Mobility and missing data: What difference does non-response make to observed patterns of intergenerational class mobility by ethnic group?

Lucinda Platt

ISER Working Paper
Number 2005-10

Institute for Social and Economic Research

The Institute for Social and Economic Research (ISER) specialises in the production and analysis of longitudinal data. ISER incorporates the following centres:

- ESRC Research Centre on Micro-social Change. Established in 1989 to identify, explain, model and forecast social change in Britain at the individual and household level, the Centre specialises in research using longitudinal data.

- ESRC UK Longitudinal Studies Centre. This national resource centre was established in October 1999 to promote the use of longitudinal data and to develop a strategy for the future of large-scale longitudinal surveys. It was responsible for the British Household Panel Survey (BHPS) and for the ESRC's interest in the National Child Development Study and the 1970 British Cohort Study

- European Centre for Analysis in the Social Sciences. ECASS is an interdisciplinary research centre which hosts major research programmes and helps researchers from the EU gain access to longitudinal data and cross-national datasets from all over Europe.

The British Household Panel Survey is one of the main instruments for measuring social change in Britain. The BHPS comprises a nationally representative sample of around 5,500 households and over 10,000 individuals who are reinterviewed each year. The questionnaire includes a constant core of items accompanied by a variable component in order to provide for the collection of initial conditions data and to allow for the subsequent inclusion of emerging research and policy concerns.

Among the main projects in ISER's research programme are: the labour market and the division of domestic responsibilities; changes in families and households; modelling households' labour force behaviour; wealth, well-being and socio-economic structure; resource distribution in the household; and modelling techniques and survey methodology.

BHPS data provide the academic community, policymakers and private sector with a unique national resource and allow for comparative research with similar studies in Europe, the United States and Canada.

BHPS data are available from the Data Archive at the University of Essex
http://www.data-archive.ac.uk

Further information about the BHPS and other longitudinal surveys can be obtained by telephoning +44 (0) 1206 873543.

The support of both the Economic and Social Research Council (ESRC) and the University of Essex is gratefully acknowledged. The work reported in this paper is part of the scientific programme of the Institute for Social and Economic Research.

Acknowledgement:
The permission of the Office for National Statistics to use the Longitudinal Study is gratefully acknowledged, as is the help provided by staff of the Centre for Longitudinal Study Information & User Support (CeLSIUS), in particular Julian Buxton. The author, however, retains responsibility for the interpretation of the data.

Readers wishing to cite this document are asked to use the following form of words:

> **Platt, Lucinda (June 2005) 'Mobility and missing data: What difference does non-response make to observed patterns of intergenerational class mobility by ethnic group?', *ISER Working Paper* 2005-10. Colchester: University of Essex.**

For an on-line version of this working paper and others in the series, please visit the Institute's website at: http://www.iser.essex.ac.uk/pubs/workpaps/

Institute for Social and Economic Research
University of Essex
Wivenhoe Park
Colchester
Essex
CO4 3SQ UK
Telephone: +44 (0) 1206 872957
Fax: +44 (0) 1206 873151
E-mail: iser@essex.ac.uk
Website: http://www.iser.essex.ac.uk

ABSTRACT

While there is an extensive literature on intergenerational class mobility in Britain there is less work on the impact of these different forms of missing data on the observed transitions between parent's class and child's class. It may be particularly pertinent to consider the impact of missing data when evaluating ethnic group differences in intergenerational mobility, since unemployment is known to differ systematically with ethnicity, as do survey and census response. Research into intergenerational class transitions as they differ by ethnic group in Britain is the subject of only a small number of studies, and there has not been any detailed examination of the impact of missing data in studying ethnic group differences in intergenerational social mobility. The contribution of this paper is to highlight the issue of missing data in studying social mobility, to outline its potential implications, and to offer some preliminary analysis of the impact of missing data on our understanding of intergenerational social mobility as it varies by ethnic group.

The paper examines three forms of missing data that we might expect to effect the measurement of ethnic group differences in intergenerational social mobility. First, where measured origins or destinations are not accommodated by class categories – e.g. the unemployed; second, those whose origins are measured but who do not survive in the study sample, because of non-response or emigration; and third, those for whom information on variables of interest that are, or might be, associated with achieved social class is missing. These three forms of missing data are explored by (a) examining the impact of operating different assumptions in relation to non-class outcomes, particularly unemployment but also 'other' outcomes; (b) investigating non-response at the point of measurement of destinations, by exploring and comparing the characteristics of those for whom we know the reason for non-response (emigration, recorded through embarkations data) and those for whom we do not know the cause of their non-response. This section of the paper goes on to look at whether there are selection effects in observed patterns of origins and destinations. In the following section, the paper examines (c) whether the treatment of missing values on variables included in models of emigration, attrition and social class outcomes affect the results obtained. Here just two of the potential treatments of missing data, listwise deletion and using dummy variables for missing values are contrasted.

The paper concludes that, despite big differences between ethnic groups in levels of non-class outcomes, emigration and non-response, the impact of these on results as far as it was investigated here was modest. Nevertheless, missing data remains an issue potentially highly pertinent for ethnically differentiated mobility research, and thus, it is argued, being self-conscious about how missing data are treated is important in itself.

Mobility and missing data: What difference does non-response make to observed patterns of intergenerational class mobility by ethnic group?

Introduction

There are potentially three types of missing data relevant to the study of intergenerational class mobility:

1. where measured origins or destinations are not accommodated by class categories – e.g. the unemployed;
2. those whose origins are measured but who do not survive in the study sample , because of non-response or emigration (case non-response or, in longitudinal data such as that used here, often termed attrition);
3. those for whom information on variables of interest that are, or might be, associated with achieved social class is missing (item non-response).

While there is an extensive literature on intergenerational class mobility in Britain (see, for just a few examples, Bottero and Prandy, 2000; Erikson and Goldthorpe, 1993; Goldthorpe et al, 1987; Halsey et al, 1980; Heath and Payne, 2000; Marshall et al, 1997; Payne and Roberts, 2002; Prandy, 1998; Savage, 2000) there is less work on the impact of these different forms of missing data on the observed transitions between parent's class and child's class (though for some exceptions see the discussion below). It can be argued to be particularly pertinent to consider the impact of missing data when evaluating ethnic group differences in intergenerational mobility, since unemployment is known to differ systematically with ethnicity, as are survey and census response. Research into intergenerational class transitions as they differ by ethnic group in Britain is the subject of only a small number of studies, such as those by Heath and Ridge (1983), Heath and McMahon (1999) and Platt (2005) ; and while Platt (2005) touched on the issue of ethnic group differences in the proportions of those whose origins were observed whose destinations were also observed, there has not been any detailed exploration of the issue of missing data in studying ethnic group differences in intergenerational social mobility. The contribution of this paper is to highlight the issue of missing data in studying social mobility, to outline its potential implications, and to offer some preliminary analysis of the impact of missing data on our understanding of intergenerational social mobility as it varies by ethnic group.

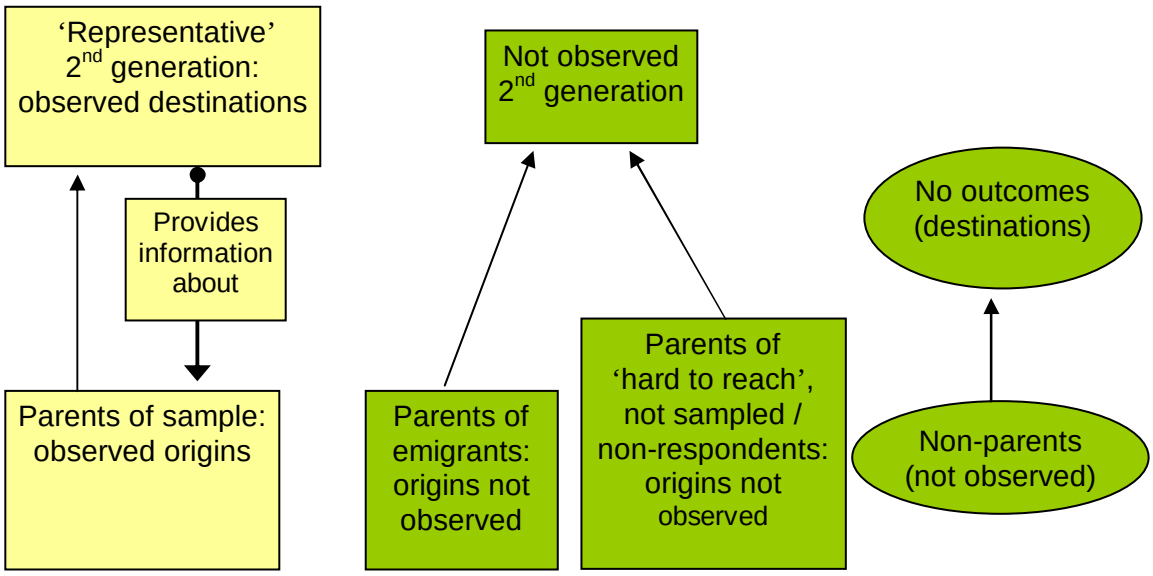
Examining the three types of missing data and their implications for mobility studies in more detail, I first consider the allocation of respondents to non-class categories. When looking at the relationship between class origins and destinations we can note the fact that some origins will not fit the main classes used for analysis. They will therefore be classified as 'other'. For some other cases there will simply be insufficient information to accord class origins. They will therefore be classified as missing. Similarly in the destination class, there will be those who fall outside the main class scheme being either unemployed or economically inactive ('other') or for whom relevant information is simply missing. These issues are relevant to all studies of class mobility and are common to both retrospective and prospective designs. Given the focus within mobility measurement on the relative chances of movement between classes for those of different origins, missing data is often not considered highly problematic – since the odds ratios of the chances of ending up in one class versus another for those from one class background compared to those of another class background remain unaffected by changes in the marginals. The solution therefore has

