

## The Contribution of Residential Mobility to Sample Loss in a Birth Cohort Study: Evidence from the First Two Waves of the UK Millennium Cohort Study

Ian Plewis<sup>1</sup>, Sosthenes C. Ketende<sup>2</sup>, Heather Joshi<sup>2</sup>, and Gareth Hughes<sup>3</sup>

Longitudinal surveys gain measurement occasions but lose cases over time. Knowing more about reasons for sample loss improves the chances of being able adequately to adjust for it. This article shows that residential mobility is an important predictor of sample loss over the first two waves of the most recent UK birth cohort study, the Millennium Cohort Study. The measure of residential mobility used takes account of moves after wave one using the information available from the administrative side of the survey operation. We find that the residentially mobile are more likely to be nonrespondents even after controlling for a range of background variables. Mobile households are, however, somewhat less likely to refuse than the nonmobile and mobility is unrelated to noncontact, provided the sample cohort member has been located. The implications of the results are discussed in terms of the assumption of data being “missing at random” in longitudinal surveys.

**Key words:** Longitudinal surveys; attrition; nonresponse; moving home; logistic regression; multilevel models; UK Millennium Cohort.

### 1. Introduction

Repeatedly measuring cases – the defining characteristic of longitudinal survey designs – generates the data that are needed to describe and explain change. It does, of course, take time to gather these repeated measurements and it is almost inevitable that a proportion of the selected cases will not be measured on every chosen occasion. Some cases are never measured either because they cannot be contacted at a known address or because they choose not to cooperate from the outset. This is the problem of unit nonresponse that affects both cross-sectional and longitudinal designs. In addition to the continuing problems of noncontact and noncooperation, cases are lost from longitudinal surveys because they cannot be located or traced. Cases can be measured on some occasions in longitudinal surveys but not on others and they are often known as wave nonrespondents. Others – often referred to as attrition cases – are part of the study initially but are, sooner

<sup>1</sup> University of Manchester, Centre for Census and Survey Research, Bridgeford Street, Manchester M13 9PL, UK. Email: ian.plewis@manchester.ac.uk

<sup>2</sup> University of London, Institute of Education, Centre for Longitudinal Studies, 20 Bedford Way, London WC1H 0AL, UK. Emails: s.ketende@ioe.ac.uk and h.joshi@ioe.ac.uk

<sup>3</sup> County Hall, Beverley HU17 9BA, UK. Email: gareth.hughes@eastridings.gov.uk

**Acknowledgments:** An earlier version of this article was presented at the Methodology of Longitudinal Surveys conference (MOLS2006) in July 2006. We would like to thank Lisa Calderwood for useful comments on a previous draft.

or later, lost never to return even though, in principle, they could be remeasured. The concern of all who analyse longitudinal data is that those who are lost from the study, either temporarily or permanently, are likely to be different from those that remain and therefore inferences based on the observed sample may differ from those that would have been made if there had been no sample loss.

The UK is exceptionally fortunate in having a series of four national longitudinal birth cohort studies, spanning the period from 1946 to 2000. These studies are following children born in 1946 (National Survey of Health and Development), 1958 (National Child Development Study or NCDS), 1970 (British Cohort Study or BCS70) and 2000/1 (Millennium Cohort Study or MCS). Details about the first three cohorts can be found in Ferri et al. (2003) and about the fourth in Smith and Joshi (2002); Plewis (2007a). This article is based on data from two waves of MCS, one when the cohort child was aged nine months and the second at age three years.

The feature of this article is its focus on the contribution of residential mobility (i.e., any change of postal address) to overall nonresponse after the first wave, and to noncooperation (i.e., refusal) and noncontact separately. (It is not possible to separate wave nonresponse from attrition after just two waves of data collection.) We use the information on mobility from the respondents to Wave 2 of the study and we also draw on data available from the survey's address database for all the measured cases at Wave 1. In other words, we are able to assess whether knowing that someone changed address between the first two waves of the study helps us to improve our predictions of nonresponse at Wave 2 beyond those that we can make just by using our knowledge of the characteristics of cases measured at Wave 1.

Lepkowski and Couper (2002) analysed data from two U.S. longitudinal studies to show that residential mobility up to Wave 1 is associated with the probability of being located or traced at Wave 2 and, for one of the studies, with the probability of cooperating at Wave 2 (after controlling for a range of other variables). Hawkes and Plewis (2006) show that family mobility is related to attrition and, to an extent, wave nonresponse in a birth cohort study (NCDS), again after controlling for other variables. However, neither Hawkes and Plewis nor Lepkowski and Couper were able to establish whether residential mobility *after* Wave  $t-1$  is related to nonresponse at Wave  $t$ .

Clearly, we would expect residential mobility after Wave  $t-1$  to be related to the probability of being a nonrespondent at Wave  $t$  merely because mobile families can be difficult to trace and locate. What is less easy to predict is the relation between mobility, and noncontact and refusal separately. The circumstances that lead families to change address might also make them more difficult to contact whereas the characteristics of families that lead them to report changes of address to the survey team might also be those that incline them to cooperate. We are interested more in predicting nonresponse than in explaining it and so the quality of the prediction will depend on the predictor variables available from the previous wave or waves, some of which will be correlated with residential mobility. Hence, we might find that although residential mobility is related to different kinds of nonresponse it is not an important predictor after controlling for other variables. It is this question – the contribution of residential mobility to nonresponse before and after controlling for other variables – which we attempt to answer here.

The period in the life course covered in this article is one where moving home is relatively common. This phenomenon was documented by Grundy and Fox (1985) for the 1970s. Indeed the Census of England and Wales in 2001 (Standard Table 8) found that infants (or their next of kin) were more likely to have moved in the year before the census than individuals in any other age group except people in their twenties<sup>4</sup>. Children aged one to five years and adults in their thirties (at an age likely to be their parents) also had relatively high rates of movement recorded by the census. Analysing by household, we can see that the households with children under five were the most likely household type to have moved together in the 12 months prior to the census, particularly if there was only one child (Census of England and Wales Standard Table 9). This mobility poses a major challenge for fieldwork in birth cohort studies.

