Institute of Education

New birth cohort study: theoretical sampling design options

CLS working paper number 2020/4

By Alice Sullivan (Director 1970 British Cohort Study),
Heather Joshi (Founding Director Millenium Cohort Study)
and James Williams (Founding Director Growing Up in Ireland)

With acknowledgements to:

Alissa Goodman, John Bynner, George Ploubidis, Richard Silverwood, Aase Villadsen, David Church, Emla Fitzsimons, Stella Tsoli

CENTRE FOR LONGITUDINAL STUDIES

Economic and Social Research Council

Contact the authors

Alice Sullivan (UCL Centre for Longitudinal Studies) alice.sullivan@ucl.ac.uk

Heather Joshi (UCL Centre for Longitudinal Studies) h.joshi@ucl.ac.uk

James Williams (Economic and Social Research Institute) james.williams@esri.ie

This working paper was first published in April 2020 by the UCL Centre for Longitudinal Studies.

UCL Institute of Education

University College London

20 Bedford Way

London WC1H 0AL

www.cls.ucl.ac.uk

The UCL Centre for Longitudinal Studies (CLS) is an Economic and Social Research Council (ESRC) Resource Centre based at the UCL Institution of Education (IOE), University College London. It manages four internationally-renowned cohort studies: the 1958 National Child Development Study, the 1970 British Cohort Study, Next Steps, and the Millennium Cohort Study. For more information, visit www.cls.ucl.ac.uk.

This document is available in alternative formats. Please contact the Centre for Longitudinal Studies.

Tel: +44 (0)20 7612 6875

Email: clsfeedback@ucl.ac.uk

Disclaimer

This working paper has not been subject to peer review.

CLS working papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

Any opinions expressed here are those of the author(s) and not those of the UCL Centre for Longitudinal Studies (CLS), the UCL Institute of Education, University College London, or the Economic and Social Research Council.

How to cite this paper

Sullivan, A., Joshi, H. and Williams, J. (2020) *New birth cohort study: theoretical sampling design options,* CLS Working Paper 2020/4. London: UCL Centre for Longitudinal Studies.

Summary

- A new nationally representative birth cohort study would provide longitudinal evidence on the development and life-chances of a new generation of children in the UK. To maximise its potential to inform science and policies of the future, its content should be broadly based and record details on as many of the main areas of the child's life as possible including, in particular: physical and emotional health and well-being; relationships; and learning.

- There are some advantages to an 'accelerated design' with two cohorts, one to be born in the early 2020s ('the infant cohort'), and one already born in the 2010s ('the child cohort'). In particular this would fill the gap on experiences in a changing context of children born since the Millennium. However with a limited budget, a dual cohort comes at the cost of smaller sample sizes in each cohort, and these considerations have to be carefully traded off. If a dual cohort is commissioned, we recommend recruiting a child cohort in one of the final years of primary school. Age 11 maximizes the period of history recovered and also has points of comparison to previous cohort studies. Surveys from 11 onwards could be at intervals of 3 to 4 years, and under 11 they should occur more frequently.

- Securing a sampling frame from which an opt-out can be operated for the infant cohort will be central to the success of the project. Securing the political will to achieve this at a high level and enthusiasm for the project in the statistical authorities will be essential. Discussions with the ONS and NHS regarding administrative and legal feasibility must be initiated at a high level. Demonstrating public interest also relates to developing public engagement, without which initial recruitment and continuing response in the surveys may not be sustained. Alternatives to an opt-out consent would need to be carefully piloted and participant consent would need to be made very easy and appealing to give.

- While large sample sizes are desirable, these need to be traded off against budgetary constraint. We have considered the case for clustering by geographical ward but have recommended a simple national sampling design would be more straightforward to analyse, and would lead to a more nationally representative, less selective sample. Efficiency savings from geographically clustered designs are likely to be limited in practice.

- The Millennium Cohort Study was geographically clustered by ward, and wards chosen were disproportionately stratified, which added complexity for users. Potential benefits of the clustering included fieldwork efficiency, however in practice these savings would have been limited and short-lived. The over-representation of ethnic minorities added to their sample size, but in a selective way. Since 2001 the proportion of national births to non-white ethnic minorities has risen substantially, meaning that numbers in a random sample would naturally be increased, and could be achieved using data on ethnicity in the national sampling frame.

