achityp	experienced violence from other students in last 12 months?
w2youbulnyp0e	how often?
w2racmotyp	do you think these were racially motivated?
w2cignowyp	do you ever smoke cigarettes at all?
w2cigfreqyp	how often/how much?
w2alceveryp	have you ever had a proper alcoholic drink?
w2alcfreqyp	how often over the last twelve months?
w2canntryyp	have you ever tried cannabis?
w2sprayyp	written on walls with spray cans in last 12 months?
w2smashyp	smashed public property in last 12 months?
w2shopyp	shoplifted since in last 12 months?
w2fightyp	fight in public in last 12 months?
w2extrtu1mp	Whether main parent has, in last twelve months, paid for private lessons also taught to young person in school
w2extrtu4mp	Whether main parent has, in last twelve months, paid for private lessons in other subjects
w2parasp2mp	What main parent would like young person to do when they reach school leaving age?
w2parasp1mp	What main parent thinks young person will do when they reach school leaving age?

B.3.3 Wave 3

Variable name	Description
w3heposs9yp	likelihood of YP applying for university
w3hlikeyp	likelihood of YP getting into university
w3emaawaryp	aware of the EMA
w3emaapyp	applied for the EMA
w3emagoyp	going to apply for the EMA
w3emasuyp	application for the EMA was successful
w3truantyp	played truant since wave 2?
w3truant1yp	longest period of time truanted
w3truant2yp	main reason for doing it (bullying, bored, didn't like school, didn't like particular teachers, didn't like particular subject, something else)
w3truant3yp	parent assisted truanting since wave 2?
w3truant4yp	how often did this happen?
w3namesyp	been called names?
w3youbulnyp0a	how often?
w3excpalyp	excluded from a group of friends or activities since wave 2?
w3youbulnyp0b	how often?
w3moneyyp	forced to hand over money since wave 2?
w3youbulnyp0c	how often?
w3thhityp	threatened with violence by other students since wave 2?
w3youbulnyp0d	how often?
w3achityp	experienced violence from other students since wave 2?
w3youbulnyp0e	how often?
w3racmotyp	do you think these were racially motivated?
w3cignowyp	do you ever smoke cigarettes at all?
w3cigfreqyp	how often/how much?
w3alceveryp	have you ever had a proper alcoholic drink?
w3alcfreqyp	how often over the last twelve months?
w3canntryyp	have you ever tried cannabis?

Variable name	Description
w3sprayyp	written on walls with spray cans since wave 2?
w3smashyp	smashed public property since wave 2?
w3shopyp	shoplifted since wave 2?
w3fightyp	fight in public since wave 2?
w3abs1myyp	off school for more than a month?
w3abs1mwyp	why? suspension, illness, other
w3abs1mnyp	how long off in total during year 11?
w3suspendyp	been suspended during year 11?
w3expelyp	been expelled during year 11?
w3expwhatyp	where did you go instead? another school, PRU, boarding school, psychologist/counsellor, never went back to school, back to same school, taught at home
w3extrtu1mp	Whether main parent has, in last twelve months, paid for private lessons also taught to young person in school
w3extrtu4mp	Whether main parent has, in last twelve months, paid for private lessons in other subjects
w3parasp2mp	What main parent would like young person to do when they reach school leaving age?
w3parasp1mp	What main parent thinks young person will do when they reach school leaving age?

