Millennium Cohort Study

First Survey:

A Guide to the SPSS Dataset

3rd Edition

Compiled by

Peter Shepherd, Kate Smith, Heather Joshi and Shirley Dex

with the assistance of
Brian Dodgeon, Denise Hawkes, Gareth Hughes, Ian Plewis and other members of the Millennium Cohort Team

2004

Centre for Longitudinal Studies

Bedford Group for Lifecourse & Statistical Studies

Institute of Education, University of London

E·S·R·C
ECONOMIC & SOCIAL RESEARCH COUNCIL

CONTENTS

ACKNOWLEDGEMENTS

We wish to acknowledge the support for the survey of our principal funders: the Economic and Social Research Council; the consortium of UK Government Departments (led by the Office for National Statistics); the governments of Wales, Scotland and Northern Ireland; and the International Centre for Child Studies.

The work could not have been carried out successfully without the involvement of over a hundred advisors drawn from researchers, policy makers and funders, who we consulted throughout the design of the surveys, particular thanks to those who have served on the MCS Advisory Committee. Further details are given in the body of this *Guide,* and in Appendix 1.

Also:

- National Centre for Social Research (NatCen) researchers, programmers, field-managers, and interviewers

- The staff of the Information Centre at Newcastle of the Department for Work and Pensions (formerly the Department of Social Security)

- The mothers, fathers and other family members of the babies who form the Millennium Birth Cohort.

PREFACE

This document has been prepared to accompany the 3rd version of the deposit (February 2004), with the UK Data Archive at the University of Essex, of data from the first survey of the Millennium Cohort Study. The dataset issued at this time contains a number of corrections that have come to light in subsequent data cleaning, and supercedes the early editions. Data checking is on-going and it is possible further corrections may be incorporated as the data and documentation are updated in the future.

The other elements of the deposit, to which reference will be made throughout this document, are identified below. Users are advised that they will need to consult all elements of the documentation to gain a full understanding of the data.

MCS Deposit: Elements

Title	Format
Millennium Cohort Study First Survey Dataset	SPSS
Millennium Cohort Study First Survey: Technical Report on Sampling	Word
Millennium Cohort Study First Survey: CAPI Questionnaire Documentation (National Centre for Social Research)	Word
Millennium Cohort Study First Survey: Technical Report on Fieldwork (National Centre For Social Research) To be deposited at a later date	Word
Millennium Cohort Study First Survey: Code Book and Edit Instructions (National Centre for Social Research)	Word
Millennium Cohort Study First Survey: Interactive Data Dictionary for the SPSS dataset (based on the SPSS Data Dictionary)	Idealist for Windows
Millennium Cohort Study First Survey: Guide to the SPSS Dataset	Word
Millennium Cohort Study First Survey: Derived Variables	Word

PART ONE

1.1 Introduction

1.1.1 History

The Millennium Cohort Study offers large-scale information about the New Century's babies, and the families who are bringing them up, for the four countries of the United Kingdom. Its First Sweep, carried out during 2001-2002 has laid the foundations for a major new longitudinal research resource, taking a new "year long" cohort of around 19,000 babies. In England and Wales they were born over the twelve months, starting in September 2000 in England and Wales, and over 13 ½ months from late November 2000 in Scotland and Northern Ireland. Information has been collected from parents when the babies were aged nine months. The sample design allowed for disproportionate representation of families living in areas of child poverty, in the smaller countries of the UK and in areas with high ethnic minority populations in England. The first survey recorded the circumstances of pregnancy and birth, as well as those of the all-important early months of life, and the social and economic background of the family into which the children have been born. This multidisciplinary baseline data will reveal the diversity of starting points from which these 'Children of the New Century' are setting out.

The study's broad objective was to create a new multi-purpose longitudinal dataset, describing the diversity of backgrounds from which children born in the New Century are setting out on life. Additional government funding led by ONS has extended the objectives to provide control cases for the National Evaluation of Sure Start, and to provide sufficient cases for intra-country analysis of Wales, Scotland and Northern Ireland, raising the target sample from 15,000 to over 20,000 children.

This report reviews the background to the data collection for the first survey of the Millennium Cohort, from its commissioning by the ESRC in early 2000 to its deposit in the ESRC Data Archive in May 2003. It covers developing the funding and design of the project, the two pilot surveys and the main fieldwork in all four UK countries and the subsequent coding and data preparation, along with unanticipated challenges and resulting modifications of original plans.

The Millennium Cohort Study is the fourth of Britain's world-renowned national longitudinal birth cohort studies. Each follows a large sample of individuals born over a limited period of time through the course of their lives, charting the effects of events and circumstances in early life on outcomes and achievements later on. They show how histories of health, wealth, education, family and employment are interwoven for individuals and vary between them. The data collected are used for many scientific and policy purposes.

1.1.2 Background to ESRC funding

The three original birth cohort studies followed cohorts born in Great Britain in one week each of 1946, 1958 and 1970. For a comparison of some of their findings see, Ferri, Bynner and Wadsworth (eds) (2003) *Changing Britain: Changing Lives*, Institute of Education, London. During the 1980s and 1990s funding for following up

the existing cohorts was precarious, and none was forthcoming for starting a new study in the series. In 1998, the ESRC established a National Strategy for Longitudinal Data Resources which put the future funding of the 1958 and 1970 cohorts on a firmer footing, alongside the British Household Panel Study. In 1999, the British government decided to mark passing of the Millennium with a new birth cohort study. The ESRC commissioned a feasibility study from Professor Jonathan Gershuny who held wide consultations with the research community. This resulted in an invitation to tender for the Principal Investigator role in a new Study and to submit expressions of interest to carry out its fieldwork. These invitations were issued when the new Millennium was already two months old on February 23 2000.

The original ESRC funding was for a study which in some ways resembled that of the previous cohorts, but which departed from the previous designs in the following respects:

- The sample was to include children born over a full twelve months rather than one week, to include births in all seasons
- The sampled birth dates should include the birth dates in the year 2000, even though it might not be possible to start with birth in January
- The geographical coverage of the study was to include the whole of the UK
- The content should emphasize the social and economic circumstances of the families
- The target sample size was to be 15000
- The target interview length was constrained by a fieldwork budget of £1.7m
- The contractors were to consider sample designs allowing for the over-representation of ethnic and national minorities
- The first survey was to be carried out when the children were at about the same age in months, as near as possible to 6 months old

A further contrast with the earlier surveys follows as concerns the mode of data collection. Most of these conditions pointed towards fieldwork being carried out by a professional fieldwork agency, with experience of probability sampling and social scientific questionnaires, administered by computer, rather than the mobilisation of health visitors that had carried out the interviews shortly after the births of the previous cohorts. Constraints on the available numbers of professional interviewers, and the consequent need to spread interviewing over time were also considerations favouring a wider span of birth dates than a single week.

1.1.3 ESRC Funding of the Centre for Longitudinal Studies (CLS)

The Principal Investigator (PI) contract for the Millennium Cohort Study, First Survey (MCS1) was awarded in May 2000, to a consortium based in three institutions:

I. Institute of Education: Centre for Longitudinal Studies (John Bynner);
II. University College London: Department of Community Epidemiology and Public Health, (Michael Marmot), and Institute of Child Health, (Catherine Peckham);
III. City University: Department of Psychology, (Dermot Bowler).

The scientific leader of the project was Heather Joshi, based at CLS in the IoE, along with the internal management team.

Heather Joshi was confirmed as Director of the Millennium Cohort Study in October 2001 when the Principal Investigator Contract for Sweep 2 was awarded to the CLS-led consortium through the Joint Centre for Longitudinal Research. At that time the study was taken into the umbrella of the portfolio of longitudinal data resources overseen by the ESRC's National Strategy Committee for Longitudinal Data. The management structure for the MCS1 can be found in Appendix 1.

1.1.4 Objectives

The objectives of the first MCS survey were laid down in the proposal to the ESRC in March 2000:

1. To chart the initial conditions of social, economic and health advantages and disadvantages facing new children in the new century, capturing information that the research community of the future will require.
2. To provide a basis for comparing patterns of development with the preceding cohorts.
3. To collect information on previously neglected topics, such as fathers' involvement in the children's care and development.
4. To focus on the children's parents as the most immediate elements of the child's 'background', charting their experience as mothers and fathers of this year's babies, to record how they (and any other children in the family) are adapting to the newcomer, and what their aspirations for her/his future may be.
5. To emphasize intergenerational links including those back to the parents' own childhood.
6. To investigate the wider social ecology of the family, including, social networks, civic engagement and community facilities and services, splicing in geo-coded data when available

1.1.5 Fieldwork Sub-contract

As there was only one expression of interest in the fieldwork contract after the February Call, ESRC required the fieldwork sub-contract to be re-advertised. It took CLS and partners just over two months to develop the design of the survey sufficiently to specify the Invitation to Tender for the fieldwork contract, and one month for three potential contractors to submit their proposals. Following post-tendering negotiations to reduce the price as instructed by the tendering panel, The National Centre for Social Research (NatCen) was appointed to carry out the fieldwork at the end of September 2000, 7 months after the original commissioning. The fieldwork in Northern Ireland was sub-contracted by NatCen to Central Survey Unit of NISRA (the Northern Ireland Statistics and Research Agency).

1.1.6 Co-funding

During the last three months of 2000 a consortium of government Departments (led by ONS) substantially supplemented resources available to the Survey. This enabled the length of the interview at Sweep 1 to be extended by 15 minutes from the 90 available with ESRC funds, and to increase the target sample size from 1500 each in the Celtic countries - by 1500 in Wales, 1000 in Scotland and 500 in

Northern Ireland. The target sample in England was also boosted by approx 2650 cases in 35 additional disadvantaged wards to provide control cases for the National Evaluation of Sure Start. The ONS funding also supported work on later sweeps and enhancement of health and neighbourhood data most of whose results will come after the first deposit described here

Substantial co-funding by government departments, over at least five years, has brought new objectives to the Study. One is for the MCS to provide controls for the national evaluation of Sure Start described above. The governments of Wales, Scotland and Northern Ireland contributed to boosting the samples for the first sweep in Celtic countries in order to achieve adequate samples for analysis in the individual countries. The ONS consortium also contributed to collecting ecological data, enhancing the health information and the preparation of reports and analysis.

There has also been small scale but timely co-funding by the ICCS (International Centre for Child Studies) to aid the attempts to reach and survey health visitors and to provide a thank-you voucher for participants in the first pilot.

1.2 Sample Design

One of the first tasks for the PI team was to select a sampling strategy and establish how it could be operationalized in the tight timescale demanded by the survey. It was decided by mid July, that the sample of a years births should be tightly clustered geographically, and disproportionately stratified to over-represent areas with high proportions of: ethnic minorities in England, areas of high child poverty and the three smaller countries of the UK respectively. It was decided to use the geography of electoral wards as a sampling frame. The Index of Deprivation 2000 (Noble et al (2000), *Measuring multiple deprivation at the small area level: The indices of deprivation, 2000.* Final report for the DETR.) used wards with 1998 boundaries in England. Its child poverty component provided as up to date as possible an indicator of wards with high proportions of children in families receiving means tested benefits.

