STUDY PROFILE
The UK Millennium Cohort Study: the making of a multi-purpose resource for social science and policy in the UK

Heather Joshi University College London Institute of Education, UK
h.joshi@ucl.ac.uk
Emla Fitzsimons University College London Institute of Education, UK

Received February 2016 Revised May 2016) http://dx.doi.org/10.14301/llcs.v7i4.416

Abstract

This paper gives an account of the origins, objectives and structure of the Millennium Cohort Study (MCS) – some 19,000 individuals born in the UK in 2000-2001 – and its use in a wide range of research on many aspects of their lives in childhood years. We highlight some of the mass of output on the first five surveys to age 11 in 2012. Topics discussed are social inequalities in child development; comparisons with other cohorts; areas not well covered by previous national cohorts: season of birth, fathers, ethnicity and childcare; parental behaviour; intergenerational links; social ecology and differences between and within UK countries. We also discuss the challenges faced by the National Evaluation of Sure Start (NESS) in drawing controls from the MCS. As the cohort marches to its seventh survey in 2018, and beyond, the potential for research across life course domains will only continue to grow.

Keywords

Millennium Cohort Survey, life course inequalities, Sure Start, neighbourhood effects, UK regions

Introduction

The UK inaugurated a new national cohort study to greet the new millennium with an investment in a multi-purpose data resource. The study has now run for six sweeps – at ages nine months, three, five, seven, 11 and 14 years – and is preparing for a seventh follow-up at age 17. The scale of the resource and its variety means that a complete story would extend well beyond the scope of a single article. Indeed a profile of the Millennium Cohort (MCS) up to its fifth sweep at age 11 has already been published, focussing on its potential use for epidemiological research (Connelly & Platt, 2014). The aim of this article is to explain to potential users how the objectives, structure and content of the study were shaped by the history of its forerunners in previous British cohorts and by the context of its funding. Its primary goal was to create a rich and easily accessible scientific resource, shaped by expert input. It aimed to provide continuity with the past and to reflect the present. The purpose of this paper is to provide information on its background for users of the resource. This should also be useful for anyone contemplating founding a cohort study in the future. In reviewing how objectives have so far been met, we hope to indicate the scope for further analysis.

Cohort studies have come to be regarded as an important way for research and policy to take account of the life course – the sequence of events and experiences in individual lives through a number or domains, in the context of historical developments at the level of society (Elder, 1985). A framework recognising the family and social

contexts that surround individual development (Bronfenbrenner, 1979) also contributed intellectual underpinning for the multi-dimensional approach taken by MCS. It was always intended to serve multiple purposes, which themselves grew as the sources of funding increased. The study's objectives, listed in box 1, emerged over its first few years. The first six were incorporated in the project as initially funded by the Economic and Social Research Council (ESRC). Co-funding by Government departments enhanced the resources available to achieve these six aims. It also extended the scope of the study to three further objectives, seven to nine in box 1.

Box 1: MCS Objectives (from MCS report to funders, 2001)

1. To chart the initial conditions facing new children in the new century in terms of social, economic and health advantages and disadvantages, building evidence for future research on individual development.
2. To provide a basis for comparing processes of development with the preceding British cohorts.
3. To collect information on previously neglected topics, such as the role of fathers, non-parental childcare and ethnicity.
4. To focus on the experience and aspiration of the children's parents as the immediate 'background', of the child's early years
5. To emphasise intergenerational links including those back to the parents' own childhood.
6. To investigate the wider social ecology of the family: social networks, civic engagement, community facilities and services, splicing in geo-coded data as available.
7. To cover the whole of the United Kingdom, providing big enough samples for analysis within Wales, Scotland and Northern Ireland.
8. To provide evidence for use in the National Evaluations of Sure Start and of the Children's Fund.
9. To enhance the content of the survey by collecting data from sources beyond survey interviews, drawing on supplementary sources of funding if necessary.

Our first section gives an account of the early history of the study around 2000-2001. The second describes the building up of the data resource since then, summarised in table 1. The third section reviews some results that have emerged in relation to the first seven objectives, and also addresses the eighth objective – the study's novel use in impact evaluations.

We leave a detailed treatment of objective nine, enhancements, beyond the scope of this paper. There have been a number of ways in which supplementary information has been drawn into the database, indicated in table 1. Its funding has tended to be 'added-on', mainly from government funders. These include an important and growing set of consented linkages to administrative records, which will continue to be an important part of the study. In due course these linkages of administrative records to the study should yield enough material for a separate account, including the challenges that have been faced in their implementation. Likewise, the collection of biomedical samples and measurements (such as height, weight and physical activity) are not discussed here. Some of them are described by Connelly and Platt (2014) and others are ongoing. Though their story is not told here, both types of enhancement add greatly to the wealth and complexity of this multi-faceted asset.

Box 2 points towards further information about the MCS datasets and research using them. An appendix provides further detail of the innovative use of MCS in the National Evaluation of Sure Start.

Table 1. UK Millennium Cohort Study (MCS): main sources of information

Born 2000/01 [a]

Fieldwork		2001/2	2003/4	2006	2008	2012	2015
Age (Sweep)	**Collection Mode (see note)**	9 months (MCS1)	3 years (MCS2)	5 years (MCS3)	7 years (MCS4)	11 years (MCS5)	14 years (MCS6)
Source	b	Mother Resident partner	Mother Resident partner	Mother Resident partner	Mother Resident partner	Mother Resident partner	Mother Resident partner
	c		Assessments	Assessments	Assessments	Assessments	Assessments
	d		Height/weight	+waist circumference	+waist circumference/body fat	+body fat	+body fat
	e				Cohort member	Cohort member	Cohort member
	f		Biological samples				Biological samples
	g				Accelerometry		Accelerometry
	h						Time use diary
	i		Older siblings	Older siblings			
	j			Teacher	Teacher	Teacher	
	k	Birth registration Hospital records[n]		School records[l]	School records [m]	School records[m]	
Size	o	18,818	15,808	15,460	14,043	13,469	11,938
Response rate [%]	p	96.4	81.0	79.2	72.0	69.0	61.2

Notes to table 1:

a. Eligible if born between 1/9/2000 and 31/8/2001 (England & Wales), or between 23/11/2000 and 11/01/2002 (Scotland and Northern Ireland), and resident in a stratified sample of electoral wards.
b. Home visit face-to-face and self-completion. Main respondent is almost always the mother.
c. Home visit, assessments of cognitive skills directly administered to cohort members.
d. Home visit, height and weight measured by interviewer at each survey + items indicated at MCS3 to MCS6.
e. Home visit, cohort member self-completion.
f. Biological samples collected at home visit: oral fluid (for immunities) at MCS2; saliva of cohort member and co-resident biological parents (for DNA) at MCS6.
g. Record of physical activity collected via accelerometer devices. In-home placement via post after the interview.
h. In-home placement; diary self-completion outside home visit.
i. Home visit, sibling self-completion.
j. Postal survey, outside England only MCS3; UK MCS4; England and Wales only MCS5.
k. Linkage to administrative records. Consent obtained for linkage to health records to age 14, education records to age 16, and parents' DWP and HMRC records. Data linkages are ongoing and further consents are planned.
l. Routine records of the Foundation Stage Profile, state schools in England.
m. England, Wales, and Scotland, state schools only.
n. Hospital episode of delivery.
o. Unweighted achieved sample of children, including 702 in new families added at sweep 2
p. 'Response rates' expressed as percentage of families responding out of the 19,244 ever interviewed – no adjustment for death or emigration of cohort.

Other supplementary data collection: Health visitor survey, births after assisted fertility, nursery observation and shed milk teeth.

Origins

By the 1990s, Britain already had three national birth cohort studies, 1946, 1958 and 1970 (Pearson, 2016). Their value as multi-purpose research resources and as documents of social change, mobility and inequality was increasingly appreciated, but their long-term continuation was not assured. It would have been hard for the research community to give priority to starting a new cohort if that diverted resources from the existing studies. Apart from the Avon Longitudinal Study of Parents and Children (ALSPAC) that had recruited children born around 1991 in the Bristol area of the south west of England, there was little up-to-date evidence on inequalities in early child development – a topic in which the New Labour government, elected in 1997, was particularly interested. There was thus delight and surprise in the research community when, as part of plans to mark the millennium, the government allocated additional funding, outside the regular research budget, for a new cohort. Two crucial factors in that decision shaped the study. One was the condition that at least some (preferably at least half) of the cohort members should be born in the year 2000, which meant that decisions about design and implementation had to be made to a very short timescale for such a large undertaking. The second factor was that the study was to be commissioned by the Economic and Social Research Council (ESRC). This implied that the study would be a multi-disciplinary research resource, in content and access, on the model of other ESRC investments – in contrast to what was then the 'medical model', applying to the Medical Research Council's 1946 cohort, with health-oriented hypotheses investigated primarily by in-house researchers.

A tender was published as late as February 23rd 2000 calling for a Principal Investigator (PI) on the basis of a scoping study commissioned in 1999 (Pearson, 2016). The ESRC called for a study that would enable comparison and continuity with the previous cohorts, but also build on them, producing a rich long-term research resource of use to social science and social policy, documenting children's early years in the first instance. A crucial

requirement was to spread the births over a year rather than, as in the three previous national birth cohorts, over a week. This reflected both scientific considerations - allowing for variations by season of birth – and practical ones. It was no longer feasible to deploy National Health Service (NHS) staff in the data collection. Cost considerations aside, they would not have the expertise to administer computer-assisted instruments (CAI), and there were not enough trained CAI interviewers in the UK to mobilise for a one-week swoop.

