CLS Cohort Studies

Working Paper 5

Mobile Families and Other Challenges in the Design of the Millennium Cohort Study

Heather Joshi
Ian Plewis
Andy Cullis
Mahmood Sadigh
Mos Mojaddad
Katie Dodd
Joanne Woods
Aliy Chapman

December 2002

Centre for Longitudinal Studies

CLS

Mobile families and other challenges in the design of the Millennium Cohort Study

Heather Joshi, Ian Plewis, Andy Cullis, Mahmood Sadigh, Mos Mojaddad,
Centre for Longitudinal Studies,
Bedford Group for Lifecourse and Statistical Studies
Institute of Education, London

and

Katie Dodd, Joanne Woods and Aliy Chapman, Analytical Services Directorate,
Department for Work and Pensions

Abstract

Mobile families and other challenges in the design of the Millennium Cohort Study

The Millennium Cohort Study continues the tradition of Britain's national longitudinal birth cohort surveys, but it has broken the design mould of being based on all births in a week. The specification that it should cover a year's births provided the opportunity for some innovative design. This paper describes the disproportionately stratified design, based on a sample of electoral wards and the practicalities of implementing it.

The need to sample offered the possibility of over representing populations of particular interest: children living in economically deprived areas, in areas where minority ethnic groups are concentrated, and in the smaller countries of the UK. Eligible families in sampled wards are selected from records held by the Department for Work and Pensions, who also subsequently administer an opt-out to families in sampled areas.

Although the final analysis of the sampling process (and the substantive results) will not be ready until 2003, it will be possible to present an interim report on the experience of drawing the sample in this way, focusing particularly on England and Wales, where the survey started earlier. The number of children identified through the administrative records will be compared with those recorded in Birth Registration records, as far as ward boundary changes allow. The cases withdrawn before a sample can be issued to field are analysed by type of ward and type of claimant.

A preliminary assessment of the coverage of the Child Benefit Register is offered in a comparison with data from vital registration. This exercise has so far only covered those wards in England where there have been no boundary changes between those used to define the sample in 1998 and those applying in 2000-2001 when the children were actually born.

The extent of mobility in and out of target areas by families with children aged 6-9 months is also to be discussed.

The strengths of using the Child Benefit register as a sampling frame have been confirmed. The difficulties have in some ways been less than anticipated, and the requirement of a geographically defined sample would have formed a challenge whatever method of sampling had been adopted.

The Millennium Cohort Study (MCS) continues the tradition of Britain's national longitudinal birth cohort surveys, a tradition that started in 1946 and has since been reinforced in 1958 and 1970. The specification that the MCS should cover a year's births, rather than just a week as was the case previously, provided the opportunity for some innovative design. Rather than trying to cover all aspects of the design in one paper, we concentrate here on the use of Child Benefit records to generate a disproportionately stratified and clustered sample of eligible children.

Although the final analysis of the sampling process (and substantive results) will not be ready until 2003, it is possible to present an interim report on the experience of drawing the sample in this way, focusing particularly on England and Wales, where the survey started earlier. We define the population for the MCS and then describe how the Child Benefit records were used to obtain a sample that could be issued to the fieldwork contractors. We focus on losses from the sample, and the characteristics of those lost, *before* fieldwork commenced. An important issue here is the extent of mobility in and out of target areas by families with children under 9 months. The numbers of children identified through the administrative records are compared with those recorded in Birth Registration records, as far as ward boundary changes allow. We conclude with some general comments about the value of the Child Benefit Register as a sampling frame.

1. The Millennium Cohort Study population

The research design for the Millennium Cohort Study (MCS) was based on the following five principles:

(1) The MCS should provide data about children living and growing up in each of the four countries of the UK.

(2) The MCS should provide usable data for sub-groups of children, in particular those living in advantaged and disadvantaged circumstances, and for children of ethnic minorities and those living in Scotland, Wales and Northern Ireland.

(3) As well as data about children, the MCS should provide data about their family circumstances and the broader socio-economic context in which the children grow up.

(4) The MCS should include children born throughout a single 12-month period.

(5) All children born as members of the MCS population should have a known and non-zero probability of being included in the selected sample.

In addition, there were a number of practical constraints that influenced the chosen design:

(1) The sample had to include a substantial proportion of children born in the year 2000 and measured in their first year of life.

(2) A planning period that was very short, starting in May 2000 to prepare for fieldwork commencing one year later.

(3) A design that facilitated comparisons with earlier UK birth cohort studies, notably those that started in 1946, 1958 and 1970, was essential.

In brief, the fundamental aim of the research design, and in particular the sample design, was to ensure a proper representation of the total population, while at the same time having sufficient numbers of key sub-groups for analysis. Probability (or random) methods of selection, combined with stratification and clustering, would achieve these aims and would enable the vagaries of sampling to be properly accounted for through the computation of theoretically sound sampling errors.

The MCS population is a population of children defined as:

> all children born between 1 September 2000 and 31 August 2001 (for England and Wales), and between 23 November 2000 and 11 January 2002 (for Scotland and Northern Ireland), alive and living in the UK at age nine months;

and, after nine months:

> for as long as they remain living in the UK, at the time of sampling.

The birth dates for Scotland and Northern Ireland are about three months later than those for England and Wales in order to avoid possibly substantial overlap with a Department of Health sponsored survey of infant feeding practices, which was sampling births between September and November 2000. The overlap between the two surveys in England and Wales was expected to be sufficiently small not to be a problem. The sampling period for Scotland and Northern Ireland was extended from the original 12 months to make up for a shortfall in the expected numbers of births.