Contents

Tables and figures

Chapter 1. Introduction: The need for new cohort evidence

This report sets out the rationale and sampling design options for a new UK birth cohort study, incorporating an accelerated longitudinal design, i.e. a dual cohort comprising a birth cohort of infants (potentially aged 9months) and a parallel older child cohort (potentially aged 9-11) years old), each starting in the early 2020s, in synchronized sequence. We consider the background and scientific case for a new dual cohort (Chapter 1), design considerations (Chapter 2), sampling frame (Chapter 3) and whether or not the sample should be clustered (Chapter 4). We conclude with a set of recommendations (Chapter 5). An annex provides details of lessons to be drawn from the sample design of the Millennium Cohort Study.

1.1 Scientific case for a new study of cohorts: social change

While the main purpose of this report is to set out design options for a new cohort study, such design needs to be guided by scientific and policy goals. We are clearly in a time of rapid social, economic and attitudinal change. Society is deeply politically divided and we are now witnessing some of the far-reaching consequences of society's longstanding inequalities (see the newly launched Deaton review on inequalities), and growing generational divides (see the final report from the Resolution Foundation's intergenerational commission). We face great uncertainty as a nation, and the past is not a reliable guide to our future path. Now more than ever we need good data and evidence how these social problems and political and economic changes will affect current and new generations. What does the future hold for Brexit babies?

Social scientific findings are contingent on time, place, and population (when, where, and who). There is now an 18 year gap in the UK's national birth cohort series, and data on previous cohorts do not enable us to understand contemporary and future trends and processes, and how these will affect a new generation of children as they grow up.

Related to this point, an important feature of a new cohort is for it to be nationally representative, so that findings apply to society as a whole, and so that it contains vulnerable groups who tend to be substantially underrepresented in studies that allow a high degree of self-selection (e.g. *UK Biobank*), and groups who are missing from routine administrative data.

Population trends and social change

Social change since 2000 has been rapid, and a new cohort will provide insights into these changes and their effects. For example:

- Population growth since 2000 has been driven by net immigration and high birth rates among immigrant groups, accompanied by increased life expectancy. While current uncertainties mean that future trends are hard to predict, the make-up of the UK in terms of cultural and national origins has changed dramatically. The proportion of births to non-UK born mothers rose from 12% in 1970 to 16% in 2000 to 28% in 2017 (ONS).
- There has been an increase in the proportion of births out of wedlock, and in single-parent and step-families. For example, the proportion of births out of wedlock rose from 8.3% in 1970 to 40% in 2000 to 48% in 2017 (ONS). Mothers are increasingly

likely to be active in the labour market – by 2017, nearly three quarters of mothers with dependent children in England were in paid employment (ONS).
- Living standards have stopped rising across generations and milestones such as leaving the parental home, partnership/marriage, home ownership, childbearing, and independent living are being reached later for more recent generations, and not necessarily in an orderly sequence.

Some of the social issues and problems faced by current youth are quite different from those faced by previous generations. A new cohort would enable analysis of the causes and consequences of current and future trends affecting youth.

- Trends in mental health have worsened, especially amongst adolescents, and especially amongst girls.
- Rates of overweight and obesity continue to grow, driven by sedentary behaviour and poor diet. Meanwhile, other risky health behaviours, such as smoking and alcohol consumption, have declined.
- Social media use has led to an increase in connectivity, but there are concerns about bullying and problematic use.
- There have been reported declines in 'real life' social and political participation, and an increasing problem of loneliness and decreased social trust.
- There are increasing concerns about a polarisation of political attitudes, and about 'post-truth' beliefs driven by social media bubbles, leading to damaging social trends, e.g. a decline in vaccination rates.
- Globalisation, automation and digitisation bring new opportunities but also risks and employment is increasingly precarious.
- Climate change poses severe policy challenges, and various forms of environmental pollution (including air, noise and light pollution) are taking a disproportionate toll on the young and the poor.