Finding, as we do, that residential mobility after Wave 1 is indeed an important predictor of nonresponse at Wave 2 has implications for the ways in which we might adjust for nonresponse in longitudinal surveys as it casts some doubt on the common assumption that missing longitudinal data are, to use Little and Rubin's (2002) term, "missing at random." We return to these issues in the final section. In the next section, we describe the design of MCS and how the sample has declined between its first two waves. In Section 3, we compare our sources of information about residential mobility and present our methods of analysis. In Section 4 we describe the correlates of nonresponse and consider whether these are also related to mobility and we then bring these findings together by building binary and multinomial logistic regression models for (i) overall nonresponse, (ii) refusals and all noncontacts separately and (iii) refusals and located noncontacts separately.

## 2. The Millennium Cohort Study (MCS)

The MCS population is that of children born between 1st September 2000 and 31st August 2001 (for England and Wales), and between 24th November 2000 and 11th January 2002 (for Scotland and Northern Ireland), alive and living in the UK at age nine months, eligible to receive Child Benefit at that age;<sup>5</sup> and then, after nine months, for as long as they remain in the UK at the time of sampling. The sample is disproportionately stratified and clustered; the smaller UK countries are over-represented as compared with England, and families living in more disadvantaged areas and in areas with high proportions of ethnic minority families are also over-represented. The productive (i.e., interviewed) sample at Wave 1 comprises 18,552 families clustered in 398 electoral wards across the nine UK strata (advantaged and disadvantaged wards in each of the four UK countries plus an ethnic minority stratum in England): the response rate at Wave 1 is 72%. Further details can be found in Plewis (2007a). Correlates of unit nonresponse in MCS – generated from auxiliary data held in the administrative records used to select the sample – are presented

<sup>4</sup> [https://www.nomisweb.co.uk/query/construct/submit.asp?menuopt=201&subcomp=&Session\\_GUID={CE4D3A81-513E-41A7-91BB-D2D302841AC4}](https://www.nomisweb.co.uk/query/construct/submit.asp?menuopt=201&subcomp=&Session_GUID={CE4D3A81-513E-41A7-91BB-D2D302841AC4})

<sup>5</sup> Child Benefit is a universal benefit available to all dependent children who are permanent residents of the UK. Children of nonresident parents, e.g., diplomats, overseas students and visitors, are not eligible, nor those of very recent international migrants and asylum seekers whose residence status has not yet been determined. These considerations are thought to account for most of the approximately three percent discrepancy between registered births and Child Benefit claims. There are also an unknown but small number of families who do not claim even though eligible.

in a companion paper (Plewis 2007b). That paper shows that movers into the sampled clusters were under-represented in the first wave.

The productive sample at Wave 1, i.e., with interview data from at least a main<sup>6</sup> respondent or partner, formed the issued sample at Wave 2 apart from cases known to have died or emigrated. In other words, Wave 1 refusals were not reissued. Substantial resources were, however, devoted both to ensuring further cooperation and to tracing cases that had changed address (see Moon 2006 for more detail). The UK response rate at Wave 2 (at age three years) was 79%. The size of the productive sample was 15,590 although this figure includes an extra 692 families in England who became part of the sample for the first time at Wave 2. These “new families” (whose addresses, though eligible, were identified too late for the first survey) are not included in the analyses presented here (see Plewis and Ketende 2006 for more details).

### **3. Methodology**

#### *3.1. Identification of Residentially Mobile Families*

Families who were mobile between Waves 1 and 2 can be identified in two ways. The first is by using the fact that the main respondents in Wave 2 were asked whether or not they had changed address since the last survey. This measure has three disadvantages:

- i) It is based only on those cases that were productive<sup>7</sup> at Wave 2.
- ii) Further information is lost because of item nonresponse (1.8%) to the question about address changes.
- iii) Recall bias might be a problem especially for moves that took place around Wave 1 (over two years before).

The second approach is to use data collected by the survey administration team for tracing purposes. This database of addresses of the cohort families is regularly updated with new addresses whenever a move of a cohort family is reported. The move might be reported by the cohort family themselves but could also come from administrative sources such as Child Benefit records. Using this database allows us to capture most moves, regardless of the family’s participation and reporting at the second wave, thereby avoiding loss of information about mobility. It also gives us more data for longitudinal analysis. One weakness of the address database is that the data on number of moves between these two waves cannot be determined accurately.

Table 1 presents the mobility rates by UK country for the two measures of mobility. We see that the second estimate, based on all eligible cases from Wave 1, is nine percentage points higher than the first, which is based just on the respondents to the survey question at Wave 2. The differences are fairly consistent across the four UK countries. All families that reported a move at Wave 2 were registered as doing so on the address database. Table 1 shows, however, that 7% of all productive families at Wave 2 did not report a

<sup>6</sup>The main respondent is almost always, but not necessarily, the child’s mother.

<sup>7</sup>A productive case is one that had some data from at least one instrument (main or partner interviews, anthropometry or cognitive assessments) other than data carried forward from Wave 1.

Table 1. Mobility rates by UK country for the two measures of mobility

UK country (Wave 1)	Main respondents self-reporting		Address database			
			Eligible		Productive	
	% Mobile	Base	% Mobile	Base	% Mobile	Base
England	31	9,172	38	11,426	36	9,172
Wales	26	2,228	35	2,744	33	2,228
Scotland	30	1,796	41	2,303	38	1,796
Northern Ireland	25	1,440	33	1,912	33	1,440
Total	29	14,636	38	18,385	36	14,636

Notes: Mobile = different address at first two waves of MCS. Weighted percentage of mobile families; observed sample numbers. Base for productive cases at Wave 2: those with a response to the survey question on moving home; 262 are missing. Base for address database: the eligible column is all eligible for Wave 2 ( $n = 18,552$ ) less deaths and emigrations ( $n = 167$ ).

move which was recorded in the address database. We can reasonably suppose, based on the disadvantages of the survey measure, that the measure of mobility generated from the database is the more accurate of the two, and we use it throughout the rest of the article. It is, however, just possible that some moves recorded on the address database arose because of changes in street names, postcodes etc. rather than changes of dwelling. We have no way of establishing whether this happened but the number of such changes is likely to be very small.

### 3.2. Estimation of the Distance of the Move

It is possible that the association between residential mobility and nonresponse depends on the distance moved between interviews at Waves 1 and 2, and this distance was calculated from the postcodes on the address database. The UK Office for National Statistics postcode lookup files gives the coordinates ("Eastings" and "Northings") of the centroids of each postcode. The distance was obtained as the squared root of the sum of the squared difference between the Eastings and the squared difference between the Northings.