The population includes:

> (a) children living in non-household situations (women's refuges, hostels, prisons etc.) at age nine months;
> (b) children not born in the UK but resident in the UK at age nine months.

> The population excludes:

> (a) children who die before age 9 months;
> (b) UK-born children who emigrate from the UK before 9 months.

This definition of the population meant that (for England and Wales) one third of the sample was born in the year 2000. Although there would have been some advantages in collecting data before the children reached the age of nine months, it was not practically possible to do so, given the need properly to pilot the fieldwork procedures.

2. Obtaining the sample

For the purposes of this paper, we merely note that the MCS sample in England and Wales is essentially a sample of electoral wards, the boundaries of these wards being those in existence on 1 April 1998. Wards with very small populations (fewer than 25 births expected in a year) were amalgamated with adjoining wards, creating some

combined data points, loosely referred to as 'wards' hereafter. The English 'wards' were grouped into three strata: an advantaged stratum, a disadvantaged stratum and an 'ethnic minority' stratum. The latter two strata were over-represented in the sample. Advantage and disadvantage were based on the ward-based Child Poverty Index (Noble et al., 2000), 'ethnic minority' on the proportion of Black and Asian families in the 1991 Census. There was no explicit ethnic minority stratum in Wales. For both countries, all eligible children in the ward were included in the sample. Further details of this part of the sampling process can be found in Plewis (2001).

Once the wards were selected, a list of all nine month old children living in those wards was required. Data would then, in principle, be collected about all these children and their families. The lists were generated from the Child Benefit (CB) register held by Analytical Services Directorate (ASD) Information Centre, part of the Department for Work and Pensions (DWP). Child Benefit is a universal provision, payable (usually to the mother) from the child's date of birth. Take-up of Child Benefit is thought to exceed 97%. The Information Centre did not, however, release the addresses for all births known to them. They were obliged to exclude a small number of so-called sensitive cases. Child Benefit data contain a number of flags that indicate that a case may be sensitive. Sensitive cases can be identified by using the correspondence flag that is recorded on the customer (parent) record, and the exclusion code that is on the non-customer (child) record.

A case was deemed sensitive if:

- there had been a child death in the family in the last five years;
- the family was in correspondence with DWP. As the type of correspondence is not identifiable, such a case was deemed sensitive unless the correspondence arose when a child approaches school leaving or there was a routine notification of change of address;
- the child was in another person's award;
- the child had been taken into care;
- the family had previously been selected for the DWP-sponsored Family and Children's Survey (FACS);
- there was an unknown exclusion code entered on the CB non-customer record.

The raw administrative data on Child Benefit (100% Generalised Matching Service (GMS)) is updated on a four weekly basis. The first scan took place to pick up children age seven months. Upon receipt of the Child Benefit GMS data, the Analytical Services Directorate (ASD) Information Centre uses a computer package called QAS (Quick Address System) to set back any postcodes listed (for England, Scotland and Wales) to their 1998 postcodes. They are set to 1998 as this was the year used to define the wards. If there is a missing postcode within the data, then QAS will attempt to match either a full 1998 postcode to the first line of the address listed, or alternatively a partial 1998 postcode to the address. Also, if there is only a partial postcode listed within the data, then QAS will attempt to match a full 1998 postcode to that address. In addition, the Information Centre was aware of some changes of address for families that had moved out of the selected wards. However, families can continue to be paid Child Benefit without notifying DWP of a change of address, especially if the benefit is paid directly into a bank or building society account. Hence, not all out-migrants (and, more relevantly for our purposes, not all in-migrants) were picked up at the second scan, taking place four weeks later when the batch of babies were aged about eight months. After sampling was under

way it was decided that these in-migrant families should be picked up when the child was aged approximately eight months and checked for sensitivity at aged nine months, four weeks later than the original samples. In-migrant families, or at least those who had a newly announced change of address between these scans, began to be picked up at Wave 8 (there were 13 waves altogether).

The CB 100% GMS database is held not by ward but by postcode. Once QAS has matched as many postcodes as possible, the Millennium Cohort wards are then added to the data using the Office for National Statistics (ONS) 1999 Central Postcode Directory (this is related to 1998 wards) via an imputation method.   Postcode areas and ward boundaries do not always align because one is a commercial Royal Mail area and the other a local government boundary. In those few cases where the ward boundary cuts across the seven character postcode boundary, all families within the postcode were included.

ASD Information Centre sent out a letter to all families with a child eligible for the study, and who had not been excluded, inviting them to participate. They also received a leaflet describing the nature and purposes of the study. Families would have received the letter when the child was about seven months old. Families had the opportunity to opt-out to DWP at this stage, either by phone or by letter. Names and addresses of all families wishing to participate at this point were passed as the children approached nine months of age, via the Centre for Longitudinal Studies (CLS), to the fieldwork team at the National Centre for Social Research (NatCen).