1.2 Themes

To maximise its potential to inform multiple policies and scientific agenda, a new birth cohort study should be broad-ranging, rather than narrowly focussed on a restricted set of specific hypotheses, in order to provide a comprehensive resource for the scientific and policy community, and to allow researchers to examine the inter-relationships between domains. Breadth of content also allows currently unanticipated research questions to be addressed in the future. The scope of data collection would be determined by the scientific leads (whoever was awarded this role), in consultation with the scientific community, but it is likely that themes for data collection would include:

- **Health and wellbeing**: The health risks facing a generation born since 2012 include the obesogenic environment, the mental health crisis, air pollution, climate change, and falling vaccination rates. They will be more likely than previous generations to be born to parents who are overweight or depressed. There is a pressing need for data which will enable researchers to examine the factors that promote healthy and unhealthy behaviours and outcomes, and the consequences of health issues for later outcomes. The rate of improvement in UK life expectancy has stalled since 2011, and it is argued that we need effective preventative approaches that address the social, behavioural and wider determinants of health. Thus, combining biomedical and social measures has the potential to produce vital insights for policy.

- **Relationships:** Family structures and relationships are the foundation for child well-being and development. Wider relationships with adults and peers also influence learning and well-being as children develop. Social capital and networks influence life choices and opportunities. Online activity increasingly influences a range of social interactions. We need data that will inform us about the differences in patterns of relationships and social behaviour for this new generation and the consequences for outcomes including health and well-being.
- **Learning**: Researchers and policymakers need to understand cognitive and educational development in childhood and beyond in order to tackle skills deficits and promote learning. This includes understanding the role of the home, early care settings, and school environments, and the importance of learning that is fostered by cultural and creative activities. We need to be able to compare learning trajectories for a new generation to previous generations, and to understand which novel factors are promoting or impeding learning in particular domains.

Across these themes, new cohort studies will contribute to our understanding of:

- **Inequalities and intergenerational transmissions**: differentials between groups according to factors such as socio-economic position, sex and ethnic group. Relationships between parental circumstances, assets, and characteristics and outcomes for offspring.
- **Cross-generational differences**: Addressing differences in experiences and outcomes for a new generation compared to older generations and examining differences in risk factors and inter-relationships.
- **Life course processes**: the consequences of earlier factors for later outcomes and the development of trajectories and identification of sensitive periods for intervention.
- **Policy development and evaluation**: With increasing devolution of government policy to a country level, a new cohort will be a vital tool in assessing policy at the country level, potentially to an even greater extent than was the case for MCS (Taylor, Rees and Davies, 2013).

1.3 Background to this report

Following the cancelation of the *Life Study* in 2015, in 2017, ESRC conducted a Longitudinal Studies Strategic Review, which highlighted the value of the UK's investment in longitudinal and birth cohort studies and recommended that the UK's investment in longitudinal surveys continue, including through the commissioning of a new birth cohort. To inform this review, a symposium was convened by CLS at the British Academy to examine 'The value and future of birth cohort studies for social science and policy', chaired by David Willetts and Jane Falkingham, and a report from this meeting was produced. The participants were from government departments, funding bodies, the third sector, and a range of academic disciplines and perspectives. A consensus was reached that while lessons needed to be learned from the *Life Study*, a new national birth cohort is feasible and should be commissioned as a matter of urgency.

With the termination *of Life Study*, continuation of the British birth cohort study series at the original 12 year intervals was lost. There was strong support at the British Academy meeting for repairing this gap using a two-cohort design such as that implemented by *Growing up in Ireland*, which started with a cohort of infants at nine months old and children at nine years old. Administrative (and retrospective) data going back to birth could be used to make up some of the data gap for the child cohort.

Understanding Society

One consideration discussed at the British Academy was the option to follow new babies born in Understanding Society, instead of commissioning a new birth cohort, and we are aware that ESRC is also commissioning work concurrent to this report, to investigate this option.

As discussed in the Longitudinal Studies strategic review, birth cohort studies are complementary to, and distinctive in design to household panel studies. One aspect that sets the cohorts apart from the household panel is their larger numbers of contemporary children all passing through a particular age, as well as their deep focus on childhood development.

Our view is that following births in Understanding Society would be valuable but does not substitute for a purposefully sampled new birth cohort. In particular, the number of births per year to UKHLS study members is relatively small, and to date these have been declining significantly each year in both numbers, and population representation as drop-out from the study naturally accumulates. If, as we believe would be optimal, a new birth cohort study is commissioned alongside an exercise to follow UKHLS births, we recommend very close coordination between these efforts so that design features and measures can be fully aligned.