Distance (kms) =

$$\frac{\sqrt{(\text{MCS2Eastings} - \text{MCS1Eastings})^2 + (\text{MCS2Northings} - \text{MCS1Northings})^2}}{1000}$$

The distance calculated is a straight line distance, as the crow flies, generally, though not exactly a guide to how far the family travelled on the land. Over half of the moves were two kilometres or less. Table 2 shows the distribution of mobile families in each UK country at Wave 1 by distance moved. The shortest distances are most common in Wales. In the other countries up to one fifth have recorded distances of more than 10 kilometres. The majority of the nine percent of cases in Table 2 where the distance is not known are untraced movers, i.e., those known to have moved but whose destination cannot be established. All these cases are unproductive.

*Table 2. Distance moved (Km) by mobile families by UK country*

Distance moved (km)	UK country (Wave 1)			
	England %	Wales %	Scotland %	Northern Ireland%
<1	29	40	37	33
1, 2	22	18	16	20
3–10	20	20	19	20
>10	19	12	20	18
Unknown	9.2	9.5	9.0	9.2
Total (N)	100 (4,432)	100 (996)	100 (964)	100 (640)

Notes: Distances are straight lines between postcode centroids. The 731 (9%) of families whose distance of move was not known are a combination of 687 untraced movers whose address at Wave 2 is not known and 44 whose address is known but for whom the information is inadequate for calculating distance.

Among the Wave 2 productive families reporting any moves, 78% of the self-reported movers changed homes only once, 16% moved twice and the remaining 6% moved more than twice.

### 3.3. Statistical Analysis

All the analyses are based on the sample of cohort families that were productive at Wave 1 and were eligible for Wave 2 (see Note 4 to Table 1). Two response variables were used:

- i) A binary variable: productive or not at Wave 2.
- ii) A variable with three categories which splits the nonrespondents into refusals and noncontacts. The noncontact category includes untraced movers in the first analysis but excludes them from the second.

Various socio-demographic and socio-economic explanatory variables were included in the analyses along with the measure of residential mobility we derived from the address database. All the explanatory variables except the distance of the move were measured at Wave 1 and they were grouped as follows:

- i) UK country and stratum (ward type).
- ii) Socio-demographic variables – ethnic group of the cohort member, number of parents/carers in the household, main respondent's age, main respondent's highest academic qualifications, number of people in the household, number of cohort member's siblings, distance of the move.
- iii) Socio-economic variables – family income, whether the family receives means-tested benefits, combined main respondent and partner labour market status, housing tenure, type of accommodation, number of rooms.
- iv) Two ward-based measures: the Child Poverty Index and an indicator of "rurality."

Logistic regression was used for the binary response, multinomial regression for the response with three categories. Because the sample design is complex, analyses were conducted in Stata 9.2 using Stata's survey commands (with appropriate sampling weights to adjust for over-sampling in disadvantaged and ethnic minority wards). In addition, a two-level model (households within wards) was fitted to establish whether the two

different types of nonresponse varied across the electoral wards sampled at Wave 1 after controlling for the effects of all the explanatory variables of importance. This model is:

$$\log \pi_{ijk}/\pi_{ijK} = b_{0jk} + \sum_{l=1}^L b_{lk}x_{lij}$$

$$b_{0jk} = b_{00k} + u_{0jk}$$

where  $\pi_{ijk}$  is the probability of not being contacted ( $k = 1$ ) or refusing ( $k = 2$ ) and  $\pi_{ijK}$  is the probability of responding (the reference category);  $i = 1 \dots n_j$  are households within wards  $j$  ( $j = 1 \dots J$ );  $x_l$  are the  $L$  explanatory variables and  $\mathbf{u}$  (the vector of level-two residuals) has a multivariate normal distribution with mean =  $\mathbf{0}$ . The *MLwiN* package (Rasbash et al. 2004) was used to fit this model.

## 4. Results

### 4.1. Mobility and Nonresponse

As already shown in Table 1, during the 27-month period between Waves 1 and 2, 38% of eligible families moved home at least once. This proportion varied by UK country and stratum (Table 3): families living in the more advantaged wards in Northern Ireland had the lowest percentage mobile (30%) and families living in the disadvantaged wards in Scotland the highest (47%). The overall proportion of eligible families who were productive at Wave 2 was 0.84 (95% CI: 0.83–0.85). Mobile families were more likely to be unproductive at Wave 2: the estimated unadjusted odds ratio was 1.47 (95% CI: 1.33–1.61).

The key question, however, is whether mobility between Waves 1 and 2 is still associated with nonresponse *at Wave 2* after allowing for explanatory variables measured *at Wave 1* that are associated both with nonresponse and with mobility. To start with, all the variables listed in Section 3.3 were included, one by one, in logistic regression models that also included mobility. The estimates of the adjusted odds ratios from these separate analyses are reported in Tables 4–7 and summarised in Section 4.1. All the estimated odds ratios are, from Wald tests, statistically significant at the 0.05 level unless otherwise stated.

Table 3. Mobility by UK country and stratum

UK country (Wave1)	Stratum						Total	
	Not disadvantaged		Disadvantaged		Minority ethnic		%	N
	%	N	%	N	%	N		
England	36	4,555	43	4,496	37	2,375	38	11,426
Wales	32	824	38	1,920	NA	NA	35	2,744
Scotland	36	1,123	47	1,180	NA	NA	41	2,303
Northern Ireland	30	718	36	1,194	NA	NA	33	1,912
Total	35	7,220	42	8,790	37	2,375	38	18,385

Note: Weighted percentages; observed sample numbers.

Table 4. Adjusted odds ratios of being unproductive: UK country, ward type and mobility

Variables		OR	(95% CI)
Mobility	Nonmover (ref.)	1	
	Mover	1.48	(1.34–1.62)
UK country <sup>a</sup>	England (ref.)	1	
	Wales	1.13	(0.99–1.29)
	Scotland	1.36	(1.15–1.61)
	Northern Ireland	1.58	(1.34–1.85)
Mobility	Nonmover (ref.)	1	
	Mover	1.43	(1.30–1.57)
Ward type	Not disadvantaged (ref.)	1	
	Disadvantaged	1.70	(1.49–1.95)
	Minority ethnic	2.30	(1.91–2.76)

Note: <sup>a</sup>Interaction between UK country and mobility ( $p < .001$ ) – see text.