Another implication of taking births from a whole year was that they would have to be sampled. The original specification aimed for a cohort of 15,000 children from a population of births expected to be around 700,000. Sampling provided the opportunity for spatial clustering to contain fieldwork costs and create scope for multi-level modelling. Stratification of the sample would permit disproportionate representation of particular groups. The sample design was left to the bidder.

Several academic teams submitted bids over the five weeks to March 31st 2000. Six weeks later the group asked to proceed was the one based in the Centre for Longitudinal Studies (CLS), home of the 1958 and 1970 cohorts at the Institute of Education. The bid was headed by John Bynner, the director of CLS, in an interdisciplinary and inter-institution partnership with the International Centre for Health and Society, University College London (UCL) (Michael Marmot), the Institute of Child Heath, UCL (Catherine Peckham), and the Department of Psychology, City University (Dermot Bowler). Heather Joshi become the scientific director of the study, supported by a number of colleagues at CLS with experience in running and analysing cohort studies, notably John Bynner, Director of Methodology Ian Plewis, and Neville Butler, with his expertise on paediatrics and his experience of the 1958 and 1970 studies. In the short time available the team had drawn on experts from a number of disciplines in health and social sciences to propose a broadly based survey of health and development in its social context. These partnerships continued to develop, as noted below.

The award of the scientific contract for sweep one was finalised August 1st 2000, well into the Millennium year. There was then a several-month process of tendering for the fieldwork contract, alongside the following developments, which proceeded in parallel:

- negotiations with government departments for substantial co-funding
- design of the data collection instruments
- finalisation of sample design
- development of a sampling frame
- setting a fieldwork timetable.

Co- funding: The Office of National Statistics (ONS) consortium

During 2000 various government departments made a financial commitment to the first two sweeps. These permitted the enlargement of the cohort, in terms of size and content, beyond the original budget. These plans were co-ordinated by the ONS, which played an important role in maintaining the scientific integrity of an extended questionnaire, and balancing the interests of the departments, which were in turn balanced against scientific considerations through the study's governance structure.

Funding from the three devolved administrations increased the target sample size from 1,500 families in each country to 3,000 in Wales, 2,500 in Scotland and 2,000 in Northern Ireland. After these additions, it was Wales, rather than the more populous Scotland, which was the 'smaller' country with the largest presence in MCS.[i] In England, the sample was boosted by approximately 2,600 families drawn from extra wards in disadvantaged areas, as potential controls in the evaluation of Sure Start, funded by the Department for Education and Skills (DfES). Altogether the target sample size rose to over 20,000.

The Department of Health funded add-ons to the first survey included the piloting of parent-held primary care records as a source of data (not taken forward), postal surveys of mothers who had used assisted fertility treatment, and of Health Visitors[ii] and linkage of survey births to birth registration and hospital episode statistics which were all successfully undertaken (respectively Hawkes, 2006; Redshaw, Hockley & Davidson, 2007; Brasset-Grundy et al. ,2007; Hockley et al., 2008).

The consortium, which also included the Department of Work and Pensions (DWP), funded an extension of interview and, crucially, the cost of producing basic reports, not covered by ESRC funding for resource creation. In later sweeps other government departments also contributed to the study – as noted in our acknowledgements.

Design of the data collection instruments

As noted above, the main data collection from parents at the first survey was determined at the outset, by ESRC, to be by computer assisted interview and self-completion (CAPI and CASI). Telephone interviews were not considered suitable for initiating a new cohort, and in any case the quantity of information to be collected would have been too much for a single telephone contact. The internet was not, at that time, an option, when only a minority of families were online. To create a new multi-purpose longitudinal dataset, with the objectives and theoretical framework described in the introduction – namely to capture the diversity of backgrounds from which the Children of the New Century were setting out on life – the content of the interviews was developed, at some speed, in discussions with an extended group of collaborators. Their expertise spanned demography, developmental psychology, economics, epidemiology, geography, midwifery, paediatrics, public health, social psychology, sociology, statistics and survey methodology. The 17 external advisers or collaborators involved for the first sweep (from nine institutions) are listed by Dex and Joshi, (2004, p 6-7) along with 23 members of the CLS internal team of academics and professionals. Although longitudinal funding was not yet fully confirmed, the scientific content was designed with a view to a long-term future. After deliberation among the internal and external team, the content of the first survey was debated by 55 potential users of the dataset, from academe and government departments, in a one-day consultative conference organised by the CLS on October 11th 2000. This led into the CAPI development and piloting by the fieldwork contractor the National Centre for Social Research (Natcen). The resulting instrument covered a rich set of information on household demographics, pregnancy and delivery, physical health of child and parents, their mental health, the family's income, education, employment and housing, neighbourhood, parenting, childcare, various aspects of lifestyle, attitudes, and relationships. The full set of questions, typically lasting 70-75 minutes went to a main informant, almost always the mother, and a shorter set of questions, taking around 30 minutes, went to the main informant's partner – almost always the child's father. This combination is reflected in the entry for the first survey in table 1.

Sample design

The population from which every child should have a known, non-zero chance of selection was defined as all those born within eligible dates (specified in table 1 note a), alive and living in the UK at age nine months, and eligible to receive Child Benefit at that age (Plewis, 2007a). The sampling strategy was to make a selection of areas of residence, and within them to recruit 100% of the children born in the eligible period. The statistical geography available for such clusters in 2000-1 was the boundaries of electoral wards as they stood before updating at the 2001 census.

Details of how these wards were sampled, and disproportionately stratified by ethnicity, area disadvantage, and UK country, are reported in the Technical Report on Sampling (Plewis, 2007a). Wards with a relatively high proportion of minority ethnic population (in England), were identified from the 1991 census. The allocation of the rest according to economic disadvantage was based on the most recent Child Poverty Index, recording the proportion of children in a ward whose families received means-tested benefits. The cut-off value to define a disadvantaged ward was 38.4% receiving such benefits, the bottom quartile of wards in England and Wales in 1998[iii]. This gave nine strata overall, 'disadvantaged' and 'advantaged' in each of England, Scotland, Wales and Northern Ireland, plus the minority ethnic wards in England. The target samples by country and the relative size of the disadvantaged stratum within them emerged from the negotiations about co-funding. Before sampling with the aim of reaching a 20,000 target, it was necessary to forecast the number of births in each of the nine strata and estimate what proportion of their parents would respond.

Sampling Frame and Recruitment

DWP's support of the study was crucial in providing access to the Child Benefit register as a sampling frame. Officials operated the initial contact with respondents, offering an opt-out, if a child of the relevant age was identified at an address of interest. The claiming of Child Benefit was then near-universal among the resident population[iv]. It was feared that going through the written opt-in approach required via birth registration would have led to considerable bias[v]. However, the use of the opt-out was questioned by the NHS Multi-Centre Research Ethics Committee (MREC). Upon incorporating a supplementary

verbal opt-in to the interviewer, the survey received ethical approval to proceed in May 2001.

Fieldwork period

As the basic shape of the survey emerged, the fieldwork was put out to tender (August 4th 2000). The fieldwork contractor, NatCen (appointed Sept 28th 2000), then joined the deliberations with advisers and funders, about the timing of fieldwork in the first sweep and embarked on programming and piloting the questionnaire.

By this point it was clear that the aspiration to survey children born in 2000 could only be met by starting cohort birth dates in September 2000 and fixing the interview age at nine months. This put the start of fieldwork back to June 2001, which would be clear of the census and, as it turned out, a general election. Although medically focused studies put priority on collecting information at (if not before) birth, from the point of view of social research, data on conditions during the child's first year are also very valuable. Some information about the birth was collected retrospectively from the mothers, and some from routine hospital statistics. On the 'millennium' timetable there was no question of collecting biological specimens at delivery, ruling out funding from the Medical Research Council. The September to August birthdays of the cohort in England and Wales coincide with a school year. However in Scotland and Northern Ireland, fieldwork was deferred to start with births in the last week of November 2000, to avoid double sampling with a survey on infant feeding. In these two countries, recruitment was extended for an extra six weeks of birth dates in the light of fewer births than expected in the months after the start date (Plewis 2007a, Appendix 1). It was decided not to extend birth dates in England and Wales, although there were also fewer births over the relevant period, which meant that the overall target sample would not have reached 20,000.

Building up the data resource

The prospects for continuation became more secure as time went on. Funding for a second sweep (MCS2) at age three was announced by ESRC in 2001 and awarded to CLS on a single tender although on competitive bidding for fieldwork a new agency was brought in. In 2004, ESCRC confirmed funding through CLS for two more surveys, MCS3 at age five and MCS4 at age seven. In 2010 the ESRC confirmed funding for the survey at age 11 (MCS5), at age 14 (MCS6) in 2015 and indicated support for age 17 (MCS7), which was confirmed in 2015[vi]. Government funders generally offered co-funding once ESRC support was established.

The structure and progress of the study over 15 years are summarised in table 1. It shows the major sources and types of information collected, by various modes, in each of the first six surveys, spanning nine months through 14 years. They come from a variety of informants, starting with both parents, and adding in data collection from the cohort children themselves, teachers, and external sources. Some of the scientific content is discussed in the next section.

