Department for Education

Missing Key Stage 2 data in LSYPE2: technical report

December 2016

Catherine Saunders, Ian Brunton-Smith, Bryn Garrod, Martin Stepanek, Anna Vignoles, Alex Sutherland

RAND EUROPE

UNIVERSITY OF CAMBRIDGE

GSR GOVERNMENT SOCIAL RESEARCH Social Science in Government

Preface

The Second Longitudinal Study of Young People in England (LSYPE2), which started in 2013, was designed in order to understand the compulsory education, school-to-work transitions, careers and lives of young people. Although it is of rich academic interest, the key purpose for this dataset is to provide a resource for evidence-based policy development. A significant barrier to achieving this purpose could be the 'missingness' present in LSYPE2, owing to a boycott of Key Stage 2 (KS2) testing in 2010. In 2010, 15,518 maintained schools were expected to administer KS2 tests, but 4,005 (26 per cent) of these schools did not administer them.

The Department for Education (DfE) commissioned RAND Europe, in collaboration with Professor Vignoles at the University of Cambridge and Professor Brunton-Smith at the University of Warwick, to explore strategies to address this missing data relating to the boycott[1] that will allow the best usage of LSYPE2 data in the future and will maximise the value of this important study.

This report presents the technical details of this work, the assessment of available strategies and statistical methods to deal with missing data as they apply to LSYPE2, and the methods and approaches taken, specifically, inverse probability weighting (IPW) and multiple imputation (MI). There are methodological strengths and limitations to all analyses; the key issues and concerns, particularly relating to multiple imputation, are therefore described. However, because the aim of this work was to balance methodological rigour with practical application, the main output is the creation of imputed KS2 variables and an IPW to allow practical use across a range of stakeholders to address the issue of missing KS2 data relating to the boycott. This report is accompanied by a separate user guide, which supports and guides analysts in the use of these variables.

Dr Alex Sutherland
Project Lead
RAND Europe
Tel: +44(0)1223 353 329
Email: alex_sutherland@rand.org

[1] There is a small amount of missing KS2 data in LSYPE2 for other reasons, in particular, relating to attendance at independent schools (490/13,100, 3.7%) and where consent for linkage of LSYPE2 survey responses to the National Pupil Database (NPD) is not given or not possible (892/13,100, 6.8%). The work presented in this report, however, focuses on the missing KS2 data that can be attributed to the boycott.

Contents

Preface 2

List of figures 5

List of tables 6

Executive summary 7

Acknowledgements 8

Abbreviations and acronyms 9

1. Introduction 10

The Key Stage 2 SATS boycott 10

LSYPE2 10

The research problem 11

2. Missing Data 12

Types of missing data 12

Addressing missing data 14

Multiple imputation 15

Full information maximum likelihood approach 16

Inverse probability weighting approach 16

Ad hoc approaches to missing data 16

Approach to missing KS2 data in this project 18

3. Predictors of missingness 20

Summary of key findings 20

Primary school–level predictors of KS2 boycott 21

School characteristics 22

School performance in the years before and after the boycott 24

Combined model for school, pupil and other predictors of missingness 25

Assessing and understanding between-school variation 30

Summary 31

4. Predictors of KS2 attainment 32

Summary of measures used 32

Results of KS2 predictors analysis 35

Sensitivity analyses 39
'Trimming' the attainment model for imputation 40
Summary and implications for missing data analysis 41
5. Methodological issues relating to MI 43
Sampling, attrition and missing KS2 data for other reasons 43
Survey weights and survey sampling 45
Selection of variables to include in the MI model 45
Predictors of missingness and predictors of attainment 45
Compatibility 45
Number of imputations 47
The clustered/multilevel nature of the dataset 47
Methodological issues around the inclusion of different variables 48
The distribution of KS2 attainment 48
Derived variables 49
Polynomials 49
KS1 attainment 49
KS4 attainment 49
Teacher assessments 49
Pupils who attended a boycott school but have a non-missing KS2 test score 50
Perfect prediction 51
Multiple imputation is a simulation-based approach 51
6. Methodological issues relating to IPW 52
7. Development of MI model 53
Variables included in the imputation 53
Assessing convergence 53
8. Development of the IPW 58
9. Summary 59
References 60
Appendix: Distribution of imputed variables 63
Appendix: Sensitivity analyses 64

List of figures

Figure 1. Data sources and time frames for LSYPE2 missing data project......................11

Figure 2. Distribution of the KS2 score being imputed 48

Figure 3. Mean KS4 attainment across burn-in iterations 55

Figure 4. SD KS4 attainment across burn-in iterations 55

Figure 5. Mean KS4 attainment-squared across burn-in iterations 55

Figure 6. SD KS4 attainment-squared across burn-in iterations 55

Figure 7. KS2 attainment distribution in complete cases (#0) vs imputed data (#13) 56

Figure 8. Comparison of IPW approaches 58

List of tables

Table 1. Types of missing data 13

Table 2. School type by boycott status (2010 data) 22

Table 3. Pupil age range by boycott status (2010 data)...22

Table 4. Special education needs and school absence by boycott status (2010-11 data) 23

Table 5. Inspection outcomes by boycott status (2011-12 data).....................................23

Table 6. Test performance by boycott status 24

Table 7: Predictors of missingness model results 27

Table 8: Predictors of missingness models compared 30

Table 9: Predictors of missingness using percent of year 11 pupils attending boycott school 31

Table 10: Summary of measures used 34

Table 11: Multilevel random intercept model of KS2 attainment results 37

Table 12: Nested random intercept models of KS2 attainment compared 39

Table 13: 'Trimmed' multilevel predictor model of KS2 attainment results.......................42

Table 14. Teacher assessments by boycott status, 2009, 2010, 2011 50

Table 15: Variables included in the multiple imputation 54

Table 16. Mean KS2 attainment by imputation 57

Executive summary

The Second Longitudinal Study of Young People in England (LSYPE2) provides a resource for evidence-based policy development. A significant barrier to achieving this purpose could be the missingness present in LSYPE2, owing to a boycott of Key Stage 2 (KS2) testing in 2010. Specifically, in 2010, 15,518 maintained schools were expected to administer KS2 tests, but around one-quarter (4,005 schools; 26 per cent) of these did not administer it. Boycotts of national tests leave gaps in pupils' attainment records and, in the case of LSYPE2, threaten to undermine a large-scale longitudinal study with substantial policy relevance. This project sought to find a way to calculate values for pupils who attended schools that boycotted KS2 tests in 2010 and/or partly mitigate the effect of the boycott on this study.

Prior attainment data is something that should be incorporated into even basic analyses of LSYPE2, and so analysts and researchers need to decide whether the missing data arising from the boycott will cause difficulties when they are making inferences and conclusions in their work, and they then need to take appropriate steps to deal with these difficulties if necessary.

The results of the analyses undertaken for this project suggest that complete-cases analyses using only pupil-level data that include a random effect for primary sampling unit (i.e. secondary school) should be unbiased. Comparing complete-case analysis with multiple imputation (MI) suggests that MI would be more efficient than the complete-case approach – i.e. standard errors (SE) would be smaller, meaning this approach should be used if statistical inference is the aim of a given analysis.

In this report we present an introduction to LSYPE2 and the KS2 Standard Assessment Tests (SATS) boycott in Chapter 1 and to statistical issues with missing data, and to methods for addressing these in Chapter 2. In Chapters 3 and 4 we explore predictors of missing KS2 test scores among LSYPE2 cohort participants, as well as predictors of KS2 attainment. Chapters 6 to 9 describe the methodological approaches and challenges to the MI and inverse probability weighting (IPW) approaches taken. Sensitivity analyses are presented in the appendix to this report. The user guide accompanying this report walks potential users of the imputed/inverse probability weighting variables through descriptive and multivariable analyses.

We have taken a pragmatic approach to this work, balancing the need for practical solutions for analysts with the desire to exhaustively explore options for dealing with missingness. This report includes an assessment of the strengths and limitations of the approaches taken, in particular, the assumptions made in the development of the MI data.

Acknowledgements

We gratefully acknowledge the helpful and insightful comments provided throughout this project by Dr. George Leckie (University of Bristol) and Dr. Gary Abel (University of Exeter). We thank responders to the consultation sent out during this work (Tom Benton, Jake Anders, Steve Strand, Andy Ross) and we thank, in particular, Professor Harvey Goldstein for his engagement in and comments on the work. We are also very grateful to Clare Baker and Sarah Lasher at the Department for Education for their support. Any errors are the responsibility of the authors.

Catherine Saunders, University of Cambridge
Ian Brunton-Smith, University of Warwick
Bryn Garrod, Martin Stepanek, RAND Europe
Anna Vignoles, University of Cambridge
Alex Sutherland, RAND Europe

Abbreviations and acronyms

Abbreviation	Acronym
CI	confidence interval
DfE	Department for Education
EAL	English as an additional language
FIML	full information maximum likelihood
FSM	free school meals
GHQ	general health questionnaire
ICC	intra-class correlation coefficient
IPW	inverse probability weight(ing)
KS1	Key Stage 1
KS2	Key Stage 2
KS4	Key Stage 4
LL	log-likelihood
LSYPE1	The First Longitudinal Study of Young People in England
LSYPE2	The Second Longitudinal Study of Young People in England
MAR	missing at random
MCAR	missing completely at random
MNAR	missing not at random
MI	multiple imputation
NPD	National Pupil Database
Ofsted	Office for Standards in Education, Children's Services and Skills
PMM	predictive mean matching
PSU	primary sampling unit
SD	standard deviation
SE	standard error
SEN	special educational needs
TA	teacher assessment

1. Introduction

In this chapter we provide an introduction to the three dimensions of this work. The Key Stage 2 SATs boycott in 2010, the Second Longitudinal Study of Young People in England (LSYPE2) study, and a longer methodological introduction to the issues of and approaches to addressing missing data.

The Key Stage 2 SATS boycott

In 2010, 15,518 maintained schools were expected to administer KS2 tests, but 4,005 (26 per cent) of these schools did not. Teaching unions cited stress for pupils and perceived unfairness of league tables for schools in more difficult areas, with harder to teach children (Curtis, 2009, Shepherd and Williams, 2010) as the key reasons for the boycott. The 2010 decision to boycott the tests (or not) was made at the school (head teacher or senior leadership) level, and therefore no pupils at boycott schools will have taken the tests.

LSYPE2

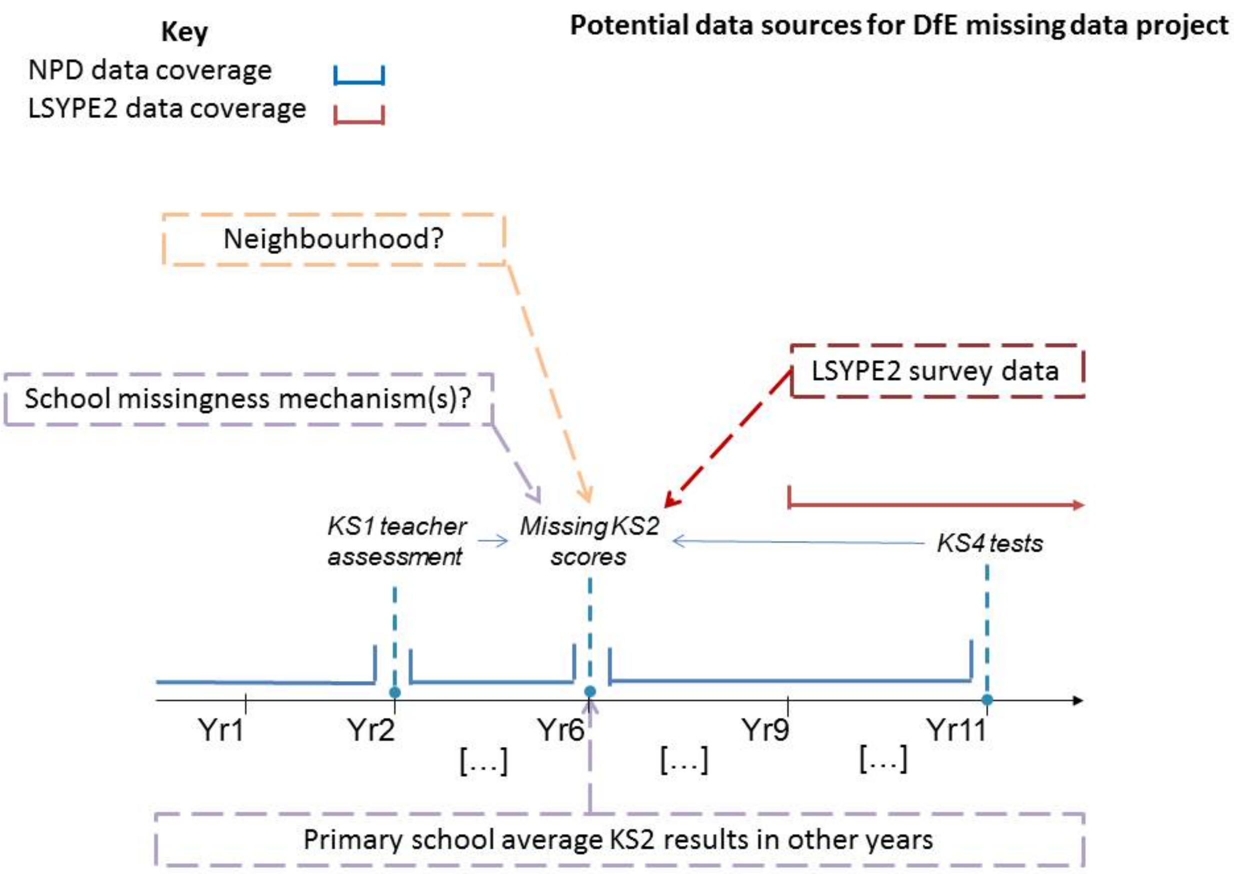
LSYPE2 follows more than 13,000 pupils as they move on from compulsory education and begin their career paths, providing detailed insights into their background, choices and lives from the ages of 13/14 to 19/20. LSYPE2 is a representative sample drawn from individuals attending year 9 between 1 September 2012 and 31 August 2013 who would be turning 14 within that time period and who would normally be resident in England at the time of sampling. This is the cohort of pupils who were affected by the SATS boycott in 2010. LSYPE2 is an incredibly rich data source. The study is intended to have seven ‘waves'. To date, Waves 1, 2 and 3 have been completed. Additional pupil-level data from the National Pupil Database (NPD), including pupil and school characteristics and pupil attainment (including KS2 results), are also available and linked to LSYPE2 survey responses.

When reading this report, it is important to keep in mind that there are two sets of schools that pupils attended that we incorporate into our analysis at different points. The first is the school that pupils were sampled in for the LSYPE2 study (the so-called primary sampling unit). This forms the unit that pupils are clustered in for the majority of our analyses, so for the most part ‘school’ should be read to mean ‘secondary school’. The second set of schools were the primary schools that LSYPE2 pupils were attending in 2010, the year of the boycott. For some analyses – and we make explicit which ones – we are using these primary schools as the unit of clustering.

The research problem

Our research was faced with two issues: the fact that missingness in KS2 data arose prior to the LSYPE2 survey and the fact that there would likely be substantial interest in using KS2 data when analysing LSYPE2. To overcome these issues, we had to combine data from different sources that relate to different time frames and units of analysis. Beyond that, and perhaps more importantly, we tried to find an approach that can incorporate measures that are able to account for missingness, as well as provide plausible estimates for the missing values. Figure 1 illustrates the combination of sources and time frames that the project had to incorporate.

Figure 1. Data sources and time frames for LSYPE2 missing data project



In the next section, we give a brief overview of types of missing data, their relevance to this project and common approaches available to deal with missingness.

2. Missing Data

Missing data is a problem common to all research, and one which can have a serious impact on the quality of the collected data (Little and Rubin, 2002). At its most benign, missing data simply reduces the achieved sample size and, consequently, the precision of estimates. However, the effects of missing data are not restricted to reductions in sample size. Critically important is that they can also lead to biased inferences about relationships between variables. Broadly speaking, if the LSYPE2 survey responses from those pupils who did not attend schools that participated in the KS2 boycott in 2010 differ systematically from those who did, then any analysis based only on those pupils with test scores runs the risk of producing biased estimates of the 'true' population value. For example, one obvious outcome of interest is attainment at Key Stage 4 (KS4), where KS2 results are often used as a predictor. If the probability of attending a boycott school is related to KS2 attainment and thus to KS4 attainment, perhaps because of the characteristics of schools which chose to participate in the boycott, naive estimates of KS4 attainment may be inaccurate and potentially misleading.

Types of missing data

In order to address problems of missing data, a careful assessment of the nature of missingness, and the reasons why it may have occurred, are needed.

The conceptual framework for handling missing data was introduced by Rubin in 1976 (Seaman et al., 2012b), who classified missing-data mechanisms into separate classes. Within a class, certain analyses are valid, and others are not. The key issue is dependence among missingness and unobserved variables, because it is the existence of such relationships that potentially undermines the validity of subsequent analyses. Table 1 gives an overview of different types of missing data.