The sample for the first sweep included babies born between September 1 2000 and August 31 2001 in England and Wales, who will form an academic year cohort. In Scotland and Northern Ireland the start date of the birthdays was delayed to November 23 2000 in order to avoid an overlap with an infant feeding survey being carried out in September and October. In the event the sampled cohort was extended to 59 weeks of births to make up for a shortfall in numbers which became apparent during fieldwork. The last eligible birth date in these countries was January 11 2002.

Children with sample birth dates were eligible for the survey if they were living in one of approximately 400 electoral wards across the whole of the UK when they were 9 months old.

The disproportionately stratified design of the survey was to ensure adequate representation of:

- All UK countries
- Areas in England with higher minority ethnic populations in 1991
- Disadvantaged areas

The estimated breakdown of the target sample for Sweep 1 is outlined in table 1 below:

Table 1 Millennium Cohort Sample Structure by stratum and country: wards and babies aged 9 months *(numbers of children in italics)*

Country	**Total wards** ***target nos of children***	***(of which, boost from original number)***	**Advantaged wards** ***expected sample***	**Disadvantaged wards** ***expected sample***	**Ethnic wards** ***expected sample***
England	**200** ***13146***	(babies in 35 wards for Sure Start *2646*)	110 *5511*	71 *5258*	19 *2606*
Wales	**73** ***3000***	(*1500 babies in disadvantaged wards*)	23 *897*	50 *2219*	Na
Scotland	**62** ***2500***	*(1000 babies)*	32 *1243*	30 *1285*	Na
Northern Ireland	**63** ***2000***	*(500 babies in disadvantaged wards*)	23 *762*	40 *1322*	Na
Total	**398** ***20646***	***5,700***	**188** ***8,413***	**191** ***10,084***	**19** ***2,606***

'wards' counts 49 amalgamations of small wards ('superwards') as units. See Technical Report on Sampling)

1.3 Sampling procedures

1.3.1 The use of Child Benefit Records

Another key decision was to take advantage of an offer by the Department of Social Security (now DWP) to make Child Benefit records available to find the children. We had originally assumed there was no alternative to birth registration records. We came to favour the CB route, because of concern about non-response to the postal opt-in operated by ONS. The representativeness of the sample could be highly compromised in relation to socially excluded groups, including those with poor literacy or poor grasp of English. DWP ask people to opt-out, which is known to be more inclusive of marginal and low literacy cases. The arrangement to use the DWP system was confirmed in September 2000

However DWP do withdraw 'sensitive cases' from the issued sample. These include parents of children who have died, or have been taken into care or where there is an investigation into benefit fraud, for example. At one point families who were in correspondence with DWP for virtually any reason were being excluded, but the

criteria for these exclusions were reviewed before MCS1 sampling started and in the event fewer than 3% of cases were affected One of the reasons for exclusion was to avoid families who had also taken part in the Survey of Low Income Families (now the Families and Children Survey, FACS). This involved only 40 cases (out of over 27,000).

1.3.2 Recruitment via Health Visitors

Because the Child Benefit records do not reveal all families who have moved into the sample wards as the child approaches 9 months of age, we attempted to find movers-in with the help of Health Visitors. These local community health professionals were expected to be aware of families transferring into these areas in the course of their duties., They were asked to see if families moving into survey wards were willing to be recruited . They were also asked to inform and reassure other families, who might have received the opt-out, about the survey.

It is difficult to quantify how important this reassurance may have been, but we do know that health visitors reported 220 cohort in-moving families with children over 6 months of age, however only 56 had not also been found by DWP. We also used our Health Visitor contacts to distribute a postal survey on local services in the sample areas.

There were several problems which may explain the rather disappointing result of this exercise. Firstly, helping with the survey was not part of the health visitor's already demanding normal duties. Secondly, Health Visitors' caseloads do not neatly coincide with electoral wards, and thirdly there is no central list of Health Visitors for easy contact. Tracking down the relevant Health Visitors or Supervisors proved time consuming, particularly in view of the reorganisation of the Community Care Trusts into Primary Care Trusts during 2001. The exercise benefited vitally from the help and experience of Neville Butler and the financial support of ICCS. It was aided by additional ESRC funds for maps. We eventually distributed 3330 packs to Health Visitors with the help of 306 supervisors.

.

1.4 Ethical Clearance

1.4.1 MREC

The process of gaining medical research ethical approval proved a major hurdle. As had been the practice with the previous cohort studies, medical research ethical clearance was sought (in February 2001). This was as a general precaution for future health data collection and was specifically required because of the proposal to involve Health Visitors. Any research using NHS staff needs to be given such clearance. We were directed to the South West Multi-Centre Research Ethics Committee in March 2001, who felt that opt-out sampling could be coercive and fail to obtain properly informed consent. They did however accept that written opt-ins would tend to exclude vulnerable people, so procedures were devised in consultation with the Committee to give potential respondents more information before they committed themselves for interview. Advance letters introducing the interviewer were sent shortly before her/his first visit and they were asked to arrange interviews

generally after their first visit, whose main purpose should be to give information. A simplified information sheet was produced, and translated into several languages.

1.4.2 Codes of Practice

In order to support our assurances of confidentiality to informants, Ethics Committees, and government agencies to whose records links are being made, the CLS extended the Cohort Studies Code of Practice to cover all those working with MCS data and developed a Data Security Policy, setting out the secure, isolated computing environment which handles any named data files within CLS.

1.5 Survey Content

1.5.1 Development and Piloting

There were barely 13 months from the ESRC letter notifying CLS of the intention to award the contract to the main survey going into the field. This was an exceptionally tight timetable to develop a survey. The imperative to catch at least some Millennium births before they leave infancy constrained the time available, and compressed operations which could well have taken more time

The process of questionnaire development, which had already begun, dominated the months following the award of the fieldwork contract. Fifty-five potential users of the dataset from academe and government departments attended a consultation meeting on October 11 2000. An instrument was initially piloted in January 2001, and redeveloped, into a shorter version for the second Dress Rehearsal Pilot in April 2001.

1.5.2 First Pilot

The first pilot in January 2001 was conducted as a paper interview and computer aided self-completion instrument (CASI) in order to assess the timing of the instrument before the major work to convert the interview schedule into CAPI format. The sample size was boosted from 30 to 60 thanks to the ONS consortium funding. Further details are in the NatCen Technical Report on Fieldwork.

The overall response by parents to the pilot survey was positive, and interviewers reported that the gift vouchers did not seem to be necessary to persuade people to talk about their babies. However the instrument was substantially over length, averaging over 2 hours, the target being 105 minutes. Reductions to the questionnaire were made in consultation with collaborators, and using the results factor analysis of the CASI responses on attitude and behaviour.

1.5.3 Dress Rehearsal Pilot

The second pilot took place during April 2001 and was fully computer based (CAPI and CASI). As a ‘dress rehearsal’ for the main stage, all the contact and administrative processes were tested as well as the near final form of the survey instruments. Thirteen wards were selected for this pilot, including one in each of

Wales & Scotland. The wards in England and Wales were chosen from those that were to be used in the main stage. As the Scottish wards had not yet been selected, a large deprived ward was purposively picked.

The DWP sampling route was tested with letters sent from the DWP at Newcastle to parents of babies born between June12th and July 22nd 2000 on the Child Benefit register in the chosen wards. The use of an advance letter sent by interviewers was also piloted.

In addition, Health Visitors were approached in the twelve English and Welsh wards in order to pilot their contribution. Two HV supervisors declined to help, as we had not received MREC approval at that time.

The main result of the survey was that respondents and interviewers had, on the whole enjoyed the experience. The interview, and particularly the self-completion however were still too long, especially where translation was needed. Of the two experiments which were part of the Department of Health financed feasibility studies, the one on transcribing data from the Personal Child Health Record ('Red Book') was not sufficiently successful to take into main fieldwork, but the pilot consent for health record linkage went well.

1.5.4 Approaching the Main Stage

There were only three weeks between receiving the report on the Dress Rehearsal Pilot and the first main survey briefing. The main instrument was cut by 8 minutes and the partner by 5 minutes. Refinements were made to individual items to improve their comprehension and the flow of the questions, routings were checked and answer codes revised. Checks were also imposed to improve the automatic edits written into the CAPI programme to ensure the quality of data collected. A few changes were made to questions for use in Scotland and Northern Ireland - on qualifications, religion and ethnicity to make them country relevant.

1.5.5 Structure and Content of Final Instrument

The content of the Sweep 1 instruments is summarized in table 2. The module lettering reflects the order of each interview with the self-completion inserted between interview questions on health and employment. The lettering of the modules appears in the CAPI document and in the labelling of variables in the SPSS dataset, described below.

Table 2 Summary of Survey Elements

Respondent	Mode	Summary of content
Mother/Father	Interview	Household Module
Mother/main		Module A: Non-resident parents Module C: Pregnancy, Labour and Delivery Module D: Baby's health and development Module E: Childcare Module F: Grandparents and Friends Module G: Parent's health
	Self-	Module H: - Baby's temperament & behaviour - Relationship with partner - Previous relationships - Domestic tasks - Previous pregnancies - Mental health - Attitudes to relationships, parenting, work, etc
	Interview	Module J: Employment and Education Module K: Housing and local area Module L: Interests and time with baby
Father/Partner	Interview	
		Module B: Father's involvement with baby Module C: Pregnancy, Labour and Delivery Module F: Grandparents and Friends Module G: Parent's health
	Self-	Module H: Self-completion - Baby's temperament & behaviour - Relationship with partner - Previous partners - Previous children - Mental health - Attitudes to marriage, parenting, work, etc
	Interview	Module J: Employment and Education Module L: Interests and time with baby

This reflects the broad topics originally envisaged, but it was not possible to collect as much detail on all the subjects that had been suggested. Psychometric inventories (listed in appendix 2) were pruned, as were details of housing history, education, employment history, income from different sources and neighbourhood facilities. It was decided not to risk sensitive questions such as illegal drug use or abortion in the first round of the survey.

The study's first three objectives - capturing intra cohort diversity, allowing for inter-cohort comparisons and covering new topics) were implemented in the instrument for Sweep 1. Objective 4, the recording of adaptation of other family members to the new baby and aspirations for the baby's future was not. Neither was the previous generation (Objective 5) well covered. Due to the limits of interview time, some questions had to be sacrificed or reserved until later sweeps to concentrate adequately on topics essential to the first data collection. Work is under way on objective 6, with the help of government co-funding for the Health Visitor Survey and the addition of Neighbourhood Statistics data, but the first release of the data does not contain a full set of geographic variables, which awaits the development of procedures to protect respondents' confidentiality.