The introduction of each of these additional variables does change the estimated coefficients for mobility, but not substantially, so mobility remains an important and statistically significant predictor throughout.

#### 4.1.1. UK Country and Ward Type

- i) Families in Northern Ireland and Scotland were more likely to be unproductive than families in England.
- ii) Families from disadvantaged and ethnic minority wards were more likely to be unproductive than families in more advantaged wards.

The adjusted odds ratios for mobility are little changed from the unadjusted estimate of 1.47 when UK country and ward type are controlled, being 1.48 and 1.43 respectively (Table 4). There was, however, a statistically significant interaction between mobility and UK country, with mobile families in Northern Ireland being more likely to be productive than mobile families in England (estimates not shown).

#### 4.1.2. Socio-demographic Variables

- i) Families where the cohort member was nonwhite were less likely to be productive.
- ii) Single parent families were more likely to be unproductive than families with two parents/carers.
- iii) The likelihood of being unproductive decreased with the main respondent's age.
- iv) Families whose main respondents had higher-level educational qualifications were less likely to be unproductive than those with lower qualifications or with no qualifications at all.
- v) There was some suggestion that households containing four people were more likely to be productive, and households with six or more people (and where there were four or more children; data not shown) were less likely to be productive than households with two or three people.
- vi) Excluding the untraced movers, households who had moved more than 10 km between waves were more likely to be productive.

The adjusted odds ratios for mobility are little changed from the unadjusted estimate after the inclusion of variables (i) to (v) in the model. The most substantial effect (from the unadjusted estimate of 1.47 down to 1.31) occurs after controlling for the main respondent's age (Table 5).

*Table 5. Adjusted odds ratios of being unproductive: socio-demographic variables and mobility*

Variables		OR	(95% CI)
Mobility	Nonmover (ref.)	1	
	Mover	1.49	(1.36–1.64)
Ethnicity, main respondent	White (ref.)	1	
	Mixed	2.28	(1.53–3.12)
	Indian	1.23	(0.88–1.70)
	Pakistani/Bangladeshi	1.79	(1.46–2.19)
	Black/Black British	2.31	(1.85–2.90)
	Other	2.37	(1.70–3.30)
Mobility	Nonmover (ref.)	1	
	Mover	1.38	(1.25–1.52)
Household type	2 parents (ref.)	1	
	1 parent	2.21	(1.97–2.48)
Mobility	Nonmover (ref.)	1	
	Mover	1.31	(1.19–1.44)
Main respondent's age	14–19	1.39	(1.18–1.65)
	20–29 (ref.)	1	
	30–39	0.51	(0.47–0.57)
	40+	0.56	(0.43–0.74)
Mobility	Nonmover (ref.)	1	
	Mover	1.44	(1.32–1.58)
Education, main respondent	Higher degree	0.56	(0.43–0.72)
	First degree	0.46	(0.38–0.56)
	Diplomas in Higher Education	0.58	(0.48–0.71)
	A/AS/S levels	0.71	(0.60–0.84)
	O-Level/GCSE grades A-C (ref.)	1	
	GCSE grades D-G	1.16	(0.98–1.38)
	Other academic qualification (including overseas)	1.75	(1.36–2.25)
Mobility	None of the above	1.89	(1.67–2.15)
	Nonmover (ref.)	1	
	Mover	1.45	(1.32–1.60)
Household size	2/3 (ref.)	1	
	4	0.85	(0.76–0.95)
	5	0.91	(0.80–1.04)
	6+	1.18	(1.00–1.38)
Distance moved (km)	Nonmover (ref.)	1	
	<1	1.01	(0.83–1.24)
	1, 2	0.84	(0.71–1.00)
	3–10	0.96	(0.80–1.15)
	>10	0.64	(0.51–0.81)

Table 6. Adjusted odds ratios of being unproductive: socio-economic variables and mobility

Variables		OR	(95% CI)
Mobility	Nonmover (ref.)	1	
	Mover	1.36	(1.24–1.49)
Family annual income	£0–£3,100	1.33	(0.99–1.78)
	£3,100–£10,400	1.72	(1.53–1.93)
	£10,400–£20,800 (ref.)	1	
	£20,800–£31,200	0.58	(0.50–0.68)
	£31,200–£52,000	0.45	(0.38–0.54)
	£52,000 +	0.47	(0.36–0.63)
	Don't know	1.24	(1.01–1.51)
	Refused	1.74	(1.31–2.30)
Mobility	Nonmover (ref.)	1	
	Mover	1.39	(1.26–1.52)
On benefits	No (ref.)	1	
	Yes	2.03	(1.83–2.25)
Mobility	Nonmover (ref.)	1	
	Mover	1.34	(1.22–1.47)
Labour market participation	Main & partner employed (ref.)	1	
	Partner employed, main not	1.39	(1.24–1.56)
	Partner not, main emp. p/t	1.70	(1.31–2.21)
	Partner not, main emp. f/t	2.60	(2.22–3.05)
	No partner, main employed	2.13	(1.71–2.65)
	No partner, main not emp	3.03	(2.67–3.45)
Mobility	Nonmover (ref.)	1	
	Mover	1.25	(1.13–1.38)
Tenure <sup>a</sup>	Own (ref.)	1	
	Rent	2.44	(2.19–2.72)
	Other	2.07	(1.69–2.54)
Mobility	Nonmover (ref.)	1	
	Mover	1.33	(1.20–1.47)
Accommodation type <sup>b</sup>	House/Bungalow (ref.)	1	
	Flat/maisonette	2.11	(1.86–2.39)
	Studio/Bed-sit etc	2.49	(1.63–3.80)
Mobility	Nonmover (ref.)	1	
	Mover	1.27	(1.14–1.41)
Number of rooms <sup>c</sup>	< = 3	2.04	(1.71–2.44)
	4	1.46	(1.29–1.66)
	5 (ref.)	1	
	6	0.77	(0.68–0.88)
	7 +	0.63	(0.53–0.74)

Notes: <sup>a</sup>Interaction between tenure and mobility ( $p < .001$ ) – see text. <sup>b</sup>Interaction between accommodation type and mobility ( $p < .001$ ) – see text. <sup>c</sup>Interaction between number of rooms and mobility ( $p < .05$ ) – see text.