Eurocohort

Funders will be aware of a proposal for a European child well-being cohort (Eurocohort). In the likely event that there is insufficient funding for two sets of accelerated cohorts in the UK, a pair of UK cohorts might serve as the UK arm of a EuroCohort, possibly with enhancements not attempted in the basic Euro design. The advantage of cross-national comparability would need to be balanced against a relative lack of UK control over content, and this trade-off would need careful consideration. It is likely that a survey for EuroCohort in the UK would not be designed to have disproportionate (boosted) samples in Scotland, Wales and Northern Ireland, but there may be opportunities, resources permitting, for a UK study to do so, providing a set of population representative national surveys for Scotland, Wales and Northern Ireland.

MCS and Next Steps Babies

It would be possible to include 'children of' MCS and Next Steps, and potentially 'grandchildren' of BCS70 and NCDS, as a stratum within the new cohorts. A separate report by Fitzsimons and Bornstein sets out projected numbers of births in each year into MCS and Next Steps, which gives an indication of the potential samples to be gained through this route. This report sets out the scientific value and design considerations for continuous data collection on babies born to MCS and Next Steps.

Chapter 2. Options for the design of a national dual cohort: overview

The rationale for having an accelerated cohort design, with a 'child' cohort starting at late primary school age is to 'accelerate' the availability of information on development in mid childhood and adolescence. There is an additional rationale that a child cohort started in the early 2020s would, at least partially fill the gap in the series of British cohorts since the last one born in 2000-1. The cost to this approach is, for a given budget constraint, a smaller sample size in each of the dual cohorts that could be achieved in a single birth cohort design. Thus, funders will need to weigh up the trade-offs between cost and sample size in deciding whether to adopt an accelerated two-cohort design or a single birth cohort.

The idea of an accelerated cohort study is to follow groups of infants and older children in parallel. This has a number of advantages compared to a single cohort design. It has been adopted in several, though not all, recently commissioned national cohort studies, including The Longitudinal Study of Australian Children, the German National Educational Panel, Growing Up In Scotland, and Growing Up in Ireland. In the case of Growing Up in Ireland, this design was chosen to accelerate the availability of policy relevant data on school age children for which policy makers would otherwise have to wait many years if data collection started in infancy only. '*This study design could yield data in a much more timely fashion on a wider range of (policy relevant) topics than would be possible from either an infant or child cohort in isolation'* (McNamara et al, 2019).The parallel design has the advantage of enabling the cohorts to be compared, and to anticipate the substantive and fieldwork problems likely to face the younger cohort as it reaches the ages already covered. The consultations taken across Europe by the team scoping a European cohort have also come out in favour of a dual cohort design (Ozan, Pollock and Fox, 2016).

In a paper commissioned for the ESRC review, Bynner and Elias (2017) argued for the new cohort to follow an 'accelerated' dual design. While more than two cohorts in an accelerated design would allow more stages of childhood, from pre-school to adolescence to be covered simultaneously, the authors considered this would be outweighed by the logistical drawbacks of following multiple cohorts simultaneously, and the dilution sample size which would likely be entailed. As Bynner and Elias (2017) point out, another option to a dual cohort design is a single cohort. The Longview report of 2006 favoured the latter, for fear that a dual design would jeopardize cohort size. Their recommendations in 2007 were for a single year birth cohort (nationally representative) augmented by a pregnancy cohort in a small number of localities. The crucial difference between 2007 and 2019 is that the desirability of filling in for a missing cohort did not apply then, but is now an important consideration. However for a limited budget, the choice of a dual cohort would have important implications for the sample sizes which could be achieved, and this trade-off needs to be carefully considered, depending on the overall budget envelope which could be achieved. (In Chapter 3 we set out power calculations for different sample sizes, and suggest that if budget only allowed a sample of 20,000 in total, a single-cohort design might be preferable).

As set out in the introduction, the aim of a single or a dual cohort should be to collect information about the context and progress of child development in the 2020s onwards to inform science and policy. In the case of a dual cohort the data should be comparable across the two new cohorts, and with previous UK cohorts, to monitor and understand change across time and age. A focus on child wellbeing, physical and mental health, as well as cognitive and educational development, should be complemented by information on the family, parental health, education, economic resources, and location. The surveys should seem relevant to the informants as well as funders, have a clear public as well as scientific interest and be reasonably non-burdensome to respondents. The commissioning of a new study must prioritise the practical feasibility of any proposed design, and avoid designs which are overly ambitious, unproven and untested. The acceptability of the proposed level of participant burden must be given due weight, and careful piloting will be essential.

