

# The Millennium Cohort Study:

## Technical Report on Sampling

4th Edition

July 2007

Edited by  
**Ian Plewis**

With contributions from Lisa Calderwood, Denise Hawkes,  
Gareth Hughes and Heather Joshi.



### **Centre for Longitudinal Studies**

Bedford Group for Lifecourse and Statistical Studies  
Institute of Education, University of London



## Contents

Acknowledgements	4
Preface to the Fourth Edition	5
1. Introduction	6
2. The Millennium Cohort Study Population	7
3. Stratifying and Clustering the Population	8
4. Target Sample Sizes	12
5. Selecting the Sample	14
6. Obtaining the Sample	20
7. Final Sample, Sweep 1	23
8. Expected and Final Samples, Sweep 1	34
9. Sampling Errors for Sweep 1	36
10. Mobile Families and Unit Non-response, Sweep 1	39
11. Weighting to Adjust for Non-response Bias	44
12. Partial, Domain and Item Non-response	46
<b>Appendix 1:</b> Births, the Disappearing Target	48
<b>Appendix 2:</b> Further Information on Sampling Errors	53
<b>Appendix 3:</b> Statistical Models for Mobile Families	71
<b>Appendix 4:</b> Statistical Models for Unit and Partner Non-response	74
References	78

Centre for Longitudinal Studies  
Bedford Group for Lifecourse and Statistical Studies  
Institute of Education, University of London  
20 Bedford Way  
London  
WC1H 0AL

website: [www.cls.ioe.ac.uk](http://www.cls.ioe.ac.uk)

© Centre for Longitudinal Studies

ISBN 1 898453 62 4

The Centre for Longitudinal Studies (CLS) is one of five centres that comprise the Bedford Group for Lifecourse and Statistical Studies ([www.ioe.ac.uk/bedfordgroup](http://www.ioe.ac.uk/bedfordgroup)).

CLS is an ESRC Resource Centre and is devoted to the collection, management and analysis of large-scale longitudinal data. The Centre houses three internationally renowned birth cohort studies: the 1958 National Child Development Study (NCDS), the 1970 British Cohort Study (BCS) and the Millennium Cohort Study (MCS).

The views expressed in this work are those of the authors and do not necessarily reflect the views of the Economic and Social Research Council.

## **Acknowledgements**

This technical report could not have been produced without the help of a number of different people and organisations:

George Smith and Mike Noble, University of Oxford;  
Peter Goldblatt and colleagues, Office for National Statistics;  
Robert Beatty and colleagues, Northern Ireland Social Research Agency;  
Jason Bradbury, Katie Dodd and colleagues, Information Centre, Analytical Services Directorate, Department for Work and Pensions (formerly DSS);  
Andrew Shaw and colleagues, National Centre for Social Research;  
past and present colleagues in the Centre for Longitudinal Studies, especially John Bynner, Andy Cullis, Kevin Lynch, Mos Mojaddad and Mahmood Sadigh.

## **Preface to the Fourth Edition**

This is the fourth and final edition of a document that describes the design and sampling outcomes of the first sweep of the Millennium Cohort Study. It includes all the material from the third edition together with some minor amendments to take account of corrections incorporated into the dataset deposited at the Data Archive in June 2006. It also includes some details on multiple births and the cohort child's age at interview that were not in the third edition. Details about response after sweep one can be found in Plewis and Ketende (2007).

## 1. Introduction

1.1 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 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 (see Section 2) should have a known and non-zero probability of being included in the selected sample.

1.2 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.

## 2. The Millennium Cohort Study Population

2.1 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 24 November 2000 and 11 January 2002 (for Scotland and Northern Ireland, see 2.2), alive and living in the UK at age nine months, and eligible to receive Child Benefit at that age;

*and, after nine months:*

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

Although Child Benefit is, in principle, a universal benefit, in fact eligibility is governed by a set of rules that mean that families whose residency status is temporary (for example, members of foreign armed forces) or uncertain (for example, asylum seekers) are ineligible (see Section 6 for further discussion of this issue).

2.2 The birth dates for Scotland and Northern Ireland are 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 fieldwork in Scotland and Northern Ireland started a week earlier than originally envisaged and was then extended by a further six weeks in order to increase the sample size and thus to compensate for a birth rate that was lower than originally expected (see 4.4, 5.10 and Appendix 1).

2.3 The population includes:

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

2.4 The population excludes:

- (a) children who died before age 9 months (see Cullis, 2007);
- (b) UK-born children who emigrated from the UK before 9 months;
- (c) children not established as resident in the UK at age nine months.

2.5 This definition of the population meant that (for England and Wales) one third of the sample were 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.

### 3. Stratifying and Clustering the Population

3.1 The population was stratified by UK country - England, Wales, Scotland and Northern Ireland.

3.2 In order to meet the principle of adequately representing disadvantaged and ethnic minority children (see 1.1(2)), stratification by these variables was needed. Ideally, this stratification would have been at the individual child/family level but data on socio-economic and demographic variables of this kind are not available for the population at the individual level. Relevant data are, however, available at the electoral ward level and so the population of wards was stratified as described in (3.3) and (3.4). (The description of how wards were selected from these strata is deferred until Section 5.)

3.3 For England, the population was stratified, via the stratification of electoral wards extant on 1 April 1998, into three strata:

- 1) The '*ethnic minority*' stratum: children living in wards which, in the 1991 Census of Population, had an ethnic minority indicator of at least 30%. In other words, at least 30% of their total population fell into the two categories 'Black' or 'Asian'.
- 2) The '*disadvantaged*' stratum: children living in wards, other than those falling into stratum (1) above, which fell into the upper quartile (i.e. the poorest 25% of wards) of the ward-based Child Poverty Index (CPI; see 3.12) for England and Wales. The cut-off for the upper quartile was 38.4%. In other words, in the poorest 25% of wards in England and Wales, the CPI was at least 38.4%.
- 3) The '*advantaged*' stratum: children living in wards, other than those falling into stratum (1) above, which were not in the top quartile of the CPI. 'Advantaged' is therefore a relative term in this context.

3.4 For Wales, Scotland and Northern Ireland, there were just two strata:

- 1) The '*disadvantaged*' stratum: children living in wards (known as Electoral Divisions in Wales) extant in 1998 (1984 in Northern Ireland) that fell into the top part of the Child Poverty Index, the cut-off for this top part being determined by the upper quartile (38.4%) of the distribution of the CPI for England and Wales (see Table 3.1).
- 2) The '*advantaged*' stratum: children living in wards that were not in the top part of the Child Poverty Index.

3.5 It was not possible to have a 'poverty' cut-off point based on the overall distribution of the Child Poverty Index for all four UK countries, as the sample for England and Wales had to be selected before the CPI data became available for Scotland and Northern Ireland. As the criterion is essentially an arbitrary one, and as our real interest was in getting an



adequate representation of disadvantaged children rather than disadvantaged wards, this is not a serious drawback.

- 3.6 In addition to the explicit stratification described in (3.3) and (3.4), implicit stratification by region and ward size was introduced as a result of systematically sampling from ward lists (see 5.1 and 5.2).
- 3.7 The ethnic minority indicator for England was based on 1991 Census data, the only data available at the ward level. The data were converted from 1991 wards to apply to 1998 wards. The categories 'Black' and 'Asian' were defined as:
  - a) 'Black': Black Caribbean, Black African and Black Other as reported on the Census form;
  - b) 'Asian': Indian, Pakistani and Bangladeshi as reported on the Census form.
- 3.8 Population movements between 1991 and 2000, and the undercount of ethnic minorities in some parts of England in the 1991 Census (OPCS/GRO(S), 1993), mean that the indicator used - at least 30% of the total population of the ward falling into the categories 'Black' or 'Asian' - was subject to an unknown degree of error. Nevertheless, the indicator offered the best way of determining which wards were likely to have (and, indeed, did have – see 8.3) substantial numbers of births to ethnic minority mothers in 2000/01.
- 3.9 Two ethnic minority indicators could have been used, one for 'Blacks' and one for 'Asians'. However, as these are both heterogeneous groups, and as there are other groups, for example 'Chinese', excluded from both, it was decided to use just a single indicator.
- 3.10 It is estimated that, in Great Britain in 1995-97, about 10% of children aged under five were from ethnic minority groups (Schuman, 1999). However, population estimates for 1997 show that, whereas 7.2% of the population for England belongs to ethnic minority groups, the corresponding figures for Wales and Scotland are 1.2% and 1.1%. The figure for Northern Ireland, although not available for that year, is likely to be smaller still. Hence, stratification by ethnic minority group was adopted only for England (although children from ethnic minority families could, of course, be selected and identified in Wales, Scotland and Northern Ireland and in the advantaged and disadvantaged strata in England).
- 3.11 The ethnic minority indicator for England is, as expected, very skewed: the median is 0.51%, the upper quartile 1.61% but four wards have a value over 70%. Only 169 out of 8412 wards (2% of the total) have a value on the indicator above the chosen cut-off of at least 30%.

3.12 The Child Poverty Index (CPI) used for stratification (see 3.3) is defined as the percentage of children under 16 in an electoral ward living in families that were, in 1998, receiving at least one of the following benefits:

- Income Support;
- Jobseekers Allowance;
- Family Credit;
- Disability Working Allowance.

The first two of the above benefits were paid to people out of work, the second two to those in work but on low incomes.

3.13 The CPI is one component of the Index of Deprivation or ID2000 (Noble et al., 2000). The CPI was chosen as the stratifying factor for disadvantage, because its focus on child poverty is more closely related to the concerns of MCS than the full ID2000. The correlation between the CPI and the full ID2000, across 8412 English wards, is, however, high (0.93). In addition, the CPI was available for all four UK countries at the time of sampling whereas there was no ID2000 for Scotland and Northern Ireland at that time.

3.14 The CPI is, however, only one of a number of indicators of children's disadvantage that could have been used. It would, for example, have been possible to restrict it to children under five rather than to children under 16. This age range would have been closer to the age of the MCS sample at first interview but, because the index would then be based on smaller numbers, it would vary more from year to year. In fact, the correlation between the under five and under 16 versions of the CPI is 0.97 for England. Another possibility would have been to construct an index based just on out-of-work benefits, perhaps a better indicator of poverty, but also an index more strongly affected by the percentage of lone parents in the ward. Again, however, the correlation between the index based on the two out-of-work benefits and one based on all four benefits, for children under five in England, is high (0.96).

3.15 The distribution of the CPI, for all four UK countries, is given in Table 3.1.

**Table 3.1: Distribution of the Child Poverty Index by UK Country**

	Mean (SD)	Median	Upper Quartile	Lower Quartile	Number (%) of Disadvantaged <sup>(1)</sup> Wards
ENGLAND	26.7 (17.0)	22.4	37.4	13.1	1853 (22) <sup>(2)</sup>
WALES	34.5 (14.9)	34.1	44.0	23.6	333 (38)
SCOTLAND	32.6 (19.0)	30.2	44.9	18.0	426 (34)
NORTHERN IRELAND	39.4 (20.0)	37.9	53.4	24.0	275 (49)

<sup>(1)</sup> CPI > 38.4%.

<sup>(2)</sup> this figure excludes the 169 wards in the ethnic minority stratum, 134 of which are disadvantaged.

We see that Wales, Scotland and Northern Ireland appear to be more disadvantaged than England, Northern Ireland markedly so. About one third of the wards in Wales and Scotland are disadvantaged on the criterion given in (3.4) and almost a half in Northern Ireland.

- 3.16 It is important to be aware that both the ethnic minority indicator and the Child Poverty Index used for stratification purposes are aggregate measures. Not all births in ethnic minority wards will be to ethnic minority women. Similarly, not all births in disadvantaged wards will be to disadvantaged families and many births in advantaged wards will also be to disadvantaged families. In England in 1998, about 37% of disadvantaged families with a child under 16 were in advantaged wards, 54% were in disadvantaged wards and 10% in ethnic minority wards. We return to this point when discussing the selected sample in (5.8).
- 3.17 The wish to bring the broader socio-economic context into the analysis (see 1.1(3)), particularly as represented by the areas or local neighbourhoods that the children live in, and the need to keep field costs down, led to the decision to cluster the sample. Moreover, the chosen method of stratification - by characteristics of electoral wards - meant that using wards, rather than alternative geographical aggregates such as postcode sectors, was the most appropriate way to implement the clustering. In addition, the issues of measuring local context and reducing fieldwork costs pointed to the advantages of including all births in selected wards in the sample, rather than sub-sampling within wards.
- 3.18 Wards do, however, vary considerably in population size and therefore in the expected numbers of births in a year. Some wards have more than 600 births in a year whereas others, usually in rural areas, have as few as three or four. The variability in ward population is most marked in

England, but is found in all four UK countries. The very small wards, in particular, posed problems for the organisation of fieldwork and so small wards were combined to create 'superwards' with at least 24 expected births in a year.

3.19 The rules for combining wards were as follows:

- (a) link two or more geographically contiguous wards within the same stratum and within the same local authority or district to create a 'superward' of at least 24 births in a year. Ideally, but not necessarily, only small wards were to be combined;
- (b) if (a) was not possible then the criterion of contiguity was dropped;
- (c) if (b) was not possible then the criterion of the same district was dropped;
- (d) under no circumstances were wards from different strata to be combined.

3.20 In England, it was only within the advantaged stratum that wards had to be combined, resulting in a reduction from 6392 wards to 5289 advantaged wards and 'superwards'. The 532 advantaged and 333 disadvantaged wards in Wales were reduced to 345 and 274 respectively. In Scotland, the reductions were from 819 to 709 in the advantaged stratum and from 426 to 409 in the disadvantaged stratum. Finally, in Northern Ireland, there were 258 advantaged wards and 'superwards' (down from 291) and 242 (down from 275) in the disadvantaged stratum.

3.21 A drawback of any form of cluster sampling is that estimates are less precise than those obtained from simple random samples of the same size (see 5.3), and this loss of precision increases as cluster sizes increase. Hence, there was a case either for dividing large wards or for sub-sampling within them. It would not, however, have been easy to put either of these two strategies into practice. We would not have known the values of the CPI or the ethnic minority indicator for sub-divisions of wards, and sub-sampling after selecting wards would not have fitted easily into DWP procedures for capturing and forwarding addresses within wards (see Section 6). As already mentioned (see 3.6), there was implicit stratification by ward size (ranking by ward size before selection). This was done to avoid the possibility of selecting many large wards by chance (which would have led both to a loss of precision and to a lack of control over sample size). Therefore, it was decided not to do anything about large wards.