Table 1. Types of missing data

Missing completely at random (MCAR) – There are no systematic differences between the missing values and the observed values. *MCAR matches intuitively the idea of random missingness. The chance that a unit is missing on an occasion does not depend on any of the missing values nor on values obtained on other occasions. Under MCAR, a complete-case analysis is valid, if potentially inefficient.* **Missing at random (MAR)** – Any systematic difference between the missing values and the observed values can be explained by differences in the observed data. For example, missing KS2 results may be lower than measured scores, but only because pupils from a certain region were more likely to attend boycott schools, and because in this region scores tend to be lower on average. *In spite of the terminology, MAR does not correspond to the intuitive notion of randomness. Under MAR, missingness may be associated with any variables, observed or not, but, conditional on the values taken by observed variables, there is no residual association with unobserved ones. This is a subtle distinction, and intuitively it corresponds to the idea that any link between missingness and unobserved variables can be 'explained' by the variables that have been observed. An important consequence of the MAR assumption is that relationships that are seen among variables for units that are observed hold as well for units that are not observed. While simple complete-case analyses are generally invalid under MAR (with unit nonresponse), adjustments can be made to correct for this that are based on the observed data. These adjustments exploit the fact that relationships among variables for missing units can be 'borrowed' from those that are observed.* **Missing not at random (MNAR)** – Even after the observed data are taken into account, systematic differences remain between the missing values and the observed values. *A situation which is not MCAR or MAR is MNAR. Even after adjusting for observed variables, there remain associations between missingness and unobserved data. A valid analysis must take account of the missing-data mechanism, but this is usually unknown.*

Adapted from Sterne et al. (2009)

A key methodological point worth considering while reading this report is that multiple imputation assumes that data are missing at random conditional on measured covariates (Table 1). When data are missing not at random, bias in analyses based on multiple imputation may be as big as or bigger than the bias in analyses of complete cases (that is, analyses excluding pupils who attended boycott schools). Unfortunately, it is impossible to determine from the data how large a problem this may be (Sterne et al., 2009). Any imputation analysis needs to consider all the possible reasons for missing data and assess the likelihood of missing not at random being a serious concern.

While the observed data can be used to show that MCAR is untenable as an assumption, the distinction between MAR and MNAR rests on the properties of the unobserved data and therefore cannot be assessed directly. In practice, contextual information and other data must be used to help make such decision. It should also be noted that it is very likely that in any given setting a range of missing-data mechanisms will be operating.

Addressing missing data

The first approach for any analyst at the start of a piece of work is to explore and understand the basic descriptive statistics of the missing data in the dataset. In this case, descriptive statistics of the extent of any missing data and the characteristics of pupils affected support an initial understanding of the implications of the missing data in the analysis.

As a rule of thumb, where less than 3–5 per cent of observations in an analysis would be excluded because of missing data, a ‘complete case’ analysis, i.e. an analysis that only includes people without missing data, is usually considered appropriate. Where a higher percentage of observations than this range of 3–5 per cent are missing, it is worth considering whether there are any variables that have a particularly high proportion of missing data, and which could be excluded from the analysis to allow more observations overall to be included. In the LSYPE2 cohort, there is missing data for KS2 attainment for about 25–30 per cent of pupils (depending on the exact analysis sample and wave). Because KS2 attainment is a key variable for many analyses using the LSYPE2 cohort, it cannot simply be excluded.

Beyond these simple considerations, which should be the starting point for any analysis where missing data is an issue, there has been much statistical research on the problem of missing data, and a range of approaches have been developed to address particularly the key concerns of efficiency (excluding missing data can lead to estimates that are not very precise) and bias (where there are systematic differences between groups with and without missing data this can lead to incorrect, or biased, results).

‘Principled methods’, meaning those using suitable and clearly stated assumptions, attempt to make the most use of the observed data to maximise the efficiency of analyses and/or correct for biases due to missingness that is in some way related to both seen and unseen observations. The assumptions can be expressed in different ways, and their particular expression is linked to the main classes of principled analysis. We discuss three main classes of principled methods in more detail below: multiple imputation (MI), full information maximum likelihood (FIML) and inverse probability weighting (IPW).

Multiple imputation

One of the most widely used and flexible approaches to handling missing data is multiple imputation (MI) (Rubin, 1987, Kenward and Carpenter, 2007), which also provides a convenient framework for sensitivity analysis (Carpenter et al., 2007). We discuss this approach in detail below, and we also note that, for example, the Programme for International Student Assessment (PISA) gives five plausible values for students' mathematics scores (equivalent to five imputed values), so the notion of giving multiple plausible/imputed values in education datasets is not new (see OECD, 2014).

Under the MAR assumption, MI gives unbiased estimates and unbiased standard errors (SE), and it is efficient. The results from MI and FIML (discussed below) should be very similar. MI has limitations, but it is a useful tool for addressing missing data.

Briefly, in multiple imputation, to allow for uncertainty in the imputations, multiple versions of an imputed dataset are generated and then analysed. These imputed datasets replace missing values with 'plausible' substitutes based on the distribution of the observed data, and MI includes randomness to reflect uncertainty. The estimates from each dataset are combined using Rubin's rules, which allow for the combination of individual estimates and standard errors from each of the imputed datasets into an overall single MI estimate and SE to provide valid statistical results (Marshall et al., 2009). This analysis can be implemented in a straightforward way using standard statistical software.

It is worth highlighting that this is not a single-stage process, i.e. the multiple imputation is done first and then the imputed data are analysed. This has the advantage that an imputed dataset can be created and can then be shared with users. However, it is also worth emphasising at this stage that the imputation model and model of interest need to be consistent, and that when they are not there is the possibility of incorrect inference. (What this means is that the MI model must be at least as general as the model of interest, and the relationships in the model of interest must be preserved in the MI model).

As with all statistical adjustments, it is important to judge the sensitivity of any results to alternative specifications. This is relatively straightforward with MI, and we present results from sensitivity analyses below.

Full information maximum likelihood approach

An alternative broad class of principled approaches rests on the formulation of a model for the missingness mechanism directly, that is, a model for the probability of an observation being missing expressed in terms of other data and, when the mechanism is assumed to be non-random, on the potentially missing observation itself. This model can be combined with the model for the data in a single analysis model; in some settings this is called full information maximum likelihood (FIML). Many such models have been developed and applied in a wide variety of settings. One particularly well known example is the Diggle-Kenward model for longitudinal data (Diggle and Kenward, 1994). Such models can be very elaborate, and several software platforms are available for their implementation. For example, Mplus uses the frequentist paradigm, and WinBUGS uses the Bayesian. So although being able to estimate results with a single model is appealing, FIML may not be accessible to many users of statistical software, and those with experience of using FIML would likely be doing this analysis themselves.

Inverse probability weighting approach

The model for the missingness mechanism can also be used in a rather different, less parametric way. If the model can be estimated consistently from the observed data, and this is not directly possible under non-random mechanisms, then it can be used to generate weights in the analysis to correct for missingness, in a so-called inverse probability weighting (IPW) approach.

IPW has a long history in survey sampling. Here a model for the probability of a unit being missing is constructed from the predictors identified at an earlier, exploratory phase. The inverse of the fitted probabilities of observing the units are then used to weight the units, and in so doing produce an analysis that represents the original sampled population. For example, if explorations of missingness identified that women were half as likely to remain in the sample as men, then the observed responses from women would each be given twice as much emphasis in the final analysis (a weight of 2 would be applied). The result of these adjustments is that the analytic survey appears, at least at face value, to accurately represent the original population.

In their most basic forms, such analyses are inefficient, but a large, rather technical literature now exists that develops more efficient estimators based on IPW, and which have potentially useful properties (Scharfstein et al., 1999).

Ad hoc approaches to missing data

Alternative, ad hoc procedures are best avoided unless the proportion of missing data is very small. Such ad hoc methods usually take the form of simple single imputation methods, such as replacing missing values by observed means, or observations carried forward through time. The resulting completed dataset is then typically analysed as though it were complete. Unless this is done in a carefully structured way, with proper account taken of the fact that some data have been ‘created’ (something that is rarely done in practice), such a procedure is invalid. The resulting dataset and subsequent analyses can be biased, which is perhaps of more concern to analysts. And the measures of precision and resulting inferences will be wrong, with the analyst mistakenly acting as though there are more data than there really are (Allison, 2000). (The latter point refers to making incorrect claims of statistical significance; the former point refers to the fact that, even if statistically significant, an estimate may be far from correct).

Creating a single imputed value for the missing KS2 data (and a flag indicating that the value has been imputed) presents one simple, ad hoc option. Our view is that this approach, although it might be considered desirable in terms of simplicity for users, is methodologically weak compared with both the IPW and MI approaches, and that the IPW approach presents a parallel simple alternative strategy. Although the approach to missing data in the First Longitudinal Study of Young People in England (LSYPE1) did include single imputed values (Piesse and Kalton, 2009), these were for variables with much less missing data than the 25–30 per cent missing KS2 test scores. There are well-established critiques of the approach taken by Piesse and Kalton, primarily that whilst estimation of analytic statistics from a dataset in which there is an imputed KS2 variable (with an appropriate flag) is straightforward, the variance will be incorrect. This again means that the standard errors will be incorrect and thus *p*-values would be misleading. If one only seeks to ‘control’ for prior attainment, because the focus is on other variables, this might be acceptable; however, we anticipate that a large proportion of analyses involving LSYPE2 would entail assessing the relationship of KS2 to outcomes, such as KS4 attainment, subject choice, and so on. Under those circumstances, the single imputed value approach – pragmatic as it may be – may not be acceptable to more technically minded researchers/statisticians. This is one motivation for pursing alternatives.

These various approaches to missing data differ greatly in their flexibility, technical complexity and current practical applicability. Most importantly, the embedded key assumptions that permit conclusions to be drawn from incomplete data differ greatly among the methods, and these assumptions can have quite different implications for the substantive setting. For such methods to be useful in any given setting, it is essential that the assumptions on which they rest can be translated directly into the concepts of the research setting, and so provide a platform for constructive debate. It is not just the technical statistical features of the methods that need to be assessed for use in the LSYPE2 setting, but also their transparency and their relevance to the questions at hand. We describe in detail the assumptions underlying the approaches taken in this report.

Approach to missing KS2 data in this project

In this work we take two approaches to addressing missing KS2 attainment data over and above a simple ‘complete-case analysis’ that ignores missingness. These approaches are the creation of an IPW and the creation of an MI dataset with imputed KS2 attainment scores. We discuss the methodological approaches taken in the following sections. Although the most appropriate approach to missing data will vary from analysis to analysis, the aim of this work is to provide a practical solution appropriate for a wide range of users of the LSYPE2 data.

Methods

Descriptive statistics and technical issues for the LSYPE2 cohort are described elsewhere (TNS BMRB, 2013b), so in this report we concentrate on issues relevant to the proposed MI and IPW approaches. As noted above, we employ a complete-case analysis as a basic comparator for both the MI and IPW approaches.

For both MI and IPW, an understanding of the predictors of missingness is important. It is important to understand the nature of the missing data not only from a descriptive point of view, but also from a practical perspective. In MI, variables associated with whether or not data are missing are important to include, in order that the assumption that data are MAR (that is, conditional on measured covariates) is appropriate. In addition, the regression model which is used to describe missingness can also be used to create the IPW. Because the decision to boycott the KS2 tests was made at the school level, rather than at the pupil level, we consider both school and pupil predictors of missingness. We present these findings in Chapter 3.

In addition, a good MI model will also include variables which are predictors of the values in the missing variable itself. We present these findings in Chapter 4.

In this work, we propose the creation of an IPW and MI dataset for KS2 data missing as a result of the boycott in 2010. For both these approaches there are several methodological questions and issues. In the following two chapters, Chapters 3 and 4, we describe these questions and the approaches taken to explore or address them in this project.

We also raised and discussed the possibility of using Key Stage (KS) 1, KS2 and KS4 teacher assessment and other pupil or school characteristics contained in the NPD to impute and/or weight the missing KS2 results for the entire 2010 cohort, over and above the LSYPE2 sample. This was not feasible in the time-frame of the study, but we believe that this could be explored as a way of offering the Department for Education (DfE) an efficient way of accounting for boycotts in national testing.

In the final two chapters, we describe the technical specifications and Stata code for the MI and IPW models, plus details of the model checking. A separate user guide with example analyses is also available for users of the MI and IPW variables. This guide leaves out the technical detail presented in this report.

3. Predictors of missingness

Summary of key findings

Using descriptive and multivariable analyses of school- and pupil-level data from NPD and Wave 1 LSYPE2 cohort responses, we explored the correlates of the KS2 2010 test boycott.

In school-level analysis on primary schools, we first considered school-level measures available in publicly reported school-level data for all primary schools in England. We found that primary schools where the boycott occurred were, on average: slightly larger, less likely to have a religious affiliation, more likely to have a 3–11 age range, more likely located in the North East and more likely to be in urban centres. The pupil populations at boycotting schools on average have higher percentages of pupils with English as an additional language (EAL), in receipt of free school meals (FSM) and with special educational needs (SEN).

KS2 school-level test performance in boycott schools in the years around the boycott, and teacher assessments (TA) in the year of the boycott, were lower on average in boycott schools compared with non-boycott schools, although the absolute differences were small. The mean of the school-level percentage of pupils achieving level 4+ in TA was higher in non-boycott schools for English (81.3 in boycott and 82.3 in non-boycott schools); for maths (81.5 in boycott and 82.8 in non-boycott schools); and for science (85.0 in boycott and 86.5 in non-boycott schools). Boycott schools also had, on average, lower Office for Standards in Education, Children's Services and Skills (Ofsted) ratings reported in 2011 compared with non-boycott schools. For example, 24.1 per cent of schools with an ‘outstanding’ Ofsted rating in 2011 were boycott schools, compared with 34.9 per cent of schools with an ‘inadequate’ rating; meaning that a higher proportion of ‘inadequate’ schools compared to ‘outstanding’ schools boycotted.

In the pupil-level analysis conducted for preliminary stages of this project, we found that there were very few pupil or household characteristics for members of the LSYPE2 cohort that predict whether the pupils attended boycott schools at KS2.

In analyses using primary school-level variables from NPD linked to the LSYPE2 cohort, school characteristics were found to be predictors of a pupil having attended a boycott school, even after adjusting for pupil and family characteristics. In further analysis, the Department for Education provided the research team with a ‘non-standard’ variable, namely, the proportion of pupils in the schools from which they were sampled, being the primary sampling unit (PSU) who attended a boycott school at KS2. Perhaps unsurprisingly, this was also a very strong predictor of missingness.

LSYPE2 pupil characteristics do not predict a pupil having attended a boycott school at KS2 particularly well. Although attending a boycott school may be related to unobserved factors that then influence pupil achievement (as this cannot be well predicted at the pupil level, except by region), this lack of prediction somewhat reduces concerns about unrepresentative KS2 missingness in this cohort. In short, the lack of a systematic relationship to missingness of pupil-level predictors reinforces that the LSYPE2 cohort is representative of the wider population from which it was drawn.

Further, the lack of strong, consistent associations between pupil characteristics and school-level decisions to boycott are also to be expected. The ‘missingness mechanism’ for KS2 test results in the LSYPE2 cohort occurred at the school level rather than at the individual pupil level, because it was the schools – or rather their head teachers – who chose to boycott the tests or not. In these preliminary analyses we find that there are relatively few pupil-level predictors of pupils having missing KS2 results. This is not to say that pupil-level characteristics are not important in the MI, because despite being less important predictors of missingness they are predictive of actual KS2 results. For the IPW, the regression model developed in this analysis includes both pupil- and school-level variables, in particular, the proportion of pupils in the PSU with missing data at KS2.

Primary school–level predictors of KS2 boycott

We carried out descriptive analyses of boycott and non-boycott primary schools using public data from 2010 (school characteristics and past performance), 2011 (some additional primary school characteristics not available in 2010 data) and 2012 (school performance in the two years after the boycott) to explore whether there are any particular features of types of primary schools that are over- or underrepresented. In 2010, 15,518 maintained schools were expected to administer KS2 tests, but 4,005 (26 per cent) of these schools did not administer them. In this section, where schools have missing data or values suppressed for a particular variable, they were not included in the analysis, and they were excluded or included on a case-by-case basis. Therefore the sample size for each variable is reported separately. Because of the large sample size, in almost all analyses the differences between boycott and non-boycott schools were statistically significant ($p<0.001$) for nearly all variables.[2] We therefore focus on the magnitude and direction of these differences in this reporting.

[2] The exceptions were Percentage of pupils absent or disapplied in science Teacher Assessment (PsciTaAd) and Percentage of pupils achieving level 4 or above in both English and mathematics in 2009–2012 (PTENGMATX09-PTENGMATX12).

School characteristics

We found that ‘voluntary aided’ (21.4 per cent compared with 24.1 per cent) and ‘voluntary controlled’ (10.8 per cent compared with 16.5 per cent) primary schools were underrepresented among boycott schools compared with non-boycott schools (Table 2), perhaps reflecting that religious schools tend to serve less-deprived pupil populations.