Response

By the end of sweep 1 fieldwork, 18,552 families had been interviewed, and the cohort included 18,818[vii] children, allowing for 246 sets of twins and 10 set of triplets. This represents a response rate of 72% of all the families with eligible children living at nine months in the sampled wards (Plewis, 2007a; table 7.4) and 81% of (presumed) eligible cases released by the DWP for issue to fieldwork. Though there were some differences in response at MCS1 according to characteristics known from the Child Benefit sample, they were not judged to be substantial[viii].

The uncertainty about exactly how many children should have been eligible arose from families moving in and out of sampled wards in the normal course of events. This could not be completely anticipated in the tight timetable between the four-weekly scanning of the Child Benefit register for children aged seven months and the issuing of batched assignments to interviewers. The age three survey in 2003-4 (MCS2) provided an opportunity to catch up with families who should have been in MCS1 but had been missed because they had only recently moved to an eligible address. This group of 'new families' was only recruited in England [ix]. There is no sample refreshment by immigrants. All the children in the Millennium Cohort were resident in the UK at nine months, and would cease to be eligible if they leave the UK[x].

The boost to the sample from the 'new families' brought the total of all families ever interviewed to 19,244 (and the number of children ever taking part up to 19,519). As the response from the newly issued addresses was somewhat lower, the overall

response rate, ever interviewed out of ever eligible, went down to 71%. Such response rates have become familiar in cross-sectional surveys, but are nowhere near as high as the recruitment of mothers to the 1958 and 1970 cohorts at maternity hospitals (Pearson, 2016).

Attrition

A key question is how many members are retained over time. The last two rows of table 1 take the story forward from the maximum of 19,244 families ever interviewed. Latest results show that MCS6 covered 11,938 14 year-old children in 11, 779 families (61% of the ever-interviewed). However the other 39% are not necessarily permanent losses to follow-up. The pattern of MCS response is not a one-way drain of drop-outs, particularly between the second and third surveys when 1,444 families returned (Mostafa, 2013). By the fifth survey, 54% of the families who had ever responded had done so on all five occasions, but another 20% had participated intermittently. Attrition bias, as Plewis (2007b) showed at MCS2, is more likely than initial response bias. One solution for analysts is to supplement the survey design weights with attrition weights (Mostafa, 2013).

The extensive efforts and procedures used to keep track of the cohort families are described by Calderwood (2013). There are also considerable efforts to maintain good relations with informants by regular feedback, and by limiting respondent burden. There have been small gifts for children as gestures of appreciation, but no cash incentive for participation.

Content and Coverage

Building on the interviews with mothers and fathers, established at the first survey, the structure and coverage of the surveys evolved as the child grew older. Table 1 shows the sources of information expanded to include direct measurements and questioning of the cohort child (neither of which could have been done by telephone). Box 2 indicates where to find more detail on the surveys' content. The tradition of consultation with the research community in the design of each survey has continued, with consultative conferences before each survey. The involvement of scientific peers in the governance of the study, as a collective resource, also continues.

BOX 2. How to find further information

For more information on the dataset readers are referred to the extensive documentation on the study website (www.cls.ioe.ac.uk/mcs) including technical reports on fieldwork, and all questionnaires. The data can be accessed at the UK Data Archive (https://www.ukdataservice.ac.uk).

Hansen (2014) provides a user guide to the structure and content of each sweep.

MCS Guides to Initial Findings accompany each sweep, and can be found on the study website, for extensive descriptive information on a number of topics.

Two sourcebooks have been published by the Policy Press: *Children of the 21st century: from birth to nine months* and *Children of the 21st century: the first 5 years* (edited respectively by Dex and Joshi, 2005, and Hansen, Joshi, and Dex, 2010). Each has chapters by many of the experts who collaborated in designing the study.

Our review in this profile of research uses and findings is far from comprehensive. We have not revisited the epidemiological material reviewed by Connelly and Platt (2014). Neither have we elaborated on the teacher surveys, mentioned in table 1. We draw mainly on material collected up to age seven, as it takes time for research to emerge in peer-reviewed form. These represent a small and unsystematic selection of the 700 items recorded for MCS in the CLS online bibliography (www.cls.ioe.ac.uk/Bibliography) as of May 2016, half of which are dated since 2012. Kneale et al. (2016) provide a more detailed and systematic review of published research on selected themes collected by the survey up to age seven (child behaviour, diet, BMI, immunisation, screen time, hobbies, and child self-reports).

The CLS website (www.cls.ioe.ac.uk/mcs) also gives news of current developments, training and dissemination events and the latest research findings

HOW HAVE THE FIRST EIGHT OBJECTIVES BEEN MET?

Objective 1 – Social Advantage and Disadvantage

A leading objective of the study was to chart the social inequalities at the baseline from which the cohort set out, and to track these inequalities as the cohort members progressed through childhood. They were doing so at a time of unprecedented political interest in equalising life chances in the early years. New Labour policies to support family incomes and early years' services included setting a target to cut child poverty, Working Family Tax Credits, Sure Start and the National Childcare Strategy. Yet from many angles the achievement gaps between children with more and less advantaged backgrounds remained. In respect of cognitive scores, Brown and Sullivan (2014) report that parental education and family income were the most important predictors across the board at age 11, as they had been at all surveys since age three. The gap between rich and poor children on vocabulary at ages three and five was roughly equivalent to one year's progress at both ages three and five. It is not clear if the income-related gaps opened further, as cognitive abilities have been measured in various ways (Dearden, Sibieta & Sylva, 2011; Sullivan, Ketende & Joshi, 2013; Waldfogel & Washbrook, 2010 for example). Children whose families have been in persistent poverty show the worst outcomes (Dickerson & Popli, 2016; Schoon, Hope, Ross & Duckworth, 2010)

Child and parental health outcomes, also show gradients by socioeconomic risk (see Connelly & Platt, 2014). Among these, the most marked are in the indicators of child mental health derived from Goodman's Strengths and Difficulties Questionnaire (SDQ). A large body of research on these scores up to age seven, systematically reviewed by Kneale et al. (2016), establishes a significant relationship between increased risk of children's emotional and behavioural problems and disadvantageous life circumstances. Various factors, such as cognitive ability and self-regulation, have been shown to moderate socioeconomic risks for SDQ scores, indicating possibilities of resilience, but again, complexity rather than any simple panacea. At age 11 there were still wide gaps by socioeconomic status in SDQ scores (Gutman, Joshi, Parsonage & Schoon, 2015).

Adverse associations of child outcomes with such factors as having a teenage mother, being born after an unintended pregnancy, living with two parents who were cohabiting rather than married, or moving house, have all been statistically explained by differences in socioeconomic background (respectively, Hawkes & Joshi, 2012; De la Rochebrochard & Joshi, 2013; Crawford, Goodman, Greaves & Joyce, 2012; Gambaro & Joshi, 2016).

These findings suggest that the study is meeting its main objective as a research resource documenting the dynamics of social advantage and disadvantage, an objective which is all the better served by the enlargement in the size and content of the data base permitted by government co-funding.

Objective 2 – Comparison with other cohorts

One of the principles behind the design of the MCS was that it should enable comparison with the previous British cohorts on prevalence, and particularly, processes. MCS has been used to update the historical picture on such phenomena as the shortening of mothers' employment breaks (Hansen, Hawkes & Joshi, 2009), the marginalisation of social tenants (Feinstein et al., 2008) and increased child adiposity (Johnson, Li, Kuh & Hardy, 2015) across all four national cohorts. Comparing MCS with the second generation studies of the 1958 and 1970 cohorts, whose offspring were surveyed in 1991 and 2004, showed that the social gradient in child development (vocabulary and total difficulties scores) had not widened since the early 1990s (Blanden & Machin, 2008). This was an important interim piece of evidence for the debate about trends in the process of social mobility, which will not be resolved until the Millennium cohort reaches adulthood. The scope for inter-cohort comparisons can only grow.

What was not envisaged, but has also proved fruitful, is comparison of MCS with other datasets beyond the national cohort studies. ALSPAC, the Longitudinal Study of Young People in England (LSYPE, now Next Steps) and the National Pupil Database were used alongside MCS and BCS70 to chart the social gap in educational attainment in a quasi-cohort (Goodman, Gregg & Washbrook, 2011). Another example finds a falling prevalence of psychological difficulties in seven year olds comparing MCS with the British Child and

Adolescent Mental Health Surveys (Sellers, Maughan, Pickles, Thapar & Collishaw, 2015).

International comparison has been an unanticipated opportunity. Using roughly contemporary cohorts in the USA, Canada and Australia, along with MCS, Bradbury, Corak, Waldfogel and Washbrook (2015) found a wider gap in achievement between 11 year olds with least and most educated parents in the UK than in Canada or Australia, but not as wide as in the US. Among other factors, this reflects the income gaps between the low and highly educated parents – widest in US and narrowest in Australia. MCS data has been used in international comparisons of maternal employment around the time of a birth (Crosby & Hawkes, 2007; Huerta et al., 2011). MCS has been compared with the Fragile Families and Child Wellbeing Study in the US on native/ migrant health differentials (Jackson, Kiernan, & McLanahan, 2012) and residential mobility (Lennon, Clark and Joshi, 2016). Another international study finds maternal education more protective to the risk of children's overweight in MCS and Sweden than in China (Lakshman et al., 2012).