1.6 Developments and adaptations

1.6.1 Timing of Cohort births and age at interview

As has been described above, it was originally envisaged to start a year's fieldwork surveying children who had been born during 2000 at an age around 6 months. Given the need to put the fieldwork contract out to tender, this timetable proved unsustainable. The resulting shifting of birth dates meant that the age of child at which the interview should take place was settled at 9 months with fieldwork starting in June 2001, postponed from May 2001 in view of the census and the anticipated General Election. This meant that deposit of the dataset was also postponed one month to the end of March 2003.

1.6.2 Intensification and Extension of the Sample

Following the September 2001 Advisory Committee, there was concern over a projected shortfall in sample numbers, but we were advised by the experts and the government departments to strive to maintain or raise response rates rather than extending the sample. To this end, NatCen re-issued some extra cases in difficult areas of Great Britain.

The DWP also supplemented their operations to look for families moving into sample addresses when the babies were between 7 and 8 months old contacting them to opt-out when the children were 9 months old and issuing them for interview at 10 months of age. This took place from November 2001 onwards covering children born after 16 March 2001. This exercise found a total of 518 families with new addresses in the survey areas, but only 293 could be issued to field, as these cases involved a disproportionate number of cases excluded by the DWP as 'sensitive', suggesting that the correspondence which provoked their exclusion from the sample was also increasing the chances of their change of address being known to DWP.

The Executives of Scotland and Northern Ireland expressed concern at the May 2002 Advisory Committee about the expected shortfall in sample numbers in these countries. Instead of the target of 2500 in Scotland and 2000 in Northern Ireland, our projections for 53 weeks of sampling were for 2173 and 1659 children respectively. DWP offered to do one more wave of sampling, covering children living in Scotland

and Northern Ireland who were born in the six weeks starting 1 December 2001, finishing on 11 January 2002.

The extension of birth dates by six weeks in samples in Scottish and Northern Irish wards, was expected, on simplified assumptions to increase the total of Scottish children towards 2419 and Northern Irish children towards 1847. It would have taken at least another two weeks of births, and extra expenditure to reach the original targets of 2500 and 2000 respectively. As it is this extension delayed the delivery of the dataset by a further 6 weeks, and increased for future sweeps the number of months of fieldwork that would be required to interview the whole cohort at the same exact age.

1.7 Fieldwork

1.7.1 Briefings

The first wave of the main stage fieldwork commenced in England and Wales in June 2001, and in Scotland and Northern Ireland in September 2001. 232 interviewers who were to work in England and Wales were briefed in 17 regional one-day meetings between May 31 and June 15 2001. 42 interviewers working in Scotland were briefed at 4 sessions between August 29 and September 6. These training sessions were conducted jointly by researchers from NatCen and CLS. In Northern Ireland, some 50 interviewers were briefed at 4 sessions between August 17 and August 28.

1.7.2 Fieldwork Timetable

The fieldwork was carried out in 17 consecutive waves. Each issued wave of fieldwork contained babies born in a 4-weekly birth cycle, with the first wave covering the births between 1/9/2000-28/9/2000 in England and Wales. This rhythm of recruiting the sample was dictated by the cycle of DWP procedures, scanning the Child Benefit database every four weeks. Interviewers arranged interviews as soon as possible after the addresses were issued, aiming to reach the families while the baby was as close as possible to 9.5 months of age. Interviews with partners could be delayed until the child's first birthday (as were some main interviews where the address had been issued late).

The process for drawing each wave of the DWP sample is as follows:

Prior to fieldwork, the DWP send opt-out letters to all parents of children with an eligible birth-date who were registered (for child benefit purposes) as living within one of the sampled wards, apart from any cases flagged as sensitive. Batches of letters, including an information leaflet, were sent every four weeks to families whose babies were approximately seven months old. The letter invited parents to take part in the study and gave them the opportunity to opt-out of the study by telephoning or writing to the DWP. Any parents who opted out of the study were then removed from the sample.

The final stage was for the DWP to remove cases which they discovered had subsequently moved out of the sampled wards and to update the addresses for cases which had moved within or between sampled wards. At this stage any late opt-outs or newly sensitive cases were also removed.

The data was sent by the DWP to CLS in two stages, a week apart, in order to ensure that any late opt-outs or change of addresses could be notified as near to the start of fieldwork as possible. After the final data was received, serial numbers were assigned to each valid case and the data was sent to NatCen, for issue to the field.

The fieldwork timetable for the project detailing the dates of birth and fieldwork is shown in Table 3.

Table 3 Fieldwork timetable

Fieldwork Wave	Baby's Date of Birth	Fieldwork Period
Wave 1	1st – 28th Sept 2000	11th June – 8th July 2001
Wave 2	29th Sept – 26th Oct 2000	9th July – 5th Aug 2001
Wave 3	27th October – 23rd Nov 2000	6th Aug – 2nd Sept 2001
Wave 4	24th Nov – 21st Dec 2000	3rd Sept – 30th Sept 2001
Wave 5	22nd Dec 2000 – 18th Jan 2001	1st Oct – 28 Oct 2001
Wave 6	19th Jan – 15th Feb 2001	29th Oct – 25th Nov 2001
Wave 7	16th Feb – 15th March 2001	26th Nov – 23rd Dec 2001
Wave 8	16th March – 12th April 2001	24th Dec 2001 – 20th Jan 2002
Wave 9	13th April – 10th May 2001	21st Jan – 17th Feb 2002
Wave 10	11th May – 7th June 2001	18th Feb – 17th March 2002
Wave 11	8th June – 5th July 2001	18th March – 14th April 2002
Wave 12	6th July – 2nd Aug 2001	15th April – 12th May 2002
Wave 13	3rd Aug – 30th Aug 2001	13th May – 9th June 2002
Wave 14	31st Aug – 27th Sept 2001	10th June – 7th July 2002
Wave 15	28th Sept – 25th Oct 2001	8th July – 4th Aug 2002
Wave 16	26th Oct – 23rd Nov 2001	5th Aug – 22nd Sept 2002
Wave 17	24th Nov 2001-11 January 2002	23rd Sept – Jan 10 2003

Note: NatCen numbered these waves 2-18 as they counted the Dress Rehearsal Pilot as Wave 1

Waves 1-13 of fieldwork took place in England and Wales from June 2001 to July 2002. The last wave in England and Wales, wave 13, which included babies born on August 31, was delayed by 4 weeks for operational reasons, so this wave contained interviews mostly conducted at ten rather than nine months for these two countries. The last wave in Scotland and Northern Ireland, wave 17, was the extended sample, spanning 7 weeks of births. The latest interview (with a partner) took place in Northern Ireland on the last but one eligible day, January 10th 2003. Fieldwork in Scotland (and with all main informants) finished before the end of 2002

The aim was that the fieldwork for each wave should be as self-contained as possible, with the minimum amount of overlap. Interviewers were briefed to interview

families when the baby was 9 months and 15 days old, ideally, in order to standardize the data being collected as far as possible. Allowing for delayed interviewing due to tracing problems, the window of opportunity to interview was brief, up to 11 months of the babies' age for the main interview and up to 12 months for the partner.

75% of main interviews took place while the baby was aged 9 months, 3579 (19%) at 10 months, with 541 (3%) at 8 months, representing babies born towards the end of the 4-week span interviewed early in the fieldwork period. 479 interviews took place late, 475 at 11 months and only 4 in month12 –13. 17 were not interviewed because the time window had expired by the time they were found. They are included in the 'other ineligible', Table 7.2 , Technical Report on Sampling.

1.7.3 Languages

In order to comply with the recommendations made by the MREC, a simplified leaflet was produced for interviewers to give to respondent families on the doorstep. This leaflet, the advance letter and the thank-you letter were translated into the most common non-English languages spoken in the 19 selected 'ethnic' wards. The languages appropriate for translation were: Bengali, Gujerati, Kurdish, Punjabi, Somali, Turkish and Urdu. The first leaflet had already been translated into Welsh. Some interviews were carried out in verbal translation (in these and other languages) by relatives or friends. In certain circumstances where no one was available to translate into English, NatCen provided translator interviewers. Other languages encountered in non-trivial numbers included Arabic, Hindi and Tamil. 226 (1%) Main interviews were carried out in a non-English language and a further 547 (3%) were done in a mix of English and another language. For partners the corresponding figures were 306 (2%) and 94 (1%).

1.7.4 In field Tracing

On the whole the addresses supplied by DWP proved to be current. Unfortunately, in a proportion of those issued to the field, the families had moved, either after the baby was aged 7 months or else before the baby reached seven months, but had not informed DWP of their move. Where a family was not living at the issued address and the interviewer could not establish a new local address, cases were returned to CLS for tracing. Where a new address was found within a selected ward, cases were re-issued to the field. Where a family had moved to a non-selected area, but were resident at their old address when the baby was aged 9 months, they could be interviewed at the new address.

1.8 Achieved Sample

Overall, the project was very well received in the field. In most cases parents have been content to participate and the experience has been a happy one for both families and interviewers alike. Nevertheless the total number of families who gave at least some information did not reach the 20,000 mark. The survey reached 18,553 families, which, after allowance for 246 sets of twins and 10 sets of triplets, amounted to nearly 19,000 children in the cohort: 18,819 to be precise. 18533 main

interviews were given, almost entirely by mothers. 3194 parents, again almost all of them mothers were living without a resident partner. In 1760 cases there was a resident (or part-time resident) father who did not give information. 338 of the partners' information was given by proxy. There was thus some information for 89% of resident partners. In 20 cases it was information from the mother that was missing. Table 4 shows how these respondents are distributed over the four countries of the UK. Further details by stratum appear in the Technical Report on Sampling.

Table 4 MCS Sample size: Clusters, Children Families, by country

	Number of sample 'wards' *	Target sample as boosted	Achieved Responses **			
			Children	Families interviewed	Partners	Single Parents
Total UK	**398**	**20,646**	**18,819**	**18,553**	**13,441**	**3,194**
ENGLAND	**200**	**13,146**	**11,695**	**11,533**	**8,485**	**1,853**
WALES	**73**	**3,000**	**2,799**	**2,761**	**1,933**	**590**
SCOTLAND	**62**	**2,500**	**2,370**	**2,336**	**1,727**	**375**
N IRELAND	**63**	**2,000**	**1,955**	**1,923**	**1,296**	**376**

Notes

* Counting 'superwards' as a single unit ** All productive contacts

In the vast majority of cases the natural mother did the main interview. The exceptions are 2 adoptive mothers, 2 foster mothers, 18 lone fathers, 2 natural fathers where the natural mothers answered the partner interview, 1 father with proxy interview for natural mother and 5 maternal grandmothers. The sex of respondents to main and partner questionnaires is given in Table 5, showing that there were exceptions to the general rule of mothers being the main respondent and partners being fathers, but that the exceptions were very few.

.

Table 5 Sex of Respondents

	Sex of Main respondent		Sex of Partner interviewed or Proxied	
	Male	**Female**	**Male**	**Female**
Main and partner respondent in person	2	13203	13200	5
Main respondent in person (no-one eligible for partner)	18	3176		
Main in person, partner by proxy	1	215	215	1
Main in person, partner eligible but no response	0	1918		
No main interview, partner interviewed in person			20	0
Total	**21**	**18512**	**13435**	**6**

NOTE: The total number of Main respondents does not equal the number of families, due to the 20 cases where the Main respondent was not interviewed.