#### **4. Target Sample Sizes**

4.1 The original target for the achieved sample of children at the first contact was 15 thousand. If this figure had been divided between the four UK countries in proportion to their numbers of births, this would have resulted in samples of 12600 in England, 750 in Wales, 1200 in Scotland

and 450 in Northern Ireland. The sample sizes for the three smaller UK countries would then have been too small for useful analysis, especially bearing in mind the effects of sample attrition over time. Consequently, each of these countries was initially allocated a sample of 1500 children, leaving 10500 for England.

- 4.2 In order adequately to represent children from disadvantaged and ethnic minority backgrounds, the initial division of the estimated achieved sample across strata was half in advantaged wards and half in disadvantaged wards in Wales, Scotland and Northern Ireland, and half in advantaged wards and a quarter each in the ethnic and disadvantaged wards in England. It must be remembered that these proportions apply to ward characteristics and not to family characteristics (see 3.16).
- 4.3 After the initial allocations had been agreed, additional resources were made available to supplement the samples in different ways in each of the four countries:
  - (a) in England, this meant an extra 35 disadvantaged wards could be selected (to act as controls for the national evaluation of the Sure Start programme);
  - (b) in Wales the sample was doubled (to 3000 children) with all the additional 1500 to be selected from disadvantaged wards;
  - (c) in Scotland the sample was boosted by 1000, to be divided equally between advantaged and disadvantaged wards;
  - (d) in Northern Ireland a further 500 children were to be selected, all from disadvantaged wards.

The outcome from all these supplements was a target sample of 20646 children, 13146 in England, 3000 in Wales, 2500 in Scotland and 2000 in Northern Ireland.

- 4.4 In order to calculate the number of wards required in each country for each stratum, two assumptions were needed. The first was the expected births per ward per stratum and here we used figures that were based on child population estimates (in turn based on Child Benefit records), essentially averaging over births between 1994 and 1998 in England and Wales and between 1995 and 1999 in Scotland and Northern Ireland. As it turned out, these numbers were too high because of a downturn in births in the late 1990s that persisted up to 2001 when fertility reached an all-time low (see Appendix 1). The second assumption was the expected response rate and here we assumed that these would vary by stratum but not by country; the assumptions were 75% in the advantaged wards, 70% in the disadvantaged wards and 65% in the ethnic minority wards. These assumed response rates combine potential losses before and during fieldwork.
- 4.5 Table 4.1 shows the number of wards that needed to be selected, and the expected sample size in each stratum, *before* the wards were

actually selected. Thus, for the stratum 'England: Advantaged', the expected sample of children was calculated as  $(110 \times 68 \times 0.75) = 5610$ . The total expected sample was, at that stage, 21076: 13411 in England, 2998 in Wales, 2529 in Scotland and 2138 in Northern Ireland. These numbers differ slightly from those given in (4.3) because of rounding error. The expected samples for the advantaged and disadvantaged strata were 8480 and 10163.

**Table 4.1: Required Number of Sample Wards by Stratum and Country**

	<b>Number of sample wards</b>	<b>Expected births per ward (SD)</b>	<b>Expected response rate</b>	<b>Expected sample size before ward selection</b>
ENGLAND: Advantaged	110	68 (43)	75%	5610
ENGLAND: Disadvantaged	71	108 (63)	70%	5368
ENGLAND: Ethnic	19	197 (117)	65%	2433
WALES: Advantaged	23	48 (27)	75%	828
WALES: Disadvantaged	50	62 (40)	70%	2170
SCOTLAND: Advantaged	32	52 (18)	75%	1248
SCOTLAND: Disadvantaged	30	61 (22)	70%	1281
N.IRELAND: Advantaged	23	46 (20)	75%	794
N.IRELAND: Disadvantaged	40	48 (27)	70%	1344

## 5. Selecting the Sample

- 5.1 The sample was selected separately in each stratum in each country. In England, the populations were ordered first by the nine standard English regions and then, within region, by the ward size in descending order. The same procedure was used for Scotland, based on four regions (South, Central, North East, North West). The smaller countries of Wales and Northern Ireland were only ordered by ward size.

- 5.2 The sample was selected systematically within each stratum and country, the sampling interval being determined by the ratio of the number of wards in the populations to the number of wards required in the samples. This method of selection, when combined with the ordering described in (5.1), is slightly more efficient than just sampling randomly, producing some gains in precision as well as controlling the sample size. Note that the sample size of *children* was not fixed in advance. Rather it is a random variable that depends on which particular wards are selected. Ranking by ward size reduces the variability in the sample size to acceptable limits.
- 5.3 The sample is a disproportionately stratified cluster sample. The disproportionality means that the sample is not self-weighting and so weighted estimates of means, variances etc. are needed (see 5.9, 5.10). The clustering implies that observations are not independent and so allowance must be made for the dependence so induced when sampling errors are computed (Section 9). It was likely that the design effects for the sample would be greater than one. In other words, the sample would be somewhat less precise than a simple random sample of the same size would have been, although this depends on how far the gains from stratification and systematic selection are offset by the losses from clustering which, in turn, would vary across measures.
- 5.4 Table 5.1 compares the original target sample (4.3) and the expected sample sizes before selection (Table 4.1) with those based on the actual wards selected. The differences between the expected sample sizes before and after ward selection are small. The expected totals are 13375 (England), 3116 (Wales), 2528 (Scotland) and 2084 (Northern Ireland).

**Table 5.1: Expected Sample Size by Stratum and Country**

	<b>Number of sample wards</b>	<b>Original target sample</b>	<b>Expected sample size before ward selection</b>	<b>Expected sample size after ward selection</b>
ENGLAND: Advantaged	110	5250	5610	5511
ENGLAND: Disadvantaged	71	5271	5368	5258
ENGLAND: Ethnic	19	2625	2433	2606
WALES: Advantaged	23	750	828	897
WALES: Disadvantaged	50	2250	2170	2219
SCOTLAND: Advantaged	32	1250	1248	1243
SCOTLAND: Disadvantaged	30	1250	1281	1285
N.IRELAND: Advantaged	23	750	794	762
N.IRELAND: Disadvantaged	40	1250	1344	1322
TOTAL	398	20646	21076	21103

5.5 Table 5.2 gives the regional breakdown for England. Out of a total of 354 Local Authority districts, 152 are represented in the sample, 36 of these twice, 3 three times and 2 four times.



**Table 5.2: Sample Wards in England: Regional Breakdown**

Region	Advantaged	Disadvantaged	Ethnic	Total
SW	15	6	0	21
SE	26	7	1	34
LONDON	8	10	11	29
EAST ANGLIA	18	5	1	24
EAST MIDLANDS	12	8	1	21
WEST MIDLANDS	10	6	2	18
YORKS	6	7	1	14
NW	11	13	2	26
NE	4	9	0	13

5.6 In Wales, all but two of the 22 Local Authority districts are represented in the sample, in Scotland all but 6 of the 36 Council Authority districts are represented, and in Northern Ireland all but four of the 26 districts. The regional breakdown for Scotland is given in Table 5.3.

**Table 5.3: Sample Wards in Scotland: Regional Breakdown**

Region	Advantaged	Disadvantaged	Total
SOUTH	6	5	11
CENTRAL	15	19	34
NE	6	4	10
NW	5	2	7

5.7 It was also possible to get an estimate of the expected number of advantaged and disadvantaged families in the sample. These estimates do, however, rest on a number of assumptions:

- (a) a family is disadvantaged if and only if it is receiving at least one of the benefits listed in (3.12);
- (b) the prevalence of families with a nine month old child receiving at least one of the benefits in (3.12) is the same as the prevalence for families with at least one child under 16;
- (c) the response rates assumed to apply to advantaged etc. areas also apply to advantaged etc. families;
- (d) the applicability of data from the 1990s on expected births, numbers of ethnic minorities and the CPI to the data collection period in 2000/2001.

5.8 Table 5.4 gives the estimates, and the percentages of disadvantaged families, separately for each country. It can be seen that the expected number of disadvantaged families is always smaller than the expected sample size in disadvantaged areas (because, for many wards defined as disadvantaged, the majority of families are not disadvantaged as defined in (5.7a)). The population estimates of disadvantaged families for England, Wales, Scotland and Northern Ireland are 34%, 39%, 35%, 40% and so the expected sample proportions (after allowing for expected non-response) given in Table 5.4 are somewhat higher in each

country, less so in Scotland where the sample boost was not concentrated in disadvantaged areas. The figures for England include both advantaged and disadvantaged Black and Asian families (because socio-economic status and ethnic status are not mutually exclusive at the family level). The expected sample size for Black and Asian families in England was 1524. This is 11.4% of the total, compared with about 9% in the population.

**Table 5.4: Expected Sample Size by Family Type**

	<b>Expected sample size</b>	<b>Percent within country</b>
ENGLAND: Adv. families	8168	61
ENGLAND: Disadv. families	5207	39
WALES: Adv. families	1806	58
WALES: Disadv. families	1310	42
SCOTLAND: Adv. families	1609	64
SCOTLAND: Disadv. families	919	36
N.IRELAND: Adv. families	1165	56
N.IRELAND: Disadv. families	919	44

5.9 The sample is not self-weighting; children born to families living in disadvantaged areas, for example, have a greater chance of selection than families living in advantaged areas. Table 5.5 shows how these selection probabilities vary across strata. The first column of Table 5.5 gives the number of wards and 'superwards' in each stratum,  $h$  ( $h = 1..9$ ). The next three columns apply to each UK country separately; comparisons across countries would be meaningless. They give the prevalence of each stratum type ( $P_h$ ); the sampling fraction applied to each stratum ( $f_h$ ); and the weight  $w_h$  (proportional to the inverse of the sampling fraction). The weights - which have been scaled to sum to the number of wards selected in each country so that the mean weight is one - apply to each ward and hence to each child living within those wards (because there is no sub-sampling within wards). The next two columns apply to the UK as a whole and can therefore be compared

across all rows of Table 5.5. Here the weights have been scaled to the number of wards selected in the UK (i.e. 398). We can see from Table 5.5 that, although disadvantaged wards are over-represented in England, English disadvantaged wards are slightly under-represented in the sample as a whole (because their weight is greater than one). All the UK-wide weights for the three smaller UK countries are less than one, reflecting the boosts to the sample to get more reliable estimates for those countries. The final two columns refer just to GB and can therefore be compared across rows, omitting Northern Ireland.

5.10 The extension of the sampling period in Scotland and Northern Ireland means that births in late November, December and early January are over-represented. It is possible to correct for this by giving all births between 24 November and 11 January a supplementary weight of 0.5 (see 10.17).

**Table 5.5: Sampling Fractions and Weights across Strata**

	Number of wards and 'superwards' ( $N_h$ )	Each UK country			UK as a whole		GB	
		$P_h$	$f_h$	$w_h$	$P_h$	$w_h$	$P_h$	$w_h$
ENGLAND: Advantaged	5289	0.723	0.0208	1.32	0.554	2.00	0.585	1.78
ENGLAND: Disadvantaged	1853	0.253	0.0383	0.71	0.194	1.09	0.205	0.97
ENGLAND: Ethnic	169	0.023	0.112	0.24	0.018	0.37	0.019	0.33
WALES: Advantaged	345	0.557	0.067	1.77	0.036	0.62	0.038	0.55
WALES: Disadvantaged	274	0.443	0.182	0.65	0.029	0.23	0.030	0.20
SCOTLAND: Advantaged	709	0.634	0.045	1.23	0.074	0.93	0.078	0.82
SCOTLAND: Disadvantaged	409	0.366	0.073	0.75	0.043	0.57	0.045	0.51
N.IRELAND: Advantaged	258	0.516	0.089	1.41	0.027	0.47	n.a.	
N.IRELAND: Disadvantaged	242	0.484	0.165	0.76	0.025	0.25	n.a.	

- 5.11 The weights should be applied when estimating a mean for the sample, say. In other words, the weighted mean for variable  $y$  is:

$$\overline{y}_w = \frac{\sum_h \sum_i w_h y_{hi}}{\sum_h w_h x_h}$$

where  $i$  ( $i = 1..x_h$ ) indexes the elements in a stratum so  $x_h$  is the sample size for stratum  $h$ . The appropriate weights  $w_h$  from Table 5.5 should be used. Calculation of standard errors for  $\overline{y}_w$ , taking into account the sample design, is discussed in Section 9.

## 6. Obtaining the Sample

- 6.1 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 then part of the Department of Social Security, later the Department for Work and Pensions (DWP), but transferred to the Inland Revenue on 1 April 2003. Child Benefit is a universal provision, payable (usually to the mother) from the child's date of birth. Child Benefit claims cover virtually all of the child population except those ineligible due to recent or temporary immigrant status. 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. Such 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:

- (i) there had been a child death in the family in the last five years;
- (ii) 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;
- (iii) the child was in another person's award;
- (iv) the child had been taken into care;
- (v) the family had previously been selected for the DWP-sponsored Families and Children Survey (FACS);
- (vi) there was an unknown exclusion code entered on the CB non-customer record.

- 6.2 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 aged seven months. Upon receipt of the Child Benefit GMS data, the ASD Information Centre used a computer package called QAS (Quick Address System) to set back any postcodes listed (for England, Scotland and Wales) to their 1998

postcodes. They were set to 1998 as this was the year used to define the wards. If there was a missing postcode within the data, then QAS attempted 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 was only a partial postcode listed within the data, then QAS attempted to match a full 1998 postcode to that address. In Northern Ireland, however, 1984 wards were used.

- 6.3 The CB 100% GMS database is held not by ward but by postcode. Once QAS had matched as many postcodes as possible, the Millennium Cohort wards were 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. This method could not be used in Northern Ireland because 1984 wards were used and so addresses without at least some postcode information could not be matched. 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.
- 6.4 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, each covering four weeks of births, in England and Wales and 14 (labelled Waves 4 to 17) in Scotland and Northern Ireland, with the last (additional) wave in Scotland and Northern Ireland covering seven weeks of births.
- 6.5 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 presumed willing 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 for the first Sweep at the National Centre for Social Research (NatCen).

6.6 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 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.