MCS was at the forefront of a new generation of child cohorts in Europe and further afield (Pirus & Leridon, 2010). The newcomers, such as Etude Longitudinal Française dès l'Enfance (ELFE) in France, the National Education Panel (NEPS) in Germany, Growing Up in Ireland, Growing Up in Scotland, and Growing up New Zealand took note of the design, content and practices of MCS. They will in due course provide more material for international comparison, making due allowance for the different context of each survey in time and space. Communication between the studies was facilitated by the European Child Cohort Network (EUCCONET), 2008-2013, and the Society for Longitudinal and Life Course Studies, founded in 2010.

Objective 3 - Previously neglected topics

Season of birth

MCS (along with ALSPAC) advanced our understanding of the differences in educational attainment by season of birth (Crawford, Dearden & Greaves, 2014). It was already well known that children born earlier in the academic year – in September or soon after – tend to perform better academically than those born later – i.e. in the following August or other summer months, but there was little evidence on whether this is driven by differences in age at which tests are taken or in the age of starting school. Crawford et al. (2014) exploit a key feature of the MCS that the children born at the start and end of the academic year are very similar in age at the survey assessments, whereas they are a year apart in sitting the national tests. They conclude that age at test is the most important factor behind the difference between the oldest and youngest children in an academic cohort, which suggests allowing for age at test is important when interpreting children's test scores.

Ethnic group

The earlier birth cohorts had few members from ethnic minorities. By 2000 there were substantial numbers of children being born in the UK to parents with an immigrant heritage. The sample of wards was boosted in areas in England with concentrations of Black or Asian population, but this did not boost the chances of families in minority groups living outside such areas being chosen, for example, very few Chinese families were selected. Neither would the cohort cover immigration after 2000-1, from, for example, Eastern Europe. There was, as anticipated, disproportionate dropout by minority groups once recruited. That said, the data gathered has been an important source of evidence on the extent of differentials in health, health behaviour, and social circumstances of families in the larger ethnic minority groups, and their experience of racism (Bécares, Nazroo & Kelly, 2015). Although ethnic minority children tended to have poorer scores on cognitive assessments, all else equal, at age three and five, by age seven there was no apparent penalty to ethnic background over and above that of any other social disadvantages (Sullivan, Ketende & Joshi, 2013; Taylor, Rees & Davies, 2013). Responses from the children themselves provide insight into how different ethnic groups feel about their lives, organise their friendship groups, and use their free time (Collingwood & Simmonds, 2010).

Fathers

Collecting data directly from resident partners as well as a main respondent was another innovation for a national cohort (although this had been done in ALSPAC). Especially at the start, the main respondent was almost invariably the child's natural mother and any resident partner was almost always the natural father. This extended information on the family's circumstances (e.g. employment, the

child's grandparents) and fathers' health and activities with the child. The 'partner' data was however limited, in that only around four in five of the mothers (in the unweighted sample) had resident partners, and about one tenth of these did not respond. It was also limited by the length of the interview and self-completion time that could be asked of the household or the fieldwork budget. The relatively little analysis of the responses from resident fathers includes: the impact of father involvement with the child on children's emotional adjustment; couples' employment patterns, fathers' mental health, and the child's paternal grandparents (respectively: McMunn, Martin, Kelly & Sacker (2015) and references therein; Kanji (2013); Carson, Redshaw, Gray & Quigley, (2014); and Moulton, Flouri, Joshi & Sullivan (2015). There is also evidence, albeit more limited, about fathers' involvement with the cohort child when they live apart (Kiernan, 2006). An attempt to survey non-resident parents was piloted at MCS3, unsuccessfully.

Childcare

There was very little about non-maternal childcare in the 1958 and 1970 studies: they had no surveys in the pre-school years and mothers' employment was then far less common. By 2000, childcare had become a government priority. There was support for maintaining employment around maternity leave, and for various ways of meeting the costs of formal care. MCS1 asked about the childcare arrangements at nine months. The age three survey had a very detailed module about the sequence and type of arrangements, including any overlaps, since the first survey. A subset of group settings identified in MCS2 was followed up to assess the quality of care they provided (Mathers, Sylva & Joshi, 2007). Unfortunately the childcare module in the MCS2 questionnaire was over-complicated and under-piloted. The use of the term 'childcare' seems to have deterred some parents from reporting attendance at the growing number of nursery- or pre-schools. The questions at age five aimed to repair the history of attending early education and day care settings, which almost all the cohort experienced, given new policy to provide free, part-day nursery education. MCS has been used to relate the arrangements at nine months to child outcomes (Coté, Doyle, Petitclerc & Timmins, 2013; Hansen & Hawkes, 2006). There is evidence of early education raising the academic achievement of seven year olds, but only among children from poor families (George, Stokes & Wilkinson, 2012). It would have been hard for an observational study of a near-universal policy change to produce a reliable estimate of the policy's impact. The data does however provide evidence of children's experience of early years provision at a historic juncture.

Objective 4 – Focus on Parents

Given repeated contacts with both parents throughout the early years, MCS provides much more information about parents' attitudes, activities and practices in bringing up their children than the previous national cohorts. All of these factors, along with behaviours like alcohol consumption, smoking and breast-feeding, are thought to contribute to child health and development, and to contribute to the social gradient in child outcomes. Although some studies are interested in parenting per se (e.g. Haux & Platt, 2015), many link parental beliefs and behaviour to child development, also factoring in differences in the socioeconomic circumstances of the family (a few examples: Connelly & Platt, 2014; Ermisch, 2008; Goodman, Gregg & Washbrook, 2011; Hartas, 2011; Kiernan & Mensah, 2011). These all find that aspects of parenting (especially the home learning environment and a warm rather than a harsh parenting style) have some positive association with children's cognitive or behavioural outcomes. The power of parenting practices to enhance development can however be exaggerated if their correlates, such as material resources and mental health, are not also considered. There has been, and remains, scope for research on the parent's employment and/or worklessness in relation to child outcomes (eg, Hope, Pearce, Whitehead & Law, 2014; Parsons, Schoon & Vignoles, 2014). The rich dataset enables social scientists to explore the complex relationships involved.

Objective 5 – Intergenerational links within the cohort

There was originally an idea to have a three-generation stratum within the MCS, recruiting millennial births to members of the 1970 cohort, and perhaps the 1958 cohort. Despite the attractions of linking into existing rich data, this met with the objection that a cohort of a cohort's offspring is not representative of others born at the same time to older and younger parents, or of the

cohort's children born in other years. The concern about generalisablity has also applied to the second generation sub-studies of the 1958 and 1970 cohorts – those who happen to have been born before the age 33 survey of NCDS in 1991 or the age 34 survey of BCS70 in 2004, respectively. Following them up has also fallen off the priority list.

Nevertheless there is material on intergenerational experience that can be taken from retrospective material on the respondents' own childhood, and about their own parents' partnerships and occupation (Hawkes & Joshi, 2012; Moulton et al., 2015), Grandmothers also feature prominently in the research on informal childcare. Two studies found similar results, linking grandparental care with higher risks of children being overweight or obese (Pearce et al., 2010; Tanskanen, 2013).

Objective 6 – Wider social ecology

The idea of contexts beyond the immediate family affecting individual development (Bronfenbrenner, 1979) is partly served by survey questions touching on social capital and perceptions of the neighbourhood. It is also addressed by linking in geo-coded data on statistically, rather than subjectively, defined local area. The sampling frame on which the survey is based became out of date almost immediately. The irregularly sized ward was replaced for statistical purposes with more uniformly sized zones (Lower Super Output Areas in England and Wales). Indices of Multiple Deprivation became available from 2004, in country-specific versions. On the whole, indices of deprivation did not show much additional predictive power in multivariate analyses of child outcomes once the material circumstance of the family were taken into account (see for example Tzavidis, Savati, Schmidt, Flouri & Midouhas, 2016, and a number of other papers by Flouri cited therein). In one study, the 'neighbourhood effects' apparent in the cognitive scores of seven year olds were largely explained as school effects, once the school attended was included in the model (Heilman, Kelly, Stafford & Watt, 2013). Besides indicators of deprivation and other measures of social composition, information has been linked on more 'ecological' variables such as water quality, rurality and urban green space (respectively, Molitor, Best, Jackson & Richardson, 2009; Taylor et al 2013; Flouri, Midouhas & Joshi, 2014).

Although clearly contributing to knowledge and enhancing the data resource, many of these geographically linked variables are available for further research use only under secure conditions. This to prevent disclosure of localities, which could indirectly identify individuals. In any social survey, it is important to protect the anonymity of informants. In a longitudinal study, needing to maintain the continued confidence of participants, it is paramount. Hence the exploitation of its geographical potential faces this additional challenge.

The original structure of the survey, tightly clustered in electoral wards, where all births were supposed to have been recruited, rapidly lost its value for analysis. Not only did the boundaries change, but the families moved. Over half had moved home by the time of the age five survey[xi]. Although most moves were relatively short distance, the cohort became geographically dispersed. This was probably not the only reason that by the time they went to school, or even pre-school, the children were not tightly clustered in classrooms with other cohort members. The original sampled areas did not map neatly into school catchment areas, which themselves overlap. A few primary school teachers in the teacher surveys had as many as ten cohort children in their class, but the majority of children were the only member of the cohort in their class. This is not the ideal sample design for investigating classroom or peer-group effects. It has however proved useful for the study of residential mobility (Lennon et al., 2016).