In common with all research of this kind, the principle of giving everyone a chance of being in the study conflicts with the practical constraints of generating a sampling frame and obtaining an interview. Eligible families could have been excluded from the study at a number of stages:

   (i)   if they did not take up Child Benefit, or were not eligible for Child Benefit, or had not claimed Child Benefit before the last scan;
  (ii)   if their address could not be matched to a ward via a postcode;
 (iii)   if they were excluded by the Information Centre as a sensitive case;
  (iv)   if they opted-out of the study;
   (v)   if they had moved into a selected ward before the child was nine months old and had not notified DWP
  (vi)   if they refused to participate when approached by the fieldwork agency;
 (vii)   if they could not ever be contacted by the fieldwork agency;

There was nothing we could do about exclusions at stages (i) and (ii). ASD's best estimate of Child Benefit take-up for 2000 in Great Britain by 7 months of age is an average of 97.18%. A working estimate of loss at stages (i) and (ii) is 3% on average for Great Britain. Every effort was made to keep exclusions at stages (iii) to (iv) to a minimum, as at (vi) and (vii). A major effort was made at stage (v) which is entailed by having a geographical condition on sample eligibility. Health Visitors responsible for children living in the selected wards were approached and asked to let CLS know about children who had moved into the areas from six months of age, as such changes of address might not have reached DWP in time to have appeared on their database, and the most recent

could not have. It was hoped to find approximately the same number of in-migrants to all the selected areas as ASD Information Centre and the fieldwork team found out-migrants.

Other routes for obtaining the sample were considered and rejected. It would have been possible to use the Office for National Statistics (ONS) birth registration records. These are essentially complete but they only have the baby's address at the time of registration, an event that must take place by six weeks after birth. The major drawback with this method of generating the sample, however, was ONS' insistence that families had to *opt-in in writing* to the study. In other words, any family approached that did not reply to a letter inviting them to participate would have been deemed to be a refusal. We feared that this method would have led to an average initial non-response of between 25 and 30%, substantially more among those sections of the population - the disadvantaged and ethnic minorities - which we particularly wanted to over-represent. Moreover, this level of non-response would only be reached with the help of follow-up letters, something that we did not have enough time for. Hence, this method of obtaining the sample was rejected.

A second route would have been to rely on Health Visitors in the selected local areas to notify us of all eligible families. However, we did not believe that the Health Visitor route would be uniformly reliable for all the selected wards, especially during a period of Health Service reorganisation, the limited capacity of Health Visitors to take on this work, the lack of correspondence between electoral wards and Health Services areas, and the lack of time to obtain ethical clearance to approach Health Service personnel. Health Visitors were, however, involved in order to find recent in-migrants to the selected wards.

Another possibility would have been for the fieldwork agency to trawl for all eligible families in the selected areas. This would have had a number of advantages in terms of non-response but would have been prohibitively expensive, as the prevalence of eligible addresses in a ward is generally low (around 2%).

3. From Child Benefit records to issued sample

Table 1 shows how DWP drew 20,806 cases for England and Wales, and after excluding 2,217 cases (11%), issued 18,589 to CLS. 7% of all the original cases were excluded because they opted-out, 1.2% because they had moved out of an MCS sample ward by the time of the second scan, and surprisingly few, 2.5% were withdrawn as 'sensitive cases'. The detailed classification of the types of these sensitive cases is given in Table 1. The largest category (1.85%) is cases involving "correspondence" but no child death. Child deaths account for 0.37% of all cases. These could be deaths of the MCS-born child themselves (if the family had other children and hence remained on the register) or any other child in the family in the past five years. Overall over 89% of the families identified at DWP were issued to the field.

Figure 1 shows how the leakage from the sample varies by stratum of the sample. There is remarkably little variation by stratum, with the percent excluded ranging from 8.9% in the advantaged wards in Wales to 11.5% in the ethnic wards of England. Moving out is a little more common in the disadvantaged (and ethnic) wards than the advantaged areas, as are sensitive cases. Opting out on the other hand is most common in the advantaged areas of England (no difference by stratum in Wales).
When expressed as correlations with the Child Poverty Index used to construct the strata, across all England and Wales wards, there is a positive correlation of 0.31 with the

proportion of sensitive cases in a ward as well a negative correlation of -0.12 with the proportion opting out. The correlation between the Child Poverty index and moving out is 0.18. The proportion moving out is not related to ward size (r=-0.01) but the larger wards do tend to be poorer (r=0.37). There is little relation between ward size and opting out (r=-0.047) and only a weak association with the proportion of sensitive cases (0.13).

Results from analyses of DWP exclusions by certain characteristics of claimants (available on the Child Benefit record) for the UK data up to wave 15 are given in Figures 2 to 5. These results are based not only on the England and Wales data described above but also on the 48 out of 59 weeks births in Scotland and Northern Ireland (where the picture is very similar to England and Wales).

Figure 2 shows exclusions by age of claimant. These do not rise above average until age 35 - 39, but cases at the level or above only account for one sixth of all cases. Teenage claimants, presumably mothers, make up 9% of all cases and have the highest rate of moving out (2.9%). Both opting out and DWP exclusion tend to rise with age.

Figure 3 shows an analysis by the "Title" of the claimant. Although mothers are normally the person claiming Child Benefit, lone fathers, and in some cases fathers in two-parent families make the claim. Hence 3.6% of all claimants appear to be men with the title 'Mr'. Other titles (e.g. 'Dr') are negligible (0.4%), and the three female titles, 55% 'Mrs.', 36% 'Miss', and 5% 'Ms', are indicative of marital status, but not conclusively. The rates of exclusion are not greatly different between the three female titles. The relatively high rate of moving out of the 'Miss' category may reflect the higher mobility of teenage mothers (most of whom are never-married) noted in Figure 2. The cases with male claimants are relatively likely to have been excluded by DWP.

Figure 4 shows that there is some association between exclusion and family size, which only really affects unusually large families (of four or more) and the major reason for this is a greater chance of a case being deemed sensitive - the more children, presumably, the greater chance of the family being some problem to be flagged.