After the extension of sample birth dates in Scotland and Northern Ireland, the achieved sample of children came within 5% and 2% respectively to their targets of 2,500 and 2,000. With only 52 weeks of births in England and Wales, the achieved sample was 10% below target. About half of this shortfall can be attributed to the drop in births which accompanied the Millennium (see Technical Report on Sampling, appendix 1) The rest was due to lower response rate than expected in some parts of the country at some stages of the multistage process.

1.8.1 Response Rates

The overall response can be thought of as the combination of the leakages to numbers between the target population in the selected wards and the sample issued to field and the success the interviewers then have at securing interviews from the issued sample. The 'leakages' between the target population and the in scope population are set out in the Technical Report on Sampling as:

- Families opting out of the survey
- Families excluded by DWP
- Families excluded from the sampling frame because of their postcode could not be allocated to a ward
- Undetected in-migrants

It is possible that the target population should also include some families who do not claim Child Benefit, but we make the simplifying assumption that numbers of such families who are permanently resident in the UK are negligible. The Technical Report on Sampling makes two alternative assumptions about how many undetected in-movers there are. The estimates quoted in Table 6 assume that there is an undetected in-comer for every detected out-mover on average in each stratum. The alternative estimate sets undetected in moves to zero, which raises all (except the Northern Ireland overall response rates) above the target or assumed response rate set in the design of the survey and shown in the first column of Table 6. This table shows that, when undetected in migration is counted as a leakage, the overall

response rate is 68% for the (unweighted) UK sample, modestly below the 71% expected. It is below target in every stratum except the advantaged wards of Wales, but only marked so in Northern Ireland, with overall response rates in the combined strata of 63% where 71% had been somewhat optimistically set, given the lack of a tradition of such surveys in that country. Another handicap in Northern Ireland is that it is the only country where inability to assign Child Benefit claimants to a ward was a significant problem. Survey work in the ethnic areas of England was also something of an unknown quantity. A cautious target of 65% was missed by 3 points.

Table 6 Response Rates by Stratum and Country

		Expected Overall Response Rate	Achieved Overall Response Rate	In-scope Response Rate Fieldwork
England	Ad.	75%	73%	86%
	Disad.	70%	68%	82%
	Ethnic.	65%	62%	76%
	Total	70%	68%	82%
Wales	Ad.	75%	78%	89%
	Disad.	70%	69%	83%
	Total	71%	72%	84%
Scotland	Ad.	75%	73%	86%
	Disad.	70%	68%	83%
	Total	71%	70%	85%
N. Ireland	Ad.	75%	65%	81%
	Disad.	70%	61%	78%
	Total	71%	63%	79%
UK	All	71%	68%	82%

Source MCS Technical Report on Sampling

Out of the cases issued to field some have been deemed ineligible because they are known or thought to have moved out of the survey area before the child reached 9 months of age. Of the remaining eligible, or 'in scope' sample, the response in fieldwork averaged 82% giving at least one interview. It varied by stratum as expected, but more so, the ethnic wards, as anticipated, had least, 76%, and the advantage areas of Wales the highest, 89%, with both strata in Northern Ireland being below the stratum average for Great Britain.

Complete evaluation of sources of the known characteristics of case of survey loss before and after the CB stage will be included in a later edition of the Technical Report on Sampling. Preliminary indications of those lost before issue to field is that they are not systematically biased. A greater propensity of families in the disadvantaged areas to be excluded by DWP is balanced by a greater propensity of the inhabitants of advantaged areas to opt out.

Overall, the project was very well received in the field. In almost all cases parents have been content to participate and the experience has been a happy one for both families and interviewers alike.

1.9 Post field activities

1.9.1 Coding and Editing

Details of coding and editing activities can be found in the Codebook and Edit Instructions prepared by NatCen, included in this deposit and their Technical Report on Fieldwork (to be deposited). Special thanks to Professor Neville Butler who was tireless in developing coding frames for the open text answers to health questions, and in supervising the ICD10 coding at CLS of responses on mothers' and fathers' longstanding illness.

PART TWO

2.1 MCS First Survey Data

The main data for the first survey of the Millennium Cohort Study (MCS) is supplied to the UK Data Archive in the form of a single SPSS dataset. This holds 2018 variables for a total of 18,553 cases (families with one or more cohort child).

As Table 7 shows, the bulk of these represent personal interviews with both mother and father. Short proxy interviews were undertaken with the main respondent where the father (-figure) was unavailable during the period of the survey or prevented from answering through incapacity.

Table 7 Full and Proxy Responses by country

	England		**Wales**		**Scotland**		**N Ireland**		**UK**	
	N	***%***	***N***	***%***	***N***	***%***	***n***	***%***	***N***	***%***
Main and partner interviewed in person	8324	72.18	1908	69.11	1704	72.95	1269	66.00	13205	71.17
All eligible respondents: in person and by proxy	149	1.30	24	0.86	19	0.81	24	1.25	216	1.16
Main in person, partner eligible in person but not interviewed	1190	10.31	237	8.58	231	9.89	241	12.53	1899	10.23
Main in person, partner eligible by proxy but not interviewed	5	0.04	1	0.04	3	0.13	10	0.52	19	0.10
Partner in person, main respondent not interviewed	12	0.10	1	0.04	4	0.17	3	0.16	20	0.11
Main respondent interviewed in person (no eligible partner)	1853	16.07	590	21.37	375	16.05	376	19.55	3194	17.22
	11533	100.00	2761	100.00	2336	100.00	1923	100.00	18553	100.00

2.1.1 Variable names

The variable names on the dataset are mainly those automatically allocated by the CAPI program (Blaise 4). Within Blaise, each question has a variable name, made up of a maximum of 8 characters, and this is used to determine the variable name on the dataset. The original allocation of variable names within the CAPI program was based on the use of mnemonics with the first character (A-L) indicating the CAPI 'module' (or 'section') within the instrument. The exception is the proxy modules for which variables begin with 'px' or 'prox'. The modules and associated characters are summarized in the following box:

Module A: Non-resident parents
Module B: Father's involvement with baby
Module C: Pregnancy, Labour and Delivery
Module D: Baby's health and development
Module E: Childcare
Module F: Grandparents and Friends
Module G: Parent's health
Module H: Self-completion
Module J: Employment and Education
Module K: Housing and local area
Module L: Interests and time with baby

Where the question is repeated (eg: the same question is asked for each person in the household, baby, illness, hospital admission, etc reported), Blaise automatically allocates a number suffix (eg: name, name2, name3, name4). Unfortunately, where the variable name in the Blaise program was originally more than 6 characters long, Blaise truncates the name to allow for the suffix. As a result, there is not always a simple match between the Blaise program documentation and the data. For example, the suffix 2 in the case of some variables indicates the second child within a set of multiple births, and in others it represents the partner's answer to a question also put to the main respondent. However the partner variables are not invariably thus named.

Details of the CAPI/CASI instruments are to be found in the following, which also accompanies the data deposit:

Millennium Cohort Study First Survey: CAPI Questionnaire Documentation

2.1.2 Variable labels

The variable labels included on the dataset were also initially derived from the CAPI program. In exporting the SPSS dataset from Blaise, labels based on the wording of questions were automatically allocated. Subsequently, these have been individually reviewed and, where necessary, modified in an effort to ensure that labels are comprehensible and accurate. In particular, prefixes have been assigned to the variable labels which indicate whether the question is from the Household Interview (HHGrid), Main interview (Main), Proxy Interview (Proxy), Partner Interview (Part) or whether the variable is a survey administration variable (Admin).

Again, particular problems occurred where a question was repeated (eg: the same question is asked for each birth, relationship, job, qualification, etc reported). When initially created, the Blaise-generated dataset had identical labels for each repeat of the question. In revising these labels, efforts have been made to indicate which variables relate to which household member, baby, illness, etc.

An example based on the repeated question: *"Thinking of the baby who does not live with you now, did you have a boy or a girl?* Child 1", "Child 2", etc identify the first, second, etc absent child reported.

Variable labels for repeated question: *"Thinking of the baby who does not live with you now, did you have a boy or a girl?*

Respondent	Child	Variable name	Label
Main respondent			
	1	Hboygrl1	Main: childs sex (child 1)
	2	Hboygrl2	Main: childs sex (child 21)
	3	Hboygrl3	Main: childs sex (child 3)
Etc up to	9	Hboygrl9	Main: childs sex (child 9)
Partner			
	1	Hboygr11	Part: childs sex (child 1)
	2	Hboygr12	Part: childs sex (child 2)
	3	Hboygr13	Part: childs sex (child 3)
Etc up to	10	Hboygr20	Part: childs sex (child 10)

2.1.3 Value labels

The value labels are also similarly derived from the Blaise program and have similarly been reviewed and, where necessary, modified in an effort to ensure that labels are comprehensible and accurate.

2.1.4 Missing values

Missing values have been declared within the deposited SPSS dataset. Unless otherwise labelled, they have been given consistent values as indicated below:

-1	Not applicable
-8	Don't Know
-9	Refusal

2.1.5 Variable order

The order in which variables appear in the dataset will broadly follow the order of sections, and of questions within sections of the survey instruments, viz:

Household Module

Mother (Main)
Module A: Non-resident parents
Module C: Pregnancy, Labour and Delivery
Module D: Baby's health and development
Module E: Childcare
Module F: Grandparents and Friends
Module G: Parent's health
Module H: Self-completion
Module J: Employment and Education
Module K: Housing and local area
Module L: Interests and time with baby
Proxy Module

Father (Partner)
Module B: Father's involvement with baby
Module C: Pregnancy, Labour and Delivery
Module F: Grandparents and Friends
Module G: Parent's health
Module H: Self-completion

Module J: Employment and Education
Module L: Interests and time with baby

With few exceptions, the order of variables is in the data is the order they appear in the CAPI questionnaire documentation and the order they were put to the respondent.

2.1.6 Values

As this was a CAPI/CASI survey, the values of the data should be within the specified range for each variable.

2.1.7 Consistency

Again, the use of CAPI/CASI should ensure that all filters have been correctly followed, and that data are consistent.

Further details of the content of the data set can be found by generating an SPSS 'data dictionary'. An interactive version of this, which facilitates key word/phrase searches of the content of the dataset, also accompanies the data deposit:

Millennium Cohort Study First Survey: Interactive Data Dictionary for the SPSS dataset (based on the SPSS Data Dictionary)

2.1.8 Excluded variables

A number of variables have been removed from the dataset originally derived from the CAPI/CASI program, namely:

- Variables which might compromise the anonymity of cohort members and their families. This includes a number of variables holding the text of open answers and dates of birth (only month and year are deposited).