Objective 7 – Analysis within and across the smaller UK countries

MCS was the first national cohort study to cover all four countries of the United Kingdom[xii]. Sample sizes in the smaller countries were boosted to yield sufficient cases for within-country analysis – of particular interest given the devolution of much domestic policy to their governments.

Analyses of MCS regularly allow for differences across Scotland, Wales and Northern Ireland (if only to address survey design). The devolved governments have commissioned several research reports from CLS describing results within countries. The four for Scotland include a report on the drivers of unhealthy weight in children (Connelly, 2011). The Welsh Assembly Government also commissioned four reports from CLS and used MCS

data in its own publications (Welsh Assembly Government, 2011). Two reports for Northern Ireland drew strength from comparisons with other countries for the relatively small cohort observed in that country (Sullivan, Joshi, Ketende & Obolenskaya, 2010).

Among academic research on 'home international comparisons', and regional differences, is a study of education-related outcomes across different jurisdictions by Taylor et al., (2013). Among regions within England, London showed unexpectedly good literacy in MCS children at primary school age, adding to other evidence of a 'London effect' in educational achievement at secondary schools. Differences in policies were expected to offer examples of 'natural experiments' that the MCS might 'exploit' to detect policy impact. The legislation banning smoking in enclosed public spaces and workplaces, introduced in Scotland in 2006, ahead of England, offered one such opportunity. Hawkins, Cole and Law (2011) found little sign that it reduced the overall level of smoking among the cohort's parents in Scotland, but some evidence that it reduced the social gradient in the practice. MCS was also used to explore health inequalities across Scotland and regions of England (Cruise & O'Reilly, 2015). While there was no evidence of regional differences at birth, these authors found some evidence that geographic health inequalities may develop cumulatively during the life course.

Due to the nature of the testing system in the UK, so far it is only from state schools in England, where Standard Attainment Tests (SATs) are carried out at the ends of Key Stages 1 and 2, that such scores are available to supplement or substitute for the cognitive tests collected in the homes (e.g. Jones, Gutman & Platt, 2013). Test score data from the other countries will be linked as and when national tests are undertaken.

Objective 8 – The use of MCS in policy evaluation

National Evaluation of Sure Start

Sure Start was an area-based intervention targeted at children under four in the most deprived areas in England, and run by local cross-sector initiatives (Sure Start Local Partnerships, SSLPS). Its roll out began with 60 SSLPs in 1999, initially planned to reach 250[xiii]. MCS was used in the impact evaluation. To evaluate effectiveness of SSLPs, a comparison (control) group was required, in other words a group of children not exposed to the programme, but living in other deprived areas. It was decided to draw controls from the MCS. This was a novel application for a general purpose cohort study, and for the evaluation, the use of such a resource was also novel. Further details are given in the Appendix.

Its use in the evaluation did not affect the MCS adversely, and indeed brought with it some benefits. One was the boost of the initial sample in disadvantaged areas in England, increasing the sample size by 35 disadvantaged wards, expected to include 2,600 families. Certain aspects of the early surveys were also strengthened through collaboration with the Sure Start evaluation team, such as measures of cognitive, behavioral and emotional adjustment, and the home learning environment. On the other hand, the joint efforts at designing the MCS2 childcare questions were not, as noted above, a triumph.

However, from the viewpoint of the impact evaluation, using a cohort study as a control group is challenging – for reasons expanded in the Appendix. We maintain that the evaluation of Sure Start was adversely affected by this design. There are many practical considerations in synchronizing two different surveys. A key tension is that the measurements embedded in a longitudinal study are guided by constructs that are important to measure through life, rather than short-term interests. Central to well-designed evaluations of early childhood development (ECD) interventions, on the other hand, is the robust estimation, in the short and long run, of causal impacts of the programme on specific outcomes it targeted, and understanding the behavioural changes contributing to these impacts. This made it challenging to align the two surveys.

Cohort studies can, however, be instrumental in the estimation of causal impacts – when combined with serendipitous events/policies affecting some cohort members and not others (for instance Fitzsimons & Vera-Hernandez, 2014; Kelly, 2011). This 'natural experiment' approach overcomes many of the issues discussed in the Appendix.

National Evaluation of the Children's Fund

Another national evaluation exercise in which the MCS played a part was the impact evaluation of the National Evaluation of the Children's Fund (NECF), (Edwards, Barnes, Plewis, Morris, et al., 2006). The programmes offered by the trans-

departmental Children's Fund, announced in 2000, for children aged five-13, were also localised and just as diffuse as those offered to children aged zero-four by Sure Start. Their emphasis was on 'prevention' – of poor behavioural and academic outcomes – and provisions typically included breakfast clubs and other out-of-school activities. The evaluation had built-in fluidity, as it was intended to help improve the services as it rolled out rather than wait until there was a clear outcome to be evaluated.

The role of MCS in the evaluation was, in the first instance, to identify take-up of Children's Fund (CF) services in selected areas, rather than to provide a set of controls. The older siblings of MCS children in England, at sweeps 2 and 3, formed a 'convenience sample' of the children within the client age group, who came ready provided with a mass of data on family background. In due course it was expected to estimate the impact of service use by comparing older sibs in CF and non-CF areas as well as those who did and did not report using CF services, on the questionnaire augmented in CF areas. But this never happened. The evaluation was terminated in 2006, with two years still to run, when the organisation of the programme and services was restructured. As often happens, the aim to evaluate an intervention was overtaken by events in the policy arena. We shall never know what the data collection might have contributed to evaluation had the programme continued.

The dowry from NECF became legacy for MCS in a considerable enhancement of the data resource: the extra 692 'new families' in England recruited at MCS2, interviewer assessments of neighbourhoods, extra data on older siblings, including, across the UK, the parent-rated SDQ for up to two older siblings at MCS2 and MCS3, and consent to link older siblings to education records in England. Without the ongoing resource from the evaluation budget to check and develop these resources, relatively little use has been made of them, but their potential remains.

Conclusion

Following the evidence already built up from preceding cohorts, it was anticipated in the CLS proposal to ESRC in 2000 that

> 'The children born at the start of the new century will be setting out with different degrees of handicap and advantage in the race ahead. They will be drawn from across the social and geographical spectrum, from diverse ethnic backgrounds. Some will be well placed to build upon advantages; others will have the cards stacked against them at birth. The follow-up sweeps will reveal who overcomes inauspicious beginnings, and who fails to capitalize on inherited advantage'.

Sixteen years on, we can claim, of course not totally impartially, that the MCS is answering these and many other questions. It has delivered a rich multi-purpose resource for cross-disciplinary research of the sort that was hoped for. On the basis of impressive teamwork, MCS is now a major data resource.

Initial recruitment through a universal register on an opt-out basis was key to getting the study off the ground. Regular contact and engagement with cohort members outside of the study sweeps has been critical. It has helped maintain high response rates and remains a key priority of the study. Innovation in modes of data collection, with potential to enhance the collection and quality of data, will become a major consideration in an increasingly digital age, along with linkage to administrative records.

Research outputs took some time to gain momentum, but a complex set of findings has been built up, and continues to grow. The prospects for the asset to accumulate in scientific and policy value are excellent. Users are exploiting the rich data from many perspectives using a wide range of methods. The scope for further analysis, including international comparisons, increases enormously as the evidence accumulates.

We note some disappointments. The perhaps underestimated dispersal of the sample from selected wards made it less useful than hoped to study effects within communities, schools or classrooms. The near universal roll-out of early education, also unanticipated, made it difficult to detect the impact of this policy on child outcomes, a problem which also faced the (near) parallel cohort constructed for the National Evaluation of Sure Start. The uses of the study as a resource for evaluating policy impacts did not meet expectations, but do provide lessons for the future. Any attempt to combine evaluation with a general observational study should beware of the pitfalls that beset the MCS links with NESS and NECF. The evaluation of Sure Start got going too late to make

the best use of MCS. Designing an impact evaluation around the use of a cohort study as a control group brought with it many challenges and may have missed the opportunity for a clean and robust impact evaluation. On the plus side, the additional data collection due to the national evaluations still enhances the long-term value of the research resource, as care was taken that they added to the scientific content rather than distorting it.

As we write, the cohort is moving through adolescence, and age 14 data will be made available at the end of 2016. The next sweep is scheduled for 2018, when the cohort members will be 17 years of age, entering an important new phase, on the threshold of adulthood, taking divergent paths that will greatly influence their future wellbeing. The seventh sweep will be the first opportunity to observe such diverging decisions and the factors influencing these. On the one hand there will be much interest in these, and other, outcomes as a function of earlier life experiences; on the other hand age 17 will provide an important baseline to collect constructs to be followed through adulthood. The contribution of the study to understanding the evolution of the life course, and life, in 21st Century Britain, will only grow as the cohort marches on.