2.1 Options for the co-ordination of a dual cohort

The following sections elaborate some of the important design issues that need to be taken into account in specification of a dual cohort. The recommendations made take into account both operational feasibility and scientific considerations, and offer one possible set of design choices. However none of these needs to be thought of as 'set in stone', and could be modified according to operational or scientific need. In setting out these design options we include the following 12 aspects of the cohorts' design, considering how each design parameter would affect each cohort in parallel: (See Table 2.1 at the end of this chapter for a summary).

1. Timing of recruitment
2. Period of births covered
3. Age at first survey
4. Synchronizing two cohorts
5. Dates/ages of follow-ups
6. Size of cohort
7. Source of sample frame
8. Sample design
9. Informants
10. Record linkage
11. Pilots
12. Tracing

Infant cohort	Child cohort
1.**Period of births covered by infant cohort** The infant cohort could be born over 12 months. This rules out confining it to a shorter period, e.g. 1 week as per 1946, 1958, 1970 or 9 months, Growing Up in Ireland (GUI), or spreading recruitment over more than a year, e.g. Life Study, Danish National Cohort. Note that MCS spread births of the cohort over more than a year - not originally intended – to avoid a fieldwork clash and to compensate in Scotland and Northern Ireland for the shortfall in expected births .	**Period of births covered by child cohort** For consistency and eventual comparison we assume that the child cohort is drawn from an 'academic year'. It need not be confined to those born in UK, The inclusion of immigrants already living in the UK would need to be borne in mind when making cohort comparisons. There is also the possibility of recruiting newly arrived immigrants with eligible birth dates to both cohorts as the studies go forward, which would require further plans not explicitly covered in this paper. Consideration should be given (in both cohorts) to the fact that the 'academic year' of birth dates does not run from September

Infant cohort	Child cohort
This suggests the advisability to plan for a reserve sample should there be a shortfall in births or the proportion recruited. There is general agreement that September-August is the definition of an academic year, from a UK perspective, though it fits less well with an academic cohort in Scotland.	to August in the whole of the UK, it is April-March in Scotland. There are some analytic advantages to having the sample straddle two academic years once they reach school age, but some analytic and practical difficulties to having different parts of the cohort at different educational stages, effectively halving the sample size at each state if they were split evenly
2.Timing of the start From which year's births should the new cohort be sampled? Allowing time to establish funding, obtain clearance, pilot recruitment and develop and pilot survey instruments, it would seem sensible to think the cohort birthdays could not start until Sept 2021. This implies mainstage interviews at 6 months of age (earliest) starting in early 2022 and lasting until early 2023. If the child cohort goes first, fieldwork this might extend to a set of birthdays stretching up to some point, say August in 2025. In that case the 'new infant cohort' would be 24 years younger than the Millennium cohort, but a 22 or 23 year gap would also be well worth having.	**Timing of the start** For the child cohort there are two choices, year of start and the age at which they start. The latter is discussed below. ESRC's commissioning letter mentions late primary school. There is particular interest in seeing how far the new child and infant cohorts can be co-ordinated with the sequence of British cohort studies. The child cohort will partly make up for a 'missing cohort'. A cohort born sometime in the 2010s would be appropriate to fill the gap, reaching late primary school age in the early 2020s. The higher age at the start of the child cohort the more of the 'missing years', would be captured, because it goes back further. The later in primary school they start the more of their primary school experience is missed, though some of the relevant information maybe backfilled from health and education records and retrospective questions to parents.
3. Dates and ages at first infant survey The target age should be measured in months, somewhere between 6 and 10 months, with a recommended tolerance of +/- 4 weeks either side (except at the start). It would provide a completed sample in 13 months. At this early stage in infancy (be it 6/9/10 months) it is obviously very important to keep a tight rein on age of baby at interview The tight tolerance might also focus the mind of field staff, field support staff and even respondents when it is explained how	**Dates and age at first child survey** The date of the first child survey may be as early as 2022 or as late as 2026. What age should the child cohort start? The age label in the last row of Figures 1-3 refers to the birthday reached in the academic year (Aug Sept). If the survey runs from Oct – July, taking oldest first, most x year olds will be that age when interviewed, a few, especially the August-born will still be aged x-1.