Table 2. School type by boycott status (2010 data)

		School type N (%)					
Boycott status	Total	Voluntary aided	Community school	Foundation school	Academy	Foundation special school	Voluntary controlled
All	14,793	3,455	8,650	425	26	26	2,211
Boycott	3,923	839 (21.4)	2,559 (65.2)	97 (2.5)	4 (0.1)	2 (0.1)	422 (10.8)
Non-boycott	10,870	2,616 (24.1)	6,091 (56.0)	328 (3.0)	22 (0.2)	24 (0.2)	1,789 (16.5)

Data from Department for Education (2010c)

Schools with 3-11 age ranges made up 46.5 per cent of boycott schools but only 33.1 per cent of non-boycott primary schools (Table 3). It is possible that relatively fewer schools with a 7-11 age range (8.4 per cent of boycott and 9.3 per cent of non-boycott schools) boycotted the KS2 tests because these schools have fewer other measures (such as KS1 results) on which pupil performance is measured.

Table 3. Pupil age range by boycott status (2010 data)

		Pupil age range N (%)				
Boycott status	Total	3-11	4-11	5-11	7-11	Other*
All	14,793	5,417	5,622	2,097	1,339	318
Boycott	3,923	1,823 (46.5)	1,316 (35.5)	431 (11.0)	329 (8.4)	24 (0.6)
Non-boycott	10,870	3,594 (33.1)	4,306 (36.9)	1,666 (15.3)	1,010 (9.3)	294 (2.7)

* Other age ranges were: 2-11, 2-16, 2-18, 2-19, 3-12, 3-13, 3-16, 3-18, 3-19, 4-12, 4-16, 4-18, 4-19, 5-12, 5-16, 5-18, 5-19, 6-19, 7-16, 7-18, 8-11, 8-12, 8-18, 9-12, 9-13, 10-13, 10-14.

Boycott primary schools had slightly higher numbers of pupils (mean of school measures 272.8, vs 247.1 in non-boycott schools), with slightly higher percentages of pupils with SEN (across all three measures, School Action, School Action Plus, or a statement), with slightly higher percentages of pupil absence (both authorised and unauthorised (Table 4)). Absolute differences between these measures across boycott and non-boycott schools are relatively small.

Using 2011 data on primary school characteristics (because publicly reported data from 2010 was not available), we see that schools where a higher proportion of pupils are eligible for free school meals, and where for a higher proportion of pupils English is not the first language, are overrepresented among boycott schools. Schools with poorer Ofsted ratings were relatively overrepresented among boycott schools, and there is little difference in the proportion of boycott schools among schools that were not inspected in the year following the boycott compared with those who were (27.7 per cent vs 25.6 per cent). Although it is possible that Ofsted 2011 ratings may have been negatively influenced for those schools which boycotted by the very fact that they boycotted, even after excluding those schools that were inspected in the year following the boycott, schools with poorer historic Ofsted ratings are still relatively overrepresented (Table 5).

Table 4. Special education needs and school absence by boycott status (2010-11 data)

Measure [variable name from Public Reporting]	Number of schools (mean of school-level measures) All	Boycott	Non-boycott
Total number of pupils (including part-time pupils) [TotPup]	14,793 (253.9)	3,923 (272.8)	10,870 (247.1)
Pupils with statements of SEN or supported at School Action Plus: percentage [PSENssap]	14,767 (8.5)	3,921 (8.7)	10,846 (8.4)
Pupils with SEN, supported at School Action: percentage [PSENsa]	14,767 (11.9)	3,921 (12.3)	10,846 (11.7)
Eligible pupils with SEN with a statement or supported at School Action Plus: percentage [PEligSENssap]	13,503 (11.0)	3,687 (11.5)	9,816 (10.8)
Eligible pupils with SEN supported at School Action: percentage [PEligSENsa]	13,503 (14.6)	3,687 (15.3)	9,816 (14.3)
Percentage of sessions missed through total unauthorised absence [PAbsUa]	14,793 (0.7)	3,923 (0.8)	10,870 (0.7)
Percentage of pupils in school with persistent absence [PAbsPa]	14,793 (1.7)	3,923 (1.9)	10,870 (1.7)

Table 5. Inspection outcomes by boycott status (2011-12 data)

	Total schools	Number boycott schools	Number non-boycott schools	Boycott	Non-boycott
Pupils eligible for free school meals	14,224	3,819	10,405	median 18.3% mean 17.9%	median 12.7% mean 22.0%
Pupils w/English as another language		3,366	9,140	median 6% mean13.6%	median 4.5% mean18.5%
				By inspection grade	
Inspection outcome at last inspection				% of schools that boycotted	% of schools non-boycott
grade 1 (outstanding)	2,240	539	1,701	24.1	75.9
grade 2 (good)	7,539	2,037	5,502	27.0	73.0
grade 3 (satisfactory / requires improvement*)	4,307	1,131	3,176	26.3	73.7

grade 4 (inadequate).	373	130	243	34.9	65.1
Inspection outcome at last inspection – only schools <u>not</u> inspected in 2011/2012					
grade 1 (outstanding)	1,617	404	1,213	25.0	75.0
grade 2 (good)	3,494	994	2,500	28.4	71.6
grade 3 (satisfactory / requires improvement*)	1,353	390	963	28.8	71.2
Total not inspected in 2011-2012	6,464	1,788	4,676	27.7	72.3
Inspected in 2011-2012 (i.e. expecting an inspection at the time of the boycott)	7,995	2,049	5,946	25.6	74.4

* Schools inspected in academic year 2011-2012 would have been expecting an inspection at the time of the boycott.

School performance in the years before and after the boycott

On average across the whole country, school performance was higher in non-boycott schools across all measures in the years before and after the boycott, and teacher assessments in 2010 were higher in non-boycott compared with boycott schools (Table 6). We therefore include a measure of previous school-level average KS2 point score in our analyses below.

Table 6. Test performance by boycott status

Measure [variable name from Public Reporting]	Number of schools (mean of school-level measures)		
	All	**Boycott**	**Non-boycott**
Average point score for English and maths 2007 [ApsEngmatTest07]	13,306 (27.6)	3,647 (27.4)	9,659 (27.6)
Average point score for English and maths 2008 [ApsEngmatTest08]	13,530 (27.5)	3,682 (27.3)	9,848 (27.6)
Average point score for English and maths 2009 [ApsEngmatTest09]	13,662 (27.6)	3,732 (27.4)	9,930 (27.6)
Percentage achieving level 4 or above in both English and maths 2008 [PEngmatTestL4p08]	13,345 (74.4)	3,648 (73.1)	9,697 (74.8)
Percentage achieving level 4 or above in both English and maths 2009 [PEngmatTestL4p09]	13,466 (73.4)	3,687 (72.4)	9,779 (73.8)
Percentage of pupils achieving level 4 or above in both English and mathematics in 2009* [PTENGMATX09]	13,990 (73.6)	3,748 (73.3)	10,242 (73.7)
Percentage of pupils achieving level 4 or above in both English and mathematics in 2011* [PTENGMATX11]	14,428 (76.9)	3,841 (76.7)	10,587 (77.0)
Percentage of pupils achieving level 4 or above in both English and mathematics in 2012* [PTENGMATX12]	14,396 (82.5)	3,829 (82.2)	10,567 (82.6)

*From 2012 publicly reported performance data.

Combined model for school, pupil and other predictors of missingness

In the previous section, we set out some descriptive KS2 primary school-level relationships to missingness. In this section, we present results for multi-level, multi-variable logit models for missingness *using the LSYPE2 secondary school as the unit clustering the data*. The main purposes of this analysis were to (i) assess which factors are predictive of missingness; in order to (ii) inform the inverse probability weighting (IPW) approach to boycott data and (iii) identify variables to include in the MI model to ensure that data could be considered MAR conditional on measured covariates. Our preparatory work presented results using discrete groups of measures covering pupil, household and school (summarised above). In Table 7 we present results of analyses combining these sources of data.

The outcome measure for this analysis was a binary measure of whether or not KS2 results were available for LSYPE2 cohort members (hence our use of a logit model for analysis). This was based on a DfE-supplied measure of whether or not a pupil previously attended a boycott school in 2010 (KS2FAAT10_CONTFLAG), resulting in 2,972 pupils having missing scores. By definition, all pupils attending such schools should not have had KS2 results, but there were 167 pupils who did have KS2 scores (e.g. through moving to another school immediately before the boycott). For the purpose of this analysis, we set those scores to missing, but we will supply the actual KS2 scores with the final dataset.[3] The analysis sample size for predictors of missingness was 6,853 pupils in all models, clustered in 724 schools.

Overall, very few measures were significantly associated with missingness (Table 7). Table 7 reports odds ratios (OR), which allow a comparison of the relative strength of association for different variables included in the model.[4] At the pupil level, consistent with the bivariate and multivariate analysis undertaken by the project team and DfE, some ethnic minority and SEN pupils were more likely to be missing. Similarly, the proportion of EAL and SEN statemented students in a secondary school were both strongly positively associated with the likelihood of missingness. Pupils attending schools in 'rural' areas were less likely to be missing KS2 attainment measures, whereas pupils in the North East were substantially more likely to be missing than pupils in the reference

[3] It is worth noting that these pupils may be a-typical because they do have KS2 scores. But it might also be where, for example, pupils had 'B' (below the level of the test) or 'T' (working at the level of the test but unable to access it) recorded. Schools supply these codes when registering pupils for the test to indicate that they will not take it so these were known before the boycott took place.

[4] Odds ratios are centred around 1.00. An odds ratio greater than one (e.g. OR 2.5) indicates that the odds of an event occurring are larger relative to the comparison category. Odds ratios less than one (e.g. OR 0.50) indicate that the odds of an event occurring are lower relative to the comparison category. So one might say for an OR of 0.50 that the odds of an event occurring were about half those of the comparison category.

region (London). Again, this fits with both our own preliminary analysis and the DfE's internal analysis.

Table 7: Predictors of missingness model results

Outcome: missing	Odds ratio	Standard error	z	P>z	95% CI LB	95% CI UB
Female	1.012	0.075	0.170	0.867	0.875	1.171
Ethnicity: White (reference)						
Ethnicity: dual/multiple	1.423	0.252	1.990	0.047	1.005	2.014
Ethnicity: Indian	1.191	0.438	0.480	0.635	0.579	2.448
Ethnicity: Pakistani	2.231	0.653	2.740	0.006	1.257	3.960
Ethnicity: Bangladeshi	1.331	0.438	0.870	0.384	0.699	2.536
Ethnicity: Black	1.260	0.204	1.430	0.153	0.917	1.730
Ethnicity: Other	1.282	0.392	0.810	0.417	0.704	2.333
QOB 1 Sep-Nov (reference)						
QOB 2 Dec-Feb	0.965	0.096	-0.350	0.723	0.794	1.174
QOB 3 Mar-May	0.942	0.095	-0.600	0.549	0.773	1.147
QOB 4 June-Aug	0.954	0.093	-0.480	0.629	0.789	1.154
Ever 6 FSM? 1=Yes	1.101	0.116	0.910	0.361	0.896	1.353
SEN No SEN (reference)						
SEN School [A]ction	1.252	0.133	2.120	0.034	1.017	1.541
SEN School Action [P]lus	0.758	0.111	-1.890	0.059	0.568	1.011
SEN [S]tatemented	0.815	0.222	-0.750	0.452	0.478	1.389
Long-term limiting illness	0.990	0.103	-0.100	0.922	0.807	1.214
Overall absence (annual)	1.004	0.003	1.330	0.183	0.998	1.009
Household income (estimated)	0.989	0.015	-0.740	0.457	0.960	1.019
NSSEC: Large emp. & higher man.	0.920	0.177	-0.430	0.666	0.631	1.342
NSSEC: Higher professional	0.749	0.104	-2.080	0.037	0.571	0.983
NSSEC: Lower professional (reference)						
NSSEC: Intermediate occ.	0.905	0.106	-0.850	0.393	0.720	1.138
NSSEC: Small employers & own a/c	0.802	0.125	-1.420	0.156	0.591	1.088
NSSEC: Lower supervisory	1.063	0.175	0.370	0.709	0.770	1.469
NSSEC: Semi-routine occ.	0.853	0.109	-1.240	0.215	0.664	1.096
NSSEC: Routine occ.	0.868	0.139	-0.880	0.377	0.635	1.188
NSSEC: Never worked/long-term unemp.	0.565	0.145	-2.230	0.026	0.342	0.934
HH ed: Degree (reference)						
HH ed: HE below degree	1.090	0.127	0.740	0.459	0.867	1.371
HH ed: A/AS levels or equiv.	0.884	0.110	-0.990	0.320	0.692	1.128
HH ed: 5+ A*-C GCSEs or equiv.	0.838	0.106	-1.400	0.162	0.654	1.074
HH ed: Some GCSE passes or equiv.	0.835	0.112	-1.350	0.179	0.642	1.086
HH ed: Entry level qualifications	0.460	0.365	-0.980	0.327	0.097	2.176
HH ed: Other qualifications	1.066	0.384	0.180	0.860	0.526	2.159
HH ed: No qualifications	1.257	0.218	1.320	0.187	0.895	1.765
# people in HH: 2	1.134	0.197	0.720	0.470	0.806	1.595

Outcome: missing	Odds ratio	Standard error	z	P>z	95% CI LB	95% CI UB
# people in HH: 3	1.025	0.111	0.230	0.821	0.829	1.266
# people in HH: 4 (reference)						
# people in HH: 5	1.014	0.097	0.150	0.885	0.841	1.223
# people in HH: 6	1.102	0.151	0.710	0.478	0.842	1.442
# people in HH: 7	0.742	0.146	-1.520	0.129	0.504	1.090
# people in HH: 8	0.911	0.271	-0.310	0.754	0.509	1.631
# people in HH: 9+	0.721	0.264	-0.890	0.372	0.352	1.478
Single parent family	1.098	0.122	0.850	0.397	0.884	1.364
Anyone in HH full-time employed?	1.048	0.099	0.490	0.622	0.870	1.261
Mother's age at birth	1.064	0.035	1.880	0.060	0.997	1.134
HH: No religion (reference)						
HH: Christian	0.927	0.078	-0.900	0.370	0.785	1.094
HH: Buddhist	0.266	0.215	-1.640	0.101	0.055	1.295
HH: Hindu	1.695	0.755	1.180	0.236	0.708	4.058
HH: Jewish	0.823	0.930	-0.170	0.863	0.090	7.534
HH: Muslim	1.001	0.234	0.000	0.996	0.633	1.583
HH: Sikh	0.933	0.479	-0.140	0.892	0.341	2.550
HH: Other	0.767	0.345	-0.590	0.556	0.318	1.852
HH: is EAL? 1=yes	0.744	0.124	-1.770	0.077	0.536	1.033
HH: Owned/mortgage (reference)						
HH: Rented from LA	0.895	0.103	-0.970	0.334	0.714	1.122
HH: Private rent	0.905	0.117	-0.770	0.440	0.703	1.165
HH: Other arrangement	0.600	0.240	-1.270	0.203	0.274	1.316
School: Community (reference)						
School: Academy	1.000		(omitted)			
School: Community special	1.000		(omitted)			
School: Foundation	0.486	0.118	-2.970	0.003	0.302	0.782
School: Voluntary aided	0.866	0.096	-1.290	0.197	0.696	1.078
School: Voluntary controlled	1.118	0.163	0.760	0.444	0.840	1.487
School: % pupils SEN statement (mc)	1.056	0.028	2.050	0.040	1.002	1.112
School: % pupils SEN no statement (mc)	1.007	0.005	1.210	0.226	0.996	1.017
School: % pupils EAL (mc)	1.008	0.003	2.710	0.007	1.002	1.013
School: % pupils FSM eligible	1.001	0.005	0.150	0.879	0.991	1.011
School: headcount of pupils (mc)	1.000	0.000	0.880	0.376	1.000	1.001
School: average pupil IDACI (mc)	0.989	0.337	-0.030	0.973	0.507	1.928
School: APS 2009 KS2 (mc)	0.954	0.030	-1.490	0.137	0.897	1.015
Rural? 1=Yes	0.663	0.090	-3.030	0.002	0.508	0.865
London region (reference)						
East Midlands	0.976	0.308	-0.080	0.937	0.525	1.811

Outcome: missing	Odds ratio	Standard error	z	P>z	95% CI LB	95% CI UB
East of England	0.487	0.149	-2.350	0.019	0.267	0.888
North East	3.095	1.156	3.030	0.002	1.489	6.434
North West	1.483	0.408	1.430	0.152	0.864	2.545
South East	0.709	0.190	-1.290	0.198	0.420	1.198
South West	0.812	0.267	-0.630	0.526	0.426	1.547
West Midlands	1.470	0.408	1.390	0.165	0.853	2.532
Yorkshire & The Humber	2.245	0.652	2.790	0.005	1.271	3.965
KS4 points (mcKS4_VAPTSC_PTQ)	1.000	0.001	-0.070	0.946	0.999	1.001
Intercept	0.199	0.064	-5.020	0.000	0.106	0.374

Notes:
CI = confidence interval
equiv. = equivalent
HH = household
higher man. = higher managerial
LA = Local Authority
LB = Lower bound
UB = Upper bound
large emp. = Large employer
mc = mean centred
occ. = occupation
unemp. = unemployed

Assessing and understanding between-school variation

We assessed the cumulative effect of adding groups of variables to the model predicting missingness. The reason for doing this was to assess whether and to what extent missingness was clustered by LSYPE2 sampling school (that is, the school that the cohort members were attending in year 9 when the pupils were sampled for LSYPE2) – or, rather, whether the imputation model would need to account for clustering. In addition, we explored whether this variation in Table 8 can be explained by measured covariates. The findings show the result of this, reporting the effect of the new groups of measures on between-school variance, measured by the intra-class correlation coefficient (ICC).[5] Ideally, we would be able to account for the between-school differences and estimate the MI via a single-level imputation, as this is less complex. However, Table 8 shows that even with a complex model (model 6), containing many predictors, there is substantial variability in the school intercepts.[6]

Table 8: Predictors of missingness models compared

Result	Outcome: Missing KS2	ICC
1	Empty – outcome only	0.460
2	Pupil	0.453
3	Pupil + household	0.458
4	Pupil + household + school	0.457
5	Pupil + household + school + region/urban	0.435
6	Pupil + household + school + region/urban + KS4 attainment	0.435
School measures only		
7	School only	0.450
8	School & % of boycott students in year 11	0.132
9	% of boycott students in year 11 only	0.131

As set out above, the missingness mechanism operated at the school/head teacher/ senior leadership level, perhaps informed by pupil ability/characteristics. This means that few measures at the pupil level would be predictive of missingness and, equally, that the more important measures would be at the school level. Specifically, this would be the *Key Stage 2* school-level rather than the LSYPE2 primary sampling unit. So in the absence of measures that capture the reasons for KS2 school boycott participation, we instead included a measure of the proportion of pupils in each secondary school who attended a boycott school at KS2. This measure, on its own, reduces between-school

[5] This is the proportion of the variation in the outcome occurring between schools. If all the variation in outcomes was between schools, the ICC would be 1.00; if none was between schools, the ICC would be 0.00.
[6] Note that model 6 in Table 8 is the summary statistic for the detailed model results presented in Table 7.

variation measured by the ICC, from 0.45 to 0.13 (result 9 in Table 8). We also estimated the percentage of between-school (PSU) variance in missingness that was explained by adjusting for the percentage of boycott students for this same model. As the ICC dropped from 0.46 to 0.13, 97 per cent of between-school variance was explained. This tells us that this measure alone would be a very good predictor of pupil-level missingness[7] and would therefore be important to include in both the IPW and MI models. However, it is worth highlighting that although the proportion of pupils in each secondary school who attended a boycott school at KS2 is an important predictor of missingness, the fact that this is so does not necessarily explain the missingness mechanism. By including this variable, the clustering of missingness at the PSU school level is substantially reduced, providing some support for MI approaches that do not include clustering in the analysis. Full results for this model are given in Table 9.