B.3.4 Wave 4

Variable name	Description
w4heposs9yp	how likely do with think it is that you will ever apply to university? (1 very likely - 4 not at all likely)
w4whenapplyyp	when do you think you'll apply? 1: next two years 2: some other time 3: never 4: not sure 5: already applied 6: other

Variable name	Description
w4whyheyp0a–w4whyheyp0v	why do you want to go to university? w4whyheim for most important one of these (only asked if more than one given)
w4anyconayp	if the financial aspects of going to university have made you think about not applying?
w4anyconbyp0a–w4anyconbyp0f	if so, which of these caused that? 1: level of tuition fees 2: living costs 3: having to borrow money / getting into debt 4: having to rely on parents for money 5: something else 6: none of these
w4hepref1yp	how would you prefer to study for a degree? - 1: living at home 2: living away 3: no preference
w4hepref2yp	how would you prefer to study for degree? - 1: full time 2: part time 3: no preference
w4hepref3yp	which type of course would you prefer? 1: standard length for my subject 2: shorter length but with fewer holiday 3: longer length but with year in industry or abroad 4: something else
w4grantelyp	do you think you'd be eligible for a grant or bursary?
w4grantel2yp	if not, why not? 1: family income too hight 2: course or training not eligible 3: something else (see grantel2yp0a-0c)
w4fundstudyp0a–w4fundstudyp0j	how do you think you'll pay for expenses at uni? (any or all of below) a: student loan b: borrowing money from bank, including credit cards or overdrafts c: sponsorship from employer d: paid work at term time e: paid work in holidays f: money from parents g: money from family friends h: own savings i: anywhere else
w4debtatt1yp	getting a degree leads to better paid jobs later (1 strongly agree - 4 strongly disagree)

Variable name	Description
w4debtatt2yp	owing money is always wrong (1 strongly agree - 4 strongly disagree)
w4debtatt3yp	borrowing money from bank is a normal part of modern lifestyle (1 strongly agree - 4 strongly disagree)
w4debtatt4yp	once you get into debt it's difficult to get out of (1 strongly agree - 4 strongly disagree)
w4debtatt5yp	student loans are a cheap way to borrow (1 strongly agree - 4 strongly disagree)
w4debtatt6yp	idea of leaving uni with big debts puts people off going there (1 strongly agree - 4 strongly disagree)
w4he1yp	i don't need to have a university degree to get the kind of job i want (1 strongly agree - 4 strongly disagree)
w4he2yp	the best jobs go to people who've been to university (1 strongly agree - 4 strongly disagree)
w4he4yp	most of my friends are planning to go to university (1 strongly agree - 4 strongly disagree)
w4he6yp	people like me don't go to university (1 strongly agree - 4 strongly disagree)
w4truantyp	played truant in year 11 and how much?
w4excludeyp	been excluded during year 10 or 11?
w4expel1yp	was that expelled?
w4expel2yp	was that suspended?
w4alceveryp	have you ever had a proper alcoholic drink?
w4alcfreqyp	how often over the last twelve months?
w4canntryyp	have you ever tried cannabis?
w4pubberyp	how often do you go to pubs and bars?
w4clubberyp	how often do you go to night clubs?
w4sprayyp	written on walls with spray cans in last 12 months?
w4smashyp	smashed public property in last 12 months?
w4shopyp	shoplifted in last 12 months?

Variable name	Description
w4spares1yp	usually find plenty of enjoyable things to do in my spare time?
w4spares2yp	spend most of my spare time at my home or a friend's house?
w4spares3yp	where i live safe for people like me out on the streets?
w4spares4yp	usually have enough money to do what i like?
w4spares5yp	often get bored?
w4spares6yp	anything better than staying at home even if nowhere special to go?
w4violentyp	experienced violence in last 12 months?
w4hurtyp	been physically injured from that violence in last 12 months?
w4threatsyp	threatened with violence in last twelve months?
w4madegiveyp	forced to hand over money or other things in last twelve months?
w4namesyp	been called names in last twelves months?
w4racismyp	someone threatened or been rude because of skin colour, race, ethnic background or religion?
w4b2knibyp	ever carried a knife out of the house in last 12 months?
w4b3knieyp	what was your main reason for this?
w4extrtu1mp	Whether main parent has, in last twelve months, paid for private lessons also taught to young person in school
w4extrtu4mp	Whether main parent has, in last twelve months, paid for private lessons in other subjects
w4hepossmp	how likely main parent thinks it is that young person will go to higher education