Figure 5 shows an analysis by a particularly revealing indication on the Child Benefit register, whether or not the benefit is paid through an order book, cashable at a post office, or through a bank account. In England (and Scotland) around two thirds of cases in advantaged areas received payment through bank accounts versus about 4 in 10 in disadvantaged and ethnic areas. In Wales (and Northern Ireland) the proportion with bank accounts was even lower in each stratum, 59% and 35% respectively in Wales but the differential remained.
 Not only can this be taken as a rough and ready indicator of family affluence, it also affects the speed with which the Child Benefit system is likely to update a moving family's address. In this sample bank accounts make up for just under half of all payments (48%) and order books correspondingly just over one half (52%). The overall rate of exclusion between these two types of payment is the reasonably similar: 10.2% of the bank account cases and 11.6% of the order book cases. The latter have relatively more moving out and excluded by DWP, while the bank account 'customers' are (somewhat) more likely to opt-out. Within stratum (data not shown) there is also a consistent excess of exclusions and movers out among cases with order books.

Thus the two sources of loss from the eligible sample - opting out and DWP exclusions - are likely to introduce opposite biases as far as the economic standing of individual families are concerned. Although the families who move out are not themselves eligible for the survey, they provide some indication of the characteristics of families likely to move in (to the extent that the unknown movers in must have moved out of somewhere). Movers as detected by DWP are disproportionately young, unmarried and holders of order books. The latter association may not reflect a greater propensity of poorer families to move as much as a greater chance of such cases reporting a change of address to DWP.

4. Finding moving families

It has already been established that having a baby is a trigger to moving home (Grundy and Fox 1984) so it is not surprising that a substantial proportion of families with children born on eligible birth dates were on the move in and out of the geographical boundaries which confirmed their eligibility for the survey. Unfortunately no administrative register may ever be perfectly tuned to keeping up to the minute track of all movers.

Amongst those who were successfully interviewed around 12% (provisional) had moved home since the child's birth 9 – 10 months earlier, but there is some evidence that that the DWP was not detecting all changes of address . In the fieldwork stage interviewers failed to interview (according to interim results) nearly five times as many movers-out as had been detected by DWP, with less difference between strata. Some of these families would have moved since the '8 month' scan, but only about six weeks elapsed between the last scan and issuing to the interviewers, compared to the four week period in which DWP detects moves, so this also suggests that some changes of addresses were being reported to DWP, but, inevitably, with some delay.

The evidence from the field that altogether up to 7% of the DWP sample had changed address (and not been interviewed) since the sample was drawn caused concern about the presumably approximately equally numerous contingent who had moved in without being recorded (although most of the movers out would be ineligible for the survey, at least if they moved before the child was aged 9 months, the corresponding movers in should ideally be included). Investigations of new arrivals on the DWP scan between the 7th and 8th month were started from Wave 8 at the end of November 2001 provided some 'in-migrant' cases but it was not as successful or informative as might have been hoped. Over waves 8 - 15 (for the UK as a whole) this exercise produced 457 families with a new address at the 8-month scan. However, 38% of these were excluded from the sample resulting, so far, in 265 cases being added to the issued total. The extra exclusions were almost all (175 out of 192) due to correspondence with the Child Benefit Centre. This could suggest that people with 'changing' circumstances were particularly likely to notify a change of address than perhaps other movers who circumstances did not warrant any correspondence with the DWP. The 457 new addresses were markedly more likely to have order books than bank accounts (68% vs. 32% compared to roughly 50:50 in the sample as a whole). They were more likely to have four or more children, not to have the title 'Mrs' and more likely than the whole sample to be over 35 or aged 20 - 24. The fact that 2.2% had moved out by the 9-month sweep (compared to 1.2% in the main sample between months 7 and 8) suggests this group contains a highly mobile minority.

Figure 6 shows that the minority of these cases with bank accounts also had almost as great a chance of exclusion by DWP as those with order books. Cases with order books

may have a higher chance of registering a new address, but so it appears may bank account payees whose changing circumstances require non-trivial correspondence with the DWP. Thus, though the extra cases from this exercise are welcome, we are still left wondering about the characteristics of in-movers who don't immediately inform DWP of their new addresses. In due course we can look at the characteristics of the cases picked up by Health Visitors - but there are unlikely to be enough of these (51 in England and Wales) to be particularly informative about possible bias.

Compared to about 600 cases thought to have moved out of eligible addresses in the last six waves of sampling in England and Wales we can match 348 movers-in found by DWP and 24 additional cases found by Health Visitors (one complication of this operation was that some of the in-movers reported by Health Visitors did in fact appear in the DWP list and are not included among the "additional cases"). The picture for the first seven waves when the only information on in-movers came from Health Visitors was, for England and Wales, 27 in-movers versus around 700 movers out.

5. Coverage of Child Benefit sampling frame

In order to estimate the size of the notional sample which should ideally have been drawn, we made a working assumption that 3% of all families in Great Britain would not be covered by the Child Benefit register primarily because of non-claiming or ineligibility for the Benefit. Births in Britain not eligible for Child Benefit include those to diplomats, foreign servicemen, some foreign students, employees of overseas employers and asylum seekers. This also included the negligible numbers for whom a postcode could not be found in Great Britain. There may be non-take-up at Child Benefit by very rich, and also by people who's application is inhibited or slowed down by literacy or language difficulties. We have no solid evidence for either of these conjectures.

We made another attempt to verify how many children might have been omitted from the Child Benefit register in each English ward by comparing the number of registered births for the relevant 12 months with the issued sample. This exercise was not straightforward, as the boundaries of some MCS wards had changed between 1998 and 2000 - 2001. This means an exact match is not possible for 34 out of 200 MCS data points in England (25 advantaged, 9 disadvantaged, none of the ethnic wards).