Table 9: Predictors of missingness using percent of year 11 pupils attending boycott school

Outcome: missing KS2	Odds ratio	Std. err.	z	P>z	95% CI LB	95% CI UB
% of boycott students in year 11	1.058	0.002	27.010	0.000	1.054	1.062
Intercept	0.062	0.005	-34.470	0.000	0.053	0.073
SD of intercept	0.704	0.056			0.602	0.824
ICC	0.131	0.018			0.099	0.171

See Table 7 for abbreviations

Summary

Our combined analysis found that few pupil-level measures were associated with missingness. We believe this is because the mechanism for schools boycotting was at the school level. School-level measures correlated with missingness were the following: the proportion of English as another language pupils, the type of school and the proportion of pupils who were ‘School Action’ in terms of SEN.

[7] It may seem obvious that such a measure is a good predictor of missingness because it is, by definition, constructed using a flag for missingness. But it should be noted that this measure is the proportion of Year 11 pupils who attended a KS2 boycott school, so it also captures in part the localised spatial clustering of missingness that the ‘region’ measures only partially tap into.

4. Predictors of KS2 attainment

As the focus of this work is to create suitable values for pupils missing KS2 because of the boycott, one key step is to understand which factors are most strongly associated with KS2 attainment, and to then use this analysis to inform analysis variable selection for the imputation. Here, 'most strongly' means both the size of the correlation and its predictive power. There are many variables that could be associated with KS2 attainment; we focused on measures that have previously been shown to be associated with KS2 attainment. The starting point for this pool of measures was work previously completed that examined the relationship of measures of socio-economic deprivation to attainment at both KS2 and KS4 (Sutherland et al., 2015b, Sutherland et al., 2015a). That research pulled together a wide range of measures at the level of the pupil, household, school, neighbourhood and beyond to assess their association with attainment at primary and secondary school using two longitudinal cohort datasets. As set out in the reports of that research, the focus was on finding variables that would likely form 'core' predictors in many analytical models and which were not endogenous with pupil attainment (an example of the latter would be that parental aspirations are partly based on parental knowledge of a child's ability). As such, these measures, which have previously been validated using similar datasets, form a useful starting point for predicting KS2 attainment.

In what follows, we (i) set out a summary of the measures included in our final analysis model for KS2 attainment; (ii) present and describe results from our analysis of KS2 predictors; (iii) review sensitivity analyses undertaken; and (iv) summarise our findings and their implications for the handling of missing data.

Summary of measures used

Table 10 below gives a description and summary of the measures used in our final KS2 analysis. For that analysis, we used a random intercept multilevel linear model with KS2 attainment as the outcome, clustering pupils by secondary school (LSYPE2 sampling school). As noted above, the majority of measures were included because they had previously been used in projects with similar aims and data. There were five broad categories of measures: pupil, household/parental, school, neighbourhood/region, other attainment.

Pupil-level measures

For pupil-level measures (Panel A), we included a mixture of demographic measures, such as gender and ethnicity, alongside a measure of pupil-level deprivation (ever6FSM) that has been shown to explain variation in attainment above and beyond household or school-level factors (Sutherland et al., 2015b). We also included whether the pupil had special education needs (SEN), according to the classifications given for SEN in the National Pupil Database (NPD), whether the pupil suffered from a limiting long-term illness, and what the pupil's overall attendance was (incorporating both authorised and unauthorised absences).

Household measures

Household measures (Panel B) lean heavily towards socioeconomic status. This is because a great deal of research has previously shown that measures of household and parental socio-economic status are often strongly associated with pupil attainment (see e.g. the references in Sutherland et al., 2015a). That said, there is a distinction between material resources and so-called 'social capital', so the measures capture both material (e.g. income) and social (e.g. parental education; occupational class) capital.[8] We also include a measure of whether English is a second language at home, owing to persistent differences between EAL and non-EAL households in terms of attainment. Finally, we also include a measure of household religious denomination, reflecting concerns about what are termed 'ethnoreligious penalties' in educational and labour markets (Khattab, 2012).[9]

School-level measures

Panel C sets out the range of school-level measures included in our analysis. These factors include a broad classification of school type, but the majority of measures relate to the school intake in terms of educational difficulties (special educational needs), EAL households and the proportion of FSM-eligible pupils, alongside an average measure of the pupil's neighbourhood IDACI score. While percent FSM and IDACI will likely overlap, evidence shows that not all those eligible for FSM claim it (Hobbs and Vignoles, 2009) and that the proportion eligible fluctuates with economic cycles (Sutherland et al., 2015b). As the focus is on understanding more about a year in which data were not collected, but knowing that school-level results are likely highly correlated, we included a measure of the average KS2 points score from 2009. Finally, to capture something of the physical

[8] One should also keep in mind that general intelligence is strongly associated with pupil attainment outcomes (DEARY, I. J., STRAND, S., SMITH, P. & FERNANDESC, C. 2007. Intelligence and educational achievement. *Intelligence,* 35**,** 13-21.); however, we do not have a measure of this in the data. Further, parent-child correlations in educational attainment may owe more to inherited traits than previously thought (e.g. LUCCHINI, M., DELLA BELLA, S. & PISATI, M. 2013. The weight of the genetic and environmental dimensions in the inter-generational transmission of educational success *European Sociological Review,* 29**,** 289–301.). This research argues that 'the traditional sociological theories used to explain individual differences in educational achievement may not be the best ones, and that it is crucial to consider both genetic and environmental influences when studying social behaviours' (p.289). (See also BURGER, K. 2016. Intergenerational transmission of education in Europe: Do more comprehensive education systems reduce social gradients in student achievement? *Research in Social Stratification and Mobility,* 44**,** 54-67.)

[9] Note that such concerns appear to be based on studies that do not account for differential selection of ethnic or religious groups in education pathways. Leaving this point aside, religion may act as a proxy measure for other cultural or ethnic factors that are salient for understanding attainment.

resources per capita, and because there is mixed evidence on the impact of school size, we included a measure of school size via the headcount for completeness. Overlaid on pupil, household and school measures are two final geographical indicators covering government region and whether the school is classified as being in an ‘urban’ or ‘rural’ area. (For a discussion of these measures, see Sutherland et al., 2015a).

As the purpose is to try to best predict values of pupil attainment at one point in time, the final set of measures cover pupil attainment at both KS2 and KS4, measured by teacher-assessed and standardised exam measures. We have included KS4 results because of the strong correlation between Key Stage results for pupils (Education Endowment Foundation, 2013). We use the KS1 teacher-assessed measures of maths and English, along with the TA measure of KS2 outcomes (although we discuss possible bias in this below). Finally, we use actual KS4 points scores for all pupils we have available data for.

Table 10: **Summary of measures used**

FSM report variables	Type	Variable used	Source
Panel A: pupil			
Gender=Female	Binary	Female	LSYPE2
Ethnicity	Categorical		LSYPE2
Quarter of birth	Categorical	Qob	NPD
FSM_ever_6	Binary	EVERFSM_6_SPR10	NPD
SEN (special educational needs)	Categorical	SEN	NPD
Limiting long-term illness	Binary	Ltlli	LSYPE2
Overall absence	Continuous	OverallAbsence_Annual	NPD
Panel B: Household			
Household yearly net income	Continuous	Inc1Est	LSYPE2
NSSEC Occupations	Categorical	BestNSSEC	LSYPE2
Household qualifications	Categorical	Hhqual	LSYPE2
Household size	Continuous	Hhsize	LSYPE2
household_single parent	Binary	Singlep	LSYPE2
household_employment	Binary	Hhftemp	LSYPE2
Age of mother at birth	Continuous	Mothageatb	LSYPE2
HH religion	Categorical	Religion	LSYPE2
HH English as another language?	Binary	Hheal	LSYPE2
HH tenure type	Categorical	Tenure	LSYPE2
Panel C: School			
School type	Categorical	School_type2	NPD?
School proportion SEN with statements	Continuous	LEA10_Pct_Pupils_SEN_Statemented	NPD
School proportion SEN without statements	Continuous	LEA10_Pct_Pupils_SEN_No_Statemen	NPD
School proportion of pupils from EAL households	Continuous	LEA10_Pct_Pupils_Language_Not_En	NPD
School proportion of pupils eligible for FSM	Continuous	LEA10_Pct_Pupils_FSM_Eligible	NPD
School size	Continuous	LEA10_Headcount_Pupils	NPD
2009 school APS	Continuous	KS2FAAT10_APS09	NPD
IDACI_score of school	Continuous	Bespoke measure	DfE
Panel D: Region/neighbourhood			
Region	Categorical	Categorical	NPD
Urban	Binary	Dummy	NPD
Panel E: Attainment measure			
KS1 TA MATHS LEVEL CATEGORICAL MEASURE	Continuous	KS2_MATLEVTA	NPD

KS1 TA ENGLISH LEVEL	Continuous	KS2_ENGLEVTA	NPD
KS**2** TA measure	Continuous		NPD
KS4 attainment (GCSE point sore)	Continuous	KS4_VAPTSC_PTQ	NPD

Results of KS2 predictors analysis

We ran several multilevel linear models for KS2 outcomes on a complete-case basis (i.e. no adjustment was made for item non-response or attrition for this analysis), and we present results from the most complex version of this analysis, which includes all the variables noted above. As before, we clustered the data by LSYPE2 sampling school. The analysis sample size was 5,138 pupils clustered in 714 secondary schools. Item non-response, missingness arising from attrition, along with the non-random nature of the boycott itself, mean that caution is required in the interpretation of these results. Table 11 sets out the results, and we describe these briefly below.

As expected based on previous literature, many of the included measures were associated with KS2 attainment. Starting with pupil-level measures, we see that females scored, on average, lower than males in KS2 results, as did pupils born later in the school year relative to those born in Q1 of the school year (September–November). Ethnicity was not associated with KS2 attainment – at least not comparing other ethnic groups to the reference category of white pupils. Similarly, 'ever six FSM' was also not associated, but that could be driven in part by the boycott, as this was differentially focused in schools with higher proportions of FSM students. Compared with pupils not regarded as having special education needs (SEN), all those categorised as 'School Action Plus' and 'statemented' attained a lower mark.[10]

There were few associations between household measures and KS2 attainment. Exceptions were lower KS2 scores for those whose parents had never worked versus lower professional households, as well as and lower scores for those where the highest parent/guardian educational attainment was A/AS Level compared with degree-educated households. The number of people living in a household – a factor perhaps partly related to material wealth but also correlated with other factors – was also associated when comparing to the reference category of four-person households. Those from three- or five-person households achieved higher KS2 scores on average than pupils from four-person households. Finally, pupils from households where English was a second

[10] We originally included a measure of total pupil absence in the model for attainment. The result for this model was a strong positive association between total absence and KS2 attainment, which runs contrary to what might be expected. When reviewing the bivariate association between total absence and attainment, we noticed that there is very high volatility in the KS2 attainment measure as overall absence increases. We believe this result is a statistical artefact of that volatility, so we excluded the overall absence measure from the main analysis.

language were also likely to score lower at KS2 than English as first language households.

As with household measures, there were few school-level measures associated with KS2 performance in this analysis. As expected, the school-level average KS2 scores from 2009 were positively correlated with the 2010 pupil-level outcomes. However, somewhat surprisingly, the proportion of un-statemented SEN pupils in a school was positively correlated with pupil-level KS2 attainment. This might represent ‘School Action’ resulting in additional support being offered to all pupils in schools, or spill-over effects from that additional support being offered to some pupils in schools.

While there was no association between rurality and KS2 outcomes, there was one regional difference. Pupils in the North West had marginally higher scores that those in London, but again the selective nature of the boycott could bias results for aggregated measures, such as region (e.g. if many weaker schools were boycotting in some regions).

Finally, prior pupil attainment, teacher-assessed KS2 ability and KS4 attainment were all strongly positively correlated with observed KS2 attainment as measured by national testing.

Overall, the results of this analysis support the view that any attempt to mitigate the effect of the boycott and to some extent ‘future-proof’ the LSYPE2 data, should consider predictors from the levels of pupil, household, school and wider geography, but the final model used could be less complex than that set out above.

Table 11: Multilevel random intercept model of KS2 attainment results

Outcome: KS2 attainment	coeff.	Standard error	z	p	95% CI LB	95% CI UB
Female	-0.195	0.049	-3.950	0.000	-0.292	-0.098
Ethnicity: White (reference)						
Ethnicity: dual/multiple	-0.132	0.128	-1.030	0.302	-0.383	0.119
Ethnicity: Indian	-0.242	0.264	-0.920	0.360	-0.760	0.276
Ethnicity: Pakistani	0.142	0.224	0.630	0.526	-0.297	0.582
Ethnicity: Bangladeshi	-0.173	0.241	-0.720	0.473	-0.646	0.300
Ethnicity: Black	-0.173	0.113	-1.530	0.125	-0.394	0.048
Ethnicity: Other	-0.256	0.202	-1.270	0.204	-0.651	0.139
QOB 1 Sep-Nov (reference)						
QOB 2 Dec-Feb	-0.144	0.067	-2.150	0.031	-0.276	-0.013
QOB 3 Mar-May	-0.147	0.067	-2.200	0.028	-0.278	-0.016
QOB 4 June-Aug	-0.190	0.065	-2.910	0.004	-0.318	-0.062
Ever 6 FSM? 1=Yes	0.021	0.071	0.300	0.767	-0.119	0.161
SEN No SEN (reference)						
SEN School [A]ction	-0.893	0.077	-11.630	0.000	-1.043	-0.742
SEN School Action [P]lus	-1.059	0.101	-10.450	0.000	-1.258	-0.861
SEN [S]tatemented	-0.909	0.190	-4.770	0.000	-1.283	-0.536
Long-term limiting illness	-0.105	0.068	-1.560	0.119	-0.238	0.027
Household income (estimated)	0.014	0.010	1.440	0.151	-0.005	0.033
NSSEC: Large emp. & higher man.	-0.044	0.124	-0.360	0.722	-0.287	0.199
NSSEC: Higher professional	-0.043	0.086	-0.500	0.615	-0.211	0.125
NSSEC: Lower professional (reference)						
NSSEC: Intermediate occ.	0.031	0.077	0.410	0.685	-0.119	0.182
NSSEC: Small employers & own a/c	0.064	0.104	0.620	0.537	-0.139	0.267
NSSEC: Lower supervisory	-0.097	0.114	-0.850	0.395	-0.319	0.126
NSSEC: Semi-routine occ.	-0.157	0.086	-1.830	0.067	-0.325	0.011
NSSEC: Routine occ.	-0.243	0.110	-2.210	0.027	-0.459	-0.027
NSSEC: Never worked/long-term unemp.	-0.352	0.177	-1.990	0.046	-0.699	-0.006
HH ed: Degree (reference)						
HH ed: HE below degree	-0.086	0.077	-1.120	0.263	-0.236	0.064
HH ed: A/AS levels or equiv.	-0.211	0.082	-2.570	0.010	-0.372	-0.050
HH ed: 5+ A*-C GCSEs or equiv.	-0.157	0.082	-1.920	0.055	-0.318	0.003
HH ed: Some GCSE passes or equiv.	-0.104	0.089	-1.170	0.242	-0.278	0.070
HH ed: Entry level qualifications	-0.410	0.533	-0.770	0.442	-1.455	0.636
HH ed: Other qualifications	-0.043	0.247	-0.170	0.862	-0.527	0.441
HH ed: No qualifications	0.125	0.125	0.990	0.320	-0.121	0.371
# people in HH: 2	0.229	0.121	1.890	0.058	-0.008	0.466
# people in HH: 3	0.183	0.072	2.540	0.011	0.042	0.323
# people in HH: 4 (reference)						
# people in HH: 5	0.140	0.063	2.230	0.026	0.017	0.264
# people in HH: 6	0.064	0.091	0.710	0.481	-0.114	0.241
# people in HH: 7	0.131	0.136	0.970	0.334	-0.135	0.398
# people in HH: 8	-0.076	0.217	-0.350	0.724	-0.501	0.348
# people in HH: 9+	0.287	0.260	1.100	0.270	-0.222	0.795
Single parent family	0.102	0.075	1.350	0.176	-0.046	0.249
Anyone in HH full-time employed?	-0.056	0.063	-0.880	0.377	-0.179	0.068
Mother's age at birth	0.015	0.022	0.700	0.482	-0.028	0.058
HH: No religion (reference)						
HH: Christian	0.001	0.053	0.020	0.985	-0.103	0.106
HH: Buddhist	-0.191	0.445	-0.430	0.668	-1.062	0.681
HH: Hindu	0.437	0.318	1.380	0.169	-0.186	1.060