Acknowledgements

We would like to thank our colleagues Morag Henderson and Dylan Kneale for comments. We are also grateful for comments received in the reviewing process. We thank the users of the MCS data for their work in analysing the resource and suggesting material for inclusion in the surveys, especially those on scientific advisory groups. We apologise that it has not been possible to refer to all the research or include all their suggestions. The ESRC has provided the main funding for the study throughout, and is gratefully acknowledged for this. The co-funding from government departments has also been very important. The original consortium supporting the study since MCS1, was led by the Office of National Statistics, with the Departments of Education and Skills, of Health, and of Work and Pensions, the Northern Ireland Executive, the Scottish Government and the Welsh Assembly Government (and their departmental successors). The Ministry of Transport, the Home Office, and Ministry of Justice joined the government funders at later stages up to and including MCS6. Other funding has been contributed by the Wellcome Trust, the Sir Siegmund Warburg Voluntary Settlement, the United States Department of Health and Human Services, National Institutes of Health, National Institute on Alcohol Abuse and Alcoholism and the International Centre for Child Studies. We are grateful to the cohort maintenance, communications, data management and survey management teams at CLS, and the staff of three fieldwork agencies. Most of all we appreciate the participation in the study by the cohort families.

References

Belsky, J., Melhuish, E., Barnes, J., Leyland, A., Romaniuk, H. & the NESS Research Team. (2006). Effects of Surestart Local Programmes on Children and Families: Early Findings. *British Medical Journal*, *332*, 1476-1478. http://dx.doi.org/10.1136/bmj.38853.451748.2F

Belsky, J., Leyland, A.H., Barnes, J., & Melhuish, E. (2009). Sure Start in England. *Lancet, 373*- 381. http://dx.doi.org/10.1016/s0140-6736(09)60132-x

Bécares, L., Nazroo, J. & Kelly, Y. (2015). A longitudinal examination of maternal, family, and area-level experiences of racism on children's socioemotional development: Patterns and possible explanations. *Social Science & Medicine*, *14,* 128–135. http://dx.doi.org/10.1016/j.socscimed.2015.08.025

Blanden, J. & Machin, S. (2008). Up and down the generational income ladder in Britain: past changes and future prospects. *National Institute Economic Review*, *205(1*), 101-116. http://dx.doi.org/10.1177/0027950108096594

Bradbury, B., Corak, M., Waldfogel, J. & Washbrook, E. (2015) *Too Many Children Left Behind. The U.S. Achievement Gap in Comparative Perspective.* New York: Russell Sage Foundation.

Brassett-Grundy, A., Joshi, H., Butler, N. & Bryar, R. (2006). Health Visitors' Awareness of Local Service Accessibility. *Community Practitioner*, 79(3), 84-88.

Bronfenbrenner, U. (1979). *The Ecology of Human Development: Experiments by Nature and Design*. Cambridge, MA: Harvard University Press.

Brown, M. & Sullivan, A. (2014). Cognitive Development. In Platt, L .(ed), *Millennium Cohort Study: Initial findings from the Age 11 survey.* London: Centre for Longitudinal Studies.

Calderwood, L. (2013). The role of respondent characteristics in tracking on longitudinal surveys: evidence from the UK Millennium Cohort Study. *Longitudinal and Life Course Studies, 4(2*), 105-118.

Carson, C., Redshaw, M., Gray, R. & Quigley, M.A. (2014). Psychological distress in the first year of fatherhood: the influence of preterm birth on father's later wellbeing. *Journal of Epidemiology and Community Health*, 68(Suppl.1), A2-A22. http://dx.doi.org/10.1136/jech-2014-204726.42

Collingwood A. & Simmonds N. (2010). Child Self-Report. In: Hansen K., Jones E., Joshi H., Budge D., (Eds.) *Millennium Cohort Study Fourth Survey: A User's Guide to Initial Findings.* Centre for Longitudinal Studies**,** Institute of Education.

Connelly, R. (2011). *Drivers of Unhealthy Weight in Childhood: Analysis of the Millennium Cohort Study.* Scottish Government Social Research Report. Edinburgh: Scottish Government.

Connelly, R. & Platt, L. (2014). Cohort Profile: UK Millennium Cohort Study (MCS). *International Journal of Epidemiology*, *43(6),* 1719-1725. http://dx.doi.org/10.1093/ije/dyu001

Coté, S.M., Doyle, O., Petitclerc, A. & Timmins, L. (2013). Child Care in Infancy and Cognitive Performance until Middle Childhood in the Millennium Cohort Study. *Child Development*, *84(4)*, 1191–1208. http://dx.doi.org/10.1111/cdev.12049

Crawford, C., Goodman, A., Greaves, E. & Joyce, R. (2012). Cohabitation, marriage and child outcomes: an empirical analysis of the relationship between marital status and child outcomes in the UK using the Millennium Cohort Study. *Child and Family Law Quarterly*, *24(2*), 176-198.

Crawford, C., L. Dearden & Greaves E. (2014). The drivers of month-of-birth differences in children's cognitive and non-cognitive skills, *Journal of the Royal Statistical Society: Series A*, *177*, 829-860. http://dx.doi.org/10.1111/rssa.12071

Crosby, D.A. & Hawkes, D.D. (2007). Cross-National Research Using Contemporary Birth Cohort Studies: A Look at Early Maternal Employment in the United Kingdom and United States. *International Journal of Social Research Methodology, Special Issue - Child Cohort Studies*, *10(5),* 379-404. http://dx.doi.org/10.1080/13645570701677151

Cruise, S.M. & O'Reilly, D. (2015). Are the differences in adulthood ill-health across the north-south divide and between Scotland and England also evident in early childhood health indicators? *Social Science & Medicine*, *130(April*), 277–283. http://dx.doi.org/10.1016/j.socscimed.2015.02.021

De La Rochebrochard, E. & Joshi, H. (2013). Children Born After Unplanned Pregnancies and Cognitive Development at 3 Years: Social Differentials in the United Kingdom Millennium Cohort. *American Journal of Epidemiology*, *178(6*), 910-920. http://dx.doi.org/10.1093/aje/kwt063

Dearden, L., Sibieta, L. & Sylva, K. (2011). The socio-economic gradient in early child outcomes: evidence from the Millennium Cohort Study. *Longitudinal and Life Course Studies*, 2(1), 19-40.

Dex, S. & Joshi, H. (2004). *Millennium Cohort Study First Survey: A Users' Guide to Initial Findings.* London: Centre for Longitudinal Studies.

Dex, S. & Joshi. H. (Eds), (2005). *Children of the 21st century: From birth to nine months.* Bristol: Policy Press

Dickerson, A. & Popli, G. (2016). Persistent poverty and children's cognitive development: Evidence from the UK Millennium Cohort Study. *Journal of the Royal Statistical Society: Series A*, 179(2), 535-558. http://dx.doi.org/10.1111/rssa.12128

Edwards, A., Barnes, M., Plewis, I., Morris, K. et al. (2006). *Working to prevent the social exclusion of children and young people: Final lessons from the National Evaluation of the Children's Fund.* Birmingham: University of Birmingham.

Eisenstadt, N. (2012). *Providing a Sure Start: How government discovered early childhood*. Bristol: Policy Press

Elder, G.H. (Ed.), (1985) *Life Course dynamics: Trajectories and transitions, 1968-1980*. Ithaca. N.Y.: Cornell University Press

Ermisch, J. (2008). Origins of social immobility and inequality: parenting and early childhood development. *National Institute Economic Review 205(1*), 62-71. http://dx.doi.org/10.1177/0027950108096589

Feinstein, L., Lupton, R., Hammond, C., Mujtaba, T., Salter, E., Sorhaindo, A., Tunstall, R., Richards, M., Kuh, D. & Johnson, J. (2008). *The Public Value of Social Housing: A longitudinal analysis of the relationship between housing and life chances.* Smith Institute Report. London: The Smith Institute.

Fitzsimons, E. & Vera-Hernandez, M. (2014). *Food for Thought? Breastfeeding and Child Development.* DoQSS Working Paper 14-04. London: Institute of Education.

Flouri, E., Midouhas, E. & Joshi, H. (2014). The Role of Urban Neighbourhood Green Space in Children's Emotional and Behavioural Resilience. *Journal of Environmental Psychology*, 40, 179–186. http://dx.doi.org/10.1016/j.jenvp.2014.06.007

Gambaro, L. & Joshi, H. (2016). Moving home in the early years: What happens to children in the UK? *Longitudinal and Life Course Studies,*7(3*)* 265-287. http://dx.doi.org/10.14301/llcs.v7i3.375

George, A., Stokes, L. & Wilkinson, D. (2012). Does Early Education Influence Key Stage 1 Attainment? Evidence for England from the Millennium Cohort Study. *National Institute Economic Review*, 222(1), R67-R80. http://dx.doi.org/10.1177/002795011222200106

Goodman, A., Gregg, P. & Washbrook, E. (2011). Children's educational attainment and the aspirations, attitudes and behaviours of parents and children through childhood in the UK. *Longitudinal and Life Course Studies*, *2(1),* 1-18.

Gutman, L.M, Joshi, H., Parsonage, M. & Schoon, I. (2015). *Children of the New Century: Mental health findings from the Millennium Cohort Study.* Centre for Mental Health Report. London: Centre for Mental Health.

Hansen, K. (ed) (2014). *MCS Guide to the datasets for the first, second third, fourth and fifth surveys.* 8th edition. London: Centre for Longitudinal Studies.

Hansen, K., Joshi, H. & Dex, S. (Eds), (2010). *Children of the 21st century: the first 5 years.* Bristol: Policy Press.