Even where the boundaries had not changed children born in a ward do not necessarily appear on the Child Benefit register at 7 months if they have died or moved out, and the Child Benefit register may be supplemented by in-movers. Some infant deaths stay on the Child Benefit register (if there are other children receiving Child Benefit in the family) but others may never have claimed or have been removed (if they were first-borns). Nationally the proportion of infant deaths before 7 months is under 6 per thousand, but there may be local variations making the excess of births over Child Benefit greater for some wards or strata than others. There is some evidence of not unexpected infant mortality variation by stratum (see Table 2), but interesting though each local variation may be, they do not affect the estimates of the MCS notional sample, which is defined only to include survivors.

Another reason for the registered births of a ward surpassing the number of families recorded on the Child Benefit Register is multiple births. Again this requires a small adjustment, multiplying the number of families by around 1.01 nationally to arrive at the

number of children. Although we do not have a record of the incidence of multiple births in the Child Benefit register, the incidence among the issued sample suggests that there is some variation in multiple births between strata, but not enough to account for much of the excess of births over Child Benefit records.

The proportions of families in the issued sample with multiple births were as follows:

England   Advantaged   1.6 %
          Disadvantaged   0.9%
          Ethnic   0.4%

These figures are consistent with the view that families with multiple births are not being differentially excluded through opt-out or correspondence with DWP.

Another artefactual reason why birth registrations and Child Benefit figures may not match is if apparently valid postcodes are mis-reported. However the accuracy of birth registration postcodes is known to be high (ONS report over 98% accuracy in their tests), and the DWP postcodes have generated extremely few cases which turned out not to be living in an eligible area.

Where the greatest divergences between Child Benefit and registered births have been found there are thought to be concentrations of asylum seekers or US services families. In other areas where births exceeded Child Benefit we suspect high population turnover. In the former two cases - asylum seekers and foreign service families, the children may not have been eligible for Child Benefit, but they should have been eligible for sampling into the Millennium Cohort Survey at Sweep 1 as residents in the UK (although they might not stay in the population at subsequent sweeps). On the other hand where there has been net migration of babies out of an area the excess of births over Child Benefit register does not indicate a shortfall in the proportion of the eligible population covered.

Some of these issues are reflected in Table 2. There is a larger difference between ONS births and DWP children in the disadvantaged and, especially, the ethnic wards than in the advantaged wards. This could be due to greater out-migration from these types of wards although this would also lead to greater in-migration into the advantaged wards (assuming zero net international migration). It could also be due to language / literacy problems slowing down Child Benefit Claims, as speculated above.

If required, further triangulation with the results of the 2001 Census to see how far the Child Benefit register should be adjusted for non-coverage of the relevant population would be possible. For the moment we note that investigations are underway but proceed with a rough allowance for the ratio of the notional population to the number on the Child Benefit register.

6. Conclusions

The Child Benefit register has proved to be a valuable resource for obtaining the sample for the MCS once the wards had been selected. Its strengths include its near universal coverage and inclusion of birth dates and postcodes. It is particularly valuable to research in that DWP has allowed it to be used with an opt-out rather than an opt-in procedure. We

note that only about 13% (3% missing, 7% opted out, 2.5% sensitive cases) of the eligible sample was lost before the sample was issued to the fieldwork agency. This is almost certainly a much smaller figure than would have been obtained with an opt-in procedure for which a figure of about 30% up to the point of issue to fieldwork might have been expected. It is, of course, possible that non-response in the field is greater when opt-out is used and so the overall difference in response rates between the two approaches might be narrower but it is most unlikely to be entirely eliminated. Indeed, indications from the fieldwork agency are that the non-response in the field averages about 15%.

Mobile families will always be moving targets and cause problems for sample design and selection whatever sampling frame is used. Using the CB register exacerbates the problem in the sense that the addresses held by DWP can be out-of-date and so the fieldwork agency found substantial numbers of families who had moved out. The compensating system of using Health Visitors to find their replacements was disappointing but had to contend with as the survey coinciding with a major reorganization of primary health care. More work needs to be done on this issue, partly to establish the quality of the data at Sweep 1 but also because mobile families at this point of the lifecourse are likely to be mobile families later and hence will be difficult to trace as the MCS moves into its longitudinal phases. This work will be aided by the data collected about mobility in the Sweep 1 interview. But it is not only for methodological reasons that mobile families are important. They might also be the kinds of families that among the most vulnerable and who miss out on services as a result of their mobility.

This paper has covered in detail one part of the sampling process for the Millennium Cohort Study. Future papers will cover topics such as the selection of the wards, sample weights and analyses of non-response. The MCS has been designed as birth as a longitudinal study from the outset. Our aim is to analyse issues of this kind in some detail so that users of the MCS data are able to take them into account in their substantive analyses, not only of Sweep 1 but, importantly, in their longitudinal analyses of the future.

References

Grundy E and Fox A.J (1984)'Changes of address in the early years of marriage, *Population Trends,* 38, 25-30.

Noble, M., Smith, G. et al. (2000) Measuring multiple deprivation at the small area level: The indices of deprivation, 2000. Oxford University for DEPR.

Plewis, I (2001) The Millennium Cohort Study: Population And Samples: Technical Report. First Edition, August 2001, Centre for Longitudinal Studies, Institute of Education, London

Table 1 Cases approached by DWP in England and Wales.