Outcome: KS2 attainment	coeff.	Standard error	z	p	95% CI LB	95% CI UB
HH: Jewish	0.769	0.454	1.690	0.090	-0.121	1.658
HH: Muslim	-0.088	0.164	-0.540	0.591	-0.410	0.233
HH: Sikh	0.287	0.368	0.780	0.435	-0.434	1.008
HH: Other	0.298	0.307	0.970	0.331	-0.304	0.900
HH: is EAL? 1=yes	-0.512	0.117	-4.390	0.000	-0.741	-0.283
HH: Owned/mortgage (reference)						
HH: Rented from LA	0.009	0.078	0.110	0.910	-0.144	0.162
HH: Private rent	0.013	0.087	0.150	0.881	-0.158	0.184
HH: Other arrangement	0.271	0.239	1.130	0.258	-0.198	0.740
School: Community (reference)						
School: Academy	-0.690	0.473	-1.460	0.144	-1.617	0.236
School: Community special	1.616	1.724	0.940	0.349	-1.763	4.996
School: Foundation	0.158	0.121	1.310	0.190	-0.079	0.395
School: Voluntary aided	0.029	0.068	0.430	0.668	-0.104	0.162
School: Voluntary controlled	0.014	0.088	0.160	0.871	-0.158	0.187
School: % pupils SEN statement (mc)	-0.012	0.017	-0.710	0.478	-0.045	0.021
School: % pupils SEN no statement (mc)	0.011	0.004	3.130	0.002	0.004	0.018
School: % pupils EAL (mc)	0.003	0.002	1.490	0.136	-0.001	0.007
School: % pupils FSM eligible	0.002	0.003	0.670	0.503	-0.004	0.009
School: headcount of pupils (mc)	0.000	0.000	1.540	0.123	0.000	0.001
School: average pupil IDACI (mc)	0.008	0.227	0.040	0.971	-0.436	0.453
School: APS 2009 KS2 (mc)	0.130	0.020	6.540	0.000	0.091	0.169
Rural? 1=Yes	-0.089	0.069	-1.290	0.198	-0.225	0.047
London region (reference)						
East Midlands	0.091	0.129	0.710	0.479	-0.161	0.343
East of England	0.003	0.121	0.020	0.983	-0.234	0.239
North East	0.032	0.160	0.200	0.841	-0.282	0.346
North West	0.257	0.120	2.140	0.032	0.022	0.493
South East	-0.044	0.115	-0.380	0.702	-0.269	0.181
South West	0.141	0.132	1.070	0.285	-0.118	0.400
West Midlands	-0.052	0.120	-0.430	0.668	-0.288	0.184
Yorkshire & The Humber	0.004	0.129	0.030	0.976	-0.250	0.257
KS2 Maths Level (teacher assessed)	2.280	0.047	48.650	0.000	2.188	2.372
KS2 English Level (teacher assessed)	1.673	0.048	34.520	0.000	1.578	1.768
KS4 attainment (mcKS4_VAPTSC_PTQ)	0.010	0.000	22.640	0.000	0.009	0.011
Intercept	11.526	0.290	39.760	0.000	10.958	12.094

Note: Analysis clustered by secondary school (primary sampling unit for LSYPE2).
See Table 7 for abbreviations.

Sensitivity analyses[11]

We ran a series of random intercept multilevel models to assess the additive impact of including different groups of measures. Table 12 below shows model comparison results, starting with the 'empty' model (i.e. just the outcome). Models are compared in terms of the between-school variation (ICC) and the difference in minus two log-likelihood (−2LL), from adding more variables, with $p \leq .05$ meaning a significant improvement in model fit (captured through a likelihood ratio test).[12] To allow for comparisons, the sample size is the same for all analyses (n=5,100). It is clear from Table 12 that the addition of pupil and household variables substantially reduces between-school variation in KS2 attainment. Adding school-level measures reduces this further, to around 2 per cent. But this is not surprising, as these characteristics are all correlated with both pupil ability, KS2 school selection decisions and KS4 school selection.

Table 12: Nested random intercept models of KS2 attainment compared

	Model	ICC	−2LL diff. to prev. model	p-value
1	Empty − outcome only	.125	--	--
2	Pupil	.790	2542.18	<0.000
3	Pupil + household	.365	516.31	<0.000
4	Pupil + household + school characteristics	.170	250.08	<0.000
5	Pupil + household + school + region/urban	.138	18.19	0.033
6	Pupil + household + school + region/urban + attainment	.360	6281.18	<0.000

Note: the large decrease in −2LL between models 5 and 6 is because of the addition of the past and future attainment measures.

Owing to concerns about bias in teacher-assessed measures, we also re-ran the final analysis model without these measures. This led to some changes in the pattern of results – with more household-level and pupil-level measures being associated with KS2 results. This is perhaps to be expected: conditioning on a measure of KS1 attainment, even if imperfectly measured, should capture many pupil-level and household-level sources of variation.

Note that in comparison with the predictors of missingness analysis presented above, for this analysis it is the pupil characteristics that explain most of the variation in KS2

[11] MI data sets and the IPW were created on Wave 3 and Wave 1 linked data respectively, and then merged with Wave 1 data for these analyses. Differences between approaches reflect differences between samples / included excluded individuals at different waves as well as methodological differences between the approaches. Users of the MI data sets / weights should consider the most appropriate analysis sample for their data.

[12] The likelihood ratio test is commonly used to assess improvements in nested models based on the same estimation sample. The difference in log-likelihoods between models is chi-square distributed with degrees of freedom equal to the number of variables added to the model. As of 29 August 2016: http://www.ats.ucla.edu/stat/stata/faq/nested_tests.htm

performance between the schools the pupils attended in Year 9 at the time of sampling for LSYPE2.

‘Trimming’ the attainment model for imputation

The results from the previous analyses illustrate that very few measures overall were related to KS2 attainment. In multiple imputation, we know that adding redundancy to such models simply adds random ‘noise’ to imputation models, making the overall imputation less reliable. Further, many of the measures of socio-economic status used in the preceding analyses are correlated with one another. Although these measures are not perfectly collinear, it makes sense to reduce the imputation model complexity if possible. To this end, we estimated a reduce form model that excluded several measures. Results from this model, and reasons for excluding other measures, are given below. Our ‘trimmed’ attainment contained the following measures:

- Female
- HH size
- Quarter of birth
- HH EAL
- SEN
- % Pupils SEN No Statement
- Ever 6 FSM
- School 2009 KS2 results
- Highest HH qualification
- KS4 points score

As per the previous set of results, we removed the teacher-assessed attainment measures owing to concerns about bias. We know from both the preliminary stages of this project and from other research that teacher-assessed measures were lower for boycott schools. Although knowing the direction of potential bias is helpful, it further complicates the MI estimation.[13] We excluded the household income measure for a similar reason. We know from previous research that household income is estimated with unknown bias and that, with both Millennium Cohort Study and LSYPE1 data, the relationship between income and attainment was negligible (Sutherland et al., 2015a; 2015b), which is at odds with other research. Further, given the overlap in the education and occupational class measures, and the wider body of research linking parental

[13] Even if lower attaining schools did boycott, their TA scores may still be upwardly biased if they were not able to revised assessments in light of actual KS2 results.

educational attainment to pupil attainment (Lucchini et al., 2013), we decided to retain highest household qualification and to exclude measures of occupational category.

Housing tenure, religion and region were dropped because these measures were largely uncorrelated with attainment in the previous models. We retained gender and month of birth measures since these were associated with attainment, and since prior research has also establish consistent associations between factors, such as gender and attainment (see references in Sutherland et al., 2015a; 2015b). Although prior evidence links ethnicity to attainment, we excluded ethnicity because it was not associated with KS2 attainment in our sample. This lack of association could be due to the range of other measures included in the analysis, in particular, both school- and pupil-level attainment, as these would capture selection into given schools by location and pupil ability.

At the school level, we retained both the percent of pupils with SEN but no statement and the 2009 KS2 average points score for the school. Both measures remained associated with attainment net of a host of other school-level measures. Finally, given the strong association between and likely predictive validity of later KS4 attainment with KS2 attainment, we retain that measure in the reduced model. Including a potential outcome as a predictor of KS2 attainment is important because we know that attainment between KS2 and GCSE is correlated at around 0.70 (Education Endowment Foundation, 2013).

Table 13 gives the detailed results from the 'trimmed' model. The empirical overlap between excluded and retained measures means that the retained measures become more strongly associated with the outcome, owing to variation they share with omitted variables being attributed to them. We do not discuss these results in detail, but overall they suggest that the measures included are correlated with KS2 attainment and would thus be informative for the MI estimation that follows.

Summary and implications for missing data analysis

Our analysis suggested a range of measures at the pupil level that would be good candidates for inclusion in the imputation model because they are strongly associated with the actual points score for KS2 in the observed sample. The untested (and largely untestable) assumption is that the same would hold true for the unobserved group whose values we are attempting to impute. To reduce the complexity of the imputation model, we reduced the predictors of the value of KS2 attainment. This is a pragmatic step guided by the results of our analysis and by what is known from the literature about the relationships between the measures included and excluded from the imputation. In the next section, we talk through some of the technical issues relating to using multiple imputation models and how we addressed these in the course of this work.

Table 13: 'Trimmed' multilevel predictor model of KS2 attainment results

Outcome: KS2 attainment	Coeff.	Standard error	z	p	95% CI LB	95% CI UB
Female	-0.629	0.072	-8.750	0.000	-0.770	-0.488
QOB 1 Sep-Nov (reference)						
QOB 2 Dec-Feb	-0.368	0.101	-3.650	0.000	-0.566	-0.171
QOB 3 Mar-May	-0.491	0.100	-4.910	0.000	-0.687	-0.295
QOB 4 June-Aug	-0.589	0.098	-6.010	0.000	-0.782	-0.397
Ever 6 FSM? 1=Yes	-2.569	0.110	-23.260	0.000	-2.785	-2.353
SEN No SEN (reference)						
SEN School [A]ction	-3.618	0.143	-25.380	0.000	-3.897	-3.338
SEN School Action [P]lus	-3.461	0.261	-13.240	0.000	-3.973	-2.948
SEN [S]tatemented	0.026	0.087	0.290	0.769	-0.146	0.197
HH ed: Degree (reference)						
HH ed: HE below degree	-0.356	0.112	-3.170	0.002	-0.576	-0.136
HH ed: A/AS levels or equiv.	-0.242	0.116	-2.080	0.037	-0.470	-0.014
HH ed: 5+ A*-C GCSEs or equiv.	-0.446	0.112	-3.980	0.000	-0.666	-0.227
HH ed: Some GCSE passes or equiv.	-0.402	0.121	-3.330	0.001	-0.639	-0.166
HH ed: Entry level qualifications	-1.690	0.799	-2.110	0.034	-3.256	-0.123
HH ed: Other qualifications	-0.325	0.367	-0.880	0.377	-1.045	0.395
HH ed: No qualifications	-0.304	0.173	-1.760	0.079	-0.642	0.035
# people in HH: 2	0.250	0.163	1.530	0.125	-0.070	0.570
# people in HH: 3	0.344	0.101	3.410	0.001	0.146	0.542
# people in HH: 4 (reference)						
# people in HH: 5	-0.009	0.093	-0.100	0.921	-0.192	0.174
# people in HH: 6	-0.040	0.133	-0.300	0.765	-0.301	0.221
# people in HH: 7	-0.113	0.200	-0.560	0.573	-0.505	0.279
# people in HH: 8	-0.302	0.320	-0.940	0.345	-0.928	0.325
# people in HH: 9+	0.236	0.384	0.610	0.540	-0.517	0.988
HH: is EAL? 1=yes	-1.521	0.152	-10.030	0.000	-1.818	-1.224
School: % pupils SEN no statement (mc)	0.031	0.005	6.640	0.000	0.022	0.041
School: APS 2009 KS2 (mc)	0.312	0.025	12.430	0.000	0.262	0.361
KS4 attainment (mcKS4_VAPTSC_PTQ)	0.028	0.001	53.320	0.000	0.027	0.029
Intercept	29.086	0.117	248.160	0.000	28.856	29.316

See Table 7 for abbreviations.

5. Methodological issues relating to MI

Although the output of this research is a dataset that can be used by a wide range of stakeholders, it is important to consider the assumptions and limitations of the approach taken, so that end users are aware of possible reasons to be cautious about this data. This chapter describes some of the methodological issues considered in this work, as well as the approaches that we have taken to address them.

We specifically consider: sampling, attrition and other reasons for missing data, variable selection, number of imputations, the clustered/multilevel nature of the dataset, methodological issues around the inclusion of different variables and issues of bias in the KS2 teacher assessments. For each we highlight the issue and describe the approach that we have taken to address it.

Sampling, attrition and missing KS2 data for other reasons

LSYPE2 is a representative sample drawn from individuals in year 9 between 1 September 2012 and 31 August 2013 who were turning 14 within that time period and normally resident in England at the time of sampling. In addition, minimum expected sample sizes at Wave 7 were set for the following groups:

- Those who are, or have been, eligible for free school meals (FSM) at some point over the preceding three years (n≥ 2000).
- Those who are, or have been, eligible for free school meals and have special educational needs of any type (n≥750).
- Each of eight ethnic groups (White British, Indian, Pakistani, Bangladeshi, Black Caribbean, Black African, Mixed and Other) (n≥150).
- Those who attend school in the independent sector (n≥300).

Sampling was clustered by the school that the pupils were attending in year 9 to maximize the ability to distinguish school-level and pupil-level effects. Response rates at Wave 7 for different groups in the First Longitudinal Study of Young People in England (LSYPE1) were used to predict expected response rates in LSYPE2.

As mentioned in the introduction, there are many reasons for missing data, with the KS2 SATs boycott in 2010 being only one missingness mechanism.

The initial sample consisted of 17,727 pupils from 738 secondary schools (and 148 reserve schools) plus 687 pupils from 30 independent schools and 31 pupils from four pupil referral units. A total of 13,100 households were interviewed, with respondents included from 769 schools.

One of the very valuable dimensions of LSYPE2 is the ability to link the data to the NPD. At Wave 1, both pupil and parental consent for linkage was required. For 892/13,100 (6.8 per cent), this consent was not obtained. For a further 385 pupils, although consent was obtained, linkage was not possible for a variety of reasons, such as attendance at a pupil referral unit or independent school.

Overall, 11,823 pupils have consent for linkage and are identified from the linked NPD school census dataset as having attended at boycott or non-boycott school at KS2. One hundred and sixty seven (167) pupils have a KS2 point score but are flagged as having attended a boycott school. We are not sure why this is so, but one plausible reason could be that pupils changed schools during the year the KS2 tests were taken. For the inverse probability weighting approach, this is the sample that we will use (i.e. the 11,823 pupils in LSYPE2 at Wave 1, who have linked information on whether or not they attended at boycott school in 2009; of these, 8,684 pupils attended a non-boycott school and will have an IPW estimated).

When pupils drop out of LSYPE2 at later waves (i.e. Wave 2 or Wave 3), they are no longer considered as consenting to linkage. The biggest practical implication here is that information on school at KS4 and on GCSE test results are not available for these pupils.

In 2014, LSYPE2 interviewed 11,166 young people in Year 10 (Wave 2) and 10,010 in Year 11 (Wave 3). In the third wave of the study, 9,531 pupils and parents consented to linkage to NPD data, and 8,882 pupils could also be linked to whether or not they attended a boycott school at KS2. This analysis sample of 8,882 (young people in LSYPE2 at Wave 3 with consent and successful linkage to NPD) is the sample used for the MI analysis. A total of 2,356/8,882 attended boycott school in this sample.