#### 4.1.3. Socio-economic Variables

- i) Low income families, those who refused to answer income questions and families who did not know their annual income were more likely to be unproductive than high income families.

Table 7. Adjusted odds ratios of being unproductive: ward-based variables and mobility

Variables		OR	(95% CI)
Mobility	Nonmover (ref.)	1	
	Mover	1.41	(1.28–1.55)
Child Poverty Index		1.02	(1.015–1.021)
	Nonmover (ref.)	1	
Mobility	Mover	1.50	(1.34–1.67)
	Urban (ref.)	1	
Rurality	Market town	0.67	(0.51–0.88)
	Rural town	0.83	(0.67–1.05)
	Village	0.50	(0.37–0.67)

- ii) Families receiving one or more means-tested benefits were more likely to be unproductive than families not receiving any of these benefits.
- iii) Families with single mothers or partners not in paid work were more likely to be unproductive.
- iv) Homeowners were more likely to be productive than renters and the small number in other kinds of housing tenure.
- v) Families living in a house (or bungalow) were more likely to be productive than families living in other types of housing.
- vi) Families with five rooms or fewer in their home were more likely to be unproductive than those with six rooms or more.

Again, the adjusted odds ratios for mobility are little changed from the unadjusted estimate after the inclusion of these six variables in the model. There were, however, statistically significant interactions between mobility and tenure, type of housing and number of rooms. Mobile families in “other” kinds of tenure arrangements (for example, living with parents or relatives), living in flats/maisonettes and in dwellings with small numbers of rooms were more likely to be productive than their nonmobile equivalents.

#### 4.1.4. Ward-based Variables

- i) A one-unit increase in the Child Poverty Index<sup>8</sup> was associated with an increase in the likelihood of being unproductive.
- ii) Families in urban areas were more likely to be unproductive than those in rural areas.

#### 4.2. Final Models

All the explanatory variables that were individually important were entered into the full model but not all were needed for the final model. We also experimented with various ways of including the distance of the move as a predictor in the model. We concluded that the best mobility indicator was the dichotomy between movers and nonmovers which we present. Terms reflecting the distance moved added very little to

<sup>8</sup> i.e., one point in the percentage of the population aged 0–15 years in families in the ward receiving means-tested benefits in 1998/99.

the explanatory power. There was, as we have seen in Table 5, a slight tendency for longer moves to be associated with higher response, which could have reflected better reporting of moves by those who intended to respond. The detection of short-distance moves relies less on families telling the survey about their new address. We should further point out that people intending to drop out of the survey, and therefore not notifying their new address, are a broader group than the untraced movers because the survey is not wholly reliant on moving families volunteering information about their address, although this helps. The tracing operation also uses other sources, such as the Department for Work and Pensions (DWP) records. Equally, there are other ways to decline to participate than withholding information on a change of address.

#### 4.2.1. Binary Response Models

Two versions of the final model were fitted, one that includes mobility and the other not (Table 8). In addition to the model estimates, Table 8 gives the total sample observed, also weighted percentages of those families which were mobile and which were productive at Wave 2 by the categories of variables in the final model.

The adjusted estimate of the odds ratio of being unproductive for a mobile family was 1.41 (95% CI: 1.19–1.66). This is not very much less than the unadjusted estimate of 1.47 given in Section 4.1. The main difference between Model 1 and Model 2 in Table 8 is for variables whose association with being unproductive was modified by mobility (i.e., where there was an interaction). These variables were UK country (in particular Northern Ireland), tenure and the type of accommodation. The average odds ratio of being unproductive for Northern Ireland families compared with families in England was 1.67 when mobility was ignored but it increased to 2.06 when mobility was accounted for. However, mobile families in Northern Ireland were less likely to be unproductive than mobile families in England, the estimate of the odds ratio being 0.55 (95% CI: 0.44–0.69) (estimates not shown). Similarly, the average estimated odds ratio of being unproductive changes from 1.42 for families living in flats/maisonettes and studios/bed-sits when mobility was not controlled for to 1.77 when it was, as compared with families that were living in houses or bungalows. Mobile families living in flats/maisonettes and studios/bed-sits were, however, less likely to be unproductive than mobile families that had been living in houses/bungalows, the estimated odds ratio being 0.62 (95% CI: 0.49–0.79) (estimates not shown). The pattern for tenure was similar: families who were neither owner occupiers nor renters (5.5%) were more likely to be productive if mobile than their nonmobile counterparts: the estimated odds ratio is 0.46 (95% CI: 0.33–0.66) (estimates not shown).

#### 4.2.2. Multinomial Response Models

Table 9 shows results from a multinomial logistic regression using the same explanatory variables (including the interactions with mobility) but where the response variable has three outcome categories, i.e., being productive (the reference category), refusal and other unproductive (noncontact, untraced movers, other nonresponse, ill/incapacitated, loss of computer data in the field, language problems).

Table 8. Adjusted odds ratios of being unproductive including/not including mobility