**All waves 1-13 by detailed reason for exclusion**

	n	%	
Total number of families approached by DWP	20806	100.00%	
Total number of families excluded	2217	10.66%	
All removed by DWP ('sensitive cases')	515	2.48%	
FACS family	35		0.17%
Child death in last 5 Years – No Correspondence	70		0.34%
Child death in last 5 Years – Correspondence	6		0.03%
Correspondence – No Child death in last 5 Years	385		1.85%
Child has unknown exclusion code	4		0.02%
Cases potentially sensitive	1		0.00%
Child in another persons award	14		0.07%
Family has moved out of the sample area	253	1.22%	
Family opted out of the project	1449	6.96%	
All families included in the issued sample	18589	89.34%	

Table 2: Comparing ONS registered births September 01 2000 – August 31 2001 and children receiving Child Benefit by stratum (England only).
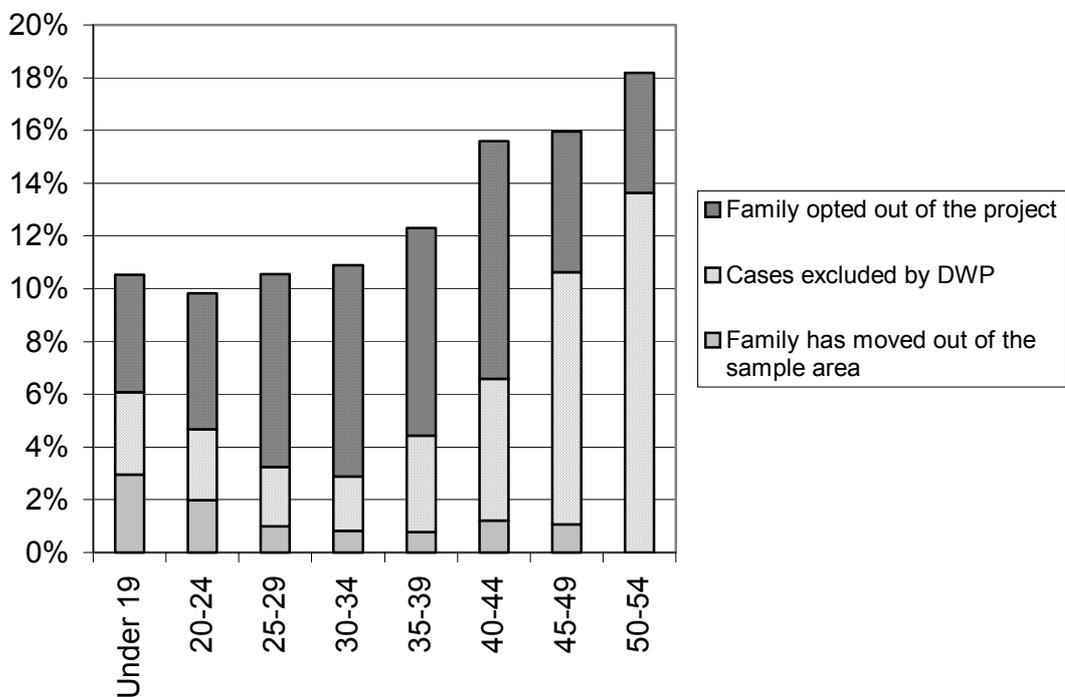
Stratum	Number of data points (1)	ONS registered births	DWP/CB families	CB children (2)	Neonatal deaths (3)	'Corrected' CB children (4)	Difference
Adv.	85	5473	5188	5271	16	5287	3.4%
Disadv.	62	6161	5676	5727	25	5752	6.6%
Ethnic	19	4465	3906	3922	26	3948	11.6%