Restricting the analysis sample for the MI to only 8,882 pupils has some limitations, as pupils with potentially relevant information are excluded. However there are two strong arguments in favour of this approach. Linked KS4 attainment is only available for this sample (and 8,722/8,882 have data available). KS4 attainment is such a strong predictor of KS2 attainment that the imputation model will be much more precise in this cohort compared with the cohort of young people without KS4 attainment available. In addition, linked NPD attainment at KS4 is one of the key outcomes of interest for the LSPYE2 cohort, and so restricting the analysis sample to only those young people with these data available is a reasonable approach from a policy perspective. Finally, focusing on young people with KS2 test scores missing through the SATs boycott, rather than through other mechanisms, simplifies the MI analysis approach required and supports attempts to address the MAR assumption by including predictors of having attended a boycott school, rather than other mechanisms, which may have different predictors.

Survey weights and survey sampling

The use of weighting in multiple imputation is a complex and evolving area.[14] The initial LSYPE2 cohort was sampled with the following characteristics:

- Those who are, or have been, eligible for free school meals (FSM) at some point over the preceding three years (n≥ 2,000).
- Those who are, or have been eligible, for free school meals and have special educational needs of any type (n≥750).
- Each of eight ethnic groups (White British, Indian, Pakistani, Bangladeshi, Black Caribbean, Black African, mixed and other) (n≥150).
- Those who attend school in the independent sector (n≥300).

The imputation model is merely trying to impute the missing values in the sample at hand. In this case, whether the sample is representative or not of the population is less important. Put differently, the imputation model is a data-generating process, not a substantive model from which we are trying to make inferences to a wider population. We therefore do not account for sampling weights in the imputation model developed. That said, the survey weight for LSYPE2 was created using the inverse of the sampling probability, which in turn was a function of the variables noted above. The models we created contained many of these variables and others, which in turn should reduce bias. Users would still be free to apply survey weights if possible, but it is not clear what effect that would have on results. Note that our IPW approach does combine MI and IPW, as recommended by (Seaman et al., 2012b).

Selection of variables to include in the MI model

Predictors of missingness and predictors of attainment

As discussed in the previous two sections, the MI model should include both predictors of missing KS2 test scores (to support, as far as possible, the MAR assumption for the MI) and predictors of KS2 attainment (to allow as good a prediction as possible for the missing values).

Compatibility

In addition to this, however, one of the key requirements for an analysis using an imputed variable is that the imputation and analysis models are 'compatible'. In general, the

[14] There are various approaches, examples of which are shown in the following documents. As of 29 August 2016:
http://www.lse.ac.uk/statistics/events/SpecialEventsandConferences/CarpenterJR.pdf
http://missingdata.lshtm.ac.uk/talks/RSS_2012_04_18_seaman.pdf

model that is used for the imputation of missing data should be at least as complex as the analysis model of interest. In addition, in order for an analysis using multiple imputation not to be biased, the ‘outcome’ variable for the analysis should be included in the imputation model. Further, if the analysis model includes interaction terms, relationships between covariates, or polynomials, those relationships should, ideally, be reflected in the imputation model as well (Seaman et al., 2012a, White et al., 2011).

In practical terms, to develop an imputed variable for KS2 test score that is useful across as wide a range of stakeholders and uses as possible, this means that we will need to include as wide a range as possible of outcome variables for which analyses with the KS2 test scores as a predictor may be carried out, plus any possible relationships between the variables. There are three possible strategies for multiple imputation: to impute the whole dataset to use for all analyses, to do a new imputation for each project or to do a separate MI for each analysis. The approach that we are taking here is broadly consistent with the first strategy, which is to impute the whole dataset to use for as many analyses as possible. This is also, therefore, one of the key challenges in this work – to provide an imputed dataset for future use without knowing in advance the analyses for which it will be used. In order for the imputation dataset to be compatible with as many future analyses as possible, it is important to include a basket of the most important future outcomes and the most important variables for analyses that may use this data, and to include a sufficiently complex imputation model.

In order to ensure that the imputed dataset is compatible with the analysis models for as many future analyses as possible, we consulted with DfE, key users and advisors on the most important outcomes they would be considering in analyses using LSYPE2 survey data.

We received seven consultation responses, with the following dimensions most consistently identified:

- Attainment
- Ambitions and future plans
- Bullying
- Mental health, well-being and non-cognitive skills
- Participation in risky behaviours

We therefore included the following five variables in the imputation model to cover these outcome dimensions: KS4 attainment (linked from NPD), having been bullied in the last 12 months (Wave 2), participation in risky behaviours (a derived variable from Wave 1), General Health Questionnaire [GHQ] (Wave 2), and future intentions to attend higher education (Wave 2). Consultation respondents also identified a large number of pupil, family and household characteristics of interest (e.g. SEN, EAL, ethnicity, region, income), and these are all additionally included in the imputation model, from the predictors of KS2 attainment and missingness analyses presented above. Users of the data should exercise caution if adjusting for KS2 using the imputed data in analyses with outcomes not added here; relationships between covariates and outcomes will likely be biased using the MI dataset if the outcome was not included in the imputation model.

Number of imputations

When selecting the number of imputations required, it is important to avoid a situation where someone else analysing the same dataset could get a substantially different result (Royston, 2004). A conservative estimate is that the number of imputations should be greater than the percentage of incomplete cases (White et al., 2011). In our MI dataset we have 8,882 young people with survey responses at Wave 3 and ongoing consent to NPD linkage; of these, 2,276 attended a boycott school at KS2 (25 per cent). We included 30 imputed KS2 variables in the final dataset for use by analysts, and we would recommend that all imputations are used in each analysis.

The clustered/multilevel nature of the dataset

An important feature of the LSYPE2 is the fact that pupils are clustered within schools. This has an impact on the analysis strategies used for these data and for the approach to addressing missing data. Specifically, relevant for this work are the following:

- Pupils are sampled from the schools that they are attending in year 9 (the primary sampling unit) – 729 schools with a mean of 17 pupils per school
- There is an earlier level of clustering, at the school attended for Key Stage 2. There are 5,404 unique identifiers matching to 11,823 members of the LSYPE2 cohort, with a mean of 2.2 pupils per school at KS2 (range 1–16).
- There is a further level of clustering, namely, the school attended for Key Stage 4 where the pupil has changed school since Year 9, although we expect this number to be low.

There are no simple approaches to MI with multilevel data, although the Stat-JR software developed by the University of Bristol presents one approach (University of Bristol, 2016). Researchers often take a pragmatic approach to this issue and ignore clustering in the imputation but allow for it in the analysis.

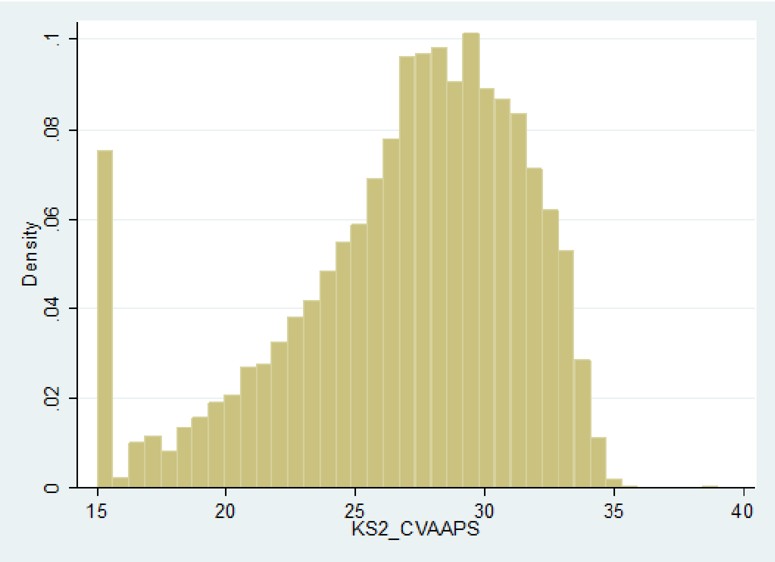
We expect that the most important level of clustering for most users is at the PSU level. In the predictors of missingness and predictors of KS2 attainment analyses described in the previous chapters, we explored the extent to which between-school clustering could be explained by other measured covariates. In addition, we requested a bespoke variable from NPD that includes the proportion of pupils in the PSU who attended a boycott school at KS2, which obviously explained almost all of between-PSU variation in missingness. By adjusting for pupil-, family-, household- and school-level characteristics, the amount of residual clustering of KS2 scores is low.

Methodological issues around the inclusion of different variables

The distribution of KS2 attainment

Figure 2 shows the APS for KS2 performance. The large number of pupils with a value of 15 reflect that 15 is the point score allocated to pupils working below the level of the test and also those who took the test but were not awarded a level. So it is possible that some pupils had an APS of 15. This is a challenge for the analysis, because, clearly, distributional assumptions associated with the normal distribution will not be met.

Figure 2. Distribution of the KS2 score being imputed



Imputation by predictive mean matching (PMM) is an approach to addressing non-normality in MI. It is an option in the Stata ‘mi’ commands that we use. Briefly, instead of imputing a value for the KS2 performance based on the normal distribution, it will select a value from the nearest matching existing member of the cohort (i.e. the distribution of KS2 scores will be maintained in the imputed datasets). This approach to non-normality does have limitations; it is an ad-hoc rather than a theory-based approach, and it has not been widely used because software has only recently incorporated it (Morris et al., 2014a).

Derived variables

Recent methodological work on approaches to multiple imputation has highlighted that, for derived variables, it is better (less biased) to impute the derived variable directly, rather than to individually impute the constituent variables with missing data (even if only the numerator or denominator has missing data) (Morris et al., 2014b). The practical implication of this point is that final 'derived' variables from the LSYPE2 datasets (TNS BMRB, 2013a) are included in the imputation models directly, rather than the variables that make up the derived variables.

Polynomials

We entered KS4 scores flexibly (e.g. up to 3rd- or 4th-order polynomials). In MI impute chained, the best (although not perfect) approach to including non-linear relationships is the 'Just another variable' approach, i.e. imputing each power-transformed variable as separate variables, ignoring their logical relationship (Seaman et al., 2012a). We did not consider non-linear forms of the KS2 variable. Note in the technical details below that the higher order polynomials caused some problems with MI convergence and that, therefore, in the final model only linear and squared terms are included.

KS1 attainment

KS1 attainment is only available as a continuous variable in NPD for those pupils with a KS2 result (it is used in the calculation of the value-added score), and levels are more complicated to incorporate/lead to a more complex model. Because KS4 attainment is such as strong predictor of KS2 attainment, KS1 is not incorporated in the MI.

KS4 attainment

There are some philosophical questions about using KS4 attainment in predicting KS2 attainment, which occurred before KS4 in time. However, from a statistical perspective it is more important to include good predictors of the missing variable, and so it is included in the MI. Similarly, KS4 attainment will likely be a commonly used outcome variable, so its inclusion in the imputation makes the imputation model consistent with analysis models.

Teacher assessments

Three separate analyses of boycott and non-boycott schools were carried out by DfE in 2010. Brief summaries of these analyses are provided in the 2010 Statistical First Releases of the KS2 assessments (Department for Education, 2010a, Department for Education, 2010b) and full details of the analyses were provided to the authors of this report by DfE. DfE analyses explored whether teacher assessment (TA) data would be a reasonable predictor/substitute for test data. This was done by looking at the improvement in share of pupils at level 4+ in both English and maths by school between 2009 and 2010 and then between 2010 and 2011. The analysis shows that there was a substantially higher (1.1 percentage points) increase in the share between 2009 and 2010 for the boycott schools than for the non-boycott schools in English, followed by 0.7 percentage point *smaller* increase between 2010 and 2011. The data are shown below. Boycott schools seem to have improved much more than non-boycott schools in the boycott year, and much less in the year after, suggesting that some boycott schools may have inflated the TA for some of their pupils. However, the authors conclude that around 98 per cent of TA in boycott schools were accurate[15] and note that the TA results are generally higher than the test results (for example, 85 per cent of pupils nationally got level 4 or above in the KS2 maths test in 2013, but 87 per cent were at level 4+ using TA).

Table 14. Teacher assessments by boycott status, 2009, 2010, 2011

	Level 4+ (%)			
	English		Maths	
Year	Boycott	Non-boycott	Boycott	Non-boycott
2009	79.4	80.6	80.7	80.9
2010	80.8	80.9	81.2	81.4
2011	81.0	81.8	81.2	82.2

In light of concerns about bias identified by the findings presented above, and the fact that KS4 attainment alone is a very good predictor of KS2 attainment, we do not include KS2 TA in the imputation model.

Pupils who attended a boycott school but have a non-missing KS2 test score

To identify pupils missing because of the boycott, we used a specific flag variable supplied by the DfE (KS2FAAT10_CONTFLAG). We did this because, as we note above, there were several reasons for pupils missing KS2 data, and we wanted to be sure we were focusing on imputing only those missing because of the boycott rather than, for example, those missing because there was no consent for linkage.

[15] For instance, one would expect English level 4+ to increase by 0.3 percentage points between 2009 and 2010 in boycott schools. However, it increased by 1.4 percentage points, suggesting that 1.1% of pupils in boycott schools may have been given a level 4 rather than a level 3. A similar percentage of pupils may have been given a level 5 rather than a level 4, so that suggests that around 2% of teacher assessment in boycott schools has been inflated.

For the IPW, we create a weight for everyone with a KS2 score – and this will include people who attended boycott schools, because they would still make it into the complete-case analyses.

For the MI, we only impute KS2 attainment for people with a missing score.

Although it is not ideal to provide missing data solutions for different samples, the MI approach requires the missingness mechanism to be modelled, and so it is provided only for pupils still in LSYPE2 at Wave 2 (i.e. the only missingness mechanism considered is from the boycott).

It would have been possible to provide an IPW for this same sample, and this would have the advantage that differences in the two methods are minimised because they are applied to the same data. The advantage to generating the IPW for all pupils with non-missing KS2 and linked data at W1 is that the weight is also available for all pupils for whom KS2 results are available.

Perfect prediction

For some categorical variables with only a few pupils in each category, the MI models run into problems with perfect prediction (this occurs when there is a category of any predictor variable in the imputation model in which the outcome is always 0 or always 1). We have taken a pragmatic approach to this issue. For some variables, we group rarer categories together. We have also used the augment option in Stata (White et al., 2010) to address this issue when it occurs during the imputation.

Multiple imputation is a simulation-based approach

If someone were to attempt to recreate the models we have generated, even if using exactly the same dataset, the values produced would vary slightly unless the same starting value (random number generation 'seed') were used. Without this same 'seed' number, there would be negligible differences in results (e.g. results could differ two or three decimal places in).

6. Methodological issues relating to IPW

Two issues were raised during the development of the IPW. The first was around variation in missingness between PSUs. In initial analyses we were able to identify that there was high variation in missingness between PSUs, but because only a few pupils were sampled per school we did not have precise estimates of this for each individual school. In consultation with our quality assurance reviewers and DfE, we therefore requested a non-standard variable from NPD, namely, the proportion of pupils who attended a boycott school in each PSU. Findings from preliminary analyses using this variable are presented in Table 9. We include this variable in the IPW model to allow a more precise weight to be estimated, and to account for between-school clustering in missingness.

The second issue was around how to incorporate covariates with missing values. For this we identified that the more robust solution would be to generate the linear predictors using an MI too and to use ‘mi’ commands at the prediction stage, before generating the weight (Seaman and White, 2014).

We have different samples for the IPW and MI calculations that follow. The disadvantage of this is that therefore have two different datasets that are not strictly comparable. But the advantage is that the IPW model then includes more pupils who did not attend a boycott school (n = 8,684 for IPW vs n = 6,606 for the MI models).

7. Development of MI model

We performed the multiple imputation using the multiple imputation with chained equations command in Stata, 'mi impute chained', because variables other than KS2 attainment also had some missing data. This command fills in missing values in multiple variables iteratively by using chained equations and allows different variable types (normal and non-normal data, as well as ordered categorical, categorical and binary variables). The number of observations with missing data for each variable is included in Table 15 below.

A good imputation model will include variables which predict whether or not the data are missing (to help support the MAR assumption) and predictors of the missing variables (in this case KS2 performance), and it should be at least as complex as the analysis model for which it is used. As discussed in Chapter 4, we include five 'outcome' variables to incorporate some relationship between the imputed data and these key dimensions. The variables included in the multiple imputation model are described in Table 15 below.

Variables included in the imputation

In Table 15 below we set out the variables included in the imputation and the source dataset as a summary of the foregoing analyses and discussion. The table sets out the variables included in the imputation model, source (in brackets), and the primary reasons why variables were selected for inclusion (as a predictor of *missingness*, a predictor of *attainment*, or an *outcome* variable included for compatibility).