6.7 There was nothing we could do about exclusions at stages (i) and (ii). ASD's best estimate of Child Benefit coverage for 2000 in Great Britain by 7 months of age is 97.2%. But, apart from the possibility of eligible families being too rich or too ill-informed to claim, most of the children not claimed for are ineligible as being the children of non-nationals with temporary or unconfirmed residence status, such as foreign armed forces, overseas students, and recent immigrants including asylum seekers. Their eligibility for a long-term survey as well as for Child Benefit is arguable, although it is noteworthy that they are excluded. If they are not considered part of the eligible population, a working estimate of loss at stages (i) and (ii) is 3% for Northern Ireland because of the difficulties of matching current postcodes to 1984 wards but zero for Great Britain. Every effort was made to keep exclusions at stages (iii) and (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. Nevertheless, a substantial number of in-migrants were not found at this stage. Many were, however, picked up at sweep 2 although only in England (see Plewis and Ketende, 2007).

6.8 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' requirement 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 substantial initial non-response among those sections of the population - the disadvantaged and ethnic minorities - that we particularly wanted to over-represent. Hence, this method of obtaining the sample was rejected.

- 6.9 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 (see 6.7).
- 6.10 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%).

## **7. Final Sample, Sweep 1**

- 7.1 There were two stages in the route from establishing the size of the initial sample in the selected wards to obtaining a sample of respondents. The details of the first stage are described in (6.1) to (6.5) and the accounts are presented in Table 7.1. It should be noted that the 'total' columns are not weighted and therefore give rates for the sample as designed rather than the rates that would have been obtained if all strata had been sampled proportionately.

**Table 7.1: From Child Benefit Sample to Issued Sample by Stratum and Country**

	England				Wales			Scotland			N. Ireland			UK
	Ad.	Disad.	Eth.	Total	Ad.	Disad.	Total	Ad.	Disad.	Total	Ad.	Disad.	Total	All
Child Benefit sample	6364	6678	3906	16948	1074	2784	3858	1579	1748	3327	1109	1959	3068	27201
Exclusions	100 (1.6%)	182 (2.7%)	149 (3.8%)	431 (2.5%)	19 (1.8%)	77 (2.8%)	96 (2.5%)	27 (1.7%)	37 (2.1%)	64 (1.9%)	41 (3.7%)	127 (6.5%)	168 (5.5%)	759 (2.8%)
Opt-outs	490 (7.7%)	455 (6.8%)	253 (6.5%)	1198 (7.1%)	70 (6.5%)	181 (6.5%)	251 (6.5%)	116 (7.3%)	111 (6.4%)	227 (6.8%)	91 (8.2%)	119 (6.1%)	210 (6.8%)	1886 (6.9%)
Moved out (DWP)	58	103	49	210	8	35	43	19	34	53	10	27	37	343
Moved in (Health Visitors)	32	8	6	46	2	4	6	2	2	4	0	0	0	56
Issued sample (to field)	5748 (91%)	5946 (90%)	3461 (90%)	15155 (90%)	979 (92%)	2495 (91%)	3474 (91%)	1419 (91%)	1568 (91%)	2987 (91%)	935 (85%)	1629 (84%)	2564 (85%)	24180 (90%)



## Notes on Table 7.1

### *Child Benefit sample*

- (i) The families picked by the Child Benefit GMS scan at seven months (6.2). The Child Benefit sample for Northern Ireland is inflated by 3% to allow for the difficulties of matching addresses to 1984 wards (see 6.3).
- (ii) It includes those 'new' in-migrants picked up at the eight month scan from Wave eight (see 6.4). There were 518 of these cases in the UK.
- (iii) The Child Benefit sample differs from the *target* sample to the extent that there is migration into or out of the sampled wards and non-claiming of Child Benefit by families eligible to be included in MCS.
- (iv) It refers to families rather than to children. There were, however, 11 families that were issued twice, once for child A in an early wave and later for a subsequent child B in a later wave. These families are therefore counted twice.

### *Exclusions*

- (i) All 'sensitive cases' as defined in (6.1).
- (ii) 43% of the 'new' in-migrants (6.4) were subsequently excluded, mostly by virtue of being in correspondence with DWP but not merely about change of address.
- (iii) Percentage based on Child Benefit sample.

### *Opt-outs*

- (i) By parent(s) in response to the letter from DWP.
- (ii) Percentage based on Child Benefit sample.

### *Moved out (DWP)*

- (i) Found by DWP to have moved out of the sampled wards between the seven and eight month scans, and are therefore ineligible for the study.
- (ii) The rate, based on the Child Benefit sample, was 1.3% for the UK as a whole.

### *Moved in (Health Visitors)*

Notifications from Health Visitors to CLS, not also found by DWP.

### *Initial sample (not shown in table)*

- (i) The initial sample is the Child Benefit sample plus 'moved in (Health Visitors)'.  
(ii) The 'initial eligible' sample is the initial sample minus those found by DWP to have moved out

### *Issued sample (to field)*

- (i) This includes 293 of the 518 'new' in-migrant cases.
- (ii) The percentage is based on the initial eligible sample interpretable as the sample loss rate prior to field operations if all migration is accounted for by the 'moved out' and 'moved in' data. If 'new' in-migrants had been included from Wave 1 in all areas then, all other things being equal, the rates of loss would have been marginally higher (as these families were more likely to be excluded by DWP) although

the total number issued to field would, of course, also have been larger, proportionately more in England and Wales than in Scotland and Northern Ireland because fieldwork in Scotland and Northern Ireland started later (see 2.2).

- 7.2 Table 7.1 shows that about 10% of the initial eligible sample were lost before field operations started. The percentage is higher in Northern Ireland, partly because of the matching problem discussed earlier but also because exclusions were higher than elsewhere in the UK apart from the ethnic wards in England. Exclusions were generally higher in disadvantaged wards but opt-outs tended to be higher in advantaged wards. Only in the English advantaged wards did the numbers found by the Health Visitors come close to the numbers of those detected by DWP as moved out.
- 7.3 The second stage – from issued sample to final sample – is set out in Table 7.2. As for Table 7.1, the totals are not weighted.

**Table 7.2: From Issued Sample to Final Sample by Stratum and Country**

	England				Wales			Scotland			N. Ireland			UK
	Ad.	Disad.	Eth.	Total	Ad.	Disad.	Total	Ad.	Disad.	Total	Ad.	Disad.	Total	All
Issued sample	5748	5946	3461	15155	979	2495	3474	1419	1568	2987	935	1629	2564	24180
Ineligible	129	90	54	273	11	42	53	39	30	69	9	13	22	417
ELIG	95%	94%	94%	95%	97%	95%	96%	95%	94%	94%	97%	96%	96%	95%
Uncertain eligibility	255	470	298	1023	34	150	184	60	108	168	50	100	150	1525
Unproductive	747	864	715	2326	102	375	477	175	239	414	153	316	469	3686
REF1	12%	13%	16%	14%	10%	14%	13%	12%	13%	13%	15%	15%	15%	14%
Productive	4617	4522	2394	11533	832	1928	2760	1145	1191	2336	723	1200	1923	18552
Complete households	4274	4075	1977	10326	757	1764	2521	1039	1059	2098	655	1014	1669	16614
Partial households	343 (7.4%)	447 (9.9%)	417 (17%)	1207 (11%)	75 (9.0%)	164 (8.5%)	239 (8.7%)	106 (9.3%)	132 (11%)	238 (10%)	68 (9.4%)	186 (16%)	254 (13%)	1938 (10%)

## Notes on Table 7.2

*Issued sample*: see Table 7.1.

### *Ineligible*

This group includes those who moved to an ineligible address, i.e. a non-sampled ward, those where there was an error in the Child Benefit record, and if the baby had died.

### *ELIG*

ELIG is the eligibility rate of the sample issued to field. This is the ratio of cases known or estimated to be eligible to all issued cases as defined by Lynn et al. (2001). Estimated eligibility rates are used for cases with uncertain eligibility as explained below.

### *Uncertain eligibility*

This includes four sub-groups of cases with their UK sizes and estimated eligibility rates in brackets:

- (i) Not issued to an interviewer (46; 98.2%).
- (ii) Unable to establish if resident at an address in a sampled ward (224; 98.2%).
- (iii) Unable to confirm residence at eligible address (5; 98.2%).
- (iv) Moved – current address could not be ascertained (1250; 33.2%).
- (v) The eligibility rates for sub-groups (i) to (iii) are determined by the ratio of cases known to be eligible to cases where eligibility is known, for the UK sample, i.e (productive + unproductive)/(productive + unproductive + ineligible).
- (vi) The eligibility rate for sub-group (iv) is an estimate based on an analysis of movers. It is half the known eligibility rate of all movers for whom an address could be ascertained.

### *Unproductive*

This includes three sub-groups of cases with their UK sizes in brackets:

- (i) Non-contact (305).
- (ii) Refusal (3097).
- (iii) Other non-response (284).

### *REF1*

REF1 is the refusal rate in the field. This is the ratio of refusals to all issued cases known or estimated to be eligible as defined by Lynn et al. (2001). Estimated eligibility rates are used for cases with uncertain eligibility as explained above.

### *Productive*

- (i) Some data obtained (beyond the household grid). A household is a group of people at the same address who either share at least one meal a day or share living accommodation.
- (ii) From the 11 families issued twice, six were included twice in the final sample and three were included once.

### *Complete households*

Main respondent plus partner where there was an eligible partner (including proxy data) plus main respondent, no eligible partner.

### *Partial households*

- (i) Final sample minus complete households.
- (ii) Percentage based on productive cases.

- 7.4 Many more movers-out were discovered in the field than were found by DWP. They are more prevalent in disadvantaged areas (this is also so for those found by DWP). This could reflect net out-migration from these areas by families with young children but it is known that there is a group of in-movers that were not picked up by the study. This group is smaller than would have been the case had the strategy introduced from Wave eight (see 6.2) not been used. Further evidence on this issue became available from sweep 2 (see Plewis and Ketende, 2007).
- 7.5 There were more unproductives in the disadvantaged areas and ethnic areas in England than in advantaged areas. On the other hand, opt-outs (another form of refusal) were more likely in advantaged areas (Table 7.1).
- 7.6 The pattern for the partial interviews is similar to the pattern for refusals, being a little more common in disadvantaged wards (except in Wales) and notably more common in ethnic wards in England and in disadvantaged wards in Northern Ireland.
- 7.7 Table 7.3 combines the information in Tables 7.1 and 7.2 and shows that the overall refusal rate is slightly higher in Northern Ireland and for the ethnic wards in England.

**Table 7.3: From Initial Sample to Final Sample by Stratum and Country**

	<b>England</b>				<b>Wales</b>			<b>Scotland</b>			<b>N. Ireland</b>			<b>UK</b>
	Ad.	Disad.	Eth.	Total	Ad.	Disad.	Total	Ad.	Disad.	Total	Ad.	Disad.	Total	All
Initial Sample	6396	6686	3912	16994	1076	2788	3864	1581	1750	3331	1109	1959	3068	27257
Ineligible	187	193	103	483	19	77	96	58	64	122	19	40	59	760
Uncertain eligibility	355	652	447	1454	53	227	280	87	145	232	123	284	407	2373
Unproductive	1237	1319	968	3524	172	556	728	291	350	641	244	435	679	5572
REF2	19%	19%	21%	20%	16%	19%	18%	19%	19%	19%	21%	22%	20%	20%
Productive	4617	4522	2394	11533	832	1928	2760	1145	1191	2336	723	1200	1923	18552
Mean cluster size	42	64	126	58	36	39	38	36	40	38	31	30	31	47

### Notes on Table 7.3

#### *Ineligible*

'Ineligible' from Table 7.2 plus 'moved out (DWP)' from Table 7.1.

#### *Uncertain eligibility*

- (i) 'Uncertain eligibility' from Table 7.2 plus 'exclusions' from Table 7.1 plus cases assumed missing from the Child Benefit sample in Northern Ireland.
- (ii) The eligibility rate for the exclusions from the Child Benefit sample plus cases assumed missing from the Child Benefit sample in Northern Ireland is defined in the same way as for Table 7.2 – see note (v) ( $= (18553 + 5571)/(18553 + 5571 + 760) = 0.969$  for the UK).

#### *Unproductive*

- (i) 'Unproductive' from Table 7.2 plus 'opt-outs' (assumed eligible) from Table 7.1.
- (ii) REF2 is the overall refusal rate, based on the initial sample and defined in the same way as REF1 using the two eligibility rates given previously.

#### *Productive*

As in Table 7.2.

#### *Mean cluster size*

The number of productive cases divided by the selected number of clusters as given in Table 5.1.

7.8 Table 7.4 gives a set of rates that describe response.

**Table 7.4: Response Rates (%) by Stratum and Country**

	<b>England</b>				<b>Wales</b>			<b>Scotland</b>			<b>N. Ireland</b>			<b>UK</b>
	Ad.	Disad.	Eth.	Total	Ad.	Disad.	Total	Ad.	Disad.	Total	Ad.	Disad.	Total	All
RR1	76	72	66	72	79	72	74	77	74	75	68	64	66	72
RR2	84	81	74	81	88	81	83	85	81	83	80	77	78	81
RR2(W)	84	81	74	82	88	81	85	85	81	83	80	77	78	82
CON	98	95	94	96	98	96	97	98	95	96	96	93	95	96
COOP	87	85	79	84	89	85	86	87	85	86	83	82	83	85



## Notes on Table 7.4

### *RR1*

- (i) This is the overall response rate to the study defined as the ratio of productive cases to all cases in the Child Benefit sample known or estimated to be eligible (Lynn et al., 2001). This is calculated from Table 7.3.
- (ii) It assumes that the numbers found by DWP at eight months equals the numbers in the wards at nine months. This is, however, unrealistic for two reasons. First, for Waves one to seven, we only know the numbers in the selected wards at seven months. If the system of finding new movers had been introduced at Wave one then the numbers in the denominator of this rate would have been higher. The numerator would also have been higher but, as we know that (a) the rate of exclusions for the new movers was higher and (b) fewer of them were interviewed, this rate is marginally too high. Second, we know from Sweep 2 that, at least for England, there are families who were living in the sampled wards at nine months that were not picked up from the Child Benefit register.

### *RR2*

This is the response rate in the field, based on Table 7.2 and defined as the ratio of productive cases to all cases known or estimated to be eligible in the issued sample (Lynn et al., 2001). This rate describes the success of the field operations.

### *RR2W*

This is the weighted version of RR2, allowing for varying selection probabilities across strata (see Table 5.5). It differs from RR2 only for the four countries and the UK as a whole.