- Variables which held no data. In general, their presence in the original CAPI/CASI data reflected the need to allow for the repetition of questions/groups of questions (eg: within the household grid). However, there are a small number of empty variables which remain on the data where it was felt that this would aid users.

2.1.9 Data Collection Errors

In a number of cases, interviewers made errors in data collection which were identified by the fieldwork agency during the data preparation stage. Where possible, the data was cleaned to correct these errors. In a small number of cases (identified below) this has not been possible and users should exercise caution when using

data for these cases. These cases mostly involved incorrect application of the proxy module and are identified on the variable 'errtype'.

	Error Type	N	Action taken
1	Proxy module done in error i.e. the proxy section of the main interview was completed about a partner who was not eligible to be interviewed by proxy.	117	Data deleted from proxy module, household outcome code re-classified to 'partial household' and partner outcome code re-classified to unproductive.
2	Partner interview done by proxy in error i.e. the main respondent has completed the partner interview on behalf of partner. Partner should have done the interview him/herself	42	Data deleted from partner interview, household outcome code re-classified as 'partial household' and partner outcome code re-classified to unproductive.
3	Partner answered proxy in person, should have done normal partner interview. i.e. the partner completed the proxy module in person (about him/herself).	6	Data transferred from proxy section to equivalent variables in partner interview, household outcome code re-classified as 'main and partner in person' and partner outcome code re-classified to 'partial interview in person'.
4	Main interview done by father, partner interview by mother i.e. the data indicates that the mother did the main interview and the father did the partner interview but the main interview was actually conducted by the father (in error) and the partner interview was actually conducted by the mother (in error).	2	**NONE**
5	Father did both main and partner interviews i.e. the data indicates that the mother completed the main interview and the father completed the partner interview but actually the father conducted both interviews (should have only done the partner interview).	1	**NONE**
6	Main interview done by partner, no other interview i.e. the data indicates that the mother completed the main questionnaire and the father did not respond to the partner questionnaire but actually the father completed the main interview (in error) and there was no partner interview.	1	**NONE**
7	Grandmother (person 1 in household) was incorrectly coded as natural mother. The actual natural mother (who was person 3) completed the main interview.	1	Relevant variables corrected.

2.1.9 Some useful variables

Variables included on the initial dataset, which may be of particular value to users are identified below.

Information	Variables
Survey elements completed	RESPONDS
Household outcome code: Main respondent outcome code: Partner respondent outcome code:	OUT3DIG MAINOUTC PARTOUTC
Main respondent:	MRESP
Partner respondent	PRESPST
Self-completion: Main Partner	HSELFC HSELFC2
Identifiers	FAMSRNO
Ethnic group (main respondent) England Wales Scotland Northern Ireland	 ETHGRE ETHGRW ETHGRS ETHGRNI
Date of interview	INTDATE
Multiple births	MULTIPLE
Baby's year and month of birth	BDOBY/BDOBM, (BDOB2Y/BDOB2Y,BDOB3Y/BDOB3M)
Baby's birth weight	BWEIGHT (BWEIGHT2, BWEIGHT3)
Baby's sex	BSEX (BSEX2, BSEX3)
Cohort member is first live born (natural mothers)	FSTLBIRTH
Number of children in household	TOTCHILD
Family Composition	FAMILYCO
Country Region	COUNTRY REGION
Type of ward/superward Ward number	PTTYPE SPRPTNO
Weights: - for within country analyses - for UK analyses	 WEIGHT1 WEIGHT2

Additional information about the self-completion, proxy interviews, identifiers and derived variables used within CAPI is given below, along with additional guidance for users.

2.1.10 Self-completion

The self-completion (CASI) was administered during the interview for both mother (main) and father (partner). In each case, the interviewer handed the laptop computer used for the interview to the respondent and explained how they should answer the questions appearing on the screen. Where the respondent was unable or reluctant to use the laptop, the interviewer assisted, and if necessary administered the self-completion as an interview.

The variables which hold the data for the self-completion are identifiable through their variable names – all begin with the character "H". These variables are identified below.

Variables which hold data for the Self-completion (CASI)
(Ordered left to right)

Mother (Main)

HHAPCHG	HUNFAM	HHAIRBR	HFEEDNG	HINJUR	HOBBATH	HWARY	HSHY
HFRETS	HSLEEP	HRGMILK	HRGSLEE	HRGNAPS	HRGSOL	HSLFUSS	HWKFUSS
HSCREAM	HHAPCHG2	HUNFAM2	HHAIRBR2	HFEEDNG2	HINJUR2	HOBBATH2	HWARY2
HSHY2	HFRETS2	HSLEEP2	HRGMILK2	HRGSLEE2	HRGNAPS2	HRGSOL2	HSLFUSS2
HWKFUSS2	HSCREAM2	HHAPCHG3	HUNFAM3	HHAIRBR3	HFEEDNG3	HINJUR3	HOBBATH3
HWARY3	HSHY3	HFRETS3	HSLEEP3	HRGMILK3	HRGSLEE3	HRGNAPS3	HRGSOL3
HSLFUSS3	HWKFUSS3	HSCREAM3	HPCKCRY	HREGPAT	HNEEDST	HTLKIMP	HCUDIMP
HANNOY	HTHINKB	HLEAVEB	HCOMPET	HPATIEN	HGIVEUP	HTIRED	HDEPRES
HWORRY	HRAGE	HSCARED	HUPSET	HKEYDUP	HNERVES	HHRTRAC	HNOSHAR
HPARTLK	HFAMHLP	HSENSIT	HLISTEN	HLONELY	HJOYEXC	HWISHAF	HBRINK
HMAKEUP	HCLOSER	HGOOUT	HHAPREL	HFORCE	HPSTREL	HPSTRNO	HFEED
HNAPPY	HGETUP	HCOOK	HCLEAN	HWASH	HDIY	HCASH	HTENDIL
HGENCAR	HPAYHLP	HHWCHANG	HOTHBAB	HNOOTHB	HBOYGRL	HYEAR	HMONTH
HWHERNW	HALIVE	HEVSEE	HSEEOFT	HEVLIV	HLIVYR	HLIVMTH	HMAINTN
HBOYGRL2	HYEAR2	HMONTH2	HWHERNW2	HALIVE2	HEVSEE2	HSEEOFT2	HEVLIV2
HLIVYR2	HLIVMTH2	HMAINTN2	HBOYGRL3	HYEAR3	HMONTH3	HWHERNW3	HALIVE3
HEVSEE3	HSEEOFT3	HEVLIV3	HLIVYR3	HLIVMTH3	HMAINTN3	HBOYGRL4	HYEAR4
HMONTH4	HWHERNW4	HEVSEE4	HSEEOFT4	HEVLIV4	HLIVYR4	HLIVMTH4	HMAINTN4
HBOYGRL5	HYEAR5	HMONTH5	HWHERNW5	HALIVE5	HEVSEE5	HSEEOFT5	HEVLIV5
HLIVYR5	HLIVMTH5	HMAINTN5	HBOYGRL6	HYEAR6	HMONTH6	HWHERNW6	HEVSEE6
HSEEOFT6	HEVLIV6	HLIVYR6	HLIVMTH6	HMAINTN6	HBOYGRL7	HYEAR7	HMONTH7
HWHERNW7	HEVSEE7	HSEEOFT7	HEVLIV7	HLIVYR7	HLIVMTH7	HMAINTN7	HBOYGRL8
HYEAR8	HMONTH8	HWHERNW8	HEVSEE8	HSEEOFT8	HEVLIV8	HLIVYR8	HLIVMTH8
HMAINTN8	HBOYGRL9	HYEAR9	HMONTH9	HWHERNW9	HEVSEE9	HEVLIV9	HMAINTN9
HNGHFRN	HWORKSUF	HFAMSUF	HMOTHHAP	HMARRYKD	HNEEDDAD	HLONEPAR	HRACESCH
HRELSCH	HRACENGH	HRELNGH	HRACEMAR	HRELMAR	HBOOST	HNOTSEP	HCOMPTRB
HEDBETPA	HSELFSAT	HNOGOOD	HDOWELL	HUSELESS	HFAILURE	HPOSATT	HEFFIC1
HEFFIC2	HEFFIC3	HLIFESAT					

Variables which hold data for the Self-completion (CASI)
(Ordered left to right)

Father (Partner)

HPCKCRY2	HREGPAT2	HNEEDST2	HTLKIMP2	HCUDIMP2	HANNOY2	HTHINKB2	HLEAVEB2
HCOMPET2	HPATIEN2	HGIVEUP2	HTIRED2	HDEPRES2	HWORRY2	HRAGE2	HSCARED2
HUPSET2	HKEYDUP2	HNERVES2	HHRTRAC2	HSENSIT2	HLISTEN2	HLONELY2	HJOYEXC2
HWISHAF2	HBRINK2	HMAKEUP2	HCLOSER2	HGOOUT2	HHAPREL2	HFORCE2	HPSTREL2
HPSTRNO2	HHWCHAN2	HOTHBAB2	HNOOTHB2	HBOYGR11	HYEAR11	HMONTH11	HWHERN11
HALIVE11	HEVSEE11	HSEEOF11	HEVLIV11	HLIVYR11	HLIVMT11	HMAINT11	HBOYGR12
HYEAR12	HMONTH12	HWHERN12	HALIVE12	HEVSEE12	HSEEOF12	HEVLIV12	HLIVYR12
HLIVMT12	HMAINT12	HBOYGR13	HYEAR13	HMONTH13	HWHERN13	HALIVE13	HEVSEE13
HSEEOF13	HEVLIV13	HLIVYR13	HLIVMT13	HMAINT13	HBOYGR14	HYEAR14	HMONTH14
HWHERN14	HALIVE14	HEVSEE14	HSEEOF14	HEVLIV14	HLIVYR14	HLIVMT14	HMAINT14
HBOYGR15	HYEAR15	HMONTH15	HWHERN15	HALIVE15	HEVSEE15	HSEEOF15	HEVLIV15
HLIVYR15	HLIVMT15	HMAINT15	HBOYGR16	HYEAR16	HMONTH16	HWHERN16	HALIVE16
HEVSEE16	HSEEOF16	HEVLIV16	HLIVYR16	HLIVMT16	HMAINT16	HBOYGR17	HYEAR17
HMONTH17	HWHERN17	HEVSEE17	HSEEOF17	HEVLIV17	HLIVYR17	HLIVMT17	HMAINT17
HBOYGR18	HYEAR18	HMONTH18	HWHERN18	HEVSEE18	HSEEOF18	HEVLIV18	HLIVYR18
HLIVMT18	HMAINT18	HBOYGR19	HYEAR19	HMONTH19	HWHERN19	HEVSEE19	HSEEOF19
HEVLIV19	HLIVYR19	HLIVMT19	HMAINT19	HBOYGR20	HYEAR20	HMONTH20	HWHERN20
HEVSEE20	HEVLIV20	HMAINT20	HNGHFRN2	HWORKSU2	HFAMSUF2	HMOTHHA2	HMARRYK2
HNEEDDA2	HLONEPA2	HRACESC2	HRELSCH2	HRACENG2	HRELNGH2	HRACEMA2	HRELMAR2
HBOOST2	HNOTSEP2	HCOMPTR2	HEDBETP2	HSELFSA2	HNOGOOD2	HDOWELL2	HUSELES2
HFAILUR2	HPOSATT2	HEFFIC4	HEFFIC5	HEFFIC6	HLIFESA2		

Further details of the self-completion are to be found in the following, which also accompany the data deposit:

Millennium Cohort Study First Survey: Technical Report on Fieldwork
Millennium Cohort Study First Survey: CAPI Questionnaire Documentation

2.1.11 Proxy interview

As noted above, where the main respondent was unavailable or unable to understand or respond to questions put by the interviewer or to the self-completion, short proxy interviews were undertaken with the main respondent. The variables which hold the data for the proxy interview are identifiable through their labels – all begin with the endorsement "Proxy:". These variables are identified below.