Assessing convergence

We explored whether the number of iterations for the 'burn-in period' for each chain (one chain per imputation) was appropriate;[16] the default in Stata is 10. The required length of the burn-in period for a chain to reach approximate stationarity or, equivalently, to converge to a stationary distribution, will depend on the starting values used and the missing data patterns observed in the data. It is important to examine the chain for convergence to determine an adequate length of the burn-in period prior to obtaining imputations. We assessed this by plotting the mean and the standard deviation (SD) of

[16] Multiple imputation by chained equations works by predicting missing values for a variable based on a model using the values of all other variables. However, there will usually be missing values in more than one variable, so, for this to work, the prediction needs values for all variables. Initially missing values are replaced by arbitrary values, such as zero. By repeating the predictions for all variables in turn many times, the predicted values gradually improve to something more likely given the other data. This can be checked by looking at, for example, how the mean and standard deviations of the variables stabilise over time. This allows us to choose a number of times for all the variables to be predicted before we settle on an imputed data set. The total number of rounds of predictions is called the 'burn in' period.

each variable against the iteration number. For all variables except KS4 performance (which included squared and cubed terms), these values converged within 10 iterations. Because it is likely that the KS4 polynomial terms were causing some problems with the convergence, we dropped the cubed term from the model and explored a burn-in period of 50 iterations. The results from this analysis are presented below (Figure 3 to Figure 6). The estimates converge within 10–20 iterations, and so we use a burn-in period of 20 iterations for each chain. We explore inclusion of KS4 polynomial terms (to the 3rd power) with orthogonal transformations in a sensitivity analysis presented in the appendix.

Table 15: Variables included in the multiple imputation

Type of variable	Reason for inclusion	Missing obs. (N)
Binary		
Free school meals at any point in the last 6 years (NPD)	Attain	2
Household non-English as a first language (Survey, Wave 1)	Missing	14
Female (Survey, Wave 1)	Attain	59
Having been bullied in the last 12 months (Survey, Wave 2)	Outcome	642
Categorical		
Ethnicity (Survey, Wave 1)	Attain	77
Special Educational Needs (NPD)	Attain	2
Ordered categorical		
Highest household qualification-level (Survey Wave 1)	Attain	59
Higher education aspirations (Survey, Wave 1)	Attain	330
Continuous		
GHQ (Survey, Wave 2)	Outcome	1535
Household size (Survey, Wave 2)	Attain	14
KS2 school mean attainment in 2009 (NPD)	Attain	561
Continuous (non-normal)		
Participation in risky behaviours (derived variable, Survey, Wave 1)	Outcome	453
KS4 average point score (NPD)	Attain, outcome	160
KS4 average point score - squared (derived)	to improve imputation	160
Included in the imputation, but no missing data		
The mean point score at KS2 of pupils at the school where they took KS4 exams (NPD)[17]	Attain	0
% EAL pupils in the KS2 school in 2010 (NPD)	Missing	0
Mean IDACI of pupils at KS2 school in 2010 (NPD)	Missing	0
% pupils with SEN in the KS2 school in 2010 (NPD)	Missing	0
% pupils at the school the pupils attended at time of sampling with missing KS2 data because of the boycott (NPD)[18]	Missing	0

[17] This is a difficult variable to try to understand – it is the mean KS2 score at the school the pupils attended for KS4. It is important to include because conceptually it relates to the clustering of KS2 attainment within the PSU/KS4 school. There are some limitations, as this is the mean for the year that the KS2 boycott occurred. We also know from the predictors of KS2 attainment analyses that, overall, pupil rather than school characteristics explained the majority of the between-PSU/KS4-school variation in performance. During the year of the boycott, the teacher assessment scores will have been included in the calculation for this variable for pupils with missing test scores.
[18] The impact of this variable on the imputation model is explored in a sensitivity analysis presented in the appendix.

Figure 3. Mean KS4 attainment across burn-in iterations

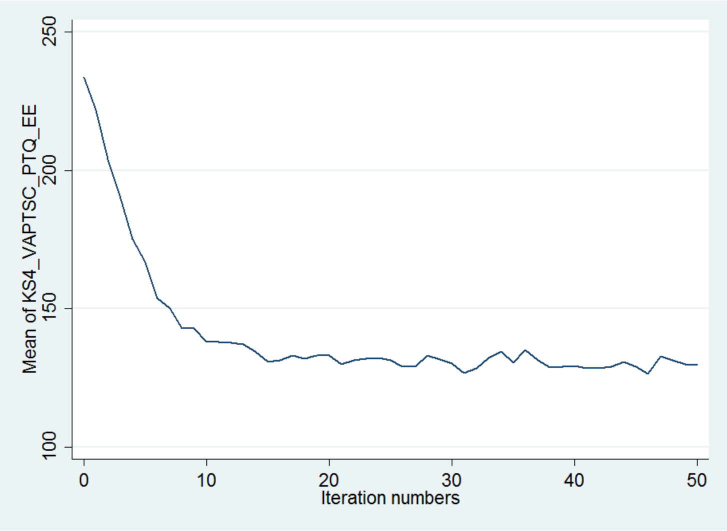


Figure 4. SD KS4 attainment across burn-in iterations

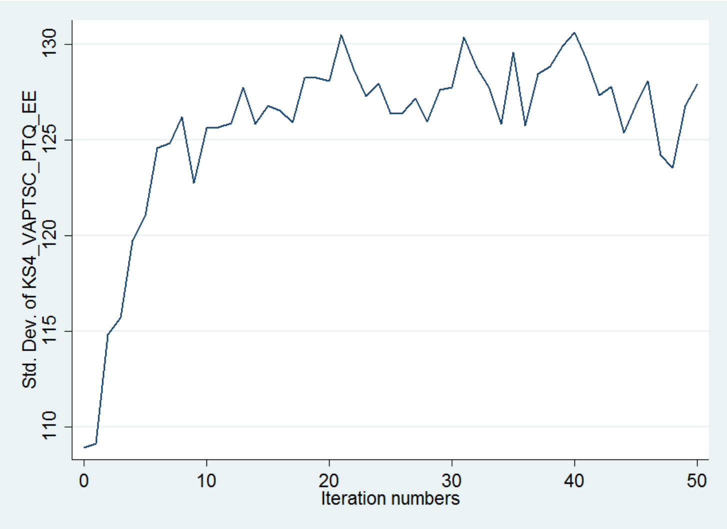


Figure 5. Mean KS4 attainment-squared across burn-in iterations

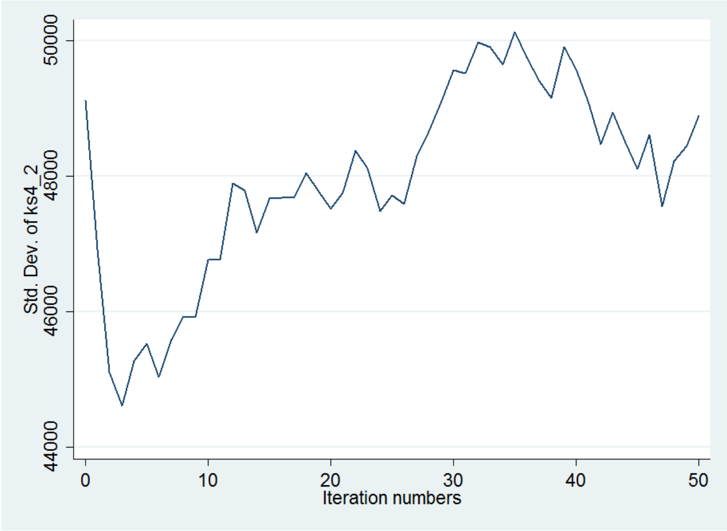
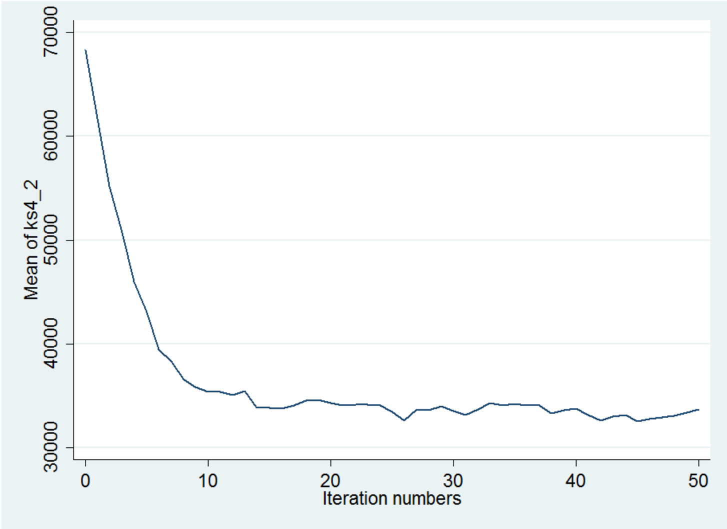
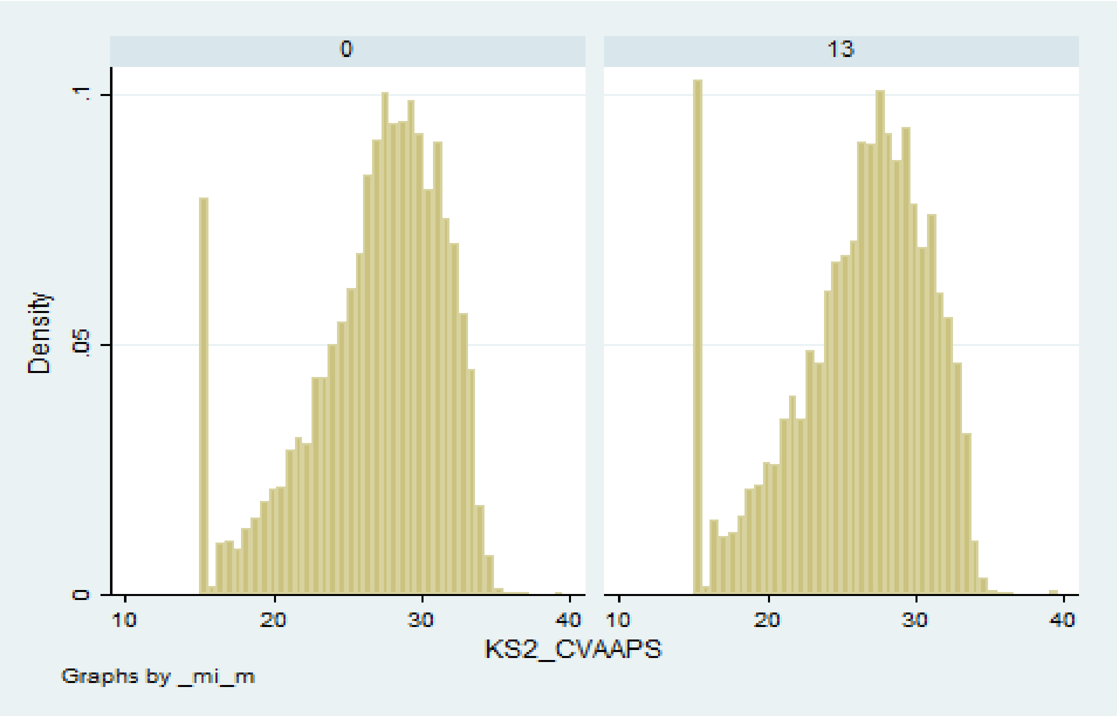


Figure 6. SD KS4 attainment-squared across burn-in iterations



We also explored the distribution of KS2 scores across each imputation. An example (from one imputation) is presented below in Figure 7, with the distributions from all imputations given in the appendix.

Figure 7. KS2 attainment distribution in complete cases (#0) vs imputed data (#13)



KS2 point scores of 15 are slightly overrepresented in the imputed data compared with the complete cases. This is also reflected in the slightly lower mean KS2 point score across all imputations compared with the complete cases (Table 16). It may be that the boycott sample contained more individuals with characteristics of those who were likely to receive a score of 15, or it may also be that the imputation has inflated the number of such cases. But the higher number of low scores is consistent with earlier analyses of school-level characteristics.

Table 16. Mean KS2 attainment by imputation

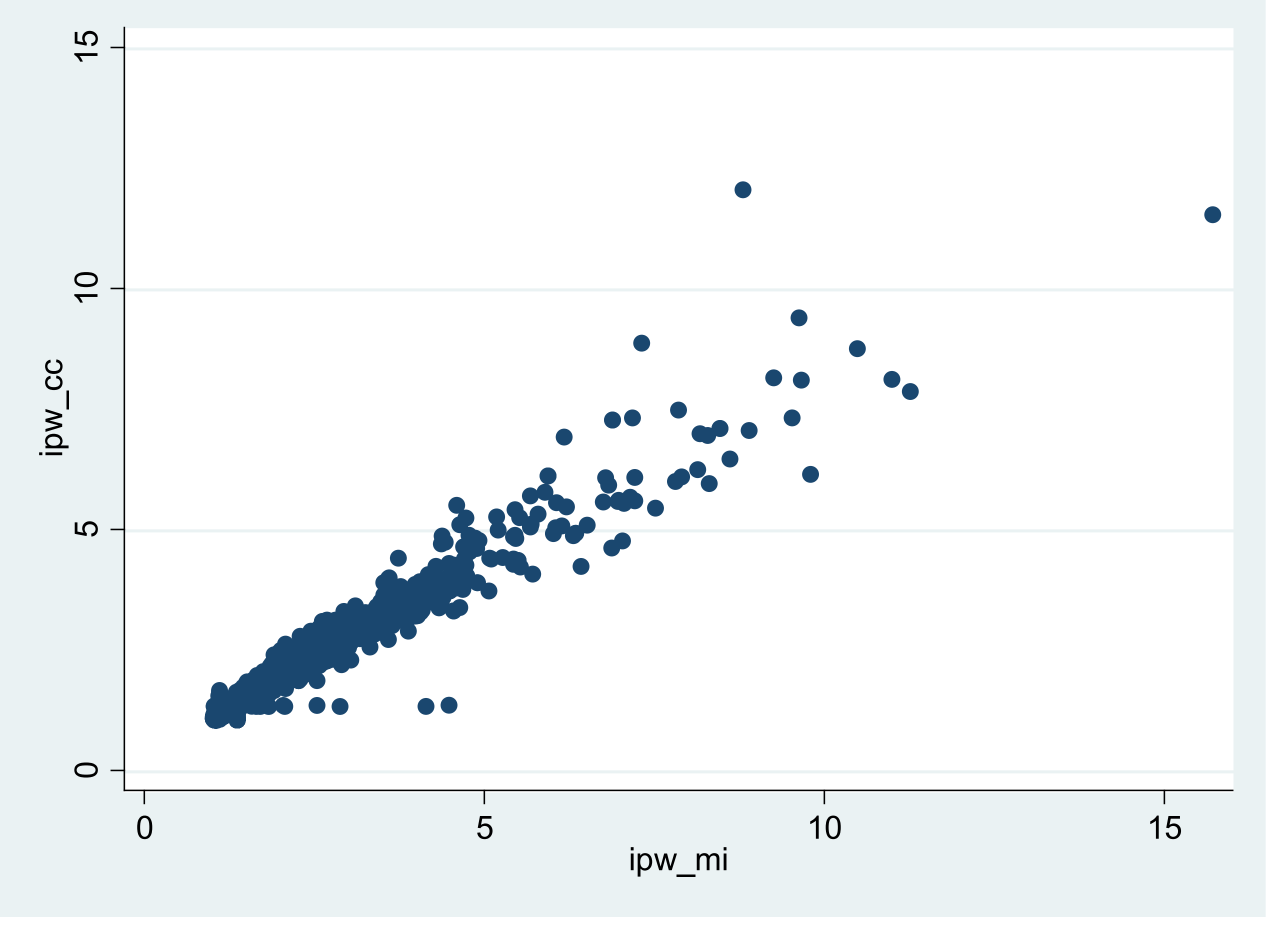
Imputation	Sample size	Mean KS2 attainment
Complete case	6,606	27.2
1	2,276	27.0
2	2,276	26.9
3	2,276	26.9
4	2,276	27.0
5	2,276	27.0
6	2,276	27.0
7	2,276	26.9
8	2,276	27.0
9	2,276	26.9
10	2,276	27.0
11	2,276	26.9
12	2,276	27.0
13	2,276	27.0
14	2,276	27.0
15	2,276	26.9
16	2,276	27.0
17	2,276	27.0
18	2,276	27.0
19	2,276	26.9
20	2,276	27.0
21	2,276	27.0
22	2,276	27.1
23	2,276	26.9
24	2,276	27.0
25	2,276	27.0
26	2,276	26.9
27	2,276	26.9
28	2,276	27.0
29	2,276	26.8
30	2,276	27.0

8. Development of the IPW

Using logistic regression with missingness as an outcome, and only including those variables identified as 'missingness' in Table 15 as predictors, we estimated the inverse probability weight using the predict option in Stata after the regression model. Specifically the IPW is the inverse probability weight, calculated as the multiplicative inverse of the probability of the observation being present – that is, non-missing – in the dataset that we used for the missingness analysis (11,823 observations).

We used MI to address the issue of missing covariates in the IPW analysis, and we explored whether the IPW created using the imputed covariates was consistent with the IPW from the complete-case analysis (Figure 8). We found that the two were very similar. We explore these similarities further in sensitivity analyses presented in the appendix to this report.

Figure 8. Comparison of IPW approaches



For four observations a missing value was estimated, and for these variables the weight was replaced with a value of 1.

9. Summary

Boycotts of national tests leave gaps in pupils' attainment records and, in the case of LSYPE2, threaten to undermine a large-scale longitudinal study with substantial policy relevance. This project sought to find a way to calculate values for pupils who attended schools that boycotted KS2 tests in 2010 and/or mitigate the effect of the boycott on this study.