Variables		% mobile Wave 2	% productive Wave 2	Base (N)	Model 1 (Mobility not included)		Model 2 (Mobility included)	
					OR	(95% CI)	Wald test	OR
Mobility	Nonmover (ref.)	NA	86	11,353	NA	NA	NA	1 <0.001
	Mover	NA	80	7,032	NA	NA	1.41 (1.19–1.66)	
UK country	England (ref.)	38	84	11,426	1	<0.001	1 (2)	
	Wales	35	83	2,744	1.14 (1.00–1.30)		1.24 (1.04–1.48)	
Family annual income	Scotland	41	80	2,303	1.41 (1.18–1.68)		1.41 (1.11–1.78)	
	Northern Ireland	33	78	1,912	1.67 (1.41–1.98)		2.06 (1.70–2.50)	
	£0–£3,100	55	77	300	1.12 (0.81–1.55)	<0.001	1.13 (0.81–1.56)	<0.001
	£3,100–£10,400	48	73	4,401	1.22 (1.08–1.38)		1.23 (1.09–1.39)	
	£10,400–£20,800 (ref.)	38	83	5,590	1		1	
	£20,800–£31,200	34	89	3,275	0.79 (0.67–0.94)		0.79 (0.67–0.94)	
	£31,200–£52,000	34	91	2,377	0.74 (0.61–0.89)		0.72 (0.60–0.88)	
	£52,000 +	38	91	847	0.92 (0.69–1.22)		0.90 (0.67–1.19)	
	Don't know	35	80	1,083	1.18 (0.96–1.45)		1.17 (0.95–1.44)	
	Refused	35	74	476	1.70 (1.23–2.34)		1.71 (1.24–2.34)	
Ethnicity, cohort member	White (ref.)	38	85	15,147	1	<0.001	1 <0.001	
	Mixed	43	79	543	1.35 (1.04–1.75)		1.35 (1.04–1.75)	
	Indian	31	82	464	1.20 (0.85–1.68)		1.18 (0.84–1.66)	
	Pakistani/Bangladeshi	31	76	1,263	1.16 (0.92–1.46)		1.18 (0.93–1.49)	
	Black/ Black British	39	70	663	1.78 (1.45–2.19)		1.81 (1.48–2.21)	
	Other	40	69	258	2.26 (1.49–3.43)		2.26 (1.47–3.47)	
Tenure	Own (ref.)	31	88	10,603	1	<0.001	1 (3)	
	Rent	48	75	6,558	1.29 (1.13–1.47)		1.17 (1.02–1.34)	
	Other	63	77	1,166	1.21 (0.97–1.50)		1.78 (1.31–2.42)	

Table 8. *Continued*

Variables		% mobile Wave 2	% productive Wave 2	Base (N)	Model 1 (Mobility not included)		Model 2 (Mobility included)	
					OR	(95% CI)	Wald test	OR
Accommodation type	House/Bungalow (ref.)	35	85	15,587	1		<0.001	1
	Flat/maisonette	59	72	2,650	1.42	(1.25–1.61)		1.77
	Studio/Bed-sit, etc	73	68	104				(1.47–2.14)
Main respondent's age	14–19 years (ref.)	62	72	1,061	1		<0.001	1
	20–29 years (ref.)	44	79	8,159				<0.001
	30–39 years	31	88	8,531	0.68	(0.61–0.76)		0.70
	40+ years	27	87	624				(0.63–0.78)
Education, main respondent	Higher degree	35	90	598	0.74	(0.56–0.98)	<0.001	0.73
	First degree	35	92	2,247	0.63	(0.52–0.76)		0.62
	Diplomas in higher education	38	90	1,540	0.69	(0.57–0.85)		0.68
	A/AS/S levels	38	87	1,701	0.77	(0.65–0.92)		0.76
	OLevel/ GCSE grades A–C (ref.)	38	83	6,144	1			1
	GCSE grades D–G (ref.)	40	81	1,976				
	Other academic qualification (including overseas)	35	74	521	1.24	(0.96–1.60)		1.25
	None of the above	40	72	3,592	1.34	(1.18–1.52)		1.35
CPI	Child Poverty Index (continuous variable)	NA	NA	NA	1.004	(1.000–1.007)	<0.04	1.003
								(1.000–1.007) <0.05

Notes: 1. Weighted percentages; observed sample numbers. 2. Interacts with mobility ( $p < .001$ ). 3. Interacts with mobility ( $p < .001$ ). 4. Interacts with mobility ( $p < .001$ ).

5. Fit statistics: Model 1:  $F(26, 364) = 35.8, p < .001, n = 18269$ . Model 2:  $F(33, 357) = 30.1, p < .001, n = 18269$ .

Table 9. Relative risks of being unproductive including untraced movers

Variables		Refusal vs. Productives		Other unproductives vs. Productives	
		RR	(95% CI)	RR	(95% CI)
Mobility	Nonmover (ref.)	1		1	
	Mover	0.79	(0.63–1.01)	3.28	(2.58–4.18)
UK country	England (ref.)	1		1	
	Wales	1.11	(0.90–1.38)	1.57	(1.17–2.12)
	Scotland	1.26	(1.00–1.59)	1.69	(1.15–2.49)
	Northern Ireland	2.28	(1.88–2.77)	1.44	(0.96–2.16)
Family annual income	£0–£3,100	1.17	(0.74–1.84)	1.06	(0.70–1.60)
	£3,100–£10,400	1.32	(1.10–1.58)	1.13	(0.96–1.32)
	£10,400–£20,800 (ref.)	1		1	
	£20,800–£31,200	0.77	(0.62–0.94)	0.82	(0.64–1.03)
	£31,200–£52,000	0.67	(0.51–0.88)	0.79	(0.60–1.05)
	£52,000 +	0.87	(0.62–1.22)	0.91	(0.58–1.43)
	Don't know	1.12	(0.85–1.48)	1.22	(0.94–1.59)
	Refused	2.09	(1.49–2.93)	1.19	(0.75–1.89)
Ethnicity, cohort member	White (ref.)	1		1	
	Mixed	0.93	(0.62–1.38)	1.80	(1.26–2.58)
	Indian	1.36	(0.91–2.02)	0.98	(0.63–1.53)
	Pakistani & Bangladeshi	1.20	(0.87–1.64)	1.19	(0.86–1.63)
	Black/ Black British	0.99	(0.70–1.41)	2.61	(2.06–3.30)
	Other	2.18	(1.13–4.22)	2.38	(1.51–3.76)
Tenure	Own (ref.)	1		1	
	Rent	1.14	(0.98–1.34)	1.35	(1.06–1.72)
	Other	2.11	(1.51–2.95)	1.17	(0.72–1.90)
Accommodation type	House/Bungalow (ref.)	1		1	
	Flat/maisonette Studio/Bed-sit, etc	1.18	(0.92–1.52)	3.11	(2.42–4.00)
Main respondent's age	14–29 years (ref.)	1		1	
	30 + years	0.81	(0.70–0.94)	0.58	(0.50–0.67)

Table 9. *Continued*

Variables	Refusal vs. Productives		Other unproductives vs. Productives		
	RR	(95% CI)	RR	(95% CI)	
Education, main respondent	Higher degree	0.69	(0.47–1.02)	0.79	(0.50–1.24)
	First degree	0.56	(0.44–0.71)	0.70	(0.53–0.92)
	Diplomas in higher education	0.77	(0.59–1.00)	0.57	(0.43–0.77)
	A/AS/S levels	0.77	(0.61–0.99)	0.74	(0.58–0.94)
	O Level/GCSE grades A-C, GCSE grades D-G (ref.)	1		1	
	Other academic qualification (including overseas)	0.96	(0.65–1.43)	1.56	(1.17–2.07)
	None of the above qualifications	1.32	(1.12–1.57)	1.38	(1.17–1.62)
	Child Poverty Index (continuous variable)	0.998	(0.994–1.002)	1.009	(1.004–1.014)

Notes: 1. Model fit:  $F(66, 324) = 26.6, p < .001, n = 18,269$ . 2. Other unproductives include noncontact, untraced movers, other nonresponse, ill/incapacitated, loss of computer data in the field, language problems.