often been to simply ignore those cases which cannot be allocated to a clear class outcome; although in the loglinear analyses that are typical of much work on intergenerational mobility some techniques do exist for including such unclassified cases (see Allison, 2002)). Additionally, as Miller (1998) has pointed out it could be important to consider unemployment as an additional category to the class positions strictly defined within mobility measurement, and he illustrates how results may change under different ways of treating unemployment.[1] He explored the impact of including or not including the unemployed in terms of goodness of fit of patterns of association as well as number of cases included or excluded from analysis. Minority ethnic groups, even the 'successful' ones such as Indians have higher risks of unemployment than the majority. Therefore, as I have argued elsewhere (Platt 2005; forthcoming 2005), it is important to consider unemployment alongside actual class destinations when considering mobility patterns as they vary by ethnic group. Equally, it is important to reflect on what the broader implications of that choice are, and how it might shape the reported findings.

The second issue is one which is dependent on genuinely longitudinal data to take account of. Measurement of mobility based on retrospective recall of origins, which is typically used in mobility studies, is, by definition, unable to take account of selection issues: the differential 'survival' up to the point of measurement of destination of those whose origins would have otherwise meant they would have been incorporated into the study. Those who have emigrated or otherwise proved hard to follow-up even though their origins were in the geographical area of interest (England, Britain, England and Wales) are, by definition, unknown quantities. This point is illustrated in Figure 1, which illustrates the way a retrospective study informs us about intergenerational mobility. From Figure 1, we can see that for such studies, the sample (of destinations) is selected according to the sample member's parent having a child in the first place (or sometimes more specifically a son), according to the child (or son) having survived long enough to be included in the sample, and having not emigrated in the meantime, and as long as the child does not belong to what is termed here a 'hard to reach' group for the purposes of being selected in a random sample survey (i.e. not being a non-respondent). From such a sample, the information about parents is then collected. Given the danger of recall error, discussed further below, the quality of the information on parent(s)'s origins is likely to be less precise than for a prospective study, and there may be a tendency to remember a better occupation than that which the parent was actually engaged in when the respondent was 'about 14'. This could be a particular issue for minority ethnic group members if their parents experienced downward mobility on migration, and could thus result in a class from a country with a very different class and occupational structure being attributed, even if migration and ensuing downward mobility had occurred well before the respondent was 14 years old.

[1] Yamaguchi (2003) has also considered techniques for accounting for outcomes not classifiable into class destinations in studying class mobility, however this paper does not make use of his approach.

Figure 1: Studying intergenerational mobility retrospectively

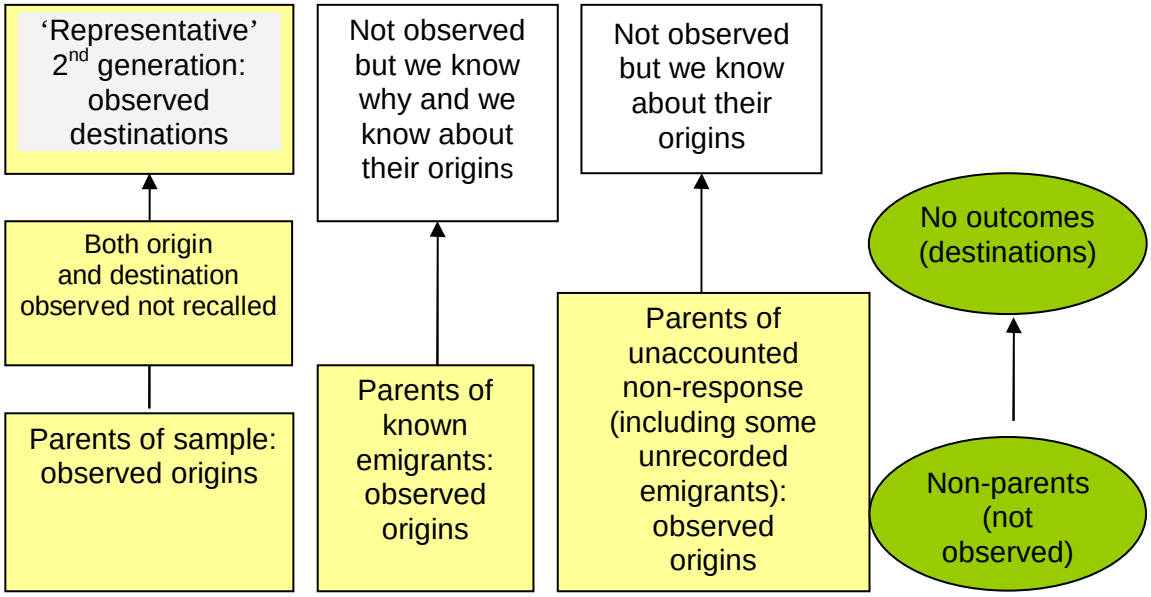


In relation to non-respondents, while there are survey methods for estimating some of the characteristics of non-respondents to sample surveys, it is not possible to extrapolate about their origins without very strong assumptions, which would have to be based on the very information which mobility studies set out to explore in the first place (i.e. the congruence between parent's and child's social position). The issue of selection into the study sample at the point of measurement of destinations is, then, one that can only be treated by longitudinal data; and few sources of British longitudinal data have sufficiently large samples of minority ethnic group members to allow examination of this issue. The exception, for England and Wales, is the source used here, the ONS Longitudinal Study (LS). Figure 2 illustrates, by contrast with Figure 1, how we know something even about those whose destinations are not measured at the outcome point. This additional information on those who fall outside the main transition sample is used in this paper. The prospective analysis of ethnic group patterns of intergenerational mobility employed here is unique to this study and thus enables the consideration of selection issues in ethnic mobility patterns that has not previously been attempted.

Figure 2 shows how the origins (i.e. the parents' characteristics) are observed when the study sample were children, and then the sample's own outcomes (the destinations) are observed twenty or thirty years later. When these destinations are recorded the information, based as it is on the census, is approximately representative of the population as a whole at that time point, apart from the differences in response rates, mentioned above (see Simpson, 1996; Simpson and Middleton 1997; Office for National Statistics, 2003). Therefore, analysis of these origin-destination patterns is approximately equivalent to that of a nationally representative sample (also subject to differential non-response -- and typically at much higher rates) of these age groups at a

point in time.[2] But, in addition, even for those whose destinations are, for whatever reason, not observed, we can say something about their background.[3]

Figure 2: Measuring social mobility prospectively



Given that minority ethnic group members are both more likely to emigrate than the white majority and are more likely not to respond to sample surveys as well as censuses (Simpson and Middleton, 1997), it seems important at least to explore whether there is any apparent impact of non-response over time in an ethnically differentiated study of intergenerational mobility.

Third, there are a range of ways of dealing with missing responses (item non-response) and an extensive literature on how to treat this form of non-response, including in longitudinal contexts However, a common way of dealing with non-response, including in mobility studies, is simply listwise deletion – i.e. ignoring all those cases for whom origin or destination information is missing. One alternative is to incorporate dummies for the missing information. Though not recommended by Allison (2002), the attractiveness of this strategy, both in its simplicity and in enabling sample sizes to be retained and consistently-sized samples to be used across different models has made it regularly used. If non-response varies by ethnicity or if ethnic

[2] It is only approximately as measuring mobility in this way does not take account of recent immigration and the fact that current destinations are swelled by such immigration. However, such recent immigrations would need to be excluded from any study which was trying to explore the effects of within-Britain experience on ethnic minority outcomes, rather than whether they may or may not have experienced mobility in relation to parents' class in country of origin.

[3] The one selection issue which has not been addressed in this study, is the requirement of the children being co-resident with their parent(s) when parental class is measured. See, for example, the discussion in Francesconi and Nicoletti (2004) of this issue.

group is itself subject to non-response then it is possible that the approach used for dealing with missing information could affect the findings.

This paper attempts to expand our understanding of the impact of missing data on the study of social mobility and ethnicity. It includes brief treatments of the first issue, where it follows Miller (1998) and explores the impact and implications of treating 'non-class' destinations (unemployment and 'other') and 'other' origins in a range of ways. For the second, it exploits the genuinely longitudinal nature of the LS to examine the impact of loss to the original sample over time, the impact of the selected nature of those whose destinations are measured and on whom observed origin and destination patterns are based. Moreover, it precedes the examination of selection with a comparison of those who are lost from the original sample but for whom we know why (known emigrants) with those for whom we do not know the reason why they are not observed in 2001: non-respondents to the Census who will also include some un-notified emigrants.