Hansen, K. & Hawkes, D. (2009) Early childcare and child development. *Journal of Social Policy*. *38*, 211-239. http://dx.doi.org/10.1017/S004727940800281X

Hansen, K., Hawkes, D. & Joshi, H. (2009). The Timing of Motherhood, Mothers' Employment and Child Outcomes. In Stillwell, J., Coast, E. & Kneale, D. (Eds), *Fertility, Living Arrangements, Care and Mobility.* Dordrecht: Springer. http://dx.doi.org/10.1007/978-1-4020-9682-2_4

Hartas, D. (2011). Families' social backgrounds matter: socio-economic factors, home learning and young children's language, literacy and social outcomes. *British Educational Research Journal*, *37(6)*, 893-914. http://dx.doi.org/10.1080/01411926.2010.506945

Haux, T. & Platt, L. (2015). *Parenting and contact before and after separation.* Nuffield Foundation Report, June 2015. London: Nuffield Foundation

Hawkes, D. (2006). *Women's Experiences of Successful Infertility Treatment.* MCS Technical Report on Response. London: Centre for Longitudinal Studies

Hawkes, D. & Joshi, H. (2012). Age at motherhood and child development: evidence from the Millennium Cohort Study. *National Institute Economic Review*, October, R52-R63. http://dx.doi.org/10.1177/002795011222200105

Heilmann, A., Kelly, Y., Stafford, M. & Watt, R. (2013). The Contribution of Neighbourhoods and Schools to Cognitive Test Performance at Age Seven - Findings from the UK Millennium Cohort Study. *Journal of Epidemiology and Community Health*, *67(1),* A28-A28. http://dx.doi.org/10.1136/jech-2013-203126.57

Hawkins, S.S, Cole, T.J & Law, C. (2011). Examining smoking behaviours among parents from the UK Millennium Cohort Study after the smoke-free legislation in Scotland. *Tobacco Control*, *20(2*), 112-118. http://dx.doi.org/10.1136/tc.2010.037028

Hockley, C., Quigley, M. A., Hughes, G., Calderwood, L., Joshi, H. & Davidson, L.L. (2008). Linking Millennium Cohort data to birth registration and hospital episode records. *Paediatric and Perinatal Epidemiology*, 22(1), 99-109.

Hope, S., Pearce, A., Whitehead, M. & Law, C. (2014). Family employment and child socioemotional behaviour: longitudinal findings from the UK Millennium Cohort Study. *Journal of Epidemiology & Community Health*, *68(10),* 950-957. http://dx.doi.org/10.1136/jech-2013-203673

Huerta, M., Adema, W., Baxter, J., Corak, M., Deding, M., Gray, M.C., Han, W-J. & Waldfogel, J. (2011). *Early Maternal Employment and Child Development in Five OECD Countries.* OECD Social, Employment and Migration Working Papers, No. 118. Paris: OECD. http://dx.doi.org/10.1787/5kg5dlmtxhvh-en

Jackson, M.I., Kiernan, K. & McLanahan, S. (2012). Immigrant–Native Differences in Child Health: Does Maternal Education Narrow or Widen the Gap? *Child Development*, *83(5*), 1501-1509. http://dx.doi.org/10.1111/j.1467-8624.2012.01811.x

Johnson, W., Li, L., Kuh, D. & Hardy, R. (2015). How Has the Age-Related Process of Overweight or Obesity Development Changed over Time? Co-ordinated Analyses of Individual Participant Data from Five United Kingdom Birth Cohorts. *PLoS Medicine*, *12(5):* e1001828. http://dx.doi.org/10.1371/journal.pmed.1001828

Jones, E., Gutman, L. Platt, L. (2013). *Family stressors and children's outcomes*. Department for Education Research Report DFE RR254, London.

Kanji, S. (2013). Do fathers work fewer paid hours when their female partner is the main or an equal earner? *Work, Employment & Society*, 27(2), 326-342. http://dx.doi.org/10.1177/0950017012460321

Kelly, E. (2011). The Scourge of Asian Flu: In utero Exposure to Pandemic Influenza and the Development of a Cohort of British Children. *Journal of Human Resources*, *46(4),* 669-694. http://dx.doi.org/10.3368/jhr.46.4.669

Kiernan, K. (2006). Non-residential fatherhood and child involvement: evidence from the Millennium Cohort Study. *Journal of Social Policy*, *35(4)*, 651-669. http://dx.doi.org/10.1017/S0047279406000122

Kiernan, K.E. & Mensah, F.K. (2011). Poverty, Family Resources and Children's Early Educational Attainment: The Mediating Role of Parenting. *British Educational Research Journal*, 37(2), 317-336. http://dx.doi.org/10.1080/01411921003596911

Kneale, D., Patalay, P., Khatwa, M., Stansfield, C., Fitzsimmons, E., Thomas, J. (2016). Piloting and producing a map of Millennium Cohort Study Data usage: Where are data underutilised and where is granularity lost? EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London. http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=3502

Lakshman, R., Zhang, J., Zhang, J., Koch, F. S., Marcus, C. Ludvigsson J., Ong, K. K., Sobko, T. (2013). Higher maternal education is associated with favourable growth of young children in different countries. *Journal of epidemiology and community health*, 67(7), 595-602. http://dx.doi.org/10.1136/jech-2012-202021

Lennon, M.C., Clark, W.A.V. & Joshi, H. (2016). Residential mobility and well-being: Exploring children's living situations and their implications for housing policy. Symposium in *Longitudinal and Life Course Studies.* 7(3) 197-200. http://dx.doi.org/10.14301/llcs.v7i3.393

Mathers, S., Sylva, K. & Joshi, H. (2007) *Quality of childcare settings in the Millennium Cohort Study.* Report for Surestart. Nottingham: HMSO.

McMunn, A., Martin, P., Kelly, Y. and Sacker, A. (2015). Father's involvement: Correlates and Consequences for Child Socioemotional Behaviour in the United Kingdom. *Journal of Family Issues.* http://dx.doi.org/10.1177/0192513X15622415

Melhuish, E., Belsky, J., Leyland, A., Barnes, J. & the Ness Research Team. (2008). Effects of fully-established Sure Start Local Programmes on 3-year-old children and their families living in England: a quasi-experimental observational study. *Lancet*, *372(9650*), 1641-1647. http://dx.doi.org/10.1016/S0140-6736(08)61687-6

Molitor, N.T., Best, N., Jackson, C. and S. Richardson. (2009). Using Bayesian graphical models to model biases in observational studies and to combine multiple sources of data: application to low birth

weight and water disinfection by-products. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(3), 615-637. http://dx.doi.org/10.1111/j.1467-985X.2008.00582.x

Mostafa, T. (2013). *Technical Report on Response in the Millennium Cohort Study Wave 5.* London: Centre for Longitudinal Studies.

Moulton, V., Flouri, E., Joshi, H. & Sullivan, A. (2015). The influence of grandparents' social class on children's aspirations. *British Journal of Sociology of Education*, advance online access.

NESS (The National Evaluation of Sure Start Team). (2012). *The impact of Sure Start Local Programmes on seven year olds and their families.* Research Report DFE-RR220. London: Department for Education. http://dx.doi.org/10.1080/01425692.2015.1093407

Parsons, S., Schoon, I. & Vignoles, A. (2014). Parental worklessness and children's early school achievement and progress. *Longitudinal and Life Course Studies*, *5(1*), 19-41. http://dx.doi.org/10.14301/llcs.v5i1.230

Pearce, A., Li, L., Abbas, J., Ferguson, B., Graham, H., Law, C. & the Millennium Cohort Study Child Health Group. (2010). Is childcare associated with the risk of overweight and obesity in the early years? Findings from the UK Millennium Cohort Study. *International Journal of Obesity,* 34, 1160-1168. http://dx.doi.org/10.1038/ijo.2010.15

Pearson, H. (2016). *The Life Project*. London: Penguin.

Pirus, C. & Leridon, H. (2010). Large Child Cohort Studies across the World. *Population*, 65(4), 671-730. http://dx.doi.org/10.3917/popu.1004.0671

Plewis, I. (Ed). (2007a). *Millennium Cohort Study First Survey: Technical Report on Sampling (4th Edition).* London: Centre for Longitudinal Studies, Institute of Education.

Plewis, I. (2007b). Non-Response in a Birth Cohort Study: The Case of the Millennium Cohort Study. *International Journal of Social Research Methodology,* 10(5), 325-334. http://dx.doi.org/10.1080/13645570701676955

Redshaw, M., Hockley, C. & Davidson, L.L. (2007). A qualitative survey of the experience of treatment for infertility among women who successfully became pregnant. *Human Reproduction*, 22(1), 295-304. http://dx.doi.org/10.1093/humrep/del344

Sellers, R., Maughan, B., Pickles, A., Thapar, A. & Collishaw, S. (2015). Trends in parent- and teacher-rated emotional, conduct and ADHD problems and their impact in prepubertal children in Great Britain: 1999-2008. *Journal of Child Psychology and Psychiatry*, 56(1), 49-57. http://dx.doi.org/10.1111/jcpp.12273

Schoon, I., Hope, S., Ross, A. & Duckworth, K. (2010). Family hardship and children's development: the early years. *Longitudinal and Life Course Studies*, *1(3*), 209-222. http://dx.doi.org/10.14301/llcs.v1i3.109

Sullivan, A., Joshi, H., Ketende, S. & Obolenskaya, P. (2010). *The consequences at age 7 of early childhood disadvantage in Northern Ireland and Great Britain.* A report to the Northern Ireland Office of the First Minister and Deputy First Minister, December 2010. London: Institute of Education.