(1) Only those wards for which there had been no boundary changes between 1998 and 2001 and excluding one advantaged ward with a large proportion of births to US service personnel.
(2) Using multiple birth prevalences (but not allowing for triplets) given on p.10
(3) Data for Jan. – Dec. 2001
(4) Some neonatal deaths will appear in column 4 (see comment on p. 10); excludes small number of deaths after 28 days.
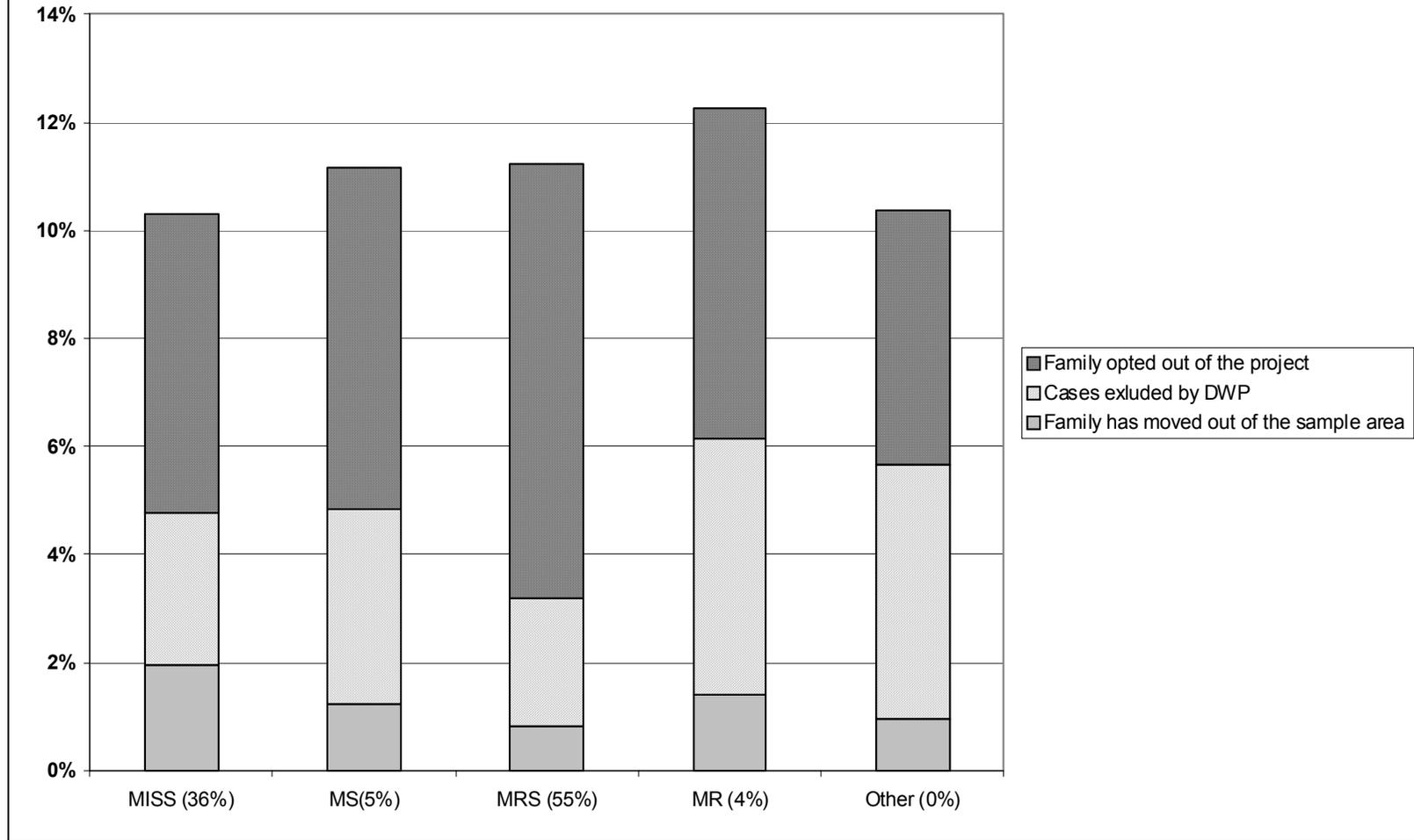
Figure 1
Loss from DWP sample by stratum
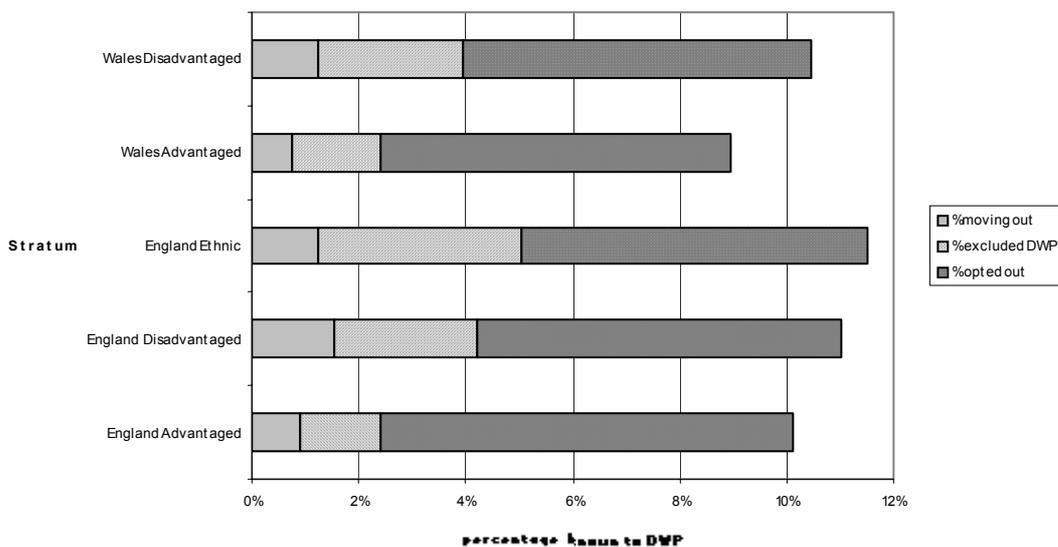
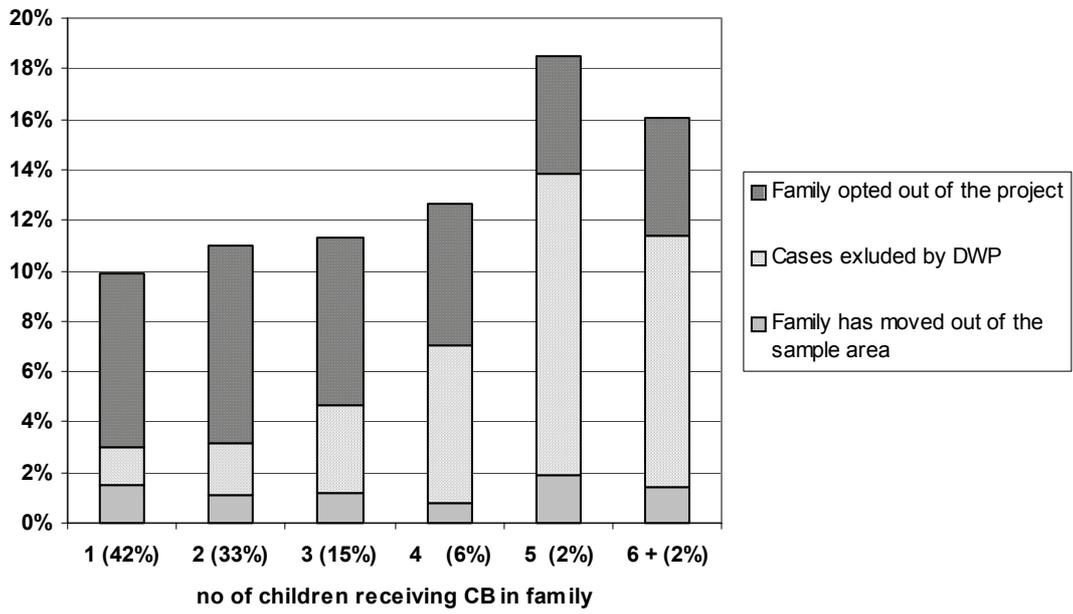**Figure 2 Exclusions from sample by age of Child Benefit claimant**