For the third issue, it briefly examines whether the coefficients vary much between models predicting the probability of both emigration and unexplained attrition and in a models of social class outcomes controlling for origins, depending on whether they use dummies for missing values or listwise deletion. In all three illustrations it explores how these issues might be particularly relevant for studies differentiated by ethnic group and how they might modify our understanding of ethnic group processes of intergenerational mobility.

The structure of the paper is thus as follows: in the next section the data source and the study design are outlined. The third section treats the impact of operating different assumptions in relation to non-class outcomes, particularly unemployment but also 'other' outcomes. The fourth section looks at non-response at the point of measurement of destinations. It explores and compares the characteristics of those for whom we know the reason for non-response (emigration, recorded through embarkations data) and those for whom we do not know the cause of their non-response. It then goes on to look at whether there are selection effects in observed patterns of origins and destinations, that is whether the predicted probabilities of ending up in professional/ managerial occupations would look different if all those whose origins were observed also remained in the study sample to have their destinations observed. The fifth section investigates whether the treatment of missing values on variables included in models illustrated in the previous section – on probabilities of emigration or attrition – affects the results obtained, examining just two of the potential treatments, listwise deletion and using dummy variables for missing values. It follows this up with a similar comparison exploring these two alternative treatments of missing values on a model of the impact of origins on destinations, examining, specifically the impact of a range of respondent's background and own characteristics (including ethnic group) on the chances of ending up in the professional / managerial classes. The final section briefly concludes.

Data and study design

To explore the role that missing data can play in mobility research, this study uses the ONS Longitudinal Study (LS). The LS is a record-linkage study of one per cent of the population of England and Wales. It was initially obtained by taking a sample of the

1971 Census, based on those born on one of four birth dates (day and month). At each subsequent Census, information from samples taken using the same sampling criteria is linked where possible. Between censuses, members are added to the study by linking information on births and immigrations, again using the same sampling criteria. Information on death and emigration from England and Wales is also linked to the study. Nor further information is lined for sample members who have died or emigrate, although linkage recommences for members who return to England and Wales.

The LS has some key advantages when it comes to exploring intergenerational mobility and ethnicity, namely the size of the study sample, which facilitates analysis by ethnic group at relatively disaggregated levels, and the longitudinal design of the study, which allows intergenerational mobility to be tracked directly. In addition to the potential of longitudinal data for studying intergenerational class transitions prospectively, the prospective design employed here also reduces the problems associated with recall error, which is known to be an issue for mobility research (see, e.g. Jacobs, 2002; Jäckle and Lynn, 2004), since parental class can be measured at the point when the children are still children and living with their parents. Recall error is potentially a particular problem for the study of migrants' mobility where children of those who suffer downward mobility may prefer to recall their parents' potential or pre-migration occupational status rather than their realised occupation on migration to Britain. The prospective approach also makes it possible to be clear that the origin information is that of the migrants' occupational position following migration to Britain rather than pre- and post-migration class being indistinguishable.

For this study, two cohorts of children aged between 4 and 15 at the point they were observed in the LS have been selected from both the 1971 and the 1981 records in the LS. They are called the '1971 cohort' and the '1981 cohort' based on the point at which they were first observed in the LS for selection into this study. Their parents' and household characteristics are measured at that point to give information about their 'origins'; and, where they continue to be observed in the LS, their own characteristics are measured in 2001 to given information about their 'destinations'. For those how are lost to follow-up by 2001, we have some information on whether they emigrated (from linked embarkations records) or died (from linked registry data), but for the majority of those who are lost to follow-up we have not further information. The 1971 and 1981 cohorts of children have been combined, or 'pooled', for the analyses in this paper. For additional information on the study design and the variables used, see the Technical Appendix to Platt (forthcoming 2005). (Following publication of the main report, the Technical Appendix will be available at www.jrf.org.uk.)

Parents' social class was measured in 1971 and 1981 respectively for the two cohorts; and, grouped in three categories – service, intermediate and working – was used to identify children's origins. The sample's own social class (their destinations), where observed, was measured in 2001 grouped into three corresponding classes (professional/ managerial, intermediate and manual/routine non-manual). Those currently unemployed were allocated to an 'unemployed' destinations category and those in other situations (looking after home and family, students etc) were allocated to an 'other' category at both origins and destinations. This information is used in the

comparison of transitions in the next section of this paper and for some of the simple tabulations in the fourth section. For the models in the fourth and fifth sections, a simple binary comparison is used to predict destinations of professional/ managerial versus any other destination, but 'other' is included alongside the other origin categories among the explanatory variables. The basis of the social class classification for origins was the CASMIN schema (sometimes known as the Goldthorpe class schema) and for destinations was the National Statistics Socio-economic Classification (NS-SeC). Both employ similar principles and approaches in class allocation and are therefore broadly comparable over time. Other variables that were included in the analyses were economic status (as measured by housing tenure and car ownership), ethnic minority concentration in ward where living when first observed, educational level of the study sample's parents, and educational level, marital status, of the sample member and economic status variables in their destination household. Parent's country of birth was used to distinguish those from migrant backgrounds from those with parents born in the UK: an aspect of the minority group members' history that was incorporated into the ethnic group classification. It also enabled the construction of a white migrant group to include among the other minority groups; and it was used to construct a proxy ethnic group for exploring the ethnic group differences in those lost to follow-up. This proxy ethnicity was validated against those for whom ethnic group was recorded at either 1991 or 2001 and against an existing proxy ethnic group variable in the 1971 LS records.

Dealing with non-class origins and destinations

This section considers the different options for dealing with non-class outcomes. Rather than simply ignoring 'missings' and positions that do not map onto standard classes, the impact of various ways of treating these is outlined and the implications for both the number of cases included and the goodness of fit of the transitions is assessed. The implications for ethnic group analysis and class allocation are further considered.

First, Table 1 illustrates the full range of origin to destination transition proportions within the study sample, while Table 2 illustrates the more limited range of 3-class by 3-class origins and destinations. The picture is obviously clearer in Table 2 and, as pointed out, if it is relative odds that are of interest, the symmetric odds ratios of, say, service class versus class working class chances of professional / managerial class position will remain the same regardless of how much extra information is included. However, in terms of showing a more comprehensive and thus in some sense 'truer' picture of the data, the more complex set of transitions may be preferred, particularly if what is of primary interest is not the transitions in themselves but how they vary between sub-groups, such as minority ethnic groups. Moreover, Table 1 illustrates that the non-class outcomes seem to vary with class position. That is, they appear to have a class gradient that may be relevant to considerations of relationships between origins and destinations.

Table 1: Complete origin to destination transitions, 1971/81-2001, row percentages

	Professional/ Managerial	Inter-mediate	Routine / Manual	Un-employed	othe r	missing	N
Service	51.5	12.4	9.4	1.4	3.2	22.0	46,513
Inter-mediate	35.8	17.8	17.3	1.8	4.8	22.5	32,111
Working	30.7	14.7	22.1	2.5	6.2	23.8	88,153
Other	21.8	10.6	19.7	3.8	9.8	34.3	10,244
Missing	20.3	12.2	17.5	3.5	7.6	38.9	606
Total	**36.5**	**14.4**	**17.8**	**2.2**	**5.4**	**23.7**	**177,627**

Source: ONS Longitudinal Study, author's analysis

Table 2: Simple version of origin to destination transitions, 1971/81-2001, row percentages

	Professional/ Managerial	Intermediate	Routine / Manual	N
Service	70.2	16.9	12.9	34,141
Intermediate	50.5	25.1	24.4	22,762
Working	45.5	21.7	32.8	59,538
Total	**53.7**	**21.0**	**25.3**	**116,441**

Source: ONS Longitudinal Study, author's analysis

I have therefore gone on to explore what might be gained by treating the class transitions in a range of different ways and operating stronger or weaker assumptions in relation to them. A set of transition matrices operating different assumptions concerning the treatment of missing, 'other' and unemployment (at both origin and destination) were tried out with their relative fit being assessed (following Miller 1998). Missing information, though not a substantial issue for origins (amounting to only 606 out of 177, 627 cases or 0.3 per cent of the sample) was a more significant issue for destinations, where it amounted to 5 823 out of the 140 900 cases who were observed at 2001 and for whom origins were not missing (4.1%). Therefore, the different transitions also incorporate different assumptions about how to treat these missing destinations. Moreover, there was also a substantial proportion of the starting sample not observed at 2001 for whom, obviously, there was no class destination information. These non-respondent missings are also included in some of the transitions under a range of assumptions. This illustrates the point that attempting to make such assumptions about non-observed is not very satisfactory. Table 3 illustrates the different options for treating unemployed and 'other' at the point of destination, 'other' at the point of origin, and missings at the point of origin and destination, with missings at destination being alternatively those with insufficient information to construct class at 2001 and these combined with those who were not observed at 2001.

Table 3 shows (in the top section) that incorporating additional destinations has, in fact, little impact on the fit of the model, though sample sizes can be increased substantially. This results from the fact that these additional destinations vary with class, as the transition table (Table 1) indicated. However, when 'other' origins are

included the fit weakens and it weakens further when all missings are included, demonstrating that missings are more evenly distributed across origins and thus do not enhance the patterns of association. This is reassuring on one level, but may indicate that the observed pattern of origins and destinations exaggerates a picture of class retention.