B.3.5 Wave 5

Variable name	Description
w5matesyp	how many of your friends in same year have applied for university? 1: all 2: most 3: half 4: a few 5: hardly any
w5anyconayp	have financial costs of uni ever made you seriously think about not applying?
w5anyconbyp0a–w5anyconbyp0e & w5anyconb2yp0a–w5anyconb2yp0j	what else you think will be a problem for you if you go to university
w5anyconcyp	which of the following most concerned about? 1: cost of tuition fees 2: living costs 3: having to borrow money or getting into debt 4: having to rely on parents for money 5: answer from w5anyconb2yp
w5whyconyp	do you think you might have any problems at uni (other than costs and finance)?
w5anyconyp	what problems?
w5considyp	have you ever seriously considered not applying to go to university?
w5othactyp	if you had seriously considered not applying, what would you have done instead if you hadn't applied? 1: started working full time 2: started learning a trade or work based learning 3: been unemployed 4: taken a full time education course at a college 5: something else
w5grantelyp	do you think you'd be eligible for a grant or bursary?
w5grantel2yp	if not, why not? 1: family income too hight 2: course or training not eligible 3: something else (see grantel2yp0a-0c)

Variable name	Description
w5fundstudyp0a–w5fundstudyp0j	how do you think you'll pay for expenses at uni? (any or all of below) a: student loan b: borrowing money from bank, including credit cards or overdrafts c: sponsorship from employer d: paid work at term time e: paid work in holidays f: money from parents g: money from family friends h: own savings i: anywhere else
w5debtatt1yp	getting a degree leads to better paid jobs later (1 strongly agree - 4 strongly disagree)
w5debtatt2yp	owing money is always wrong (1 strongly agree - 4 strongly disagree)
w5debtatt3yp	borrowing money from bank is a normal part of modern lifestyle (1 strongly agree - 4 strongly disagree)
w5debtatt4yp	once you get into debt it's difficult to get out of (1 strongly agree - 4 strongly disagree)
w5debtatt5yp	student loans are a cheap way to borrow (1 strongly agree - 4 strongly disagree)
w5debtatt6yp	idea of leaving uni with big debts puts people off going there (1 strongly agree - 4 strongly disagree)

B.3.6 Wave 6

Variable name	Description
	Attitudes to Uni (1 strong agree to 4 strongly disagree)
w6heweighyp	weighed up financial costs against long term benefits
w6heelseyp	wasn't sure what else to do
w6unibetyp	1 uni has been better than you thought, 2 uni has been worse than you thought, 3 uni has been about the same as you thought
w6unipictyp	had a good picture of what uni would be like (1 strongly agree - 4 strongly disagree) - if didn't agree then reasons given in w6unipict2yp0a-n

Variable name	Description
w6decisyp	going to uni has been the right decision (1 definitely, 2 probably, 3 probably not, 4 definitely not)
	Extra support
w6grantrecyp0a	receives a grant (from SLC)
w6grantrecyp0b	receives a bursary (from institution)
w6grantrecyp0c	receives a scholarship
	Financial support
w6fundstudyp0a	student loan
w6fundstudyp0b	borrowing from bank (inc. overdraft)
w6fundstudyp0c	sponsorship from employer etc.
w6fundstudyp0d	paid work in term time
w6fundstudyp0e	paid work in holidays
w6fundstudyp0f	family support
w6fundstudyp0g	friends
w6fundstudyp0h	savings
w6fundstudyp0i	anything else
w6diffyp	how getting on at managing finances (1 very well, 2 quite well, 3 not very well, 4 not at all well)
w6debtatt1yp	getting a degree leads to better paid jobs later (1 strongly agree - 4 strongly disagree)
w6debtatt2yp	owing money is always wrong (1 strongly agree - 4 strongly disagree)
w6debtatt3yp	borrowing money from bank is a normal part of modern lifestyle (1 strongly agree - 4 strongly disagree)
w6debtatt4yp	once you get into debt it's difficult to get out of (1 strongly agree - 4 strongly disagree)
w6debtatt5yp	student loans are a cheap way to borrow (1 strongly agree - 4 strongly disagree)
w6debtatt6yp	idea of leaving uni with big debts puts people off going there (1 strongly agree - 4 strongly disagree)