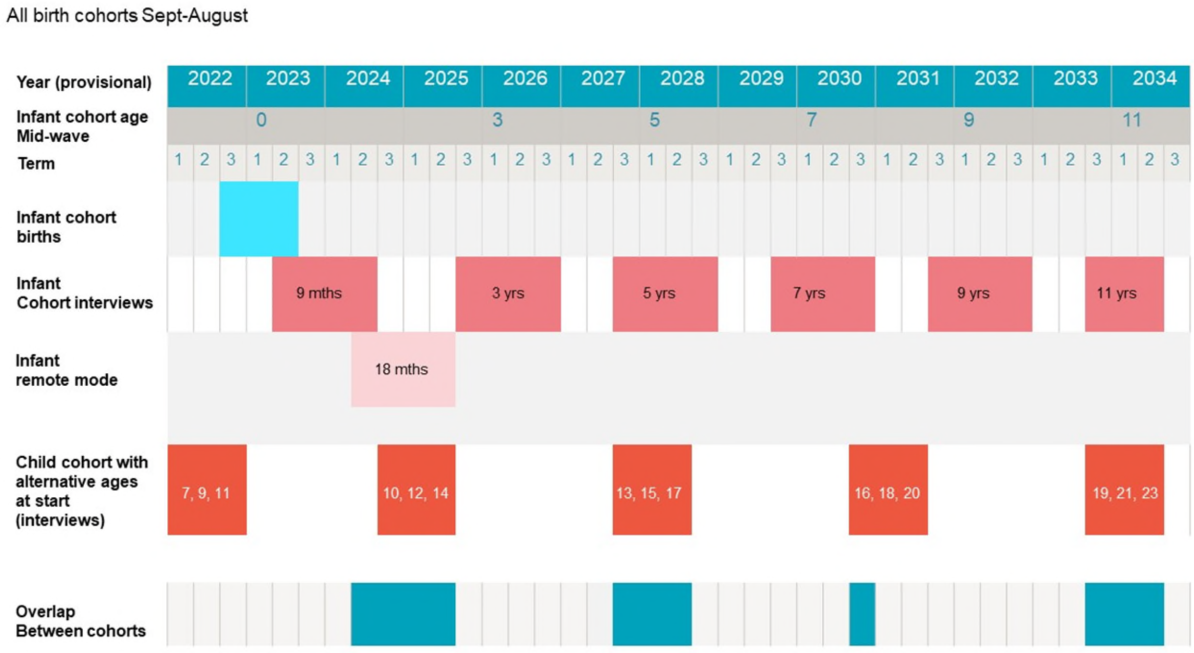
Infant cohort	Child cohort
important it is to interview between a set of given dates. If the sampling frame includes gestational data, the age at first survey could be adapted for those born prematurely. *Months of fieldwork* The interviews should be rolled out at the target months over a period of approximately 14 months (allowing minimally for some slippage in fieldwork. For comparison the first wave of infant fieldwork in GUI with 7 months of births took place over 8 months, when cohort aged 9 months. If the target month were also 9 months there would be comparability between the new cohort and MCS as well as GUI. Growing Up in Scotland interviewed at 10 months.	Assume the starting age of the child cohort is confined to later primary school years we have the following alternatives • Age 7: comparison with NCDS and MCS • Age 8: no direct comparison in British cohort studies • Age 9: no comparison in British Studies, but it formed a suitable starting point for Growing Up in Ireland. Participants in the BA symposium remarked that age 9 represents a period in the school cycle of 'relative calm' and forms a relatively stable benchmark age. • Age 10: comparison with BCS70 • Age 11: last year of primary school, parallels in NCDS and MCS We suggest that ages 10 or 11 look attractive in terms of comparability and filling the 'missing cohort' gap *Months of fieldwork* It will be important to allow sufficient time, especially if there is two-stage sampling. The experience of GUI suggests that school terms can present additional constraints. This points to the importance of configuring in a relatively lengthy period for school-based recruitment, even before the home-based fieldwork phase – possibly the best part of a school year. This applies particularly to the stratum of independent schools that would be needed in England, though in state schools the central listing of the pupil census would make some of these stages irrelevant, especially if the opportunity to collect data from teachers or head teachers was forgone. Home-based fieldwork is clearly not constrained by the school year. Although it has many difficulties and challenges it is actually much easier than school based fieldwork to schedule and implement.