The second section of the Table 3 shows the impact not just of including additional destinations but including them within the class framework, i.e. allocating 'other' destinations and unemployed to different classes. The different treatments involve allocating both unemployment and unemployment-plus-other-destinations to the routine / manual class. This treatment is based on the assumption that those who would otherwise be in the working class are more likely to be unemployed and out of the labour market in other ways. This is an approach that is not without precedent, though Miller (1998) found it not entirely satisfactory. Here, it has little effect on the fit of the model, suggesting it would not be a perverse approach. The model fit is improved substantially (and unsurprisingly) under the assumption of complete congruence of destination with origin for those who do not otherwise fit a into the strict class schemes. Given that this rigid assumption about class congruence is not supported by the levels of observed mobility and by the measure of association provided for option 1, this is not really an appropriate way of dealing with the 'problematic' responses. Including this treatment does, however, reveal the way in which it is out of line with the other treatments, even though it is an assumption that is often employed in when extrapolating an 'expected' class in relation to migrants' experience of downward mobility (see, for example, Heath and Ridge 1983).

The final two sections of Table 3 deal with allocating missing class information to the transition matrix in various ways, both those missings resulting from all causes including loss to follow-up at 2001 and only those missings resulting from insufficient information to allocate a class among those observed at 2001. Including those lost to follow-up means that the whole starting sample can be considered, thus plotting the entire cohort rather than the 65-80 per cent that would otherwise be included. Including all those who are not observed at 2001 is, as mentioned above, not very satisfactory. The large proportion of the original sample who are not-observed at 2001 also means that there is a danger that any system for allocating them to the class distribution will be likely to bias results.

Among those who were observed at 2001 but for whom there was insufficient information to allocate a class, we can see that model fit is improved when they are allocated to 'other' destinations. This suggests, then, a possible strategy if it is found desirable to both maximise sample sizes and to include a residual 'other' category to reflect the diversity of outcomes and the fact that 'other' destinations vary with ethnicity.

Table 3: Different ways of dealing with non-class origins and destinations

Option	Description	N	Cramér's V
1	Only three 'proper' class positions included for both origins and destinations	116,441	0.17
2	Unemployed and 'other' destinations combined as fourth destination, origin 'other' to missing	128,352	0.17
3	Unemployed as fourth destination, origin and destination 'other's to missing	119,858	0.17
4	Unemployed and 'other' destinations as separate destinations, origin 'other' to missing	128,352	0.17
5	Unemployed and 'other' destinations as separate destinations, 'other' origin as separate origin	135,077	0.15
6	Unemployed and 'other' destinations as separate destinations, 'other' origin as separate origin, missings as separate origins and destinations	177,627	0.12
7	Unemployed destinations to routine / manual class, 'other' origins and 'other' destinations to missing (i.e. excluded along with all missing origins and destinations)	119, 858	0.17
8	Unemployed and 'other' destinations to routine / manual class, 'other' origins to missing	128, 352	0.17
9	Unemployed and 'other' destinations to routine / manual class, retain 'other' origins	135, 077	0.18
10	Unemployed and 'other' destinations to class corresponding to origin, missing origins and destinations excluded	128, 352	0.23
11	Missing destinations (including missing due to not observed at 2001) to professional / managerial class, missing origins excluded	177, 021	0.11
12	Missing destinations (including missing due to not observed at 2001) to routine / manual class, missing origins excluded	177, 021	0.12
13	Missing destinations (including missing due to not observed at 2001) to 'other', missing origins excluded	177, 021	0.13
14	Missing destinations (including missing due to not observed at 2001) to destination class corresponding to origin class, missing origins excluded	177, 021	0.34
15	Missing destinations where respondent is observed at 2001 to professional / managerial class, missing origins excluded	140, 900	0.14
16	Missing destinations where respondent is observed at 2001 to routine / manual class, missing origins excluded	140, 900	0.15
18	Missing destinations where respondent is observed at 2001 to 'other', missing origins excluded	140, 900	0.18

Source: ONS Longitudinal Study, author's analysis

Turning to looking at such treatments broken down by ethnic group, Table 4 shows the 'basic' transition matrix of three origin classes by three destination classes for each of four ethnic groups for whom sufficient sample sizes allowed investigation of various transition matrices, without producing disclosive results (deemed to be counts of 3, 2 or 1 in any given cell). These four ethnic groups are those who defined themselves as white in 1991 and whose parents were both born in the UK, 'white non-migrants'; those who defined themselves as Caribbean or Black Other in 1991 and at least one of whose parents was born outside the UK, 'Caribbeans'; those who defined themselves as Indian in 1991 and at least one of whose parents was born outside the UK, 'Indians'; and those who defined themselves as white in 1991 and where neither (co-resident) parent was born in the UK, 'white migrants'.

Table 4: Fit of transitions by ethnic group

Type of transition matrix	Ethnic group	Proportion of group (excluding those not observed at 2001) included	Cramér's V
Only three 'proper' class positions included for both origins and destinations (1)	White non-migrant	0.83	0.17
	Caribbean	0.63	0.06
	Indian	0.79	0.13
	White migrant	0.75	0.12

Source: ONS Longitudinal Study, author's analysis

Other options from top set in Table 3 produced little variation in this pattern when distinguished by ethnic group and are thus not supplied. The different treatments only produced differences in the proportions included, with high rates of unemployment among Caribbeans and to a lesser extent among the other minority groups. The issue here is that there does not seem to be an optimal approach for treating non-class destinations across minority ethnic groups, and that the variation between ethnic groups in terms of patterns of relationships between origins and destinations would seem to be the crucial issue here. However, the consistency of the measure of association across different treatments might justify the inclusion of the extra information since the impact on sample sizes of excluding 'non-class' outcomes varies substantially by ethnic group with over a third of Caribbeans falling into such outcomes. Moreover, this becomes not simply an issue of maximising sample sizes but also of acknowledging the importance of alternative destinations, in particular unemployment, as important elements of the story of social class transitions for particular ethnic groups.

Moreover, the illustration of ethnic group differentiation reveals that the assumptions about association between origin and destination that are found for the majority are much weaker for ethnic minority groups. As a result we might question how appropriate traditional mobility approaches are for different ethnic groups.

Drop-out, death and departure: selection effects in mobility patterns

This section focuses upon a second form of missing information: case non-response, or the fact that the patterns of origins and destinations modelled are dependent on those who 'survive' in the data from the point at which their origin is measured to the point at which their outcome is measured. It is therefore of interest to explore the impact of those who disappear by the time destinations are measured, also termed attrition. Given the way the LS is constructed, tracking people through the Censuses rather than following them up personally, attrition here is rather different from attrition in other longitudinal sources, such as cohort or panel surveys. Instead, the population who survive are selected from those who are taken to be the full (or at least enumerated) population at each decennial Census. While immigration means they will not be fully representative of their age group at that point, instead they have the advantage that they are less likely to be subject to the sort of drop-out, through difficulties in locating movers, making contact or through choice not to participate further that you would find in other longitudinal sources, except in so far as there is Census under-enumeration, which is clearly an issue here.

This section, therefore, considers this issue of attrition, factors associated with it and possible selection effects impacting on observed outcomes. While the very fact of non-response means that we cannot observe the characteristics at 2001 of those who do not remain in the study up to that point, as we saw in Figure 2, we can make use of various sorts of information that is available. We have information on the backgrounds of all those who are lost to follow-up regardless of the reason (including on parents' country of birth from which I constructed a proxy ethnic group variable). Moreover, from the LS we can also identify those from the original cohorts who died before their destinations were measured; those from the original cohorts who emigrated, and those who were not known to have died or emigrated, but are 'lost' nevertheless (including those who join the armed forces; and those who between 1971 and 1983 move into long-stay psychiatric units).

Given the age range of the sample, the number of deaths is small (967 by 2001 – or 0.5 per cent of the starting sample), as is that of those joining the armed forces, and the category of those who moved to long stay psychiatric hospitals is barely relevant to this study. The number of known emigrants is more substantial (4 258 members by 2001 – or 2.4 per cent of the original sample and 11.7 per cent of all attriters by 2001). Based as it is on embarkations data and expressed emigration, it is estimated only to capture about half of actual emigration (Blackwell et al 2003). But the biggest group is those for whom their reason for non-response in 2001 cannot be identified (31 099 by 2001 or 17.5 per cent of the starting sample).

Table 5 shows the total size of the two cohorts, both at their starting point and at the point at which destinations are measured. It shows that, overall, approximately 20 per cent of the original cohort are lost to follow-up by 2001, though the proportion is slightly higher for the later cohort, of whom a smaller proportion are found at 2001 even though they have had ten years less to die or emigrate. This emphasises the age-related aspects of census non-response (Simpson, 1996; Office for National Statistics, 2003**).**

Table 5: The size of the two cohorts and the pooled sample

	Total size of cohort	Number retained after 2 decades	Number retained after 2 decades as % of total	Number retained at 2001	Number retained at 2001 as percent of total
1971 Cohort	90702	75571	83.3	73120	80.6
1981 Cohort	86925	68183	78.4	68183	78.4
Pooled Sample	177627	143754	80.9	141303	79.6

Source: ONS Longitudinal Study, author's analysis

Given that minority ethnic group members are both more likely to emigrate and not to respond to the Census, it makes sense initially to use the information on known emigrants to examine whether their characteristics are similar to those for whom we don't know about their outcomes. This gives us some handle on the selection processes at work. We can assume that some of the unaccounted for non-respondents are emigrants, but to the extent that the two groups are different we can assume that there are different processes operating in relation to the observed outcomes of our selected sample.