### *CON*

This is the contact rate – the proportion of all cases in which a household member was reached by the interviewer with whom there was contact in person.

### *COOP*

This is the cooperation rate – the number of productive cases as a proportion of cases who were contacted during the fieldwork period.

7.9 The expected response rates (4.4) are probably best compared with RR1. We see that the expectations were a little pessimistic in Great Britain but noticeably optimistic in Northern Ireland where DWP exclusions and refusal rates were higher. The field response rates, represented by RR2 and RR2W, generally exceed 80% except in the ethnic wards in England and the disadvantaged wards in Northern Ireland.

7.10 The cooperation rates are generally about 3% higher than RR2. The contact rates are very high and vary only a little from stratum to stratum.

## 8. Expected and Final Samples, Sweep 1

8.1 Table 8.1 reproduces Table 5.1 but with an extra column showing the final sample of children (not families) in each stratum. The percentages are of the original target size. For the UK as a whole, the final sample is 91% of the original target, the difference mainly due to the decline in fertility discussed in Appendix 1. There is, however, some variation by stratum, with a shortfall of 13% in the disadvantaged wards of England and Wales but an excess of 13% in the Welsh advantaged wards. The figures for Scotland and Northern Ireland are close to the targets as a result of extending the period of fieldwork.

**Table 8.1: Expected and Final Sample Sizes of Children by Stratum and Country**

	Number of sample wards	Original target sample size	Expected sample size before ward selection	Expected sample size after ward selection	Final sample size <sup>(1)</sup>
ENGLAND: Advantaged	110	5250	5610	5511	4687 (89%)
ENGLAND: Disadvantaged	71	5271	5368	5258	4592 (87%)
ENGLAND: Ethnic	19	2625	2433	2606	2416 (92%)
WALES: Advantaged	23	750	828	897	844 (113%)
WALES: Disadvantaged	50	2250	2170	2219	1954 (87%)
SCOTLAND: Advantaged	32	1250	1248	1243	1163 (93%)
SCOTLAND: Disadvantaged	30	1250	1281	1285	1207 (97%)
N.IRELAND: Advantaged	23	750	794	762	735 (98%)
N.IRELAND: Disadvantaged	40	1250	1344	1322	1220 (98%)
TOTAL	398	20646	21076	21103	18818 (91%)

Note

<sup>(1)</sup> The sample of children includes 246 sets of twins and 10 sets of triplets.

8.2 Table 8.2 reproduces Table 5.4 but with an extra column showing the final sample sizes by family type. Here, 'disadvantaged' means receiving at least one of:

- (i) Working Families Tax Credit
- (ii) Disabled Persons Tax Credit
- (iii) Income Support
- (iv) Jobseekers Allowance

Note that the first two of these are different from the corresponding benefits (Family Credit and Disability Working Allowance) used to define the Child Poverty Index (3.12) and so the data are not directly comparable (the eligibility criteria for (iii) and (iv) might also have changed). The numbers in each category in Table 8.2 are lower for the reasons already discussed but the proportions of disadvantaged families in each country are somewhat higher than expected reflecting the relatively greater fertility decline in advantaged wards (see Appendix 1). Note that these are unweighted data so inter-country comparisons of the proportions are not valid. (See Table A2.11 for more detail.)

**Table 8.2: Expected and Final Sample Sizes by Family Type**

	<b>Expected sample size</b>	<b>Final sample size<sup>(1)</sup></b>
ENGLAND: Adv. Families	8168 (61%)	6680 (58%)
ENGLAND: Disadv. families	5207 (39%)	4820 (42%)
WALES: Adv. Families	1806 (58%)	1386 (50%)
WALES: Disadv. families	1310 (42%)	1371 (50%)
SCOTLAND: Adv. Families	1609 (64%)	1437 (62%)
SCOTLAND: Disadv. families	919 (36%)	892 (38%)
N.IRELAND: Adv. Families	1165 (56%)	1050 (55%)
N.IRELAND: Disadv. families	919 (44%)	855 (45%)

Note

<sup>(1)</sup> This table excludes 61 families for whom benefit status was not known.

- 8.3 There are 3104 cohort members in England who are non-white, 27% of the total. This is a much larger proportion than the expected 11% (see 5.8). The percentage of cohort members in the ethnic wards who are non-white is 83%, far greater than the 45% expected, based on 1991 Census data for all families, although that number does only refer to Black and Asian families. Also, it covers all age groups rather than infants among whose numbers there have been divergent trends by ethnic group.
- 8.4 The great majority of cohort members (75%) were age nine months at the time of interview; 2.9% were age eight months; 19% were 10 months old and 2.6% were age 11 or 12 months.

## 9. Sampling Errors for Sweep 1

- 9.1 As described in Section 3, the Sweep 1 sample is stratified and clustered and these design features must be reflected in the calculation of sampling errors. Sampling errors are presented for a range of variables for each of the nine strata (when clustering is relevant but the sampling weights are not), for each of the four countries of the UK (when stratification and clustering are relevant and the left hand column of weights,  $w_h$ , in Table 5.5 should be used) and for the UK as a whole (when the second column of weights,  $w_h$ , in Table 5.5 should be used).
- 9.2 The following formulae were used for computing sampling errors for any variable  $y$ . They are based on Kish (1967) and computed using the *svy:mean* procedures in STATA (StataCorp, 2005).

### (a) Within stratum h

Stratum mean:

$$\bar{y}_h = r_h = y_h / x_h$$

where  $y_h$  is total for  $y$  in stratum  $h$  and  $x_h$  is sample size for stratum  $h$ .

Note that  $x_h$  is not fixed. It is a random variable that depends on which clusters were selected and so  $r_h$  is a ratio estimator.

Stratum sampling variance:

$$s_h^2 = ((1 - f_h) / x_h^2) (m_h / (m_h - 1)) \sum_{a=1}^{m_h} z_a^2$$

where there are  $m_h$  clusters within a stratum  $h$ ,  $(1 - f_h) = (1 - (n_h / N_h))$  is the finite population correction with  $n_h$  the number of sampled clusters and  $N_h$  the number of clusters in the population in stratum  $h$ ;  $z_a = y_a - r_h x_a$  and where  $y_a$  is total for  $y$  in cluster  $a$  in stratum  $h$  and  $x_a$  is sample size in cluster  $a$  in stratum  $h$ .

(b) Between strata within UK country

Country mean:

$$\bar{y} = r = \sum_h w_h y_h / \sum_h w_h x_h$$

where  $w_h$  is sampling weight, inversely proportional to  $f_h$ .

Country sampling variance:

$$s^2 = 1 / \sum_h (w_h x_h)^2 \sum_h [(1 - f_h) / (m_h - 1) \{ m_h \sum_a z_a^2 - z_h^2 \}]$$

where  $z_h = \sum_a z_a$  and  $z_a = (w_h y_a - r w_h x_a)$

(c) UK

As for (b) with appropriate  $w_h$ .

Note that these formulae do not take into account the implicit stratification introduced by ordering by region within stratum, and by size within region (5.1). However, the introduction of region as an additional stratifying variable had only a marginal effect on the results given below and in Appendix 2.

- 9.3 Tables 9.1 and 9.2 give the results for two variables – the proportion of cohort members regarded as white and the proportion of main respondents who do not have a longstanding illness - with different patterns of sampling errors. Results for a selection of other variables are given in Appendix 2.

**Table 9.1: Sampling Errors, Proportion of Cohort Members Regarded as Non-white by Stratum and Country**

	<b>Mean</b>	<b>95% Confidence Interval</b>	<b>Design Effect</b>	<b>Observations</b>
England Advantaged	0.07	0.05 - 0.10	9.4	4608
England Disadvantaged	0.17	0.11 - 0.22	23.1	4504
England Ethnic	0.83	0.78 - 0.88	11.5	2384
Wales Advantaged	0.02	0.01 - 0.03	1.5	829
Wales Disadvantaged	0.05	0.02 - 0.07	5.8	1928
Scotland Advantaged	0.03	0.02 - 0.04	1.5	1141
Scotland Disadvantaged	0.03	0.02 - 0.05	2.5	1189
Northern Ireland Advantaged	0.01	0.00 - 0.02	1.5	723
Northern Ireland Disadvantaged	0.01	0.00 - 0.01	0.8	1198
England	0.15	0.12 - 0.17	10.3	11496
Wales	0.03	0.02 - 0.04	2.9	2757
Scotland	0.03	0.02 - 0.04	1.8	2330
Northern Ireland	0.01	0.00 - 0.01	1.4	1921
UK	0.13	0.10 - 0.15	15.6	18504

Note

The binary variable is based on *adc06e00* (1 = non-white)..

We see from Table 9.1 that, for example, the best estimate of the number of non-white children aged nine months in the UK is 13% (the weighted mean) but that this figure could be as low as 10% and as high as 15%, using the 95% confidence limits. The design effect summarises the effects of the sample design compared with a simple random sample of the same size. The value of 15.6 means that the sampling variance is 15.6 times higher than it would have been with a simple random sample of 18505. Data from a simple random sample of this size would, of course, have been much more expensive to collect. A value of 15.6 does indicate a high degree of clustering that is not offset by the efficiencies introduced by stratification. It is not, of course, surprising to find such a high value for this variable because there is a lot of variation between clusters (that is, wards) for race/ethnicity, certainly in England. In Northern Ireland, however, and in the advantaged wards of Wales and Scotland, the design effects are small, essentially because there are so few non-white children in these areas and hence very little between cluster variation.

**Table 9.2: Sampling Errors, Proportion of Natural Mothers who do not have a Longstanding Illness by Stratum and Country**

	Mean	95% Confidence Interval	Design Effect	Observations
England Advantaged	0.78	0.77 - 0.80	1.9	4614
England Disadvantaged	0.78	0.76 - 0.80	2.4	4517
England Ethnic	0.83	0.81 - 0.86	2.5	2384
Wales Advantaged	0.80	0.77 - 0.82	1.1	831
Wales Disadvantaged	0.77	0.75 - 0.79	0.9	1927
Scotland Advantaged	0.81	0.78 - 0.85	1.9	1143
Scotland Disadvantaged	0.77	0.75 - 0.79	0.9	1189
Northern Ireland Advantaged	0.81	0.78 - 0.84	1.1	723
Northern Ireland Disadvantaged	0.78	0.76 - 0.80	1.1	1196
England	0.79	0.77 - 0.80	2.6	11491
Wales	0.78	0.77 - 0.80	1.1	2752
Scotland	0.80	0.78 - 0.82	1.6	2327
Northern Ireland	0.80	0.78 - 0.82	1.1	1917
UK	0.79	0.78 - 0.80	2.9	18487

**Note**

Based on *amloil00* (1 = not ill); restricted to *admres00=1*.

The design effects in Table 9.2 are generally much smaller than in Table 9.1 and show less variation from stratum to stratum and from country to country. In other words, the prevalence of longstanding illness in natural mothers of nine-month old children shows relatively little variation across wards. Consequently, the 95% confidence intervals are more tightly grouped around the mean.

9.4 Tables 9.1 and 9.2, together with the tables in Appendix 2, provide users with some guidance as to what kinds of sampling errors to expect. The highest design effects are generally to be found in the English ethnic wards, partly because these wards are larger on average (Table 7.3). Socio-economic variables like income, being on benefits and educational qualifications tend to have large design effects whereas those for health variables like longstanding illness and ‘malaise’ are much smaller. Despite the large design effects for some variables in some strata, the overall size of the sample leads to 95% confidence intervals that are generally narrow and so the means are estimated with a high degree of precision.

**10. Mobile Families and Unit Non-response, Sweep 1**

**Introduction**

10.1 There are three components that make up an overall picture of unit non-response in MCS1. These are:

- (i) The under-representation of 'new movers' into the selected wards before nine months (6.4).
- (ii) The loss between the Child Benefit sample and the sample issued to the field (Table 7.1).
- (iii) The loss between the issued sample and the final sample (Table 7.2).  
In addition, there is a further loss of data when there is a response from the main respondent but not from an eligible partner (and, albeit very rarely, from a partner and not the main respondent).

10.2 As well as unit and partner non-response, there are also respondents who give incomplete responses, either because they choose to end the interview before the end of the schedule has been reached or, more commonly, they are unwilling or unable to answer particular questions. These two kinds of non-response contribute to domain (as defined by the section of the interview schedule) and item non-response and are discussed in Section 12.

10.3 We are in an unusually favourable position to understand some aspects of unit non-response in Sweep1 because we have, from information made available from Child Benefit records, some data about families that did not participate. We know:

- (i) Whether the claimant is paid by order book or directly into a bank or building society account. This is a proxy for income: 59% of those paid by order book were on one of the benefits listed in (8.2) compared with 18% paid into an account.
- (ii) The age of the claimant: under 19 and then five year age bands to 60+.
- (iii) Number of children in the household.
- (iv) Claimant title: Miss, Ms., Mrs., Mr., other. This is a proxy for single parent status: 33% of those labelled 'Miss' and 21% labelled 'Ms.' were lone parents compared with 3% labelled 'Mrs.'
- (v) The ward (and therefore the stratum and country) the claimant was living in.

### ***New Movers***

10.4 As described in (6.4), so-called new mover families were picked up from Wave 8 onwards. This means that this group is under-represented in the sample, more so in England and Wales than in Scotland and Northern Ireland where fieldwork started later. Although the group is small in relation to the sample as a whole (293 new movers were issued to the field), we do find that more geographically mobile families have different characteristics from those who are less mobile.

10.5 Considering the variables associated with being a new mover, we find that they are more likely to be paid by order book, somewhat more likely to be titled 'Miss', and a little more likely to be in an ethnic area of England, a disadvantaged area of Scotland and in Northern Ireland. There is no association, however, with mother's age and number of children (the details of the analysis are given in Appendix 3). We also know that these new movers were more likely to have refused in the field (see 10.9). We can also say



something about the *responding* new movers (n = 198). Unsurprisingly, they have spent less time at their current address - the median is 5 months compared with 36 months for the rest of the sample – although the mean is 22 months (SD = 37) compared with 49 (SD = 50) months, indicating that some so-called new movers have actually been at the current address for a long time. The main finding is that they are more likely than the rest of the responding sample to be on one of the benefits listed in (8.2); the odds ratio is 1.61.