This dichotomy among nonrespondents is somewhat arbitrary. We know that some refusals are more definitive than others ("hard" vs. "circumstantial") and it is suspected that some noncontacts are disguised refusals in cases deliberately avoiding being contacted. Nevertheless the association of the two types of nonresponse with mobility is in opposite directions. Holding all the other variables in the model constant, the relative risk ratio comparing movers to nonmovers for refusals relative to productives is 0.79 (95% CI: 0.63–1.01). For other unproductives relative to productives it is 3.28 (95% CI: 2.58–4.18).

We do, however, know that families who have moved are inevitably nonrespondents at the next survey if they cannot be traced. To see if this feature is responsible for the lower response rate of movers, we excluded such untraced movers from the sample analysed in Table 10. This shows results from the same model as in Table 9 but with untraced movers excluded from the sample. Since none of the excluded cases could be classified as refusals, the refusal part of the model remains very similar. The changes between Tables 9 and 10 are confined to the "other unproductive" side of the model. What was not expected is that virtually all of the change occurs in the relationship with mover status, not anything else. We can conclude that the link between moving and overall noncontact for this pair of surveys works entirely through the survey data being incomplete in terms of information about changes of address.

The interaction effects with mobility described in Section 4.2.1 operate differently according to the kind of nonresponse. The interaction with UK country is mostly through refusal for Northern Ireland – mobile families there are relatively less likely to refuse – whereas mobile families in Scotland are relatively more likely not to be contacted. The interaction with type of accommodation operates through noncontact – people moving out of a flat are relatively more likely to be contacted. On the other hand, the interaction with tenure operates through refusal with those in the "other" tenure category less likely to refuse if mobile.

#### 4.2.3. Multilevel Models

The results from the multilevel multinomial model set out in Section 3.3 do not change the conclusions from the single level model in any important way and so the model estimates are not presented here (but are available from the authors on request). We do, however, find that there is statistically significant variation from ward to ward in the probabilities of refusing and being "other unproductive," only some of which is accounted for by the Child Poverty Index. The residual between-ward variation is about 4% of the total and does not differ significantly between refusals and other unproductives. There is a residual correlation of 0.39 between the proportions of refusals and other unproductives at the ward level, which could be related to unmeasured characteristics of the wards themselves, of households within wards or to the interviewers working within the wards.

## 5. Discussion

The major contribution of this article is to demonstrate the importance of residential mobility as a predictor of future nonresponse. At one level, this result is tautological: families who have moved and cannot be traced are inevitably nonrespondents at the next survey. In fact, only 3.7% of the eligible sample at Wave 2 were untraced whereas 9.5%

Table 10. Relative risks of being unproductive excluding untraced movers

382

*Journal of Official Statistics*

Variables	Refusal vs. Productives		Other unproductives vs. Productives	
	RR	(95% CI)	RR	(95% CI)
Mobility	Nonmover (ref.)	1	1	
	Mover	0.80	(0.63–1.01)	0.96 (0.68–1.34)
UK country	England (ref.)	1	1	
	Wales	1.11	(0.90–1.38)	1.57 (1.16–2.12)
	Scotland	1.26	(1.00–1.58)	1.68 (1.14–2.48)
	Northern Ireland	2.28	(1.88–2.77)	1.44 (0.96–2.17)
Family annual income	£0–£3,100	1.16	(0.74–1.83)	0.94 (0.50–1.76)
	£3,100–£10,400	1.32	(1.10–1.58)	1.30 (1.06–1.59)
	£10,400–£20,800 (ref.)	1		1
	£20,800–£31,200	0.76	(0.62–0.94)	0.95 (0.72–1.26)
	£31,200–£52,000	0.67	(0.51–0.88)	0.90 (0.64–1.28)
	£52,000 +	0.87	(0.62–1.23)	0.96 (0.54–1.71)
	Don't know	1.12	(0.85–1.47)	1.32 (0.97–1.80)
	Refused	2.09	(1.49–2.93)	1.41 (0.81–2.45)
Ethnicity, cohort member	White (ref.)	1	1	
	Mixed	0.92	(0.62–1.38)	1.92 (1.23–2.98)
	Indian	1.34	(0.91–2.00)	0.91 (0.53–1.58)
	Pakistani & Bangladeshi	1.20	(0.87–1.65)	1.37 (0.92–2.04)
	Black/ Black British	0.99	(0.70–1.41)	2.47 (1.75–3.49)
	Other	2.11	(1.08–4.09)	1.59 (0.88–2.87)
Tenure	Own (ref.)	1	1	
	Rent	1.15	(0.98–1.34)	1.41 (1.09–1.83)
	Other	2.12	(1.51–2.96)	1.17 (0.72–1.93)
Accommodation type	House/Bungalow (ref.)	1	1	
	Flat/maisonette Studio/Bed-sit, etc	1.18	(0.92–1.52)	3.13 (2.45–4.00)
Main respondent's age	14–29 years (ref.)	1	1	
	30 + years	0.81	(0.70–0.94)	0.54 (0.45–0.65)

Table 10. Continued

Variables		Refusal vs. Productives		Other unproductives vs. Productives	
		RR	(95% CI)	RR	(95% CI)
Education, main respondent	Higher degree	0.70	(0.47–1.02)	1.04	(0.60–1.82)
	First degree	0.56	(0.44–0.71)	0.72	(0.51–1.02)
	Diplomas in higher education	0.77	(0.59–1.00)	0.48	(0.31–0.75)
	A/AS/S levels	0.77	(0.61–0.99)	0.68	(0.51–0.89)
	O Level/GCSE grades A-C, GCSE grades D-G (ref.)	1		1	
	Other academic qualification (including overseas)	0.96	(0.65–1.42)	1.43	(1.02–2.01)
	None of the above qualifications	1.32	(1.11–1.56)	1.13	(0.92–1.39)
CPI	Child Poverty Index (continuous variable)	0.998	(0.994–1.002)	1.010	(1.003–1.016)

Notes: 1. Model fit:  $F(66, 324) = 19.4, p < .001, n = 17,595$ . 2. Other unproductives include noncontact, other nonresponse, ill/incapacitated, loss of computer data in the field, language problems.

were refusals and 5.1% were traced but noncontacts. Moreover, the importance of mobility remains after controlling for a number of other socio-economic and housing variables that are related both to nonresponse and to mobility: the odds of a nonmover remaining in the sample are still 1.4 times higher than those of a mover's doing so. This effect is as strong as, for example, the effect of having median level educational qualifications compared with having no educational qualifications.