First, then, I compare and contrast those who are known to emigrate with those who are lost to follow-up for unknown reasons by 2001. Specifically, I ask

- What are the characteristics of those who emigrate?
- What are the characteristics of those whose non-response is not accounted for?
- Do the two look similar or different?

In this section, I then go on to investigate how the selected nature of the sample at 2001 might be contributing to reported patterns of intergenerational mobility based on those who are observed. Specifically, I ask

- Is there any evidence of selection effects?
- And if so, how much impact does that have on the destination patterns reported?

For exploring these questions there are a number of ways in which we can measure attrition at 2001. We can measure it as (a) unexplained absence in 2001 (i.e. not known to have died or emigrated = 31 099 cases) compared to present in 2001 (i.e. the 141303 cases who 'survive'); or as (b) all absence in 2001 (i.e. 36 324 cases) compared to presence in 2001 (i.e. 141 303 cases); or as (c) unexplained absence in 2001 (i.e. 31099 cases) compared to both presence and explained absence in 2001 (i.e. 141 303 plus the 5225 deaths and emigration=146528 cases).

For the purposes of the questions asked is makes sense to use approach (a) when comparing unexplained non-response with emigration, that is the two categories of emigration and non-response (attrition) are mutually exclusive, and approach (b) when looking at selection effects overall, i.e. attrition includes all those not observed at 2001 even if the reason why they are not observed is known. This is therefore the approach used for the two parts of this section.

Tables 6-8 compare unexplained attrition and emigration across various characteristics. They show that men are more likely to both emigrate and attrit; those from higher social class backgrounds are also more likely to emigrate but unexplained attrition rates seem to vary little according to class background, with the exception of the increased proportion from 'other' backgrounds who attrit (while a smaller proportion than expected from such backgrounds emigrate). Both emigration and unexplained attrition are greater for those from minority ethnic group backgrounds and, conversely, the white non-migrant group is much less likely to do either. However, there are some differences between groups with those from Caribbean, and to a lesser extent Indian backgrounds showing high rates of unexplained attrition, while those from white migrant backgrounds show particularly high rates of emigration.

Table 6: Unexplained attrition and emigration by sex, column %

	Emigration	Unexplained attrition	All at origin
Male	57.3	61.2	51.0
Female	42.7	38.8	49.0

Source: ONS Longitudinal Study, author's analysis

Table 7: Unexplained attrition and emigration by class of origin, column %

	Emigration	Unexplained attrition	All at origin
Service	36.3	25.2	26.3
Intermediate	16.7	17.4	18.1
Working	42.5	49.0	49.8
Other	4.5	8.4	5.8

Source: ONS Longitudinal Study, author's analysis

Table 8: Unexplained attrition and emigration by selected proxy ethnic groups, column %

	Emigration	Unexplained attrition	All at origin
UK White	74.8	77.9	86.5
Caribbean	1.7	3.8	1.6
Indian	1.7	2.1	1.4
Pakistani	1.2	1.4	0.7
White migrant	8.5	5.5	3.2

Source: ONS Longitudinal Study, author's analysis
Note: columns do not sum to 100 as not all proxy ethnic groups have been included

These apparent similarities and differences between emigrants and (unexplained) attriters were pursued further with models looking at the factors associated with either outcome controlling for other relevant background characteristics (including age and cohort). The estimates from these models are shown in Table 9. Again, the focus is on attrition by 2001. On the one hand it might be argued that attrition increases with time, so attrition should be measured after the same length of time for both cohorts. On the other hand, attrition and emigration are themselves likely to be affected by wider structural and economic factors, and non-response rates vary systematically with

the different censuses. From this point of view, it makes sense to measure attrition as it appeared by a single, particular point in time, i.e. 2001, as here.

In relation to the question 'do emigrants and attriters look the same?' the answer would seem to be yes and no. From Table 9, we can see that, other things being equal, emigrants are more likely to come from the older cohort – supporting the view that the longer the time to emigrate the more likely it is to happen. Controlling for cohort, younger age groups are more likely to attrit while age appears to enhance the probability of emigration. However unexplained attrition is more likely for the younger cohort, other things being equal. Emigration would thus appear to be a feature of lifestage, while attrition is more a characteristic of youth, associating it with the higher census non-response among younger people. (However, it should be noted that few of the sample members fit the highest ages for Census non-response, which fall in the 20-25 age band.) Ethnic group concentration in ward of origin tends to increase both emigration and attrition, but is more evident for attrition. Similarly – and unsurprisingly – being born abroad increases chances of emigration, and also increases chances of attrition. Men are more likely to attrit and women to emigrate, other things being equal.

Turning to class origins – service class origins increase chances of emigration (thus those with higher chances of success would seem to be opting to leave England and Wales) and less strongly increase chances of attrition. For attrition, however, 'other' origins have a much greater effect, suggesting there may be something about unstable class origins, which increases probability of attrition. The role of instability in increasing attrition is also supported by the positive coefficients for lone parent families – though it can also be noted that children from such families are also more likely to emigrate. Father's (higher) qualifications and service class origins, as well as birth abroad increase the chances of both attrition and emigration.

When looking at proxy ethnicity it is perhaps unsurprising to find that all ethnic minority groups have a greater probability of emigrating, and the effects are fairly strong (though the effect is not significant for those with Indian (and East African) backgrounds). Comparable effects are found in relation to attrition, indicating that there may be some correspondence between emigrants and attriters.

Finally, the economic variables work in opposite directions for emigration and attrition: car ownership decreases the chances of attrition but increases the chances of emigration, while local authority housing decreases the chances of emigration but increases the chances of attrition. Thus greater higher levels of assets at origin would appear to promote chances of emigration but decrease chances of unexplained attrition.

Table 9: Logistic regressions of probabilities of attrition and emigration by 2001 controlling for background variables

	Emigration	**Unexplained attrition**
Explanatory variables	Coefficient (SE)	
Cohort (base is 1971)	**-.826 (.035)**	**.261 (.013)**
Age group (base is 12-15 when first observed)		
Aged 4-7 when first observed	*-.045 (.036)*	**.238 (.015)**
Aged 8-11 when first observed	**-.148 (.039)**	**.116 (.016)**
Ward minority grp concentration at origin (base=0)		
0.1-1%	**.234 (.058)**	**.94 (.025)**
1-5%	**.181 (.066)**	**.156 (.028)**
5-10%	*.061 (.092)*	**.296 (.037)**
More than 10%	*-.037 (.090)*	**.362 (.036)**
If sample member born abroad	**.765 (.065)**	**.548 (.034)**
Parents' class (base is working)		
Service	**.197 (.046)**	**.058 (.020)**
Intermediate	*.008 (.048)*	**.057 (.019)**
Other	*-.096 (.093)*	**.234 (.031)**
Male	**-.159 (.033)**	**.532 (.014)**
Mother's qualifications (base no higher quals)		
No co-resident mother	*-.221 (.138)*	**.384 (.041)**
Mother with higher qualifications	**.175 (.052)**	**.069 (.026)**
Father's qualifications (base no higher quals)		
No co-resident father	**.179 (.078)**	**.177 (.028)**
Father with higher qualifications	**.328 (.049)**	**.088 (.024)**
Proxy ethnic group (base is both parents UK born)		
Caribbean	**.341 (.134)**	**1.104 (.046)**
Black African	**1.138 (.249)**	**1.239 (.131)**
Indian (and East African)	*.269 (.150)*	**.339 (.056)**
Pakistani	**.699 (.161)**	**.636 (.069)**
Bangladeshi	**.841 (.355)**	**.405 (.152)**
Chinese and Other	**1.022 (.131)**	**1.105 (.070)**
White migrant	**1.007 (.067)**	**.678 (.034)**
housing tenure at origin (base is owner occupation)		
Local authority	**-.430 (.046)**	**.182 (.017)**
Private rented	**.159 (.049)**	**.235 (.022)**
Car ownership at origin (base is none)		
1 car	**.174 (.044)**	**-.128 (.017)**
2 or more cars	**.174 (.059)**	*.039 (.023)*
Constant	**-3.775 (.073)**	**-2.419 (.032)**
N	**176907**	**170261**
Wald chi2 (df)	**2246.7 (31)**	**20997 (32)**

Source: ONS Longitudinal Study, author's analysis

Notes: Coefficients in **bold** are significant at the 0.05 level; coefficients in *italics* are not significant at the 0.05 level. Bangladeshi proxy ethnicity is only available for 1981 cohort. Standard errors are adjusted for repeat observations on persons. Dummies are included for missing values on housing tenure at origin, mother's and father's qualifications and minority group concentration in area of residence, but coefficients are not quoted here for brevity.

It would seem, then, that while it may well be from these findings that many attriters may be unobserved emigrants, it is not completely clear that this is the case. Indeed it seems highly possible that while more privileged origins tend to promote emigration, the attriters are a mixture of the more advantaged (who emigrate) and the less advantaged who get lost and form the 'hard to reach' group posited above in Figures 1 and 2.