- 10.6 Looking at time at current address for all respondents, we find that older mothers, families with more children and lone parents not on benefit are more likely to have been there longer. Compared with white families, Indian, Pakistani and Bangladeshi families are likely to have been at the current address longer but Black and Mixed families for a shorter time. Appendix 3 (Table A3.2) provides more details.
- 10.7 In addition, we know that there were 1667 families that had moved away from the selected wards before they were contacted by an interviewer. These families were more likely to have had fewer children, less likely to be an older mother (over 33), less likely to be labelled 'Mrs.' on Child Benefit records, more likely to have been found by DWP in an ethnic ward in England and more likely to have been a new mover (see Appendix 3 and Table A3.3). Combining these different sources of data, it appears that older married mothers with more children are less likely to move than poorer and Black mothers.

### ***Losses from Child Benefit sample to issued sample***

10.8 The reasons for losses from the Child Benefit sample are described in (6.1) and (6.5). Table 7.1 gives the extent of the problem. A full analysis of non-response for this component is not possible as we were only provided with aggregate data by DWP, in order to protect confidentiality. Consequently we can relate each of the five variables listed above to the probability of being issued to the field – their marginal effects within stratum - but we cannot look at conditional effects. Details of the analyses are given in Appendix 4 but the main findings are as follows:

- (i) All the differences between the nine strata can be subsumed into the contrast between Northern Ireland and GB – the odds of a GB inclusion in the issued sample are 1.3 times greater than for Northern Ireland. Part of this explanation lies in the fact that matching postcodes to wards was more difficult in Northern Ireland (see 6.7).
- (ii) Claimants paid through a bank account are 1.13 times more likely to be included than those paid by order book.
- (iii) Claimants under 24 and claimants 24 - 33 are 1.4 and 1.2 times more likely to be included than mothers 34+.
- (iv) There are no differences by claimant title.
- (v) Households with one child, and households with 2 or 3 children are 1.6 and 1.4 times more likely to be included than households with 4+ children.
- (vi) There is no support for stratum by variable interactions.
- (vii) There is no evidence of variability by ward.

### ***Losses from the issued sample during fieldwork***

10.9 This refers to losses of eligible cases issued to the field (Table 7.2). For these cases, we have data on the five Child Benefit variables at the individual level. Based on a two level logistic regression model (respondents within wards - see Appendix 4), we find that:

- (i) Families in minority ethnic wards in England and in Northern Irish advantaged wards are less likely to respond (respective odds are 0.68 and 0.70), and families in Welsh advantaged wards are more likely (odds = 1.32), conditional on other variables.
- (ii) Where the claimant has the title 'Miss', they are less likely to respond (odds = 0.86).
- (iii) There is an interaction between method of payment and mother's age such that mothers paid into a bank account and over age 33 are more likely to respond (odds = 1.46).
- (iv) The new movers are less likely to respond (odds = 0.71).
- (v) There is variation in response from ward to ward, some of which is explained by the Child Poverty Index (see 2.12): an increase in 10% in the CPI reduces the odds of a response by 0.04.

### ***Response by partners***

10.10 Table 10.1 gives the response accounts for the partners. We see that partner response rates are generally high except for the ethnic wards in England and the disadvantaged wards in Northern Ireland.

#### Notes on Table 10.1

##### *Final sample*

From Table 7.2 (productive).

##### *Ineligible*

No partner in the household. Eligible cases are those with a co-resident partner, including part-time resident partners

##### *Unproductive*

No partner interview or proxy partner interview carried out.

##### *RR(P)*

The partner response rate, defined as productive/(final sample – ineligible)

##### *Productive*

Includes proxy interviews with main respondent.

**Table 10.1: From Final Sample to Partner Sample by Stratum and Country**

	<b>England</b>				<b>Wales</b>			<b>Scotland</b>			<b>N. Ireland</b>			<b>UK</b>
	Ad.	Disad.	Eth.	Total	Ad.	Disad.	Total	Ad.	Disad.	Total	Ad.	Disad.	Total	All
Final sample	4617	4522	2394	11533	832	1928	2760	1145	1191	2336	723	1200	1923	18552
Ineligible	387	978	488	1853	89	501	590	103	272	375	59	317	376	3194
Unproductive	342	444	409	1195	74	163	237	104	130	234	68	183	251	1917
RR(P)	92%	87%	79%	88%	90%	89%	89%	90%	86%	88%	90%	79%	84%	88%
Productive	3888	3100	1497	8485	669	1264	1933	938	789	1727	596	700	1296	13441
Proxy	42	58	49	149	7	17	24	13	6	19	9	15	24	216

10.11 We can look at the correlates of partner non-response by combining the Child Benefit variables with variables relating to the family or to the mother that are provided by the main respondent even if the partner data are missing. We find that:

- (i) Partners in ethnic wards in England and disadvantaged wards in Northern Ireland are less likely to respond (odds = 0.59 and 0.51 respectively).
- (ii) Claimants paid by order book are less likely to respond (odds = 0.75).
- (iii) The odds of not responding increase by 1.09 for each extra child in the family.
- (iv) Partners of unmarried main respondents are less likely to respond (odds = 0.67).
- (v) Partners of Asian and Black main respondents are less likely to respond (odds = 0.64 and 0.47 respectively).
- (vi) Partners where the main respondent's educational qualifications are below degree level are less likely to respond (odds = 0.82).
- (vii) There is variation from ward to ward in partner non-response that is not explained by the Child Poverty Index.

See Appendix 4 for more details.

## **11. Weighting to Adjust for Non-response Bias**

11.1 One way of adjusting for possible biases generated by the systematic unit non-response established by the analyses in Appendices 3 and 4 is to use non-response weights. These non-response weights can be combined with the sample weights (Table 5.5) that are needed for analyses of MCS1 data.

11.2 To compensate for the under-representation of new movers (10.4), responding new mover families in England and Wales (119 in England, 20 in Wales) are given a weight of 1.86 (i.e. 13/7); in Scotland and Northern Ireland (31 in Scotland and 28 in NI) the weight is 1.4 (i.e. 14/10). All non-movers are given a weight of 1. In the light of the analyses reported above (10.5 to 10.7), it is difficult to know whether these suggested weights will take account of all mobility.

11.3 Turning to losses from the Child Benefit sample up to being issued to the field, we know from the issued sample (10.8) that mothers paid by order book have more children on average, 2.2, compared with 1.8 for those paid into a bank account. We also know that older mothers have more children on average. Consequently, the weights for the sample loss at this stage are based on a logistic regression that includes (i) a variable contrasting Northern Ireland with GB and (ii) number of children (1, 2, 3+). Although these two variables are both statistically significant (see Appendix 4) the difference in the mean predicted probabilities of exclusion between those actually excluded and included is small (0.113 vs. 0.110) and so the weights show little variability, ranging from 0.98 to 1.04 in GB and from 1.01 to 1.09 in Northern Ireland.

- 11.4 The model for losses from the issued sample (10.9) does not discriminate substantially between respondents and non-respondents – the predicted probabilities of responding for the two groups are 0.834 and 0.801. Consequently, there is not a wide variation in the weights - from 0.88 to 1.43.
- 11.5 There is an additional set of weights that can be included to allow for the over-representation of ‘winter’ births in Scotland and Northern Ireland (2.2). Responding families in Scotland and Northern Ireland with children born between 24 November and 11 January for the periods 2000/2001 and 2001/2002 (562 in Scotland, 441 in Northern Ireland) are given a weight of 0.5, 1 otherwise.
- 11.6 All of these four sets of weights can be standardised such that their mean is one and then multiplied together (because the probabilities of being excluded at each stage are independent) to give an overall unit non-response weight. This overall weight varies from 0.49 to 2.58 although the difference between the 10<sup>th</sup> and 90<sup>th</sup> centiles is only 0.27. When combined with the sample weights for each UK country (weight1 in Table 5.5) and re-standardised, the range is from 0.25 to 3.73 with a difference of 1.2 across the middle 80% of the distribution. (The range for weight1 on its own is 1.53 – see Table 5.5). When combined with the sample weights for the UK as a whole (weight2 in Table 5.5) and re-standardised, the range is wider - from 0.13 to 4.48 - with a difference of 1.7 across the middle 80% of the distribution. (The range for weight2 on its own is 1.77 – see Table 5.5).
- 11.7 Table 11.1 contrasts the overall weights with the sample weights for each stratum. It shows that the mean overall weights are not substantially different from the sample weights, both for the separate countries and, especially, for the UK as a whole. This suggests that there is unlikely to be a substantial gain in accuracy from using the overall weights rather than the (necessary) sample weights but further investigation of this issue is needed.

**Table 11.1: Comparing Overall and Sample Weights by Stratum**

	Each UK country		UK as a whole	
	Overall weight Mean (SD)	Sample weight (weight1)	Overall weight Mean (SD)	Sample weight (weight2)
ENGLAND: Advantaged	1.42 (0.46)	1.32	1.97 (0.19)	2.00
ENGLAND: Disadvantaged	0.79 (0.45)	0.71	1.12 (0.14)	1.09
ENGLAND: Ethnic	0.30 (0.24)	0.24	0.42 (0.06)	0.37
WALES: Advantaged	1.84 (0.18)	1.77	0.59 (0.06)	0.62
WALES: Disadvantaged	0.72 (0.07)	0.65	0.24 (0.02)	0.23
SCOTLAND: Advantaged	1.32 (0.32)	1.23	0.92 (0.22)	0.93
SCOTLAND: Disadvantaged	0.83 (0.22)	0.75	0.58 (0.16)	0.57
N.IRELAND: Advantaged	1.66 (0.43)	1.41	0.51 (0.13)	0.47
N.IRELAND: Disadvantaged	0.92 (0.23)	0.76	0.28 (0.07)	0.25

## 12. Partial, Domain and Item Non-response

12.1 For the main respondent, 3.4% of the interviews were incomplete in the sense that the interview was ended before the end of the questionnaire had been reached. There were, in addition, 20 cases with no data for the main respondent. Of the eligible partners who were interviewed (see 10.10 for details of those who were not), 3.5% were incomplete.

12.2 The interview schedule divided into a set of modules or domains – 10 for the main respondent and six for the partner. Completion rates were computed for each module, based on the number of main respondents ( $n = 18532$ ) and partners ( $n = 13225$ ) who responded in person to at least part of the schedule. For the interviewer-administered sections of the questionnaire, 'refusal' and 'don't know' (as recorded by the interviewers) are not counted as valid answers. The self-completion section (H) included a number of sensitive questions and 'can't say' – which was offered as an explicit response category for all questions in this section - is not treated as a valid response. Table 12.1 gives the

proportions who provided valid responses to all the questions that were applicable to all respondents. It shows that the domain completion rates are very high except for the self-completion modules and, to a lesser extent, the housing and local area module for the main respondents. For the self-completion modules, a slightly less stringent criterion than 100% completion quickly raises the completion rates – to 60% for the main respondent and 78% for the partner if just one question remains unanswered for example.

**Table 12.1 Domain Completion Rates for Main Respondent and Partner**

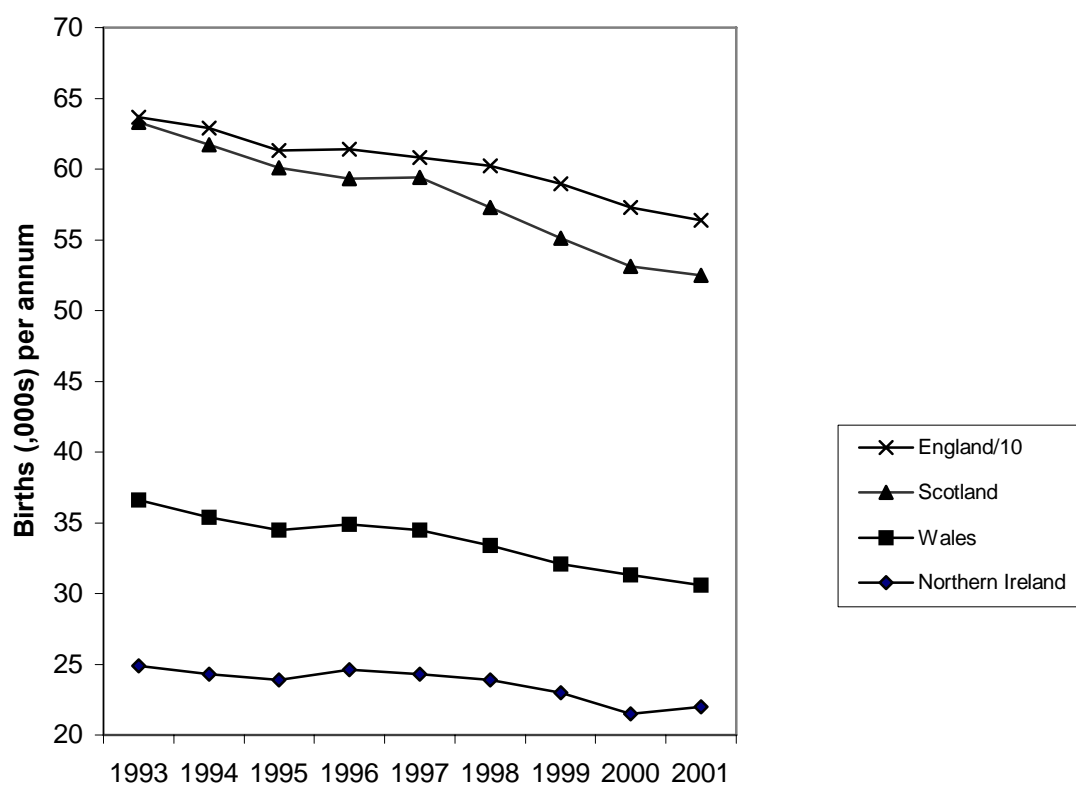
	<b>Interview schedule code</b>	<b>Number of questions to main (partner)</b>	<b>Main respondent</b>	<b>Partner</b>
Non-resident parents	A	4	99.8%	n.a.
Father's involvement	B	4	n.a.	99.9%
Pregnancy, labour and delivery	C	14	97.8%	n.a.
Baby's health and development	D	31	96.9%	n.a.
Childcare	E	2	99.9%	n.a.
Grandparents and friends	F	7 (4)	94.1%	98.4%
Parent's health	G	17 (16)	98.6%	99.3%
Self-completion	H	58 (38)	37.5%	59.4%
Employment and education	J	15 (9)	98.6%	99.5%
Housing and local area	K	23	85.6%	n.a.
Interests, Time with baby	L	5 (5)	89.3%	n.a.