The results of the analyses undertaken for this project suggest that complete-cases analyses using only pupil-level data that include a random effect for primary sampling unit (i.e. secondary school) should be unbiased. Comparing complete-case analysis with MI suggests that MI would be more efficient – i.e. standard errors would be smaller – meaning this approach should be used if statistical inference is the aim of a given analysis.

We believe that the results from our analysis will allow analysts to make more use of the LSYPE2 dataset, and that the absence of a key source of prior attainment data is something that should be incorporated into even basic analyses of LSYPE2 data. It is for analysts to decide whether the missing data arising from the boycott will cause difficulties regarding inferences and conclusions and to take appropriate steps to deal with these. Using multiple imputation and inverse probability weighting, we have been able to produce plausible values for KS2 scores (via MI) and analytical weights (via IPW) for pupils missing data due to the boycott, thus giving analysts two options when deciding how to deal with missingness.

References

ALLISON, P. D. 2000. Multiple imputation for missing data - A cautionary tale. *Sociological Methods & Research,* 28**,** 301-309.

BURGER, K. 2016. Intergenerational transmission of education in Europe: Do more comprehensive education systems reduce social gradients in student achievement? *Research in Social Stratification and Mobility,* 44**,** 54-67.

CARPENTER, J. R., KENWARD, M. G. & WHITE, I. R. 2007. Sensitivity analysis after multiple imputation under missing at random: a weighting approach. *Statistical Methods in Medical Research,* 16**,** 259-275.

CURTIS, P. 2009. Education unions plan 2010 Sats boycott. The Guardian. https://www.theguardian.com/education/2009/mar/26/education-unions-exams.

DEARY, I. J., STRAND, S., SMITH, P. & FERNANDESC, C. 2007. Intelligence and educational achievement. *Intelligence,* 35**,** 13-21.

DEPARTMENT FOR EDUCATION 2010a. National curriculum assessments at key stage 2 & 3 in England - academic year 2009 to 2010 (provisional) https://www.gov.uk/government/statistics/national-curriculum-assessments-at-key-stage-2-and-3-in-england-academic-year-2009-to-2010-provisional.

DEPARTMENT FOR EDUCATION 2010b. National curriculum assessments at key stage 2: England - academic year 2009 to 2010 (revised) https://www.gov.uk/government/statistics/national-curriculum-assessments-at-key-stage-2-england-academic-year-2009-to-2010-revised.

DEPARTMENT FOR EDUCATION 2010c. Primary School (Key Stage 2) - The Tables http://www.education.gov.uk/schools/performance/archive/primary_10.shtml.

DIGGLE, P. & KENWARD, M. G. 1994. Informative Drop-out in Longitudinal Data-Analysis. *Journal of the Royal Statistical Society Series C-Applied Statistics,* 43**,** 49-93.

EDUCATION ENDOWMENT FOUNDATION 2013. Pre-testing in EEF evaluations. London: Education Endowment Foundation.

HOBBS, G. & VIGNOLES, A. 2009. Is children's free school meal 'eligibility' a good proxy for family income? *British Educational Research Journal,* 36**,** 1469-3518.

KENWARD, M. G. & CARPENTER, J. 2007. Multiple imputation: current perspectives. *Stat Methods Med Res,* 16**,** 199-218.

KHATTAB, N. 2012. 'Winners' and 'losers': the impact of education, ethnicity and gender on Muslims in the British labour market. *Work, Employment & Society,* 26**,** 556-573.

LITTLE, R. J. A. & RUBIN, D. B. 2002. Statistical analysis with missing data. *Wiley Series in Probability and Statistics.* Second edition. ed. Hoboken: Wiley.

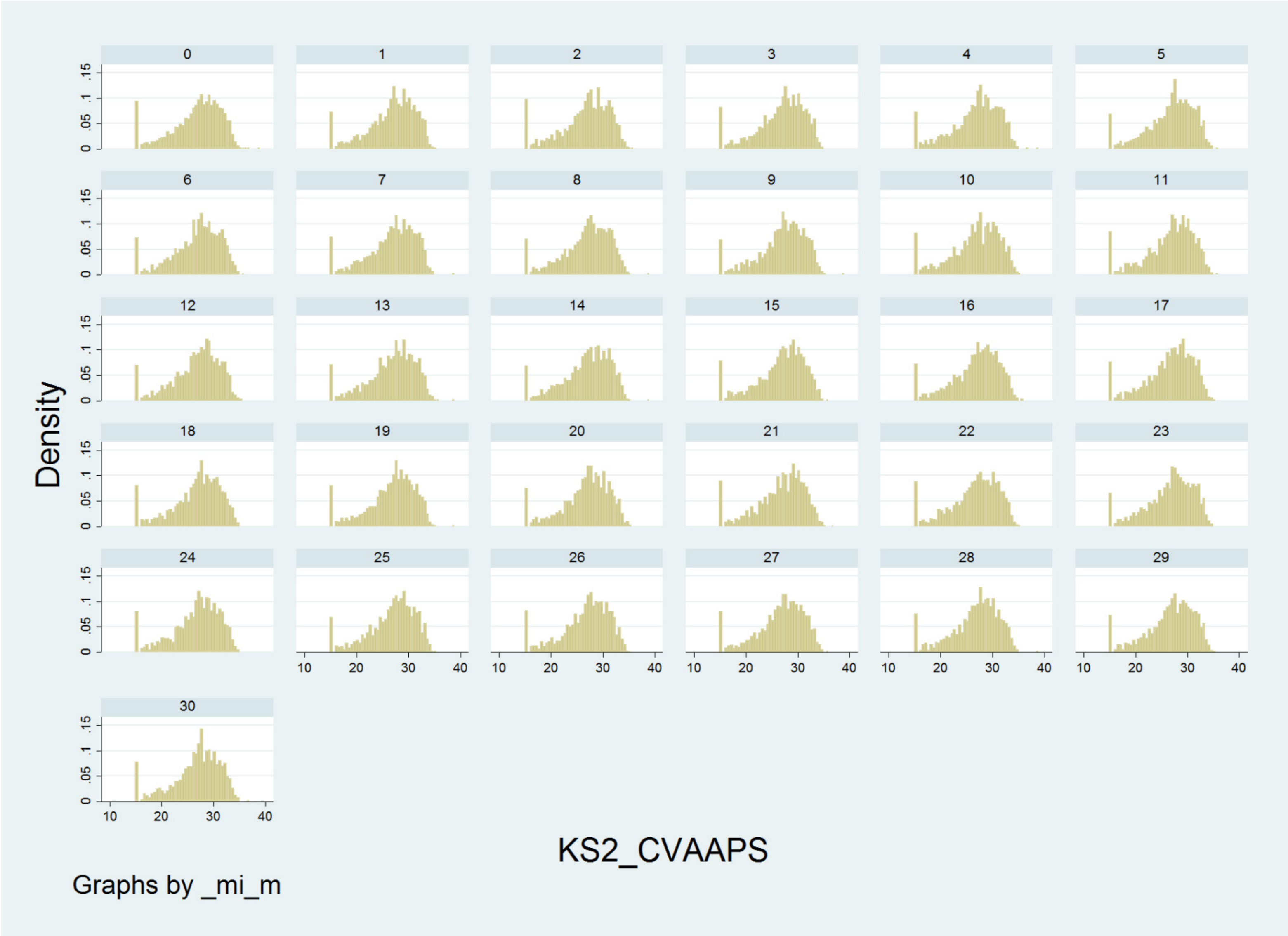
LUCCHINI, M., DELLA BELLA, S. & PISATI, M. 2013. The weight of the genetic and environmental dimensions in the inter-generational transmission of educational success *European Sociological Review,* 29**,** 289–301.

MARSHALL, A., ALTMAN, D. G., HOLDER, R. L. & ROYSTON, P. 2009. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC Medical Research Methodology,* 9**,** 57.

MORRIS, T. P., WHITE, I. R. & ROYSTON, P. 2014a. Tuning multiple imputation by predictive mean matching and local residual draws. *BMC Medical Research Methodology,* 14.

MORRIS, T. P., WHITE, I. R., ROYSTON, P., SEAMAN, S. R. & WOOD, A. M. 2014b. Multiple imputation for an incomplete covariate that is a ratio. *Stat Med,* 33**,** 88-104.
OECD (Organisation for Economic Co-operation and Development). 2014. *PISA 2012 Technical Report*. Paris: OECD. http://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf.
PIESSE, A. & KALTON, G. 2009. A Strategy for Handling Missing Data in the Longitudinal Study of Young People in England (LSYPE). London: Institute of Education.
ROYSTON, P. 2004. Multiple Imputation of missing values. *Stata Journal,* 4**,** 227-241.
RUBIN, D. B. 1987. Multiple imputation for nonresponse in surveys. *Wiley series in probability and mathematical statistics. Applied probability and statistics,.* New York: John Wiley & Sons.
SCHARFSTEIN, D. O., ROTNITZKY, A. & ROBINS, J. M. 1999. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association,* 94**,** 1096-1120.
SEAMAN, S. & WHITE, I. 2014. Inverse probability weighting with missing predictors of treatment assignment or missingness. *Communications in Statistics-Theory and Methods,* 43**,** 3499-3515.
SEAMAN, S. R., BARTLETT, J. W. & WHITE, I. R. 2012a. Multiple imputation of missing covariates with non-linear effects and interactions: an evaluation of statistical methods. *BMC Medical Research Methodology,* 12**,** 46.
SEAMAN, S. R., WHITE, I. R., COPAS, A. J. & LI, L. 2012b. Combining Multiple Imputation and Inverse-Probability Weighting. *Biometrics,* 68**,** 129-137.
SHEPHERD, J. & WILLIAMS, R. 2010. Primary heads scupper league tables with out-of-date Sats tests. The Guardian. https://www.theguardian.com/education/2010/may/10/sats-boycott-by-headteachers.
STERNE, J. A. C., WHITE, I. R., CARLIN, J. B., SPRATT, M., ROYSTON, P., KENWARD, M. G., WOOD, A. M. & CARPENTER, J. R. 2009. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *British Medical Journal,* 339.
SUTHERLAND, A., ILIE, S. & VIGNOLES, A. 2015a. Factors associated with achievement at Key Stage 4 https://www.gov.uk/government/publications/factors-associated-with-achievement-key-stages-2-and-4.
SUTHERLAND, A., ILIE, S. & VIGNOLES, A. 2015b. Factors associated with achievement: key stages 2 https://www.gov.uk/government/publications/factors-associated-with-achievement-key-stages-2-and-4.
TNS BMRB 2013a. Derived variable documentation for the 2nd cohort of the Longitudinal Study of Young People in England (LSYPE2): Wave 1.
TNS BMRB 2013b. Technical report for the 2nd cohort of the Longitudinal Study of Young People in England (LSYPE2): Wave 1.
UNIVERSITY OF BRISTOL 2016. Stat-JR: a software environment for promoting interactive complex statistical modelling. Bristol: University of Bristol.
WHITE, I. R., DANIEL, R. M. & ROYSTON, P. 2010. Avoiding bias due to perfect prediction in multiple imputation of incomplete categorical data. *Computational Statistics & Data Analysis* 54**,** 2267-2275.
WHITE, I. R., ROYSTON, P. & WOOD, A. M. 2011. Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine,* 30**,** 377-399.

Appendix: Distribution of imputed variables



Appendix: Sensitivity analyses

We compared three sample analyses (the estimated mean KS2 score, a stratified analysis, and a regression analysis predicting KS4 performance) comparing complete case and imputed data from different MI models, and weights from different IPW approaches. The sensitivity analyses for the MI models, first, excluded the variable measuring the percentage of pupils at the school the pupils attended at time of sampling with missing KS2 data because of the boycott (NPD) and, second, considered including KS4 performance as orthogonalised polynomials. The IPW sensitivity analysis includes the complete case and MI approaches to missing data in creating the weight, as well as a third IPW variable, which additionally included school type, region and area in the model. The results of these sensitivity analyses are presented below.

The estimates from the analyses using different MI model specifications and different IPW models were very similar to each other, with slightly larger differences between the MI and IPW estimates (but still very small, around 0.06 difference in mean KS2 point scores between the approaches). The orthogonal transformations for the KS4 polynomials had fewer problems with convergence than the untransformed variables. For the model excluding the variable measuring the percentage of pupils at the school the pupils attended at time of sampling with missing KS2 data because of the boycott (NPD), the standard deviation of the imputed KS4^2 variables did not stabilise, suggesting that the MI approach including this variable is preferred.

Sensitivity analyses – example analysis outputs from different MI and IPW model specifications

		Analyses using imputed data from different MI approaches			Analyses using imputed data from different IPW approaches		
	Complete case	**Imputed variable**	**Without the percentage missing at the KS4 school**	**Including orthogonalised KS4 polynomials**	**With MI for missing data in the IPW**	**Including complete cases only**	**Including region, school type and urban/rural mix**
Mean							
Sample size	8,795	11,071	11,071	11,071	8,628	8,628	8,628
KS2 point score (KS2_CVAAPS)	26.88 (26.78 to 26.97)	26.88 (26.79 to 26.98)	26.90 (26.81 to 26.99)	26.88 (26.79 to 26.97)	26.96 (26.85 to 27.07)	26.96 (26.85 to 27.06)	26.95 (26.84 to 27.05)
Stratified analysis (by EVERFSM6_SPR10)							
Mean KS2 point score (KS2_CVAAPS) - sample size	8,793	11,069	11,069	11,069	8,627	8,627	8,627
Children not in receipt of FSM	27.96 (27.85 to 28.07)	27.98 (27.87 to 28.08)	27.99 (27.89 to 28.09)	27.97 (27.87 to 28.08)	28.04 (27.91 to 28.17)	28.03 (27.90 to 28.15)	28.03 (27.91 to 28.15)
Children in receipt of FSM	25.07 (24.91 to 25.24)	25.11 (24.95 to 25.26)	25.13 (24.98 to 25.28)	25.10 (24.94 to 25.26)	25.29 (25.10 to 25.48)	25.27 (25.09 to 25.45)	25.27 (25.09 to 25.45)
Regression model (predicting KS4 performance, KS4_VAPTSC_PTQ_EE)							
Coefficients - sample size	6,434	8,652	8,652	8,652	6,328	6,328	6,328
KS2 point score (KS2_CVAAPS)	14.39 (14.06 to 14.71)	14.33 (14.03 to 14.63)	14.34 (14.02 to 14.65)	14.30 (13.98 to 14.63)	14.54 (14.10 to 14.98)	14.53 (14.10 to 14.96)	14.51 (14.07 to 14.94)
Female (reference male)	17.33 (14.55 to 20.10)	17.38 (14.84 to 19.92)	17.30 (14.74 to 19.86)	17.15 (14.59 to 19.71)	16.71 (13.48 to 19.94)	16.72 (13.60 to 19.85)	16.96 (13.78 to 20.13)
EVERFSM_6_SPR10 (reference, not in receipt of FSM)	-29.03 (-32.16 to -25.90)	-29.54 (-32.39 to -26.69)	-29.78 (-32.68 to -26.88)	-29.57 (-32.59 to -26.55)	-29.52 (-33.65 to -25.38)	-29.21 (-33.16 to -25.26)	-29.64 (-33.68 to -25.61)
Ethnicity (reference White)							
Dual/multiple	12.12 (4.79 to 19.46)	10.03 (3.20 to 16.86)	10.04 (3.16 to 16.92)	10.34 (3.67 to 17.01)	8.63 (-0.29 to 17.56)	8.72 (-0.10 to 17.53)	8.68 (-0.14 to 17.49)
Indian	19.71 (9.69 to 29.73)	24.57 (15.52 to 33.63)	23.92 (15.10 to 32.74)	25.01 (15.87 to 34.14)	23.30 (14.49 to 32.12)	22.87 (14.25 to 31.48)	23.17 (14.46 to 31.89)
Pakistani	15.71 (6.85 to 24.57)	18.02 (9.75 to 26.30)	17.19 (8.94 to 25.44)	17.59 (9.79 to 25.39)	18.68 (7.81 to 29.55)	17.99 (7.15 to 28.82)	18.55 (7.39 to 29.72)
Bangladeshi	38.64 (28.67 to 48.62)	38.20 (28.95 to 47.44)	38.47 (30.07 to 46.88)	38.60 (29.77 to 47.42)	36.37 (26.98 to 45.76)	36.49 (27.11 to 45.88)	36.57 (27.17 to 45.97)
Black	18.30 (12.84 to 23.76)	18.30 (13.35 to 23.25)	18.69 (13.59 to 23.79)	18.86 (13.75 to 23.97)	18.47 (11.65 to 25.29)	18.33 (11.95 to 24.71)	18.42 (11.75 to 25.09)
Chinese/Other	39.44 (29.68 to 49.21)	40.43 (31.37 to 49.50)	40.65 (31.83 to 49.47)	40.14 (31.26 to 49.01)	41.89 (31.60 to 52.18)	41.39 (31.05 to 51.73)	41.46 (31.20 to 51.71)

Reference: DFE- RR608

ISBN: 978-1-78105-669-1

The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Education.

Any enquiries regarding this publication should be sent to us at: team.longitudinal@education.gov.uk or www.education.gov.uk/contactus

This document is available for download at www.gov.uk/government/publications