Moving on to consider selection issues, first I consider the characteristics of all those absent at 2001 (attriters, emigrants and dead combined) with those who survive to this point. Tables 10-12 show how those whose 2001 destinations were measured in sections 4-6, above, might form a selected sample in relation to sex (an under-representation of men), class (an under-representation of those from 'other' origins) and ethnic group (a substantial under-representation of all minority ethnicities, but, in particular Caribbeans).

Table 10: All attrition by sex

	Present 2001	Absent 2001
Men	75.7	24.3
Women	83.6	16.4
All	79.6	20.5

Source: ONS Longitudinal Study, author's analysis

Table 11: All attrition by class of origin

	Present 2001	Absent 2001
Service	79.9	20.1
Intermediate	80.5	19.5
Working	80.0	20.0
Other	71.4	28.6
All	79.6	20.4

Source: ONS Longitudinal Study, author's analysis

Table 12: All attrition by proxy ethnicity

Parent's country of birth/ proxy ethnicity	Present in 2001 %	Absent in 2001 %
UK white	81.5	18.5
Caribbean	54.8	45.7
Indian	69.4	30.6
Pakistani	61.7	38.3
White migrant	63.9	36.1
All	80.1	19.9

Source: ONS Longitudinal Study, author's analysis
Note: The proportions for 'All' in this table vary slightly from those in the previous two as it was based only on those for whom a proxy ethnicity could be created.

Here, then we ask what difference does – or might – all attrition make to reported measures of intergenerational social mobility based on those whose outcomes are observed? We may identify more (or less) upward mobility among certain groups than others. However, the presence of high (and differential) levels of attrition raises the

question as to whether those patterns are driven by more of the least (or most) successful from that group being unobserved at 2001. To examine this possibility, a Heckman selection model was estimated (Van de Ven and Van Pragg, 1981), which attempted to take account of all those who were not observed at 2001 for whatever reason, and to estimate the effect of this loss on the probabilities of a professional or managerial class outcome (as opposed to any other class outcome). Table 13 shows both this selection model and a simple probit run without taking into account the selection issues, i.e. by treating those observed at 2001 as if they constitute the entire sample.

For identification of the selection model, presence at the Census 10 years after the cohort was initially sampled was employed. That is, this variable was expected to affect the probability of presence at 2001, but not professional outcomes *per se*. If this identification variable is to stand in for unobserved differences in propensity to attrit, which are revealed over time, a variable which captures a part of that history unfolding can be argued to be appropriate for this purpose. Given the possibility of one-time LS members returning to the study, there were, moreover, sample members who were not present after 10 years but had returned by 2001, as well as those who were present after 10 years but had attrited by 2001. This variable had a highly significant effect. The rho for the selection model had a positive value and the test for independence of the selection and the class outcome models was statistically significant. This suggests that there are indeed selection effects and that the errors of the two models are positively correlated. That is, chances of selection (being observed) at 2001 and of being in a higher social class are positively related

Comparison between the coefficients for the model taking account of selection and the uncorrected probit show some differences, though many of the coefficients look quite similar. Focusing on the ethnic group effects, the selection model shows a greater magnitude in the negative effects among those effects that are statistically significant. That is, Caribbeans, Bangladeshis and Pakistanis look even less likely to achieve professional / managerial class outcomes than their white non-migrant counterparts, controlling for characteristics (including education) when selection is taken account of. Moreover, not only is the negative coefficient for Caribbeans stronger when selection is taken account of, it is only at that point that it becomes significant. Conversely, results from a non-adjusted model show slightly better outcomes than would have occurred had there been no loss of the sample between the time they were observed as children (in 1971 or 1981) and 2001.

Table 13: Selection model of probability of being in the Professional /Managerial Class at 2001 and comparison with Probit Model without Selection effects

	(1) Selection model: Coefficient (SE)	(2) Simple Probit: Coefficient (SE)
Probability of being in professional managerial class		
Cohort (base is 1971)	**-.156 (.010)**	**-.139 (.008)**
Age group (base=older)		
younger	**-.125 (.009)**	**-.113 (.008)**
Middle	**-.036 (.009)**	**-.029 (.009)**
Male (base is female)	**.026 (.011)**	**.046 (.008)**
Partnered	**.671 (.001)**	**.676 (.009)**
Ward minority grp concentration at origin (base=0)		
0-1% minorities	**.109 (.013)**	**.112 (.013)**
1-5%	**.200 (.016)**	**.205 (.015)**
5-10%	**.146 (.022)**	**.158 (.022)**
10+%	**.141 (.023)**	**.157 (.022)**
Social class origin (base=working class)		
Service	**.191 (.011)**	**.194 (.011)**
Intermediate	*.011 (.011)*	*.012 (.011)*
Other	**-.067 (.021)**	**-.060 (.021)**
Mother's qualifications (base no higher quals)		
No co-resident mother	**-.084 (.029)**	**-.074 (.029)**
Mother with higher qualifications	**.064 (.016)**	**.067 (.016)**
Father's qualifications (base no higher quals)		
No co-resident father	**.074 (.018)**	**.080 (.018)**
Father with higher qualifications	**.125 (.014)**	**.129 (.014)**
Tenure of origin (base= owner occupation)		
Local Authority	**-.172 (.010)**	**-.166 (.010)**
Private Rented	**-.099 (.014)**	**-.095 (.014)**
No of cars at origin (base=none)		
1 car	**.109 (.010)**	**.104 (.010)**
2+ cars	**.175 (.014)**	**.172 (.014)**
Ethnic Group (base=white non migrant)		
Caribbean	**-.092 (.045)**	*-.051 (.043)*
Black African	*-.037 (.135)*	*.006 (.135)*
Indian	*.044 (.037)*	*.048 (.037)*
Pakistani	**-.545 (.059)**	**-.519 (.058)**
Bangladeshi	**-.359 (.128)**	**-.329 (.128)**
Chinese and Other	*.021 (.063)*	*.058 (.062)*
White migrant	*.024 (.027)*	*.041 (.026)*
Own qualifications (base=none)		
Level 1	**.594 (.015)**	**.598 (.015)**
Level 2	**.873 (.016)**	**.879 (.015)**
Level 3+	**1.646 (.018)**	**1.658 (.016)**
Constant	**-1.573 (.021)**	**-1.566 (.022)**

Selection (to presence in 2001)		
Presence after one decade	**-.407 (.006)**	
Cohort (base is 1971)	**-.276 (.003)**	
Agegroup (base is older)		
younger	**-.178 (.003)**	
middle	**-.101 (.004)**	
Male	**-.304 (.003)**	
Ward minority grp concentration at origin (base=0)		
0-1% minorities	**-.039 (.005)**	
1-5%	**-.069 (.006)**	
5-10%	**-.170 (.009)**	
10+%	**-.210 (.009)**	
Origin class (base=working)		
Service	**-.025 (.005)**	
Intermediate	**-.009 (.004)**	
Other	**-.106 (.009)**	
Mother's qualifications (base no higher quals)		
No co-resident mother	**-.160 (.012)**	
Mother with higher qualifications	**-.032 (.006)**	
Father's qualifications (base no higher quals)		
No co-resident father	**-.084 (.008)**	
Father with higher qualifications	**-.038 (.006)**	
Proxy ethnicity (base is both parents UK born)		
Caribbean	**-.516 (.019)**	
Black African	**-.552 (.058)**	
Indian	**-.062 (.016)**	
Pakistani	**-.377 (.024)**	
Bangladeshi	**-.393 (.058)**	
Chinese and Other	**-.454 (.028)**	
White migrant	**-.229 (.011)**	
Tenure at origin (base=owner occupation)		
Local authority	**-.107 (.004)**	
Private rented	**-.067 (.006)**	
Qualified mother	**-.033 (.006)**	
Qualified father	**-.037 (.006)**	
Number of cars (base=0)		
1 car	**.090 (.004)**	
2+ cars	**.059 (.006)**	
Constant	**2.03 (.009)**	
rho	**.212 (.078)**	
Wald test of rho=0: chi2(df)	**6.97(1)**	
Wald chi2 (df)	**24055 (36)**	**27151 (36)**
N	**155220**	**134992**

Source: ONS Longitudinal Study, author's analysis

Notes: Coefficients in **bold** are significant at the 0.05 level; coefficients in *italics* are not significant at the 0.05 level.
Standard errors are adjusted for repeat observations on persons
Dummies are included for missing values on housing tenure at origin, mother's and father's qualifications and minority group concentration in area of residence, but coefficients are not quoted here for brevity.

The general comparability between the selection and non-selection models is born out when predicted probabilities constructed on the basis of either model are compared, as is shown in Table 14. Thus, despite the fact that the sample that survives to have its destinations measured is strikingly different from the cohorts which start out in 1971 and 1981, selection does not appear to have a huge impact on the overall observed patterns of origins and destinations.