12.3 The majority of the items in MCS1 have some, generally very small amounts of missing data. The items that respondents are usually most reluctant to answer are to do with their own income. We find, however, that, among those eligible to respond, only 6% of main respondents and 6% of partners are unwilling or unable to reveal their earnings. Hawkes and Plewis (2006) provide more discussion of missing income data.

## APPENDIX 1: BIRTHS, THE DISAPPEARING TARGET

A1.1 The sample was drawn by selecting electoral wards in each stratum assuming that the population of children under 5 in 1998 (or 1999 in Scotland and Northern Ireland) would give a good guide to the numbers of children to be found in those areas in 2001-2 aged 9 months. The mid year population estimate for each ward was divided by 5. At the time of sampling this seemed a reasonable assumption to make, as the number of births nationally had been more or less level over the mid 1990s. However, as shown in Fig.A1.1, births had begun to fall around this time and fertility rates reached an all-time low in 2001. Table A1.1 shows that in the country as a whole, births in the period eligible for membership of MCS were about 9% down on the numbers that had informed the selection of the sample. Thus, the size of the sample of children yielded by the sample of wards turned out to be below target.

Fig. A1.1: Births in the Countries of the UK, 1993-2001





**Table A1.1: Actual and Expected Births**

	England	Wales	N. Ireland	Scotland	UK
Period of births used in original estimate	1993-4 to 1997-8		1994-5 to 1998-9		
Original sample birth dates	Sep 2000 - Aug 2001		Dec 2000 - Nov 2001		
Actual births in sample months ('000)	567	31	22	52	672
Annual average in 5-year period ('000)	617	35	24	59	735
Ratio sample to expected	91.9%	88.2%	91.0%	88.4%	91.4%

Source: *Population Trends*, quarterly data

A1.2 Table A1.2 shows the sample size as originally intended, and the numbers achieved. It was originally estimated that the eligible child population of the sampled wards was 29632, and that there would be stratum-specific response rates of 75%, 70% and 65% in the advantaged, disadvantaged and ethnic areas respectively. This led to the original target of 20646. Actually, the survey has yielded 18819 children, in 18553 families. Part of the reason for this shortfall is that births fell during recent years in the UK as a whole as noted above, which will have affected the number of families available to be sampled, although the drop in births was not uniform across the country. The fourth column of Table A1.2 shows the number of births registered, in the sampled wards, as far as can be told, over the survey birth dates. Comparison of columns 3 and 4 suggest that the fall in births was not uniform and they appear to have risen in our ethnic wards. The estimates of registered births are subject to some margin of error in England and Wales because of boundary changes to wards after 1998, but there are no such problems with the wards in the ethnic stratum. The reverse trend in their numbers of births is plausibly due to a combination of the high fertility rates of some of the ethnic groups, and the increase in numbers of childbearing age through international migration, so a 'birth dearth' is not the explanation for the small shortfall of achieved cohort members in these areas. In the advantaged areas of England, the shortfall in the sample is almost exactly the same as the overestimate of births. This would account for the shortfall completely if all the 'unborn' would have responded to the survey, or 75% of the shortfall if the assumed response rate for such areas applies. In the disadvantaged wards of England the shortfall in the sample is about twice that in the numbers born, implying 33% if the sample shortfall can be attributed to falling births. In the disadvantaged wards of Wales, the difference in population exceeds the sample shortfall, and accounts for 87% of it. After including the extra 7 weeks births in Scotland and Northern Ireland, the number of births actually registered in the sample wards over the sample dates is about the same as the expected number, so we can say that the extended sample compensated for the demographic trend in those countries. As the survey actually exceeded its target in the advantaged areas of Wales, there are just three strata in England and Wales, with a combined sample shortfall of 1537 children of which about

900 (59%) can be attributed to the fertility decline. This is about half the overall shortfall of 1827 children below target.

A1.3 These calculations ignore differences between the child population at 9 months and registered births. Some of these registered births would not have been eligible for the survey if they died or moved away, including abroad, before reaching 9-months-old. The DWP detected 27201 families. The margin here arises because of: people being ineligible for Child Benefit as well as for the survey; poorly defined postcodes (only a problem in Northern Ireland); and, any non- or late-claiming of Child Benefit. There is also a drop of about 1% moving from a count of babies to families because of multiple births.

A1.4 We have also compared the registered births in each ward (with unchanged boundaries) with the number of families reported by DWP 7-8 months later. The matches were not expected to be exact, and there are a few wards where there seem to be significant numbers of out-movers or families ineligible for Child Benefit, for reasons connected with international migration.

**Table A1.2: Estimates of Population of Selected Wards, Samples Found and Issued by DWP and Interviews Achieved**

<b>Stratum</b>	<b>Original target sample of children</b>	<b>Expected child population of survey areas</b>	<b>Registered births <sup>a</sup></b>	<b>Registered births after extension</b>	<b>Child Benefit sample</b>	<b>Issued sample (to field)</b>	<b>Final sample (families)</b>	<b>Final sample (children)</b>	<b>Percentage shortfall due to drop in births</b>
ENGLAND: Advantaged	5250	7348	6779 <sup>b</sup>		6364	5748	4617	4687	75
ENGLAND Disadvantaged	5271	7511	7185 <sup>b</sup>		6678	5946	4522	4592	33
ENGLAND: Ethnic	2625	4009	4465		3906	3461	2394	2416	n.a
WALES: Advantaged	750	1196	1103 <sup>c</sup>		1074	979	832	844	n.a
WALES: Disadvantaged	2250	3170	2805 <sup>d</sup>		2784	2495	1928	1954	87
SCOTLAND: Advantaged	1250	1657	1480	1635	1579	1419	1145	1163	n.a
SCOTLAND: Disadvantaged	1250	1836	1687	1873	1748	1568	1191	1207	n.a
N.IRELAND: Advantaged	750	1016	959	1062	1109	935	723	735	n.a
N.IRELAND: Disadvantaged	1250	1889	1711	1892	1959	1629	1200	1220	n.a
TOTAL UK	20646	29632	28174	28799	27201	24180	18552	18818	

Notes:

- a) including net outmigration, children who died before 9 months, and cases not eligible for Child Benefit
- b) based on 88% of births in wards unaffected by boundary changes;
- c) based on 71% of births in wards unaffected by boundary changes;
- d) based on 87% of births in wards unaffected by boundary changes.

## APPENDIX 2: FURTHER INFORMATION ON SAMPLING ERRORS

This Appendix expands on the results presented in Section 9 for a range of variables from different parts of the questionnaire. The variable names used on the deposited dataset are provided in the notes to each table.

**Table A2.1: Sampling Errors, Proportion of Natural Mothers who were not currently in Paid Work by Stratum and Country**

	Mean	95% Confidence Interval	Design Effect	Observations
England Advantaged	0.45	0.41 - 0.45	1.9	4617
England Disadvantaged	0.57	0.55 - 0.60	3.1	4522
England Ethnic	0.77	0.72 - 0.82	9.8	2394
Wales Advantaged	0.38	0.34 - 0.42	1.6	832
Wales Disadvantaged	0.58	0.53 - 0.62	3.7	1928
Scotland Advantaged	0.41	0.37 - 0.44	1.6	1145
Scotland Disadvantaged	0.49	0.45 - 0.54	2.7	1191
Northern Ireland Advantaged	0.34	0.29 - 0.38	1.5	723
Northern Ireland Disadvantaged	0.52	0.48 - 0.57	2.1	1200
England	0.50	0.48 - 0.52	3.6	11496
Wales	0.47	0.44 - 0.50	2.5	2753
Scotland	0.44	0.41 - 0.47	2.1	2327
Northern Ireland	0.42	0.39 - 0.46	2.1	1918
UK	0.49	0.47 - 0.50	4.0	18494

### Note

Based on *admwrk00* (1 = not currently in paid work); restricted to *admres00* =1.

**Table A2.2: Sampling Errors, Average Household Size by Stratum and Country**

	<b>Mean</b>	<b>95% Confidence Interval</b>	<b>Design Effect</b>	<b>Observations</b>
England Advantaged	3.88	3.85 - 3.93	1.9	4617
England Disadvantaged	3.93	3.86 - 4.00	3.8	4522
England Ethnic	4.80	4.51 - 5.10	15.4	2394
Wales Advantaged	3.88	3.80 - 3.97	1.4	832
Wales Disadvantaged	3.88	3.82 - 3.95	1.3	1928
Scotland Advantaged	3.85	3.77 - 3.92	1.6	1145
Scotland Disadvantaged	3.84	3.74 - 3.94	2.2	1191
Northern Ireland Advantaged	4.05	3.95 - 4.16	1.5	723
Northern Ireland Disadvantaged	4.18	4.08 - 4.28	1.5	1200
England	3.96	3.91 - 4.00	3.3	11533
Wales	3.88	3.83 - 3.94	1.5	2760
Scotland	3.84	3.78 - 3.90	1.8	2336
Northern Ireland	4.11	4.04 - 4.18	1.5	1923
UK	3.95	3.91 - 3.99	4.1	18552

Note

Based on *adtotp00*

**Table A2.3: Sampling Errors, Proportion of Lone Parents by Stratum and Country**

	<b>Mean</b>	<b>95% Confidence Interval</b>	<b>Design Effect</b>	<b>Observations</b>
England Advantaged	0.08	0.07 - 0.09	1.7	4617
England Disadvantaged	0.22	0.20 - 0.24	3.0	4522
England Ethnic	0.20	0.15 - 0.26	12.2	2394
Wales Advantaged	0.11	0.08 - 0.13	1.6	832
Wales Disadvantaged	0.26	0.23 - 0.29	3.0	1928
Scotland Advantaged	0.09	0.07 - 0.11	2.0	1145
Scotland Disadvantaged	0.23	0.19 - 0.26	2.1	1191
Northern Ireland Advantaged	0.08	0.05 - 0.11	1.9	723
Northern Ireland Disadvantaged	0.26	0.23 - 0.30	2.3	1200
England	0.13	0.12 - 0.14	2.6	11533
Wales	0.18	0.15 - 0.20	2.1	2760
Scotland	0.14	0.12 - 0.16	2.0	2336
Northern Ireland	0.17	0.14 - 0.19	2.0	1923
UK	0.14	0.13 - 0.15	2.9	18552

Note

Based on *adhtys00* (1 = no resident partner at interview).

**Table A2.4: Sampling Errors, Proportion of Households where Languages other than English are spoken by Stratum and Country**

	<b>Mean</b>	<b>95% Confidence Interval</b>	<b>Design Effect</b>	<b>Observations</b>
England Advantaged	0.06	0.04 - 0.07	6.0	4617
England Disadvantaged	0.11	0.07 - 0.15	19.3	4522
England Ethnic	0.69	0.59 - 0.78	27.1	2394
Wales Advantaged	0.11	0.06 - 0.16	6.0	832
Wales Disadvantaged	0.11	0.07 - 0.15	8.6	1928
Scotland Advantaged	0.01	0.01 - 0.02	1.2	1145
Scotland Disadvantaged	0.03	0.01 - 0.04	3.4	1191
Northern Ireland Advantaged	0.01	0.01 - 0.02	0.4	723
Northern Ireland Disadvantaged	0.02	0.01 - 0.03	1.1	1200
England	0.11	0.09 - 0.13	8.4	11533
Wales	0.11	0.08 - 0.14	8.4	2760
Scotland	0.02	0.01 - 0.03	1.9	2336
Northern Ireland	0.02	0.01 - 0.02	0.7	1923
UK	0.10	0.08 - 0.12	12.1	18552

Note

Based on *ahlang00* (1 = speaks English and/or other language at home).



**Table A2.5: Sampling Errors, Proportion of Natural Mothers Receiving Child Care for which they do not Pay by Stratum and Country**

	<b>Mean</b>	<b>95% Confidence Interval</b>	<b>Design Effect</b>	<b>Observations</b>
England Advantaged	0.53	0.50 - 0.57	3.7	2719
England Disadvantaged	0.74	0.70 - 0.77	4.4	2422
England Ethnic	0.83	0.78 - 0.89	5.0	970
Wales Advantaged	0.56	0.49 - 0.63	2.5	552
Wales Disadvantaged	0.77	0.73 - 0.81	2.6	1084
Scotland Advantaged	0.57	0.50 - 0.64	3.7	703
Scotland Disadvantaged	0.76	0.70 - 0.81	3.3	722
Northern Ireland Advantaged	0.45	0.37 - 0.52	2.9	488
Northern Ireland Disadvantaged	0.63	0.57 - 0.69	2.5	677
England	0.61	0.58 - 0.64	5.1	6111
Wales	0.65	0.60 - 0.69	3.6	1636
Scotland	0.64	0.59 - 0.69	4.2	1425
Northern Ireland	0.53	0.47 - 0.58	3.9	1165
UK	0.61	0.59 - 0.63	6.0	10337

**Note**

Based on *amcpay00* (1 = no); restricted to *admres00* = 1 and to the sub-sample of natural mothers using child care of any form.

**Table A2.6: Sampling Errors, Contact with Friends, Natural Mothers by Stratum and Country**

	<b>Mean</b>	<b>95% Confidence Interval</b>	<b>Design Effect</b>	<b>Observations</b>
England Advantaged	2.84	2.79 - 2.89	3.3	4610
England Disadvantaged	2.96	2.91 - 3.01	2.7	4505
England Ethnic	3.21	3.15 - 3.26	1.7	2378
Wales Advantaged	2.89	2.83 - 2.95	1.0	831
Wales Disadvantaged	2.85	2.79 - 2.91	1.9	1921
Scotland Advantaged	2.98	2.92 - 3.04	1.5	1141
Scotland Disadvantaged	3.05	2.99 - 3.11	1.3	1186
Northern Ireland Advantaged	3.01	2.93 - 3.09	1.6	722
Northern Ireland Disadvantaged	2.99	2.92 - 3.06	1.6	1195
England	2.90	2.86 - 2.93	3.6	11493
Wales	2.87	2.83 - 2.92	1.5	2752
Scotland	3.01	2.96 - 3.05	1.4	2327
Northern Ireland	3.00	2.95 - 3.05	1.6	1917
UK	2.91	2.88 - 2.94	4.2	18489

Note

Based on *amfrti00*; restricted to *admres00=1*.