Variables which hold data for the Proxy Interview
(Ordered left to right)

PXNUM	PXSEX	PXAGE	PXDOB	PXPMS	PXPNUM	PROXREA	PROXMAR
PROXEE	PROXEW	PROXES	PROXENI	XPROXE	PROXMUM	PROXDAD	PROXSEEM
PROXSEED	MPROXH01	MPROXH02	MPROXH03	MPROXH04	PROXHLTH	PROXILL	PROXPRB
PROXHGHT	PROXHGFT	PROXHGIN	PROXHGCM	PROXWGHT	PROXWGST	PROXWGPO	PROXWGKG
PROXACT	PROXWRKB	PROXEVJB	PROXLEAV	PROXTYP1	PROXTYP2	PROXTYP3	PROXPAYL
PROXEMPL	PROXMANG	PROXMNGN	PROXJSTY	PROXJSTM	PROXEMPN	PROXJENY	PROXJENM
PROXSEMN	PROXHRS	PROXOVTI	PROXDAYS	PROXWHNA	PROXWHNB	PROXWHNC	PROXWHND
PROXPAYN	PROXPRN	PROXPAYG	PROXPRG	PROXSEPY	PROXLEFT	PQUALAC	PQUALVOC
PROXKIDS	PROXSEEK	PROXOFTS	PROXMAIN	SOC2000X	NSSEC2		

Further details of the proxy interview are to be found in the following, which also accompany the data deposit:

Millennium Cohort Study First Survey: Technical Report on Fieldwork
Millennium Cohort Study First Survey: CAPI Questionnaire Documentation

2.1.12 Identifiers

The unique identifier which appears on the dataset is "FAMSRNO". This identifies families with one or more cohort children for whom data is available for the first survey.

2.1.13 Multiple births

	Number	***%***
Singletons	18297	98.6
Twins	246	1.3
Triplets	10	0.1
	18553	100.00

2.1.14 CAPI Derived Variables

Routing respondents through the interview required the derivation of a number of summary variables within CAPI program. These variables are indicated as such in the questionnaire documentation. These variables may be of value to users, in particular those in the household module relating to individual demographic characteristics and those in Module J on employment status.

2.1.15 Other Derived Variables

A number of derived variables have been developed by the CLS team and others. Those derived variables which are considered to be of general value to users have been included on the dataset. These variables are detailed in the accompanying document '*Millennium Cohort Study First Survey: Derived Variables'*

2.1.16 Weights

As noted above, the sample of births selected for the first survey of the Millennium Cohort Study was clustered, geographically, and disproportionately stratified to over-represent areas with high proportions of ethnic minorities in England, residents of areas of high child poverty and residents of the three smaller countries of the UK respectively. The distribution of the cases in the dataset across strata for each country is given in Table 8 below.

Table 8

	England		Wales		Scotland		N Ireland		UK	
Strata	***n***	***%***	***n***	***%***	***n***	***%***	***n***	***%***	***n***	***%***
Advantaged	4617	39.97	832	30.13	1145	49.04	723	37.69	7317	39.41
Disadvantaged	4522	39.18	1929	69.87	1191	50.96	1200	62.31	8842	47.63
Ethnic	2394	20.85	Na	na	Na	na	Na	na	2394	12.96
	11553	100.00	2761	100.00	2336	100.00	1923	100.00	18553	100.00

The use of this sampling strategy means that weighting is necessary where analysis is:

- Confined to data relating to a single country.

  The weights are:

Strata	**England**	**Wales**	**Scotland**	**N Ireland**
Advantaged	1.32	1.77	1.23	1.41
Disadvantaged	0.71	0.65	0.75	0.76
Ethnic	0.24			

- Covering all countries of the UK

  The weights are:

Strata	**England**	**Wales**	**Scotland**	**N Ireland**
Advantaged	2.00	0.62	0.93	0.47
Disadvantaged	1.09	0.23	0.57	0.25
Please Ethnic	0.37			

Further details are included in the Technical Report on Sampling section 5). Two variables have been included on the dataset to facilitate such weighting by providing the sample weights attached to each case. These are:

- **WEIGHT1 -** This variable should be used when your analysis is within one country only.

- **WEIGHT2 -** This variable should be used when your analysis covers the whole of the UK.

Within SPSS these variables should be used to weight the data. This can be achieved by:

- Selecting Menu item ‘Data’, and then ‘Weight Cases’, or

- Using SPSS syntax: command: WEIGHT {BY varname}

NB: It is not necessary (or correct) to weight analyses that are broken down by strata. Analyses by Stratum (PTTYPE) should not be weighted. Analyses by COUNTRY should be weighted

Further details of the sampling are to be found in the following, which also accompanies the data deposit:

Millennium Cohort Study First Survey: Technical Report on Sampling

Queries should be directed to the CLS (mcs@cls.ioe.ac.uk)

2.2 Assessment of Data Quality

In addition to the checks that were built into the CAPI code, or undertaken immediately after fieldwork, members of the CLS MCS team have carried out further checks. In this task they have been greatly assisted by a number of researchers who are members of the MCS Advisory Groups and/or their associates.

Checking of a large and complex dataset such as MCS is an ongoing process and is checking is continuing.

2.2.1 Household Module

Note. Person 1 is the person who answered the household grid and may not be either of the natural parents of the cohort baby.
Note. There can be up to 11 resident people in the household in addition to the cohort baby (or babies).
Note. Person 12, where present, refers to part time residents. Variables that also refer to part-time residents are ptpsex, ptpdob, ptpage, ptrelb, prelr111, prelr112 thru prelr121, ptpjob.

Natpno1. This will be missing if baby does not live with either natural parent (10 cases).

Month and Year of birth of people in household (other than cohort baby and person1 who was providing the information) have a higher number of missing values than the corresponding age at interview of each person in household.

2.2.2 Module A: Non-resident parents

Some variables labelled without prefix 'A' in this section.
Marstat, Ethnic origin codes of main respondent and baby.

Ethnic origin of main respondent: This is stored in separate variables for each country's sample (EthnicE, EthnicW, EthnicS and EthnicNI for England, Wales, Scotland and Northern Ireland respectively). Each country's ethnic origin variable records the other country samples as missing. The baby's ethnic origin and partners ethnic origins are coded separately in the same way.

England has a higher number of respondents who did not give an ethnic origin code for themselves or the baby (n=26 compared with 3, or 5 missing for other countries).

Derived variables xethnic (main) and xbbeth (baby) which contains back coding for the codes 95 ='any other' have 2 cases coded as –1 (not applicable) which should be coded –3 (missing).

The CAPI Questionnaire Documentation indicates that Section A is to be answered by Main and Partner respondent. The majority of questions in this module are answered by Main respondents only. The variables available to partner respondents are: PNum2 to Xethnic2.

2.2.3 Module B: Father's involvement with baby

Note. Questions in this section most questions were asked of all partner respondents (n=13263). The majority of these were natural fathers, but a few were natural mothers, foster mothers, adoptive, step or foster fathers.

Blookaff, BchgNap, Bfeed, Bgetup, Blon par - Around 39-45 cases have missing responses.

BbrthBb1 This question asked of Natural Parents only (not part-time resident). There is a much lower base for this variable.

2.2.4 Module C: Pregnancy, Labour and Delivery

This section has quite high numbers of Don't Knows about some of the retrospective details of pregnancy.

CPregpln. Variable has high missing values.

CTryUnit. 95 answers = Don't Know.

AskC – This variable is not in the documentation. It is a frequency of all those who have given a (multicoded) response to CHosp (CHosp01 to CHsop10).

CFertrt – multicoded with up to 4 codes (CFertrt1 to Cfertrt4) plus text variable CFTrtO where CFertrt =95. These text codes have been backcoded into CMFertR1 to CMFertR4.

CPregWk. Variable has very high Don't Know answers=454.

CProb – multicoded with up to 7 codes (CProb01 to CProb07) plus text variable CPrbOth where CProb=95. These text codes have been backcoded into CMProb01 to CMProb07.

CHosp –multlicoded into up to 10 codes (CHosp01 to CHosp10).

CdayCHos, CBHosp , CICUNam – These variables have been removed for confidentiality reasons.

CDuedate. A small number of cases were clearly incorrect by one year and have been corrected. A small number of other cases return implausible values. There are also 257 missing cases, for which the gestation period has been set at 280 days in GESTATN.

CBPart – multicoded into up to 4 variables (CBPart01 to CBPart04).

CDelT - multicoded into 3 variables (CDelT1 to CdelT3) plus an 'other'code=95 stored as text which has been backcoded into variables CMDelT01 to CMDelt03.

CPain - multicoded into 6 variables (CPain1 to Cpain6) plus an 'other' code=95 stored as text variable and backcoded into CMPain01 to CMPain06)

Clabtim 112 answer= Don't Know

CWgtpnd In 5 cases, the ounces weight (CWgtoun) is missing (-8) when the pound weight is valid.

CProbs - multicoded into up to 4 variables for Baby1 (CProbs1 to Cprobs4) plus another 'other' code=95 stored as text and backcoded into variables CMProb32 to CMProbs36.
For Baby2 the equivalent variables are Cprobs8 to Cprobs10 and CMProb61 and CMProb63.
For Baby3 the equivalent variables are: Cprobs15 and CMProb90.

CWrong – multicoded into up to 6 variables for Baby1 (Cwrong1 to Cwrong6) plus an 'other' code=95 stored as text and backcoded into variables CMWron01 to CMWron06).
For Baby 2 the equivalent variables are CWrong7 to CWrong11 and CMWron35 to CMWron39.
For Baby 3 the equivalent variables are CWrong13 to CWrong14 and CMWron69 to CMWron70.

Other Responses for multiple births coded in
For twin: CBTim2, CwgtMeas2, CwgtKilo2, CWgtPnd2, CWgtOun2 CProbs8 onwards CMProb61 onwards, CWrong7 onwards, CICU2, CICUwh2, CMWron35 onwards, CDisage2, CDisD2, CDisM2, CDisW2,
For triplet: CBTim3, CWgtMeas3, CwgtKilo3, CWgtPnd3, CWgtOun3, CProbs15 onwards, CMProb91 onwards, Cwrong13 onwards, CMWron69 onwards, CICU3, CICUwh3, CDisage3, CDisD3, CDisW3, CDisM3.