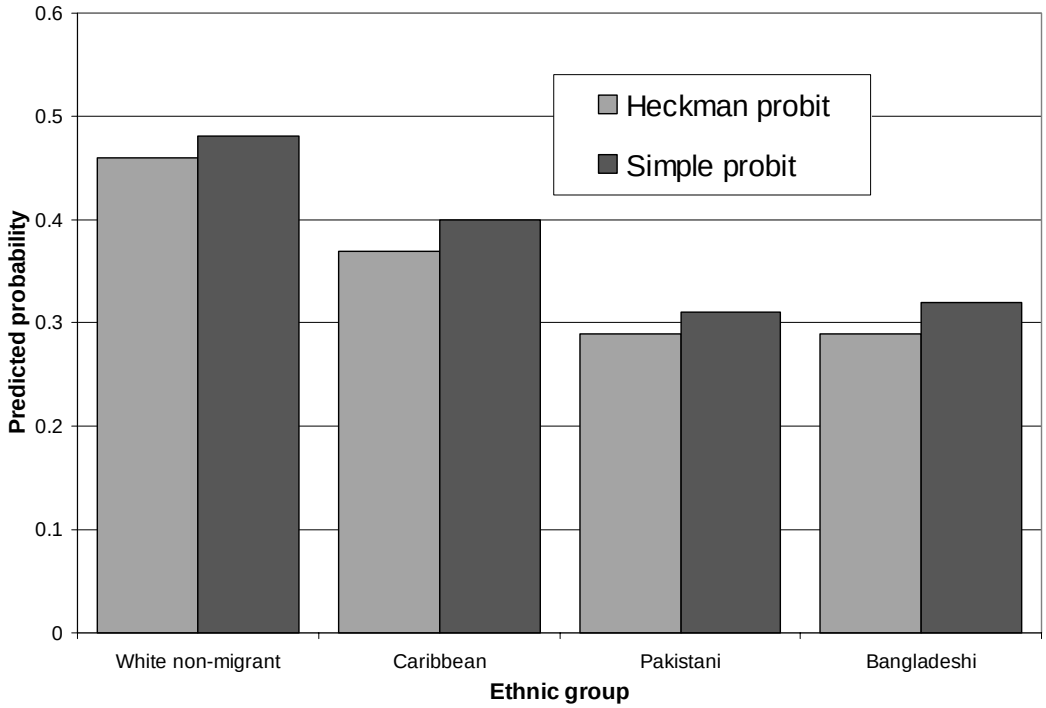
Table 14: Predicted probabilities of proportion in the professional managerial class: Selection model and simple model

	Heckman probit	Simple probit
White non-migrant	.46	.48
Caribbean	.37	.40
Pakistani	.29	.31
Bangladeshi	.29	.32

Source: ONS Longitudinal Study, author's analysis
Note: probabilities are based on group's characteristics

This point is further pursued in Figure 3 which illustrates predicted probabilities of professional or managerial class outcomes for different ethnic groups based on the simple probit and the selection model. There would appear to be impacts of the changing shape of different ethnic groups over time on what we observe in terms of patterns of social mobility, but the differences between what we observe and the counterfactual of what would have been observed without attrition (including differential emigration) are not large. In Figure 3, the predictions of professional / managerial class outcome are based on the combinations of the characteristics of the members of the different groups and the corresponding coefficients from the different models – that taking account of selection and that not. It shows that for all the four groups illustrated their average predicted chances of a professional / managerial class outcome are higher with the simple probit than they would be if all those from the original cohorts were still observed at 2001. This is not to say the simple model is necessarily 'wrong', in that the observed patterns may be what we are interested in. But it does suggest that it is those least likely to succeed who are disproportionately lost, and that therefore we may have an enhanced picture of social mobility across the different groups by failing to take account of selection.

Figure 3: Predicted probabilities of professional / managerial class outcome: selection model and simple probit compared



Source: ONS Longitudinal Study, author's analysis

The impact of different treatments of item non-response

Finally, this paper compares the use of dummy variables for missing values and of listwise deletion as ways of dealing with missing information (item non-response) on variables of interest in investigating mobility patterns. To do this, in Tables 15 and 16, it compares the two models employed to explore emigration and unaccounted non-response, with versions contrasting the use of listwise deletion with the version used in Section 3, above, utilising dummy values for missings.

Table 15: Logistic regressions of probabilities of emigration by 2001 controlling for background variables, alternative versions using listwise deletion and dummies in treatment of missing values

Explanatory variables	Listwise deletion	Dummies
	Coefficient (SE)	
Cohort (base is 1971)	**-.825 (.035)**	**-.826 (.035)**
Age group (base is aged 12-15 when first observed))		
Aged 4-7 when first observed	*-.068 (.037)*	*-.045 (.036)*
Aged 8-11 when first observed	**-.154 (.040)**	**-.148 (.039)**
Minority concentration of ward of origin (base is 0%)		
0-1%	**.235 (.060)**	**.234 (.058)**
1-5%	**.172 (.068)**	**.181 (.066)**
5-10%	*.075 (.094)*	*.061 (.092)*
More than 10%	*-.016 (.092)*	*-.037 (.090)*
If sample member born abroad	**.780 (.067)**	**.765 (.065)**
Parents' class (base is working)		
Service	**.191 (.048)**	**.197 (.046)**
Intermediate	*.008 (.049)*	*.008 (.048)*
Other	*-.052 (.097)*	*-.096 (.093)*
Male	**-.164 (.034)**	**-.159 (.033)**
Mother's qualifications (base no higher quals)		
No co-resident mother	*-.208 (.140)*	*-.221 (.138)*
Mother with higher qualifications	**.198 (.053)**	**.175 (.052)**
Father's qualifications (base no higher quals)		
No co-resident father	*.149 (.081)*	**.179 (.078)**
Father with higher qualifications	**.322 (.050)**	**.328 (.049)**
Proxy ethnic group (base is both parents British-born)		
Caribbean	**.325 (.142)**	**.341 (.134)**
Black African	**1.090 (.262)**	**1.138 (.249)**
Indian (and East African)	*.245 (.155)*	*.269 (.150)*
Pakistani	**.701 (.163)**	**.699 (.161)**
Bangladeshi	**.873 (.353)**	**.841 (.355)**
Chinese and Other	**1.019 (.137)**	**1.022 (.131)**
White migrant	**1.015 (.070)**	**1.007 (.067)**
housing tenure at origin (base is owner occupation)		
Local authority	**-.417 (.047)**	**-.430 (.046)**
Private rented	**.118 (.051)**	**.159 (.049)**
Car ownership at origin (base is none)		
1 car	**.154 (.045)**	**.174 (.044)**
2 or more cars	**.158 (.060)**	**.174 (.059)**
Constant	**-3.752 (.075)**	**-3.775 (.073)**
N	**172373**	**176907**
Wald chi2 (df)	**2053.4 (28)**	**2246.7 (31)**

Source: ONS Longitudinal Study, author's analysis

Notes: Coefficients in **bold** are significant at the 0.05 level; coefficients in *italics* are not significant at the 0.05 level. Bangladeshi proxy ethnicity is only available for 1981 cohort. Standard errors are adjusted for repeat observations on persons

Table 16: Logistic regressions of probabilities of unaccounted non-response at 2001 controlling for background variables, alternative versions using listwise deletion or dummies for missings in treatment of missing values

Explanatory variables	Listwise deletion	Dummies
	Coefficient (SE)	
Cohort (base is 1971)	**.263 (.013)**	**.261 (.013)**
Age group (base is aged 12-15 when first observed)		
Aged 4-7 when first observed	**.240 (.015)**	**.238 (.015)**
Aged 8-11 when first observed	**.114 (.016)**	**.116 (.016)**
Ward minority grp concentration at origin (base=0)		
0-1%	**.095 (.026)**	**.94 (.025)**
1-5%	**.161 (.029)**	**.156 (.028)**
5-10%	**.299 (.037)**	**.296 (.037)**
More than 10%	**.368 (.037)**	**.362 (.036)**
If sample member born abroad	**.547 (.034)**	**.548 (.034)**
Parents' class (base is working)		
Service	**.067 (.020)**	**.058 (.020)**
Intermediate	**.057 (.019)**	**.057 (.019)**
Other	**.239 (.032)**	**.234 (.031)**
Male	**.530 (.014)**	**.532 (.014)**
Mother's qualifications (base no higher quals)		
No co-resident mother	**.387 (.042)**	**.384 (.041)**
Mother with higher qualifications	**.073 (.026)**	**.069 (.026)**
Father's qualifications (base no higher quals)		
No co-resident father	**.175 (.029)**	**.177 (.028)**
Father with higher qualifications	**.088 (.024)**	**.088 (.024)**
Proxy ethnic group (base is both parents British-born)		
Caribbean	**1.104 (.048)**	**1.104 (.046)**
Black African	**1.274 (.132)**	**1.239 (.131)**
Indian (and East African)	**.340 (.057)**	**.339 (.056)**
Pakistani	**.641 (.070)**	**.636 (.069)**
Bangladeshi	**.408 (.153)**	**.405 (.152)**
Chinese and Other	**1.101 (.072)**	**1.105 (.070)**
White migrant	**.694 (.035)**	**.678 (.034)**
housing tenure at origin (base is owner occupation)		
Local authority	**.188 (.017)**	**.182 (.017)**
Private rented	**.239 (.023)**	**.235 (.022)**
Car ownership at origin (base is none)		
1 car	**-.131 (.017)**	**-.128 (.017)**
2 or more cars	*.041 (.024)*	*.039 (.023)*
Constant	**-2.426 (.032)**	**-2.419 (.032)**
N	**165946**	**170261**
Wald chi2 (df)	**5602.8 (28)**	**20997 (32)**

Source: ONS Longitudinal Study, author's analysis

Notes: Coefficients in **bold** are significant at at least the 0.05 level; coefficients in *italics* are not significant at the 0.05 level. Bangladeshi proxy ethnicity is only available for 1981 cohort. Standard errors are adjusted for repeat observations on persons

Across these two comparisons, there are very few differences. The tendency in the emigration models is for the dummies version to produce slightly bigger coefficients, but this is not true across all the variables, and it does not hold at all for the unexplained attrition models were the differences between the two are minimal and show no trend. Across both models there is only one coefficient that changes in terms of its statistical significance at conventional levels. This is the coefficient on having an absent father compared to having a father without higher qualifications (highlighted). One other variable, being in a privately tenanted household at origin compared with owner occupation (also highlighted), increases a certain amount between the two models and its statistical significance rises from the 0.05 level to the 0.001 level, but its direction, in terms of making emigration more likely remains and the change is not dramatic.