**Table A2.7: Sampling Errors, Weekly Alcohol Units, Natural Mothers by Stratum and Country**

	<b>Mean</b>	<b>95% Confidence Interval</b>	<b>Design Effect</b>	<b>Observations</b>
England Advantaged	2.96	2.68 - 3.23	4.1	4611
England Disadvantaged	2.25	2.00 - 2.50	2.9	4501
England Ethnic	0.53	0.28 - 0.77	6.6	2384
Wales Advantaged	3.19	2.78 - 3.61	1.5	831
Wales Disadvantaged	2.24	2.04 - 2.45	1.0	1923
Scotland Advantaged	2.73	2.31 - 3.16	2.7	1139
Scotland Disadvantaged	1.88	1.50 - 2.26	2.9	1182
Northern Ireland Advantaged	1.97	1.54 - 2.40	3.1	722
Northern Ireland Disadvantaged	1.71	1.44 - 1.99	2.1	1196
England	2.59	2.39 - 2.78	5.5	11496
Wales	2.76	2.48 - 3.03	2.4	2754
Scotland	2.40	2.11 - 2.70	3.0	2321
Northern Ireland	1.85	1.58 - 2.11	3.1	1918
UK	2.55	2.39 - 2.71	6.4	18489

Notes

1. Based on *amaldr00* and *amunwk00*; restricted to *admres00=1*.
2. This derived variable has not been deposited at the Data Archive.

**Table A2.8: Sampling Errors, Malaise Score, Natural Mothers by Stratum and Country**

	<b>Mean</b>	<b>95% Confidence Interval</b>	<b>Design Effect</b>	<b>Observations</b>
England Advantaged	7.53	7.47 - 7.59	1.6	4564
England Disadvantaged	7.25	7.19 - 7.30	1.2	4351
England Ethnic	7.09	6.97 - 7.22	2.0	2007
Wales Advantaged	7.44	7.32 - 7.56	1.1	824
Wales Disadvantaged	7.09	7.00 - 7.19	1.3	1891
Scotland Advantaged	7.54	7.44 - 7.65	1.1	1106
Scotland Disadvantaged	7.17	7.07 - 7.27	0.8	1145
Northern Ireland Advantaged	7.60	7.47 - 7.74	1.2	714
Northern Ireland Disadvantaged	7.17	7.03 - 7.32	1.6	1174
England	7.42	7.37 - 7.46	1.7	10922
Wales	7.28	7.20 - 7.37	1.5	2715
Scotland	7.40	7.32 - 7.48	1.1	2251
Northern Ireland	7.40	7.30 - 7.50	1.4	1888
UK	7.41	7.37 - 7.44	1.9	17776

Notes

1. Based on sum of binary variables *amtire00*, *amdepr00*, *amworr00*, *amrage00*, *amscar00*, *amupse00*, *amkeyd00*, *amnerv00* and *amhera00* to give a scale from 0 ('complete malaise') to 9 ('complete non-malaise'); restricted to *admres00*=1.
2. This derived variable has not been deposited at the Data Archive.

**Table A2.9: Sampling Errors, Friendliness of Neighbours, Natural Mothers by Stratum and Country**

	<b>Mean</b>	<b>95% Confidence Interval</b>	<b>Design Effect</b>	<b>Observations</b>
England Advantaged	1.84	1.81 - 1.87	1.8	4494
England Disadvantaged	1.94	1.91 - 1.98	1.8	4230
England Ethnic	1.94	1.88 - 2.00	2.6	1976
Wales Advantaged	1.75	1.68 - 1.82	2.0	815
Wales Disadvantaged	1.85	1.81 - 1.89	1.1	1851
Scotland Advantaged	1.80	1.74 - 1.86	1.8	1091
Scotland Disadvantaged	1.93	1.90 - 1.97	0.5	1116
Northern Ireland Advantaged	1.84	1.79 - 1.89	0.9	704
Northern Ireland Disadvantaged	1.78	1.74 - 1.83	1.3	1147
England	1.88	1.86 - 1.90	2.0	10700
Wales	1.79	1.75 - 1.84	2.1	2666
Scotland	1.85	1.81 - 1.89	1.3	2207
Northern Ireland	1.81	1.78 - 1.85	1.2	1851
UK	1.87	1.85 - 1.89	2.4	17424

Note

Based on *amngfe00*; restricted to *admres00=1*.

**Table A2.10: Sampling Errors, Family Income, Natural Mothers by Stratum and Country**

	<b>Mean</b>	<b>95% Confidence Interval</b>	<b>Design Effect</b>	<b>Observations</b>
England Advantaged	29547	27438 - 31655	11.3	4345
England Disadvantaged	17665	16502 - 18827	7.1	4168
England Ethnic	14654	12641 - 16667	13.0	1973
Wales Advantaged	25896	23629 - 28163	2.9	780
Wales Disadvantaged	16696	15359 - 18032	4.7	1841
Scotland Advantaged	27424	24093 - 30754	8.3	1014
Scotland Disadvantaged	18023	16032 - 20014	6.0	1132
Northern Ireland Advantaged	27110	24443 - 29778	3.5	648
Northern Ireland Disadvantaged	15747	14797 - 16697	1.8	1041
England	24942	23496 - 26389	16.9	10486
Wales	21624	20092 - 23155	6.3	2621
Scotland	23616	21458 - 25774	8.9	2146
Northern Ireland	21837	20012 - 23662	5.9	1689
UK	24505	23300 - 25710	20.6	16942

Notes

1. Based on *amnilp00* and *amnico00*, using the middle point in £ per year but £65K for the top group of *jnetinlp* and £100K for the top group of *jnetinco* and not adjusting for size of household; restricted to *admres00=1*.
2. This derived variable has not been deposited at the Data Archive.

**Table A2.11: Sampling Errors, Proportion of Families Receiving Benefits by Stratum and Country**

	<b>Mean</b>	<b>95% Confidence Interval</b>	<b>Design Effect</b>	<b>Observations</b>
England Advantaged	0.23	0.20 - 0.26	5.1	4613
England Disadvantaged	0.52	0.49 - 0.56	5.0	4511
England Ethnic	0.59	0.51 - 0.67	17.2	2376
Wales Advantaged	0.30	0.26 - 0.34	1.8	831
Wales Disadvantaged	0.58	0.54 - 0.63	3.9	1926
Scotland Advantaged	0.26	0.21 - 0.32	4.1	1142
Scotland Disadvantaged	0.50	0.44 - 0.56	4.5	1187
Northern Ireland Advantaged	0.24	0.19 - 0.30	2.9	718
Northern Ireland Disadvantaged	0.57	0.53 - 0.62	2.5	1187
England	0.35	0.32 - 0.37	6.1	11500
Wales	0.43	0.39 - 0.47	3.9	2757
Scotland	0.35	0.31 - 0.39	4.0	2329
Northern Ireland	0.40	0.36 - 0.44	3.7	1905
UK	0.35	0.33 - 0.37	6.9	18491

Note

Based on *adonbe00* (1 = receiving one or more of benefits listed in 8.2).

**Table A2.12: Sampling Errors, Highest Academic Qualification, Natural Mothers by Stratum and Country**

	<b>Mean</b>	<b>95% Confidence Interval</b>	<b>Design Effect</b>	<b>Observations</b>
England Advantaged	2.84	2.76 – 2.93	6.1	4608
England Disadvantaged	2.20	2.10 – 2.31	8.5	4499
England Ethnic	1.87	1.68 – 2.06	12.5	2372
Wales Advantaged	2.88	2.74 – 3.02	3.1	831
Wales Disadvantaged	2.18	2.07 – 2.29	4.3	1918
Scotland Advantaged	3.05	2.91 – 3.19	4.1	1140
Scotland Disadvantaged	2.44	2.28 – 2.60	5.6	1184
Northern Ireland Advantaged	2.99	2.82 – 3.16	3.5	719
Northern Ireland Disadvantaged	2.22	2.11 – 2.34	2.8	1191
England	2.58	2.51 – 2.65	8.2	11479
Wales	2.56	2.45 – 2.67	6.0	2749
Scotland	2.81	2.71 – 2.92	4.5	2324
Northern Ireland	2.63	2.50 – 2.75	4.7	1910
UK	2.60	2.54 – 2.66	9.7	18462

Note

Based on *adnvqm00* (0 = none...5 = higher degree; overseas and other qualifications coded 1); restricted to *admres00* = 1.



**Table A2.13: Sampling Errors, Proportion of Births Requiring Assistance beyond that of a Normal Delivery by Stratum and Country**

	<b>Mean</b>	<b>95% Confidence Interval</b>	<b>Design Effect</b>	<b>Observations</b>
England Advantaged	0.34	0.32 - 0.36	1.5	4611
England Disadvantaged	0.28	0.27 - 0.30	2.0	4505
England Ethnic	0.26	0.24 - 0.29	2.1	2384
Wales Advantaged	0.34	0.31 - 0.38	1.4	831
Wales Disadvantaged	0.33	0.30 - 0.36	1.8	1922
Scotland Advantaged	0.34	0.31 - 0.37	1.2	1141
Scotland Disadvantaged	0.33	0.30 - 0.37	1.4	1186
Northern Ireland Advantaged	0.35	0.32 - 0.39	1.1	722
Northern Ireland Disadvantaged	0.34	0.31 - 0.36	0.7	1197
England	0.32	0.31 - 0.33	2.2	11500
Wales	0.34	0.31 - 0.36	1.9	2753
Scotland	0.34	0.31 - 0.36	1.3	2327
Northern Ireland	0.35	0.32 - 0.37	1.1	1919
UK	0.32	0.31 - 0.33	2.5	18499

Note

Binary variable derived from *amdewm00* (1 = not normal).

**Table A2.14: Sampling Errors, Birth Weight (Kg.) by Stratum and Country**

	<b>Mean</b>	<b>95% Confidence Interval</b>	<b>Design Effect</b>	<b>Observations</b>
England Advantaged	3.40	3.38 - 3.41	0.9	4607
England Disadvantaged	3.31	3.29 - 3.33	1.7	4501
England Ethnic	3.17	3.14 - 3.20	1.7	2377
Wales Advantaged	3.41	3.37 - 3.45	1.0	831
Wales Disadvantaged	3.32	3.30 - 3.35	0.9	1921
Scotland Advantaged	3.44	3.40 - 3.48	1.1	1141
Scotland Disadvantaged	3.37	3.34 - 3.41	1.1	1187
Northern Ireland Advantaged	3.50	3.46 - 3.53	0.6	722
Northern Ireland Disadvantaged	3.40	3.36 - 3.43	1.3	1197
England	3.36	3.34 - 3.37	1.5	11485
Wales	3.37	3.34 - 3.39	1.2	2752
Scotland	3.42	3.39 - 3.44	1.2	2328
Northern Ireland	3.45	3.42 - 3.47	0.9	1919
UK	3.36	3.35 - 3.38	1.8	18484

Note

Based on *adbwgta00* with 1 Kg. = 2.2 lbs.

**Table A2.15: Sampling Errors, Proportion of Natural Mothers who did not Attempt to Breastfeed by Stratum and Country**

	<b>Mean</b>	<b>95% Confidence Interval</b>	<b>Design Effect</b>	<b>Observations</b>
England Advantaged	0.21	0.19 - 0.24	3.8	4610
England Disadvantaged	0.38	0.34 - 0.43	9.3	4504
England Ethnic	0.19	0.16 - 0.22	4.6	2379
Wales Advantaged	0.29	0.23 - 0.34	3.3	831
Wales Disadvantaged	0.45	0.42 - 0.49	2.4	1921
Scotland Advantaged	0.27	0.22 - 0.32	3.7	1141
Scotland Disadvantaged	0.46	0.42 - 0.51	2.5	1186
Northern Ireland Advantaged	0.37	0.32 - 0.41	1.8	722
Northern Ireland Disadvantaged	0.60	0.56 - 0.64	2.0	1196
England	0.27	0.25 - 0.29	6.1	11493
Wales	0.36	0.33 - 0.40	4.4	2752
Scotland	0.35	0.31 - 0.38	2.9	2327
Northern Ireland	0.48	0.44 - 0.51	2.7	1918
UK	0.29	0.27 - 0.30	6.3	18490

**Note**

Based on *ambfeva0* (1 = no); restricted to *admres00* = 1.

**Table A2.16: Sampling Errors, Proportion of Households without Regular Use of Car/Van by Stratum and Country**

	<b>Mean</b>	<b>95% Confidence Interval</b>	<b>Design Effect</b>	<b>Observations</b>
England Advantaged	0.07	0.06 - 0.08	2.8	4614
England Disadvantaged	0.25	0.23 - 0.27	3.0	4512
England Ethnic	0.33	0.28 - 0.39	7.2	2377
Wales Advantaged	0.07	0.05 - 0.09	1.6	831
Wales Disadvantaged	0.22	0.19 - 0.26	3.4	1926
Scotland Advantaged	0.08	0.06 - 0.11	2.9	1142
Scotland Disadvantaged	0.23	0.18 - 0.28	4.1	1187
Northern Ireland Advantaged	0.06	0.03 - 0.08	2.2	721
Northern Ireland Disadvantaged	0.23	0.18 - 0.28	3.7	1192
England	0.15	0.13 - 0.16	3.0	11503
Wales	0.14	0.12 - 0.16	2.3	2757
Scotland	0.14	0.12 - 0.17	2.9	2329
Northern Ireland	0.14	0.11 - 0.17	3.0	1913
UK	0.15	0.13 - 0.16	3.7	18502

Note

Based on *amcaru00* (1 = no).

**Table A2.17: Sampling Errors, Proportion of Main Respondents who did not Vote in Last Election by Stratum and Country**

	<b>Mean</b>	<b>95% Confidence Interval</b>	<b>Design Effect</b>	<b>Observations</b>
England Advantaged	0.44	0.41 - 0.46	2.5	4610
England Disadvantaged	0.58	0.56 - 0.61	2.8	4506
England Ethnic	0.50	0.45 - 0.55	6.4	2375
Wales Advantaged	0.40	0.36 - 0.44	1.4	831
Wales Disadvantaged	0.58	0.54 - 0.61	2.5	1923
Scotland Advantaged	0.39	0.35 - 0.44	2.8	1137
Scotland Disadvantaged	0.53	0.49 - 0.58	2.4	1183
Northern Ireland Advantaged	0.41	0.37 - 0.45	1.5	712
Northern Ireland Disadvantaged	0.41	0.37 - 0.46	2.5	1181
England	0.49	0.47 - 0.51	3.5	11491
Wales	0.48	0.45 - 0.51	2.6	2754
Scotland	0.45	0.41 - 0.48	2.9	2320
Northern Ireland	0.41	0.38 - 0.44	2.0	1893
UK	0.48	0.47 - 0.50	4.0	18458

Note

Based on *amvote00* (1 = no).