Figure 3
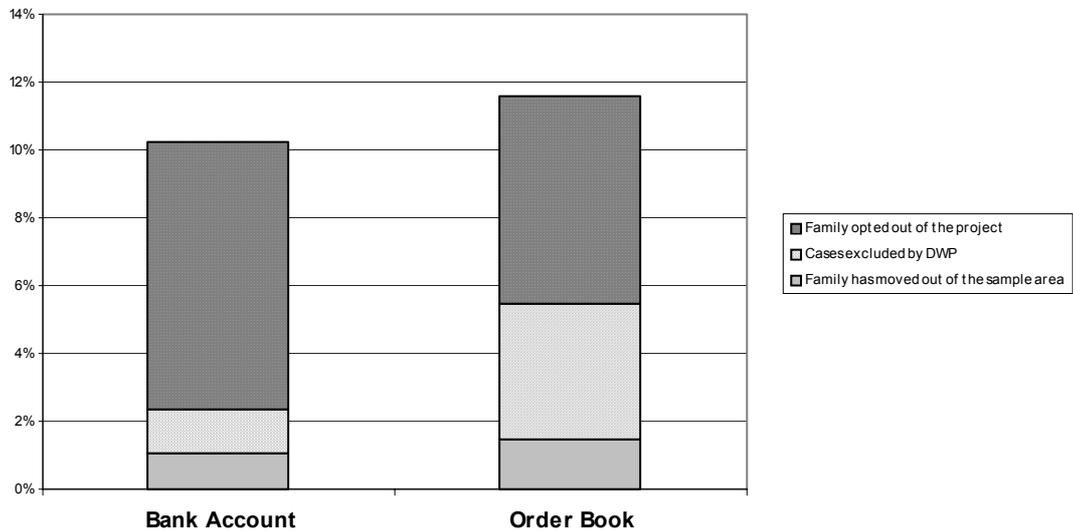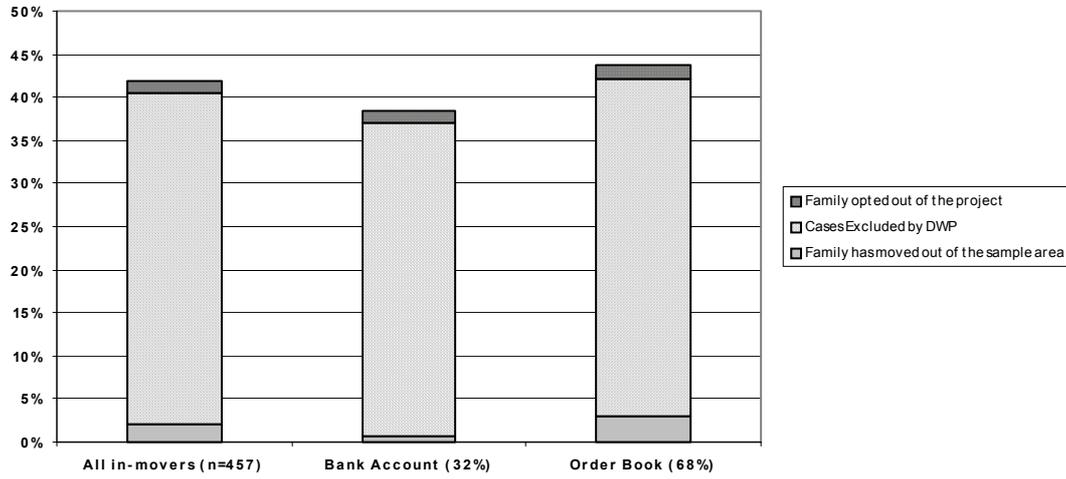**Exclusions from sample by title of CB claimant**

Legend:
- Family opted out of the project
- Cases exluded by DWP
- Family has moved out of the sample area

Categories (x-axis): MISS (36%), MS(5%), MRS (55%), MR (4%), Other (0%)

Figure 4
Exclusions by number of children in family

Legend:
- Family opted out of the project
- Cases exluded by DWP
- Family has moved out of the sample area

x-axis: no of children receiving CB in family
Categories: 1 (42%), 2 (33%), 3 (15%), 4 (6%), 5 (2%), 6 + (2%)

Figure 5
Exclusions by method child benefit paid

Legend:
- Family opted out of the project
- Cases excluded by DWP
- Family has moved out of the sample area

Categories: Bank Account, Order Book

Figure 6
Exclusions among in-movers detected by DWP between 7 and 8 months of age,
UK waves 8-15

Legend:
- Family opted out of the project
- Cases Excluded by DWP
- Family has moved out of the sample area

X-axis categories: All in-movers (n=457), Bank Account (32%), Order Book (68%)