Variable name	Description
w6heposs9yp	(asked if not applying this year and not currently in he) how likely ever to apply to university? (1 very likely - 4 not at all likely)
w6whenapplyyp	if likely when? 1: next two years 2: beyond that 3: never
w6nonhe0a–w6nonheaa	coded reasons for why never likely to apply

B.3.7 Wave 7

Variable name	Description
w7grantrecyp0a	receives a grant (from SLC)
w7grantrecyp0b	receives a bursary (from institution)
w7grantrecyp0c	receives a scholarship
	Financial support
w7fundstudyp0a	borrowing from bank (inc. overdraft)
w7fundstudyp0b	sponsorship from employer etc.
w7fundstudyp0c	paid work in term time
w7fundstudyp0d	paid work in holidays
w7fundstudyp0e	family support
w7fundstudyp0f	friends
w7fundstudyp0g	savings
w7fundstudyp0h	parental support
w7fundstudyp0k	anything else
w7diffyp	how getting on at managing finances (1 very well, 2 quite well, 3 not very well, 4 not at all well)

B.4 Cognitive Ability and Prior Attainment

B.4.1 In NPD

Variable name	Description
pass_aaa	Total number of GCSE/GNVQ qualifications at grades A*-A (GCSE equivalencies).
pass_ac	Total number of GCSE/GNVQ qualifications at grades A*-C (GCSE equivalencies).
pass_ag	Total number of GCSE qualifications at grades A*-G (GCSE equivalencies).
ptstnewe	Total GCSE and equivalents new style point score.
ptscnewe	Capped GCSE and equivalents new style point score.
fiveac	Achieved 5 or more GCSE/GNVQs at grades A*-C.
level2	Achieved Level 2 threshold (5 or more GCSE and equivalents at grades A*-C).
fiveag	Achieved 5 or more GCSE/GNVQs at grades A*-G.
level1	Achieved Level 1 threshold (5 or more GCSE and equivalents at grades A*-G).
levl2em	Achieved 5 or more GCSE and equivalents at grades A*-C including GCSE English and Maths
levl2fem	Achieved 5 or more GCSE and equivalents at grades A*-C including GCSE and equivalents in English and Maths
levl1em	Achieved 5 or more GCSE and equivalents at grades A*-G including GCSE English and Maths.
levl2fem	Achieved 5 or more GCSE and equivalents at grades A*-G including GCSE and equivalents in English and Maths
lev2eng	Achieved A*-C in GCSE English.
lev2feng	Achieved A*-C in GCSE and equivalents English (Functional English)
lev1eng	Achieved A*-G in GCSE English.
lev1feng	Achieved A*-G in GCSE and equivalents English (Functional English)
lev2mat	Achieved A*-C in GCSE Maths.

Variable name	Description
lev2fmat	Achieved A*-C in GCSE and equivalents Maths (Functional Maths).
lev1mat	Achieved A*-G in GCSE Maths.
lev1fmat	Achieved A*-G in GCSE and equivalents Maths (Functional Maths).
lev2em	Achieved A*-C in GCSE English and Maths.
lev2fem	Achieved A*-C in GCSE and equivalents English and Maths (Functional English and Maths).
lev1em	Achieved A*-G in GCSE English and Maths.
lev1fem	Achieved A*-G in GCSE and equivalents English and Maths (Functional English and Maths).
cvap2aps	KS2 average point score for contextual value added
cvap2eng	KS2 english point score for contextual value added
cvap2mat	KS2 maths point score for contextual value added
cvap2sci	KS2 science point score for contextual value added
cva_ks2	Contextual Value Added KS2 to KS4
cvap3eng	KS3 english point score for contextual value added
cvap3mat	KS3 maths point score for contextual value added
cvap3sci	KS3 science point score for contextual value added
cvap3aps	KS3 average point score for contextual value added
cva_ks3	Contextual Value added KS3 to KS4
fsm_06	Free school meal eligibility
sen_06	SEN provision type
k4_age	Age at start of academic year (presumably when this data is collected)
k4_yob	Year of Birth
k4_mob	Month of Birth