2.2.5 Module D: Baby's health and development

Dwglsa – 177 missing cases

Dwglsb - large number of cases are missing (438)

Dwglsd This variable is meant to be 3 digits but often entered with the leading zeros truncated. These have been adjusted by assuming that if the value < 9, then multiply by 100, else if the value <99 multiply by 10. There are some obvious outliers (a derived variable is under construction to identify these with a flag). However, removing them does not regularise the distribution – there are still far more values than expected below the 1st or above the 99th centile for age. The same problem applies to the twins and triplets (see variables dwglsd2, dwglsd3). As most respondents gave imperial weights (pounds and ounces) and 30% of those giving metric responses gave grammes in 3 digits, less than 20% of the sample are affected by more than 900g.

Note. Birthweight. These are all plausible values but may not be accurate at the gestation time that can be derived from interview information. Both birthweight and

gestation will be able to be checked against hospital records when they are linked to the data.

Note. The best and simplest way to regularise baby weights is to restrict them to those obtained from the red book (Where Redbook is coded '1').

Many questions in this section have separate responses for babies that were multiple births (either twins or triplets.) The naming of these variables does not always follow the same convention.
For example:
(**Note**: variable name Dact8 and Dact9 are not in expected sequence)

Variables referring to Baby1
Dact1 Dact2 Dact3 Dact4 Dact5 Dact6 Dact7 Dact9
The equivalent variables for second baby in twins are:
Dact8 Dact10 Dact11 Dact12 Dact13 Dact14 Dact15 Dact16
The equivalent variables for the third baby in triplets are:
Dact17 Dact18 Dact19 Dact20 Dact21 Dact22 Dact23 Dact24

Variables referring to Baby1 (Between 52 and 61 are missing):
Dgest1 Dgest2 Dgest3 Dgest4.
The equivalent variables for second baby in twins are:
Dgest5 Dgest6 Dgest7 Dgest8
The equivalent variables for the third baby in triplets are:
Dgest9 Dgest10 Dgest11 Dgest12

Larger than usual number of respondents with single birth baby responded 'Don't know for Dact6.

2.2.6 Module E: Childcare

ECare multicoded, has up to 7 types of care recorded plus 'other=code95' (ECare01 to ECare07) for the cohort baby.
ECare=95. These other types of care, written as text, have been backcoded into EMCare variables (EMCare01 to EMCare06).

EmainC – Although routing suggests EmainC only contains those who have more than one type of care, this is not the case. Most of the babies cared for in one way only are integrated into EmainC

EOthWho Multicoded types of care. Up to 6 variables record the different types of care (EOthWh01 to EOthWh06) plus an 'other'=95 code. Text answers are backcoded into variables EmothW01 to EmothW06.

Care for twins and triplets who are not cared for together are recorded in variables : ECare20 ECare21,EMainC2, EOthKid2, EOthAd2, ECarHrs2, ECarAge2, EAgeWk2, EAgeMth2, EcarHrs3, EcarAge3, EAgeWk3, EAgeMth3

2.2.7 Module F: Grandparents and Friends

FDadalv. – A large number of Don't Knows (Main 356. Partner=189)

FParhlp for main respondent multicoded into variables FParhlp1 to FParhlp6. 'Other' code=95 is backcoded into variables FMParh01 to FMParh06

FParhlp for partner respondent multicoded into variables Fparhlp7 to Fparhlp11. 'Other' code=95 is backcoded into variables FMParh11 to FMParh16

2.2.8 Module G: Parent's health

Note. Questions on specific illnesses (GMigrai to GDigest) are **NOT** about the current prevalence of these conditions. Questions asking about specific illnesses are all high since they asked about ***ever*** having been told by the doctor. It cannot be assumed (except in the case of diabetes), that the respondent has the condition at the interview (or when pregnant).

Main

GTypDia - multicoded answers are stored in GtyDia1 to GtypDia3. The code 'other'=95 has not been specified and therefore is not backcoded.

GTypCan is multicoded in variables GTypCa01 to GTypCa02 plus 'other' code=95 and backcoded into variables GMTypc01 to GMTypc02.

GtypDgs is multicoded into variables GTypDgs1 to GTypDgs3. The code 'other'=95 has not been specified and therefore is not backcoded.

Ghigh, Gb4wght have large missing values.

Note. Ghighin. – CAPI allowed height answers of 12 inches (n=5)

Note. Gwtb4po – CAPI allowed weight answers of 14 pounds (n=18)

Partner

GTypDia - multicoded answers are stored in GtyDia5 to GtypDia6. The code 'other'=95 has not been specified and therefore is not backcoded.

GTypCan is multicoded in variables GTypCa11 to GtypCa12 plus 'other' code=95 and backcoded into variables GMTypc13 to GMTypc14.

GTypDgs is multicoded into variables GTypDgs6 to GtypDgs8. The code 'other'=95 has not been specified and therefore is not backcoded.

2.2.9 Module H: Self-completion

Questions in this section have lower response rates from partners than from main respondents.

Note. Some main (n=607) and partner (n=462) respondents refused to complete this Self Completion section (see HSelfC, HselfC2). Others did not complete it themselves but were assisted by the interviewer (main=1277; Partner=663. This explains some of the difference in size of sample who started out filling in the questionnaire (Intro) and higher number who ended the Self Completion (HendQ)

Hrelsch, HrelNgh, HrelMar. Variables are in data set but not in the documentation. Northern Ireland only variables.

The following variables are in a different order in the data set from in the hard copy of the CAPI Questionnaire:
HFamSuf HMothHap HmarryKd HneedDad HlonePar HraceSch HraceNgh HraceMar Hboost HnotSep HcompTrb HEdBetPa

2.2.10 Module J: Employment and Education

Main

JemSe and JWrkPrg have a small number of inconsistent results about whether main respondent has ever had a paid job.

JWrkLsM Large number of months missing (Main- 197 Don't know)

JWkRtn. Some respondents (Main -24) said earlier (EAct) they were still on leave. This is inconsistent with stated routing.

JRtnWy – multicoded up to 5 coded (JRtnWy1 to JRtnWy5) plus 'other reason' code=95 backcoded into a set of variables (JMRtnWy1 to JMRtnWy6).

JNoLkWy– multicoded up to 6 coded (JNoLkWy1 to JNoLkWy6) plus 'other reason' code=95 backcoded into a set of variables (JMNoLk01 to JMNoLk06).

JEmpStM – has large missing (Main-413)

JFlxOp – multicoded up to 8 codes (JFlxOp1 to JFlxOp8).

JFlxU – multicoded up to 6 codes (JFlxU1 to JFlxU6)

JChCOp multicoded up to 6 codes (JChCOp1 to JChCOp6)

JChCU multicoded up to 6 codes (JchCU1 to JchCU6)

JWkNum. High number of Don't Know (Main -122)

JWkGndr High number of Don't Knows (Main- 167)

JWkSym High number of Don't Knows (Main- 185)

JWkWhy – multicoded into up to 6 variables (JWkWhy1 to JwkWhy66) plus 'other' code=95, backcoded into JwkmWhy1 to JwkmWhy6.

JJbChg multicoded into up to 6 variables (JJbChg1 to JJbChg6) plus 'other' code=95, backcoded into JMJbCh01 to JMjbCh06.

JTakea has high missing value/Don't Know, (Main -1821)

JGrossa has high missing values/Don't Know (Main -4391

JSETake. Missing?Don't Know values = 235 from 894.

JBen - multicoded (8) answers (Jben1 to JBen8) plus 'other' benefit backcoded into JMBen01 to JMBen08.

Note. Small number say they did not receive any Child Benefit (Most of the sample were included from Child Benefit register. A small number of cases were included by Health Visitors as in-movers to the sample Ward).

JOth – multicoded into one of 4 codes (JOth1 to JOth4) plus 'other source' code=95, backcoded into set JMOth01 to JMOth04.

JNetInLP– within valid codes for Main are 147- DK, 101- Refuse

JNetInCo Within the valid codes for Main are 945-DK and 381-Refuse.

Note: Different income bands were offered to lone parents and couples in order to approximately equivalise between these different household types, and to differentiate more finely the low incomes of lone parents.

JBank multicoded in two codes JBank1 and JBank2 (Answers are both from main respondent).

JFinChY. Text variable is not included in data but reasons coded into JXFinw01 to JXFinw07

Education variables. 70 main respondent cases are missing throughout this section.

Partner

JWkGndr2, JwkSym2 have high missing/DK (362, 335 respectively)

JTakea2, JGrossa2 have high missing/DK and refuse (720, 1896 respectively.

JSETake2 has high missing/DK Refuse (424 out of 2151).

Occupations codes (main and partner) include.
nssec – Main NS-SEC socio-economic classification (long version)
SOC2000m – this is coded without points between the numbers.
Other derived variables have also been constructed.

2.2.11 Module K: Housing and local area

KmoveMt. The month is missing for a high number of cases.

KPrvTwn, KPrvPC,KprvCty These variables have been removed from the data set for reasons of confidentiality.

KRacePb. This has high value for Don't Know (200), and this tends to be higher in disadvantaged wards.

KTransPb. This has a very high value of Don't Knows (1874). Only 3% of DK have no car.

KspacPb, KpollPb also have high values of DK (187, 250)

2.2.12 Module L: Interests and time with baby

Lrelig contains only Great Britain sample answers. Northern Ireland answers on religion are stored in LreligNI. Codes 10 and 16 were backcoded into the same variables Lrelig and LreligNI. A derived variable has been constructed to combine the responses from GB and NI (see MCS1: Code Book and Edit Instructions and MCS1: Derived Variables)

2.2.13 Proxy Module

Users should make their own assessment of quality of data collected by proxy.

Queries should be directed to the CLS (mcs@cls.ioe.ac.uk)

PART THREE

3.1 Future activities

3.1.1 Ecological Data - Neighbourhood Statistics

In order to fulfil the objective to enhance the survey data with information on the wider social ecology of the family, work has started on creating look up tables containing Neighbourhood Statistics and other data including metadata, which can be matched into the MCS files at a later stage. Thus far, official Neighbourhood Statistics data have been matched into files containing the English, Welsh and Scottish MCS wards. Once data has been added in from Northern Ireland and procedures to protect confidentiality, and to respect the copyright of other possible contextual material have been developed, ecological files will be made available to users at a future date.

3.1.2 Health Visitor Survey

Another data enhancement funded by ONS has been a postal survey of Health Visitors operating in the MCS1 wards with the aim to provide information about family related services in the survey areas. This survey took place from June 2002 onwards and provided information for 233 survey wards. These results will be incorporated in the ecological data.