An important implication of the finding about mobility is that the assumption that data are "missing at random" that is often invoked when estimating nonresponse weights for longitudinal studies is perhaps too strong. In other words, we should not assume that events or changes after the previous wave are unrelated to the probability of nonresponse even after controlling for other variables. Basing nonresponse weights for Wave  $t$  just on variables measured at earlier Waves  $t-k$  and ignoring mobility between Waves  $t-1$  and  $t$  can produce biased weights. We can get some idea of the potential bias by comparing the odds ratios for the variables in the model before and after the inclusion of mobility (Table 8). We see that, especially for Northern Ireland, tenure and accommodation type the odds ratios are different after allowing for mobility. The interaction between mobility and Northern Ireland is a little difficult to explain without more information about where families have moved to but it could be linked to movement out of areas having high percentages of one religious faith. The interactions with tenure and type of accommodation are, however, consistent with the possibility that mobile families previously living in cramped conditions lacking privacy were moving up the housing ladder into more advantageous conditions, perhaps because their economic circumstances had improved. This, in turn, suggests that changes in economic circumstances between waves might be related to response.

The importance of mobility as a predictor of nonresponse points in turn to the importance of collecting accurate data on mobility between waves and incorporating that information into the datasets released for analysis. It is unlikely to be sufficient just to rely on data from earlier waves (the number of moves between Waves  $t-2$  and  $t-1$  for example as used by Hawkes and Plewis (2006)) and so data from the tracing systems associated with the administration of the survey, sometimes known as "paradata" (Groves and Heeringa 2006), need to be brought into the main dataset, always noting the need to safeguard confidentiality.

We note that the predictors of other unproductive (nearly all of which are noncontacts and untraced movers) and refusal are different. We find, for example, that the level of family income and the failure to provide a figure for income are better predictors of refusal than they are of noncontact. On the other hand, the effects of ethnic group, tenure, type of accommodation, mother's age, mother's educational level and the area deprivation score are stronger for noncontact than they are for refusal. These issues are considered in much more detail in Plewis (2007b), that focuses separately on predictions of not being located, not being contacted and refusal.

The estimates for mobility are strikingly different for the different kinds of nonresponse: compared with productive families, noncontact is much more likely for mobile families (after controlling for other variables) whereas refusal is marginally more likely for nonmovers. But once we remove the untraced movers from the analysis (where noncontact and moving are synonymous) there is no difference between movers and nonmovers in the probability of being unproductive but not a refusal. Of course, we do not know whether the noncontacts would have refused if they had been contacted and it is

possible that the traced movers include a high proportion of families who have informed the survey team that they have moved and are therefore more committed to the study.

Finally, we attach some caveats to our findings. The findings are based on the first two waves of one UK birth cohort study and might not generalise to, say, panel studies in other countries nor even to later waves of the Millennium Cohort Study. The estimates for mobility do depend on the assumption that the model for nonresponse is properly specified. In other words, if there are omitted variables associated both with mobility and with response then the strength of the mobility estimates could be reduced. There are other variables – for example, whether or not the partner chose to be a respondent at Wave 1 – that might be associated with subsequent nonresponse but this variable is unlikely to be associated with mobility. It is also important to realise that the findings are based on the productive sample at Wave 1 and that was only 72% of the target sample then. We do not know whether the estimates for mobility are affected by selection at this stage and whether longitudinal studies with higher initial response rates would show a different pattern of results subsequently. What does seem clear, however, is that residential mobility – which is such a feature of the lives of UK families with young children – also has important methodological implications for analyses of nonresponse.

## 6. References

- Ferri, E., Bynner, J., and Wadsworth, M. (eds) (2003). *Changing Britain, Changing Lives: Three Generations at the Turn of the Century*. London: Institute of Education.
- Groves, R.M. and Heeringa, S.G. (2006). Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs. *Journal of the Royal Statistical Society, Series A*, 169, 439–457.
- Grundy, E. and Fox, J. (1985). Migration During Early Married Life. *European Journal of Population*, 1, 237–263.
- Hawkes, D. and Plewis, I. (2006). Modelling Nonresponse in the National Child Development Study. *Journal of the Royal Statistical Society, Series A*, 169, 479–491.
- Lepkowski, J.M. and Couper, M.P. (2002). Nonresponse in the Second Wave of Longitudinal Household Surveys. *Survey Nonresponse*, R.M. Groves et al. (eds). New York: John Wiley.
- Little, R.J.A. and Rubin, D.B. (2002). *Statistical Analysis with Missing Data*, (2nd ed.). New York: John Wiley.
- Plewis, I. (ed.) (2007a). *Millennium Cohort Study First Survey: Technical Report on Sampling* (4th ed.). London: Centre for Longitudinal Studies.
- Plewis, I. (2007b). Nonresponse in a Birth Cohort Study: The Case of the Millennium Cohort Study. *International Journal of Social Research Methodology*, 10, 325–334.
- Plewis, I. and Ketende, S. (eds) (2006b). *Millennium Cohort Study: Technical Report on Response* (1st ed.). London: Centre for Longitudinal Studies.
- Rasbash, J., Steele, F., Browne, W. and Prosser, B. (2004). *A User's Guide to MLWiN, Version 2.0*. Centre for Multilevel Modelling, University of Bristol.
- Smith, K. and Joshi, H. (2002). The Millennium Cohort Study. *Population Trends*, 107, 30–34.

Received December 2006

Revised December 2007