B.5 School and Neighbourhood Effects

B.5.1 Wave 1

Variable name	Description
	How much do you agree...
w1yys1yp	happy at school?
w1yys2yp	school was a waste of time?
w1yys3yp	school work was worth doing?
w1yys4yp	most of time time i didn't want to go to school?
w1yys5yp	people think my school is a good school?
w1yys6yp	on the whole i liked being at school?
w1yys7yp	worked as hard as i could in school?
w1yys8yp	counted the minutes until the end of lessons?
w1yys9yp	i was bored in lessons?
w1yys10yp	work i did in lessons was a waste of time?
w1yys11yp	work i did in lessons was interesting?
w1yys12yp	i got good marks?
w1yys13yp	school is clean and tidy?
w1yys14yp	teachers make sure homework set is done?
w1yys15yp	teachers make it clear how we should behave?
w1yys16yp	teachers take action when they see rules broken?
w1yys17yp	teachers praise me when i do work well?
w1yys18yp	i like my teachers?
w1yys19yp	teachers can keep order in class?
w1yys20yp	teachers care how stretched i am
w1yys21yp	teachers mark my work
w1yys22yp	i am good at school work?
w1yys23yp	teachers think i am good at school work?

B.5.2 Wave 2

Variable name	Description
	How much do you agree...
w2yys1yp	happy at school?
w2yys2yp	school was a waste of time?
w2yys3yp	school work was worth doing?
w2yys4yp	most of time time i didn't want to go to school?
w2yys5yp	people think my school is a good school?
w2yys6yp	on the whole i liked being at school?
w2yys7yp	worked as hard as i could in school?
w2yys8yp	counted the minutes until the end of lessons?
w2yys9yp	i was bored in lessons?
w2yys10yp	work i did in lessons was a waste of time?
w2yys11yp	work i did in lessons was interesting?
w2yys12yp	i got good marks?
w2yys13yp	school is clean and tidy?
w2yys14yp	teachers make sure homework set is done?
w2yys15yp	teachers make it clear how we should behave?
w2yys16yp	teachers take action when they see rules broken?
w2yys17yp	teachers praise me when i do work well?
w2yys18yp	i like my teachers?
w2yys19yp	teachers can keep order in class?
w2yys24yp	teachers treat everyone the same regardless or skin colour or cultural background?
w2yys25yp	teachers don't really listen to what i say in class?
w2yys26yp	get treated unfairly by my teachers?

B.5.3 Wave 3

Variable name	Description
	How much do you agree...
w3yys1yp	happy at school?

Variable name	Description
w3yys2yp	school was a waste of time?
w3yys3yp	school work was worth doing?
w3yys4yp	most of time time i didn't want to go to school?
w3yys5yp	people think my school is a good school?
w3yys6yp	on the whole i liked being at school?
w3yys7yp	worked as hard as i could in school?
w3yys8yp	counted the minutes until the end of lessons?
w3yys9yp	i was bored in lessons?
w3yys10yp	work i did in lessons was a waste of time?
w3yys11yp	work i did in lessons was interesting?
w3yys12yp	i got good marks?

B.5.4 Wave 4

Variable name	Description
	How much do you agree...
w4yys24yp	teachers treat everyone the same regardless or skin colour or cultural background?
w4yys25yp	teachers don't really listen to what i say in class?
w4yys26yp	get treated unfairly by my teachers?