3.1.3 Flagging

Cohort members will be flagged at the NHS Central Register (NHSCR), for the purposes of future tracing and in order that we will be informed of deaths and cancer registration. This process is being carried out after the survey has been completed, subject to the mother having given written permission at the end of the interview. They were asked to sign a form that permits the flagging of the study child in National Health Service central records. Parents were left with a copy of the form they had signed in the event that they had questions about the precise nature of the permission or wished to withdraw their permission after the interview.

3.1.4 Record Linkage

ONS funding has enabled us to link all cases (subject also to written consent) to hospital records of their delivery.

3.1.5 Postal survey of assisted fertility treatment

ONS funding has also allowed for an additional postal survey to mothers who received fertility treatment for the pregnancy with the MCS baby. Mothers were asked during the interview if they would be willing to answer further questions about their experience of fertility treatment after fieldwork had ended. This additional survey is about to be put into the field to around 430 MCS mothers. The survey is being carried out in conjunction with colleagues at the National Perinatal Epidemiology Unit (Oxford University).

3.1.6 Feedback

As part of the panel maintenance activities vital to retaining the cohort sample, a feedback card and document are to be sent to all the families who participated in MCS1. This will serve the dual purposes of thanking the families for their involvement in MCS1 and relaying some results from the first survey and also of checking the contact information in preparation for sweep 2.

3.1.7 Tracing

In addition to the exercise mentioned above, work has begun at CLS to update the sample information ready for MCS2.

3.1.8 Second Sweep

The second Millennium Cohort survey will be launched into the field in September 2003 when the first-born cohort children will be three years old. The fieldwork for this survey is being carried out by NOP.

Further follow-up into adulthood is planned.

3.1.9 Relations with Users

The CLS hopes that users will find the dataset of the first sweep interesting and enjoyable to use. We welcome feedback from users, and look forward to learning of their published results.

PUBLICATIONS

An entry for the MCS appears in the ONS/ISER website 'Keeping Track' A Guide to Longitudinal Resources
http://www.iser.essex.ac.uk/ulsc/projects/ldr4ss/index.php

Kate Smith & Heather Joshi (Spring 2002), 'The Millennium Cohort Study', *Population Trends* (Spring 2002) No: 107. (The article is available in downloadable format from the following Internet address: http://www.statistics.gov.uk/downloads/theme_population/PT107.pdf)

Heather Joshi, Ian Plewis, Andy Cullis, Mahmood Sadigh, Mos Mojaddad (CLS) and Katie Dodd, Joanne Woods and Aliy Chapman (Analytical Services Directorate, Department for Work and Pensions) 'Mobile families and other challenges in the design of the Millennium Cohort Study'. British Society for Population Studies Annual Conference, Newcastle. (September 2002)

APPENDICES

Appendix 1: MCS1 Management and Advisory Structure

The partner institutions
Centre for Longitudinal Studies (CLS), Institute of Education:
Professor John Bynner

International Centre for Health and Society, University College:
Professor Sir Michael Marmot

Institute of Child Health, University College:
Professor Catherine Peckham

Department of Psychology, City University:
Dr. Dermot Bowler

***The Millennium Cohort Study Management Team (CLS*)**:
Professor Heather Joshi (Project Director)
Kate Smith
Ian Plewis
Peter Shepherd
Professor Shirley Dex
Dr Elsa Ferri (Now retired)
Professor Neville Butler
Mahmood Sadigh
Denise Brown
Gareth Hughes
Denise Hawkes
Andrew Cullis (Now University of St Andrews)
Brian Dodgeon

Members of the Research Coordinating Team
(Including contributing authors to Sweep 2):*
Professor Mel Bartley (ICH)*
Dr Helen Bedford (ICH)*
Dr Leslie Davidson (University of Oxford)*
Professor Carol Dezateux (ICH)*
Professor Harvey Goldstein (IoE)
Dr Yvonne Kelly (ICHS)*
Professor Kath Kiernan (LSE)
Professor Alison Macfarlane (City University)
Dr Barbara Maughan (Inst. of Psychiatry)*
Professor Chris Power (ICH)*
Professor Ingrid Schoon (City University)*

Scientific Consultants
Dr Leon Feinstein (Bedford Group)
Dr Marjorie Smith (TCRU IoE)
Professor Tim Cole (ICH)
Professor Julia Brannen (TCRU IoE)
Charlie Owen (TCRU IoE)

Appendix 2: Psychological Inventories

Contributed by Ingrid Schoon, City University

A. Child Development

***Developmental Milestones**:* A selection of 13 questions taken from the *Denver Developmental Screening Test* were used to assess social and communication skills, as well as fine and gross motor coordination typical for a 9 months old child (see Frankenburg, W.K., Dodds, J.B. Denver (1967). Developmental Screening Test. *J. Paediatrics*, 71, 181-191).

baby 1: dact1 dact2 dact3 dact4 dact5 dact6 dact7 dact9 dgest1 dgest2 dgest3 dgest4 dmove
baby 2: dact8 dact10 dact11 dact12 dact13 dact14 dact15 dact16 dgest5 dgest6 dgest7 dgest8 dmove2

baby 3: dact17 dact18 dact19 dact20 dact21 dact22 dact23 dact24 dgest9 dgest10 dgest11 dgest12 dmove3

***Temperament and Behaviour**:* 14 questions from the *Carey Infant Temperament Scale* were used to assess temperament and behaviour of the child (see: Carey, W.B., McDevitt, S.C. (1977). *Infant Temperament Questionnaire. Dept. Educational Psychology*, Temple University, Philadelphia; Carey, W.B., McDevitt, S.C. (1995). *Revised Infant Temperament Questionnaire*. Scottsdale, AZ: Behavioural-Developmental Initiatives). The questions tap into four areas: regularity (4 items), approach-withdrawal (3 items), adaptability (2 items), mood (5 items)

baby 1: hhapchg hunfam hhairbr hfeedng hinjur hobbath hwary hshy hfrets hsleep hrgmilk hrgslee hrgnaps hrgsol.

baby 2: hhapchg2 hunfam2 hhairbr2 hfeedng2 hinjur2 hobbath2 hwary2 hshy2 hfrets2 hsleep2 hrgmilk2 hrgslee2 hrgnaps2 hrgsol2

baby 3: hhapchg3 hunfam3 hhairbr3 hfeedng3 hinjur3 hobbath3 hwary3 hshy3 hfrets3 hsleep3 hrgmilk3 hrgslee3 hrgnaps3 hrgsol3

Infant Control: Adopted from ALSPAC (Avon Longitudinal Study of Parents and Children): includes questions about sleeping, feeding, and crying (including excessive patterns)

baby 1: dwake dmslprm1 dmslprm2 dmslprm3 dmslprm4 dmslprm5 dmslpbe1 dmslpbe2 dmslpbe3 dmslpbe4 dmslpbe5 dcrybby dcryprob.

baby 2: dwake2 dmslprm6 dmslprm7 dmslprm8 dmslprm9 dmslpr10dmslpbe6 dmslpbe7 dmslpbe8 dmslpbe9 dmslpb10 dcrybby2 dcrypro2.

baby 3: dmove3 dwake3 dmslpr11 dmslpr12 dmslpr13 dmslpr14 dmslpr15 dmslpb11 dmslpb12 dmslpb13 dmslpb14 dmslpb15 dcrybby3 dcrypro3.

B. Psychological Assessment of Parents

***Maternal Attachment**:* A selection of 6 items of the Condon Maternal Attachment Questionnaire have been used to assess mother-to-infant attachment (see: Condon, J.T. & Corkindale, C.J. (1998). J*ournal of Reproductive and Infant Psychology*, 16, 57-76; Condon.J.T.: *Nature and Determinants of Parent-to-Infant Attachment* Dunn.D.J: the Early Postnatal Period. *J. Am. Acad. Child Adolesc. Psychiatry*. 1988.27.3. 293-299).

main (natural mother): hannoy hthinkb hleaveb hcompet hpatien hgiveup.

partner (natural mother).: hannoy2 hthinkb2 hleaveb2 hcompet2 hpatien2 hgiveup2.

Parenting beliefs: A selection of 5 questions originally devised by the ALSPAC team to determine mother's attitudes towards a baby were used to assess whether mothers think that babies should be brought up in a structured or in a laissez-faire environment.

main: hpckcry hregpat hneedst htlkimp hcudimp.

partner: hpckcry2 hregpat2 hneedst2 htlkimp2 hcudimp2.

***Parental Relationship**:*

Grims: A modified version of the Golombok Rust Inventory of Marital State (7 items) has been used to assess overall quality of a couple's relationship (see: Rust, J, Bennun, I., Crowe, M. & Golombok, S. (1990). The Grims: a psychometric instrument for the assessment of marital discord. *Journal of Family Therapy*, 12, 45-57).

main: hsensit hlisten hlonely hjoyexc hwishaf hbrink hmakeup.

partner: hsensit2 hlisten2 hlonely2 hjoyexc2 hwishaf2 hbrink2 hmakeup2.

***Parental Psychosocial Distress (Depression)**:* Assessed using a modified version of the Malaise Inventory (9 selected items only) (see: Rutter, M., Tizard, J. & Whitmore, K. (1970). *Education, Health and Behaviour* . London: Longmans).

main: htired hdepres hworry hrage hscared hupset hkeydup hnerves hhrtrac.

partner: htired2 hdepres2 hworry2 hrage2 hscared2 hupset2 hkeydup2 hnerves2 hhrtrac2.

***Parental Self Esteem**:* A revised version of the Rosenberg Self Esteem Inventory (6 items) was used to measure perceived self-worth (In the selection we refer to the shortened version used by Bachman and Cobb in the longitudinal study of young American men and in ALSPAC (see: Bachman, J.G., O'Malley, P.M., Johnston, J. (1978). *Adolescence to Adulthood: Changes and Stability in the lives of young men*. Ann Arbor, MI: Institute for Social Research, University of Michigan;

Cobb, S, Brooks, G.H., Kasl, S.V., Conelly, W.E. (1966). The health of people changing jobs: a description of a longitudinal study. *American Journal of Public Health*, 56, 1476-1481; Rosenberg, M. (1965). *Society and the Adolescent self-Image*. Middletown, CT: Wesleyan University Press

main: hselfsat hnogood hdowell huseless hfailure hposatt.

partner: hselfsa2 hnogood2 hdowell2 huseles2 hfailur2 hposatt2.

***Life Satisfaction*.** A measure of general *satisfaction with life* was obtained: 'On a scale from 0 to 10 how satisfied are you about the way your life has turned out so far?' (0=extremely unsatisfied to 10=completely satisfied),

main: hlifesat

partner: hlifesa2

Locus of Control: Has been assessed, as in previous sweeps of NCDS and BCS70, by three items:

- I usually have a free choice and control over my life
- I never really seem to get what I want out of live
- Usually I can run my life more or less as I want to.

main heffic1 heffic2 heffic3

partner: heffic4 heffic5 heffic6

Social Support: 3 items relating to emotional, financial and instrumental support were selected from the social support scale developed by ALSPAC team.

hnoshar hpartlk hfamhlp.