Sullivan, A., Ketende, S. & Joshi, H. (2013). Social Class and Inequalities in Early Cognitive Scores. *Sociology*, 47(6), 1187-1206. http://dx.doi.org/10.1177/0038038512461861

Tanskanen, A.O. (2013). The Association between Grandmaternal Investment and Early Years Overweight in the UK. *Evolutionary Psychology*, *11(2*), 417-425. http://dx.doi.org/10.1177/147470491301100212

Taylor, C., Rees, G. & Davies, R. (2013). Devolution and geographies of education: the use of the Millennium Cohort Study for 'home international' comparisons across the UK. *Comparative Education*, *49(3*), 290-316. http://dx.doi.org/10.1080/03050068.2013.802927

Tzavidis, N, Salvati, N, Schmidt, T, Flouri, E and Midouhas, E. (2016). Longitudinal analysis of the Strengths and Difficulties Questionnaire scores of the Millennium Cohort Study children in England using Quantile and M-quantile multilevel models *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 179(2), 427–452. http://dx.doi.org/10.1111/rssa.12126

Waldfogel, J. & Washbrook, E. (2010). *Low Income and Early Cognitive Development in the UK.* London: Sutton Trust.

Welsh Assembly Government (2011). *2011 Children and Young People's Wellbeing Monitor for Wales*. Report from the Welsh Assembly, March 2011. Cardiff: Welsh Assembly Government.

Appendix

The role of MCS in the national evaluation of Sure Start

Sure Start was a set of area-based programmes of services for children under four in the most deprived local areas in England, rolled out between 1999 and 2003. It started in 60 Sure Start localities in 1999, initially intended to rise to 250 in 2003 (Eisenstadt, 2012). Areas serving about 600 children, were conceptualised in terms of 'pram-pushing' distance from their centre. They were chosen for intervention on the basis of proposals from local cross-sector partnerships, (SSLPs). The impact evaluation, National Evaluation of Sure Start (NESS), was led by Edward Melhuish (see Belsky et al. 2006; Melhuish, Belsky, Leyland, Barnes & The NESS Research Team, 2008; Belsky, Leyland, Barnes & Melhuish, 2009; and the reports contained in http://www.ness.bbk.ac.uk/). The MCS provided a control group for the impact evaluation.

The evaluation finished in 2012, with a report on the impact on seven year olds. This concluded that the SSLPs had made some beneficial impact on four out of fifteen target outcomes, mostly on mothers' parenting and well being (NESS, 2012). However they also report considerable methodological difficulties, on which we comment below.

MCS was incorporated in the evaluation strategy because a randomised controlled trial (RCT) was ruled out, mainly for political reasons. The timing of the MCS, in following a cohort of children born in 2000/01, as Sure Start was getting under way, appeared to match well with the requirements of the evaluation sample. Furthermore, the Department for Education and Skills was funding the evaluation of Sure Start, alongside contributing funds to MCS. It was a condition on their contract that the evaluators work together with the MCS team. It was also a condition for MCS to balance the needs of NESS against those of other stakeholders.

In relying on a longitudinal study rather than a RCT to provide a control group (to measure what might have happened in the absence of the intervention), a number of methodological complexities arose. Despite the best efforts on the part of the evaluation team to overcome them, these challenges were often insurmountable and ultimately detrimental to the impact evaluation. As a result, the estimates of the programmes' impacts came with a series of caveats, undermining their interpretation as causal effects and, thereby, their utility for policymaking.

Choice of control groups

The critical bedrock of an impact evaluation is the choice of a suitable control group. The rapid expansion of the SSLPs limited the availability of possible control areas. In 2002, before any results were available, the target number of SSLPs for 2004 was doubled. One challenge was that Sure Start areas were more disadvantaged than potential control areas. The NESS team selected areas in England in which children from the MCS lived but which did not have Sure Start, using well-executed propensity score matching. However, matching did not eliminate demographic differences between the NESS sample and those selected from MCS. This exacerbates concerns about unmeasured confounding factors, and means that it cannot be ruled out that the 'impact' of the programme reflects the influence of other differences between the areas (Melhuish et al., 2008). Showing comparability across treatment and control areas in pre-programme trends in outcomes would have been reassuring, though has not, to our knowledge, been done. Furthermore, it was not possible to find suitable control areas for the 57 most deprived of the Sure Start areas (of a total of 150 available for evaluation from rounds one to four of Sure Start), so over one third of them had to be excluded from the main evaluation. The final samples consisted of 5,883 three-year-old children in 93 SSLP areas (out of the 150) and 1,879 children in 72 non-SSLP areas.

Data collection

For two separate studies with distinct objectives to track each other is extremely challenging, even in the short-term. Coordination is required on the timing and content of surveys, consistent measurement of constructs across them, preservation of confidentiality, tracing of respondents and attrition. Moreover, coordination is difficult to sustain and likely to fade out in medium to long term, rendering the estimation of long run effects impossible. The NESS and MCS research teams coordinated closely from the outset, though some problems, some unforeseen, could not be overcome.

In both studies, families were visited when the child was nine months old, three, five and seven years of

age. However, given the challenges of implementing the innovation on an unprecedented scale, it was a necessary for a SSLP to have 'bedded down' for at least three years, before surveying its areas. This meant that the NESS sample was born two years after the MCS, and the timing of data collections was not aligned. NESS surveys started in 2003. This two-year gap had major repercussions. It can never be ruled out that any NESS-MCS differences are due to time effects (e.g. a national policy may apply to just one of them), making it difficult to defend the estimated impacts as *causal* - the central aim of the evaluation. Furthermore, NESS was constrained to follow the timing of MCS rounds in terms of the children's ages whereas other ages might have been more suited to its own purposes. The surveys had common elements, but they were not the same. For example MCS interviewed partners while NESS did not, and NESS asked more questions about social support in the neighbourhood. The fact that interviewers conducting the two surveys came from different agencies and had different training also raised concerns about the comparability of data they collected, for example on cognitive assessments. Using linked administrative data helped to some extent, but this raises another possible source of bias in differential consent to data linkage across two studies.

The long-term impacts of an early years programme are of central policy importance. However, coordination on data collection across two study teams becomes increasingly difficult in the medium and long term. Whilst administrative education records were used up to age seven in NESS, we know of no plans to follow these into the future.

Other considerations

A well-designed impact evaluation should not only measure the impact of the programme on targeted outcomes, but also understand behavioural changes that contribute to measured impacts. For instance, in the case of Sure Start, how did the intervention change mothers' behaviour with children in their homes? Why were the poorest families initially least likely to benefit? Data collection and measurement must be designed accordingly, which can be challenging on two different surveys. A forced marriage has its drawbacks.

Sure Start in England took on a new form in 2004 when it was substantially increased in scale, extended to the under five age group and changed the emphasis of content. Its organisation transferred to Children's Centres run by Local Authorities, more integrated with existing services. It was a popular initiative that Labour, and subsequently the Coalition Government, sought to bring to as many places as possible. The target of 3,500 Children's Centres was achieved in 2010. Their survival as the legacy of the Sure Start continues, subject to local decisions, and less fiscal protection.

Endnotes

[i] The Welsh boost was drawn from disadvantaged wards.

[ii] Health visitors are community nurses specialising in support to young families. The charity founded by Neville Butler, the International Centre for Child Studies, also contributed to the Health Visitor survey.

[iii] The indices for Scotland and Northern Ireland were not available in time to draw a line based on a UK distribution. The proportions of wards in these countries covered by this cut-off were higher, especially in Northern Ireland.

[iv] It may be worth noting that the few families not claiming Child Benefit for children born in the UK at the time could well have been classified as 'non-resident' if they were diplomats, members of foreign armed forces or newly arrived asylum seekers.

[v] There were some restrictions put on the use of the Child Benefit register, in particular a minority (under 3%) of families were excluded if deemed 'sensitive cases' by the DWP, or if they had been approached for another survey (Plewis, 2007a). In the nature of things little is known about the various circumstances of the 'sensitive cases'. Their exclusion means that children who had already been taken into social care for example would not have been recruited, but they were not numerous. As far as can be told the social bias towards more advantaged families opting out offset the less advantaged cases being withheld as 'sensitive cases' (Plewis, 2007b).

[vi] Lucinda Platt took over from Heather Joshi as PI in 2011, during the early stages of preparations for MCS5, Emla Fitzsimons succeeded her in 2013 as MCS6 was being developed to go into the field in 2015.
[vii] This total differs from the 18,827 quoted by Connelly and Platt (2014) that appears to include twins and triplets in the 'new families' not recruited till sweep 2. A total of 18,819 children is also sometimes quoted. This reflects the N that was thought to have been collected at MCS1 before elimination of one invalid case.
[viii] Plewis (2007b) estimated non-response weights for MCS1 that made little change to the sampling weights.
[ix] The funding came from the English Department for Education and Skills.
[x] However, at age 11, there was some contact with families known to have emigrated.
[xi] 58% unweighted and including non-respondents at sweep 3.
[xii] The 1946 and 1958 cohorts had been confined to Great Britain, and the small number recruited in Northern Ireland in 1970 were excluded from follow-up because of the political 'troubles'.
[xiii] The roll out actually accelerated after the evaluation started, to reach 550 areas by 2004 when the programme's scope was further extended and transformed (see Appendix).