**Table A2.18: Sampling Errors, Proportion of Partners spending some Time Each Day with the Baby by Stratum and Country**

	<b>Mean</b>	<b>95% Confidence Interval</b>	<b>Design Effect</b>	<b>Observations</b>
England Advantaged	0.28	0.26 - 0.30	1.2	3848
England Disadvantaged	0.37	0.34 - 0.40	3.0	3042
England Ethnic	0.42	0.38 - 0.46	2.5	1444
Wales Advantaged	0.34	0.30 - 0.38	1.3	662
Wales Disadvantaged	0.35	0.32 - 0.38	1.1	1246
Scotland Advantaged	0.30	0.27 - 0.34	1.4	924
Scotland Disadvantaged	0.34	0.30 - 0.38	1.7	783
Northern Ireland Advantaged	0.32	0.28 - 0.37	1.3	587
Northern Ireland Disadvantaged	0.39	0.35 - 0.43	1.2	684
England	0.31	0.30 - 0.33	1.9	8334
Wales	0.34	0.32 - 0.37	1.6	1908
Scotland	0.32	0.29 - 0.34	1.6	1707
Northern Ireland	0.35	0.32 - 0.38	1.3	1271
UK	0.31	0.30 - 0.33	2.2	13220

Note

Binary variable derived from *aploaf00* (1 = yes).

### APPENDIX 3: STATISTICAL MODELS FOR MOBILE FAMILIES

A3.1 The model for the new movers is based on the issued sample (n = 24180) minus those known to be ineligible (n = 417, Table 7.2) and those known to have moved (n = 1250, Table 7.2, note (iv)). A further 132 cases are omitted because of item non-response. The analysis is therefore based on 22381 cases, of which just 293 were new movers. Because the number of new movers is small relative to the number of sampled wards there is little point in correcting for clustering within wards. The estimates from the logistic regression – predicting the probability of being a new mover - are given in Table A3.1. Positive estimates (and hence expected changes in odds greater than one) indicate that the new movers are relatively more prevalent in those categories.

**Table A3.1: New Movers: Estimates from Logistic Regression**

	Estimate	s.e.	$\chi^2$	d.f.	p	Change in odds
England, Ad.	0	Fixed	17.9	8	0.022	n.a.
England, Disad.	0.25	0.20				1.3
England, Ethnic	0.54	0.22				1.7
Wales, Ad.	0.16	0.37				1.2
Wales, Disad.	-0.29	0.29				0.75
Scotland, Ad.	0.11	0.33				1.1
Scotland, Disad.	0.62	0.26				1.9
N. Ireland, Ad.	0.67	0.31				1.9
N. Ireland, Disad.	0.49	0.26				1.6
Claimant: Mr.	0	Fixed				10.3
Claimant, Mrs.	0.18	0.33	1.2			
Claimant: Ms.	-0.15	0.14	0.86			
Claimant, Miss	0.59	0.24	1.8			
Paid: order book	0.57	0.14	16.5	1	<0.001	1.8

A3.2 The model for time at current address for the main respondent (n = 18418) takes account of variation from ward to ward in a two level model (families within wards) using a square root transformation of the response as time at current address has a long tail to the right. It was estimated using *MLwiN* (Rasbash et al., 2000) and the results are given in Table A3.2. The explanatory variables come from those collected in the interview. The level-two (between ward) variation is small after allowing for the explanatory variables; the intra-ward correlation is 0.012.

**Table A3.2: Main Respondents' Time at Current Address: Estimates from Two Level Regression**

	<b>Estimate</b>	<b>s.e.</b>	<b><math>\chi^2</math></b>	<b>d.f.</b>	<b>p</b>
England, Ad.	0	Fixed	72.9	8	<0.001
England, Disad.	0.20	0.084			
England, Ethnic	0.42	0.13			
Wales, Ad.	0.24	0.14			
Wales, Disad.	0.18	0.10			
Scotland, Ad.	-0.10	0.12			
Scotland, Disad.	0.23	0.12			
N. Ireland, Ad.	0.66	0.14			
N. Ireland, Disad.	0.81	0.12			
Ethnic Gp.: White	0	Fixed			
Ethnic Gp. : Indian	0.42	0.16			
Ethnic Gp.: Pakistani	0.34	0.13			
Ethnic Gp.: Bangladeshi	0.61	0.19			
Ethnic Gp. : Black	-0.29	0.14			
Ethnic Gp.: Mixed & Other	-0.36	0.14			
Mother's age	0.14	0.004	942	1	<0.001
Number of children	0.27	0.023	144	1	<0.001
Lone parent	1.1	0.17	88.2	3	<0.001
On benefits	-0.25	0.056			
Lone parent*on benefits	-0.61	0.18			
Between ward variance	0.10	0.021	24.1	1	<0.001
Within ward variance	8.6	0.090	n.a.		

A3.3 The model for moved before interview (see 10.7) (n = 24014) takes account of variation from ward to ward in a two level logistic regression model (Table A3.3). The explanatory variables come from the Child Benefit system. The between ward variation is small after allowing for the explanatory variables; the intra-ward correlation is 0.035 assuming an underlying logistic distribution at level one.



**Table A3.3: Movers Before Interview: Estimates (PQL(2)) from Two Level Logistic Regression**

	<b>Estimate</b>	<b>s.e.</b>	<b><math>\chi^2</math></b>	<b>d.f.</b>	<b>p</b>	<b>Change in odds</b>
England, Ad.	0	Fixed	27.8	8	<0.001	n.a.
England, Disad.	0.13	0.10				1.1
England, Ethnic	0.42	0.13				1.5
Wales, Ad.	-0.40	0.20				0.67
Wales, Disad.	-0.068	0.13				0.93
Scotland, Ad.	0.069	0.15				1.1
Scotland, Disad.	0.20	0.14				1.2
N. Ireland, Ad.	0.22	0.19				1.2
N. Ireland, Disad.	-0.18	0.15				0.84
Claimant: Mr.	0	Fixed	75.4	3	<0.001	n.a.
Claimant, Mrs.	-0.56	0.13				0.57
Claimant: Ms.	-0.32	0.17				0.73
Claimant, Miss.	-0.060	0.13				0.94
Number of chn.	-0.089	0.026	11.7	1	<0.001	0.91
In-mover	0.40	0.20	4.0	1	0.045	1.5
Mother's age: <24	0	Fixed	24.2	2	<0.001	n.a.
Mother's age: 24-33	-0.24	0.10				0.79
Mother's age: >33	-0.46	0.11				0.63
Between ward variance	0.12	0.029	17.8	1	<0.001	n.a.

## APPENDIX 4: STATISTICAL MODELS FOR UNIT AND PARTNER NON-RESPONSE

A4.1 The models for the losses from the Child Benefit sample are based on a series of logistic regressions with the proportion of those lost as the response and with stratum as an explanatory variable in all models together with (separately) each of the variables listed in (10.3) apart from claimant title as this variable did not predict the probability of being included (Table A4.1). The analyses exclude the data from Wave 6 for England as these contained errors in the file supplied to us.

**Table A4.1: Losses from Child Benefit Sample: Estimates from Logistic Regression**

		Estimate	s.e.	$\chi^2$	d.f.	p	Change in odds
(1): n = 25920	Paid: order book	0.14	0.040	12.3	1	<0.001	1.2
	Northern Ireland	0.30	0.057	26.6	1	<0.001	1.3
(2): n = 25907	Mother's age: <24	-0.10	0.052	29.7	2	<0.001	0.91
	Mother's age: 24-33	-0.30	0.059				0.74
	Mother's age: >33	0	Fixed				n.a.
	Northern Ireland	0.31	0.057	28.3	1	<0.001	1.4
(3): n = 25840	No. of children.: 1	-0.49	0.065	58.5	2		0.61
	No. of children.: 2,3	-0.34	0.063				0.71
	No. of children.: 4+	0	Fixed				n.a.
	Northern Ireland	0.29	0.057	26.0	1	<0.001	1.3

A4.2 The model for losses from the issued sample takes account of variation from ward to ward in a two level logistic regression model. Data from one ward were omitted from the model because fieldwork problems in that ward made it an outlier in terms its high non-response, leaving a sample of 22319 cases. The explanatory variables come from the Child Benefit system measured at the individual level, plus the Child Poverty Index at the ward level. The between ward variation is small after allowing for the explanatory variables; the intra-ward correlation is 0.036 assuming an underlying logistic distribution at level one (Table A4.2).

**Table A4.2: Losses from the Issued Sample: Estimates (PQL(2)) from Two Level Logistic Regression**

	<b>Estimate</b>	<b>s.e.</b>
England, Ad.	0	Fixed
England, Disad.	-0.10	0.11
England, Ethnic	0.33	0.14
Wales, Ad.	-0.31	0.14
Wales, Disad.	-0.10	0.12
Scotland, Ad.	-0.070	0.12
Scotland, Disad.	-0.11	0.14
N. Ireland, Ad.	0.35	0.13
N. Ireland, Disad.	0.14	0.14
Claimant: Mr.	0	Fixed
Claimant, Mrs.	-0.29	0.095
Claimant: Ms.	-0.26	0.13
Claimant, Miss.	-0.12	0.099
Order book	0.30	0.040
In-mover	0.34	0.15
Mother's age: <24	0	Fixed
Mother's age: 24-33	0.082	0.081
Mother's age: >33	-0.10	0.084
Child Poverty Index	0.0052	0.0025
Between ward variance	0.12	0.019

A4.3 A simpler version of the model corresponding to Table A4.2 was fitted and used in the construction of non-response weights. This model (Table A4.3) compresses the nine strata into four groups, contrasts mothers over 33 with those 33 and under and contrasts mothers with the title 'Miss' against the rest. In this model there is also an interaction between payment method and mother's age such that the odds of non-response are greater for mothers paid by order book if they are over 33 than if they are 33 and under. This model was estimated using the Markov Chain Monte Carlo (MCMC) method incorporated into *MLwiN* with 15 thousand iterations as this method gives slightly more accurate estimates than PQL(2) which was used for Table A4.2.

**Table A4.3: Losses from the Issued Sample: Estimates (MCMC) from Two Level Logistic Regression**

	<b>Estimate</b>	<b>s.e.</b>	<b><math>\chi^2</math></b>	<b>d.f.</b>	<b>p</b>	<b>Change in odds</b>
All other	0	Fixed	27.3	3	<0.001	n.a.
England, Ethnic	0.39	0.11				1.5
Wales, Ad.	-0.28	0.14				0.76
N. Ireland, Ad.	0.36	0.13				1.4
Claimant, Miss	0.15	0.041	13.3	1	<0.001	1.2
In-mover	0.34	0.15	4.98	1	0.026	1.4
CPI	0.0044	0.0017	6.90	1	0.009	1.0044
Mother's age	-0.38	0.059	94.0	3	<0.001	n.a.
Order book	0.12	0.054				
Age*order book	0.35	0.074				
Between ward variance	0.14	0.023	36.1	1	<0.001	n.a.

A4.4 The model for partner non-response is based on those 'at risk' (n = 14538) and therefore omits families where there is no partner. Again, a two level logistic regression model is used. The between ward variation is more substantial than it is for overall field non-response possibly because of interviewer effects and the intra-ward correlation is 0.13 (Table A4.4).

**Table A4.4: Partner Non-response: Estimates (PQL(2)) from Two Level Logistic Regression**

	<b>Estimate</b>	<b>s.e.</b>	<b><math>\chi^2</math></b>	<b>d.f.</b>	<b>p</b>	<b>Change in odds</b>
England, Ad.	0	Fixed	21.5	8	0.006	n.a.
England, Disad.	0.11	0.19				1.1
England, Ethnic	0.53	0.25				1.7
Wales, Ad.	0.29	0.22				1.3
Wales, Disad.	-0.06	0.22				0.94
Scotland, Ad.	0.27	0.19				1.3
Scotland, Disad.	0.28	0.24				1.3
N. Ireland, Ad.	0.34	0.23				1.4
N. Ireland, Disad.	0.68	0.24				2.0
Number of chn.	0.082	0.024				11.8
Main respondent: married	-0.40	0.064	40.2	1	<0.001	0.67
Educational level: 4+	-0.20	0.067	8.5	1	0.004	0.82
Ethnic group: white & other	0	Fixed	35.7	2	<0.001	n.a.
Ethnic group: Asian	0.45	0.17				1.6
Ethnic group: Black	0.75	0.13				2.1
Order book	0.29	0.061	22.2	1	<0.001	1.3
Between ward variance	0.44	0.058	57.9	1	<0.001	n.a.

## References

- Cullis, A. (2007), *Infant Mortality in the Millennium Cohort Study (MCS) Sample Areas*. London: Centre for Longitudinal Studies.
- Hawkes, D. and Plewis, I. (2006), Missing income data in the Millennium Cohort Study: Evidence from the first two sweeps. Paper presented to conference of International Association for Research into Income and Wealth (IARIW), Joensuu, Finland.  
<http://www.iariw.org/papers/2006/HawkesPlewis.pdf>
- Kish, L. (1967), *Survey Sampling*. New York: Wiley.
- Lynn, P., Beerten, R., Laiho, J. and Martin, J. (2001), *Recommended Standard Final Outcome Categories and Standard Definitions of Response Rate for Social Surveys*. Working papers of the Institute for Social and Economic Research, paper 2001-23. Colchester: University of Essex.
- Noble, M. et al. (2000), *Measuring multiple deprivation at the small area level: The indices of deprivation, 2000*. Final report for the DETR.
- OPCS/GRO(S) (1993), *Ethnic Group and Country of Birth, Great Britain two volumes*, London: HMSO.
- Plewis, I. and Ketende, S. (Eds.) (2007), *Millennium Cohort Study Technical Report on Response (2<sup>nd</sup> Ed.)*. London: Centre for Longitudinal Studies.
- Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., Woodhouse, G., Draper, D., Langford, I. and Lewis, T. (2000), *A user's guide to MLwiN, version 2.1*. London: Institute of Education.
- Schuman, J. (1999), The ethnic minority populations of Great Britain – latest estimates. *Population Trends*, Summer 1999.
- StataCorp (2005), *Stata Statistical Software: Release 9.0*. College Station, TX: Stata Corporation.