Table 17 goes on to examine whether the different treatment of missing values has a more substantial impact when measuring intergenerational mobility itself. This compares two models estimating the effects of various characteristics on the chances of ending up in the professional / managerial class at 2001. Once again the differences in coefficients are minimal where they occur at all.

Table 17: Logistic regressions of probabilities of ending up in the professional or managerial classes at 2001 controlling for respondents' own and background characteristics, alternative versions using listwise deletion or dummies for missings in treatment of missing values

Explanatory variables	**Listwise deletion**	**Dummies**
	Coefficient (SE)	
Cohort (baseline is 1971 cohort)	**-.232 (.014)**	**-.236 (.013)**
Age (baseline is aged 12-15 when first observed))		
Aged 4-7 when first observed	**-.190 (.015)**	**-.192 (.014)**
Aged 8-11 when first observed	**-.049 (.016)**	**-.049 (.015)**
Male	**.084 (.014)**	**.076 (.014)**
Partnered	**1.139 (.016)**	**1.135 (.016)**
Area concentration of minorities (baseline 0%)		
Up to 1%	**.188 (.024)**	**.191 (.023)**
1 to 5%	**.351 (.027)**	**.349 (.026)**
5-10%	**.264 (.039)**	**.268 (.037)**
More than 10%	**.260 (.039)**	**.266 (.038)**
Class of origin (baseline is working class)		
Service class	**.320 (.019)**	**.322 (.019)**
Intermediate	*.019 (.018)*	*.019 (.018)*
Other	**-.118 (.039)**	**-.096 (.036)**
Mother's qualifications		

(base no higher qualifications		
No co-resident mother	**-.126 (.052)**	**-.123 (.050)**
Mother with higher qualifications	**.117 (.027)**	**.115 (.027)**
Father's qualifications (base no higher qualifications		
No co-resident father	**.128 (.032)**	**.135 (.031)**
Father with higher qualifications	**.211 (.024)**	**.215 (.023)**
Tenure at origin (baseline is owner occupation)		
Local authority	**-.274 (.017)**	**-.278 (.016)**
Private rented	**-.161 (.024)**	**-.159 (.023)**
Car ownership at origin (baseline is no cars)		
1 car	**.185 (.017)**	**.173 (.017)**
2 or more cars	**.301 (.024)**	**.290 (.023)**
Ethnic group (baseline is white non-migrant)		
Caribbean	*-.095 (.078)*	*-.088 (.073)*
Black African	*.062 (.244)*	*.050 (.232)*
Indian	*.092 (.065)*	*.078 (.062)*
Pakistani	**-.897 (.102)**	**-.885 (.098)**
Bangladeshi	**-.525 (.231)**	**-.536 (.219)**
Chinese and other	*.141 (.107)*	*.095 (.103)*
White migrant	*.077 (.046)*	*.070 (.044)*
Sample member's qualifications (base is 0)		
Lower	**1.02 (.027)**	**1.02 (.026)**
Middle	**1.48 (.027)**	**1.48 (.027)**
Further	**2.78 (.028)**	**2.78 (.028)**
Car ownership at destination (base is 0)		
Constant	**-2.64 (.038)**	**-2.63 (.038)**
N	124619	134992

Source: ONS Longitudinal Study, author's analysis

Notes: Coefficients in **bold** are significant at at least the 0.05 level; coefficients in *italics* are not significant at the 0.05 level.
Standard errors are adjusted for repeat observations on persons

From this, admittedly brief and only indicative, comparison, it might be possible to infer that employing dummy variables for missing values is not such a dangerous strategy after all. This potentially particularly important for ethnically differentiated research since small sample sizes for many minority groups can mean that any loss of sample numbers can create difficulties for analysis.

Conclusion

This paper has raised three issues around missing data that have implications for mobility research and might be particularly pertinent for ethnically differentiated mobility research. In fact, despite big differences between groups in levels of non-class outcomes, emigration and non-response, the impact of these on results as far as it was investigated here was modest. This could be reassuring to those conducting mobility research based on retrospective designs, or attempting to retain sample sizes through the use of dummy variables. On the other hand, these conclusions remain tentative, given limitations in the data the analysis conducted here. Moreover, the use of Heckman models for testing selection effects is based on (untestable) assumptions which have been subject to some recent criticism. Alternative approaches might therefore be worth exploring. Similarly the treatment of missing values only explored two alternative treatments. Other options might be worth examining for their impact on reported results. The issue of missing data remains, then, an area for further investigation and one aim of this paper is simply to raise the issues and to introduce self-consciousness about how they are treated.

References

Allison, P. (2002) *Missing Data*, Thousand Oaks: Sage.

Blackwell, L., Lynch, K., Smith, J. and Goldblatt, P. (2003) *Longitudinal Study 1971-2001: Completeness of Census Linkage*, London: HMSO.

Bottero, W. and Prandy, K. (2000) 'Social reproduction and mobility in Britain and Ireland in the nineteenth and early twentieth centuries', *Sociology* 34: 265-281.

Erikson, R and Goldthorpe, J.H. (1993) *The Constant Flux: A Study of Class Mobility in Industrial Societies*, Oxford: Clarendon.

Francesconi, M. and Nicoletti, C. (2004) 'Intergenerational mobility and sample selection in short panels', Working Papers of the Institute for Social and Economic Research 2004-17, Colchester: University of Essex.

Goldthorpe, J.H. with Llewellyn, C. and Payne, C. (1987) *Social Mobility and Class Structure in Modern Britain.* Second Edition, Oxford: Clarendon Press.

Halsey, A.H., Health, A.F. and Ridge, J.M. (1980) *Origins and Destinations: Family, Class and Education in Modern Britain*, Oxford: Clarendon Press.

Heath, A. and McMahon, D. (1999) 'Ethnic differences in the labour market: the role of education and social class origins', CREST Working Paper 69.

Heath, A. and Payne, C. (2000) 'Social mobility' in A.H. Halsey with J. Webb (eds.) *Twentieth Century British Social Trends*. Basingstoke: Macmillan, pp. 254-278.

Heath, A. and Ridge, J. (1983) 'Social mobility of ethnic minorities', *Journal of Biosocial Science Supplement* 8: 169-84.

Jäckle, A. and Lynn, P. (2004) 'Dependent interviewing and seam effects in work history data', Working Papers of the Institute for Social and Economic Research 2004-24, Colchester: University of Essex.

Jacobs, S. (2002) 'Reliability and recall of unemployment events using retrospective data', *Work, Employment and Society* 16.3: 537-548.

Marshall, G., Swift, A. and Roberts, S. (1997) *Against the Odds? Social Class and Social Justice in Industrial Societies*, Oxford: Clarendon Press.

Miller, R. 1998. 'Unemployment as a mobility status', *Work, Employment and Society*, 12.4: 695-711.

Office for National Statistics (2003) 'Census One Number Census imputation rates by key variables'. Available at http://www.statistics.gov.uk/census2001/imputation_rates_by_variable.asp

Payne, G. and Roberts, J. (2002) 'Opening and closing the gates: Recent developments in male social mobility in Britain', *Sociological Research Online* 6.4.

Platt, L. (2005) 'The intergenerational social mobility of minority ethnic groups', *Sociology* 39.3.

Platt, L. (forthcoming 2005) *Migration and Mobility: Exploring the Impact of Origins and Achievement on the Life Chances of Britain's Ethnic Minorities*, Joseph Rowntree Foundation.

Prandy, K. (1998) 'Class and continuity in the analysis of social reproduction', *Sociological Review* 46: 154-178.

Savage, M. (2000) *Class Analysis and Social Transformation*, Buckingham: Open University Press.

Simpson, S (1996) 'Non-response to the 1991 Census: the effect on ethnic group enumeration', in D. Coleman and J. Salt (eds) *Ethnicity in the 1991 Census: Demographic Characteristics of the Ethnic Minority Population*, London: HMSO, pp.63-79.

Simpson, L and Middleton, E (1997) 'Who is missed by a national census? A review of empirical results from Australia, Britain, Canada, and the USA', Manchester: Cathie Marsh Centre for Census and Survey Research, University of Manchester. Available at http://www.ccsr.ac.uk/publications/working/missed.htm

Van de Ven, W.P.M.M. and Van Pragg, B.M.S. (1981) 'The demand for deductibles in private health insurance: a probit model with sample selection', *Journal of Econometrics* 17: 229-252.

Yamaguchi, K. 2003. 'Black-White differences in social mobility in the past 30 years: a multinomial logit latent-class regression analysis', unpublished paper: Universities of Chicago and Utrecht.