4. Synchronization of a dual cohort

Figures 2.1 -2.3 plot timelines for the first 13 years of a dual study, in blocks of 4 months, 3 per year. Near the top left of each the turquoise square represents the 12 month period when the infant cohort is born. This is placed provisionally between September 2022 and

August 2023. The reddish blocks represent periods of time when mainstage fieldwork would occur, more compressed at older ages of child than in early years because it is less crucial to interview at exact ages. It is assumed that the child cohort starts first, before the births of the infants. Figure 2,1 represents one possible time line, making the optimistic assumption that there will be sufficient resources for frequent and overlapping fieldwork, and sufficient buy-in from participants to achieve a scientifically optimal coverage.

Figure 2.1 Time line for fieldwork in a dual cohort design: 3-year intervals for child cohort



Such a scientifically optimal plan would make sure the early years were well covered, with gaps between face to face interviews becoming larger as the children got older. Frequency of contact in the early years would be constrained by the burden on families, the budget and the availability of fieldwork and survey management resource. The illustration in Figure 2.1 suggests that data is collected around 18 months, between face-to- face interviews at 9 months and 3 years, this could potentially be in some other mode or using remote forms of data collection (see section on Mode, further below), [pink block]. The other ages for a face to face interview as the children grow older in this scenario are based on replicating the Millennium Cohort, at ages 5, 7, and 11, with the addition of another survey at age 9, which would form a point of comparison for the child cohort should it be decided to start that off at 9. The figure also illustrates that this sequence of interviews in the infant cohort would generate a fair amount of overlap in fieldwork periods with the child cohort (dark blue-green blocks on the bottom row). The extent of fieldwork overlap is not affected by the age at which the child cohort starts.

Figure 2.2 Time line for fieldwork in a dual cohort design: 4 year intervals for child cohort

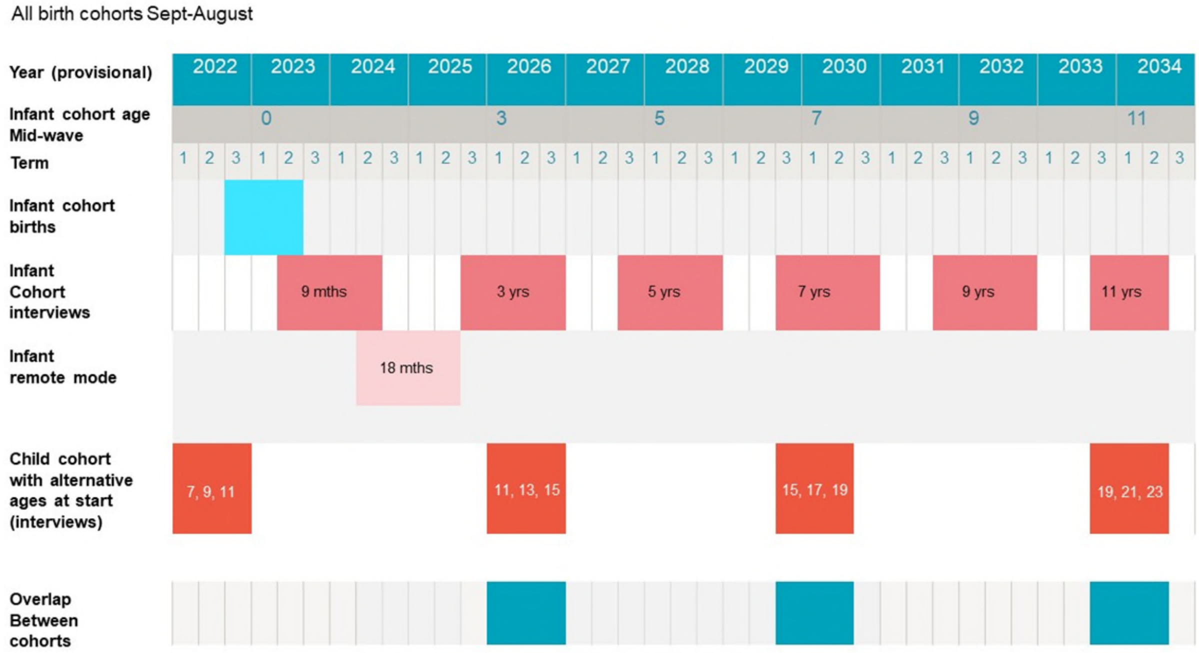
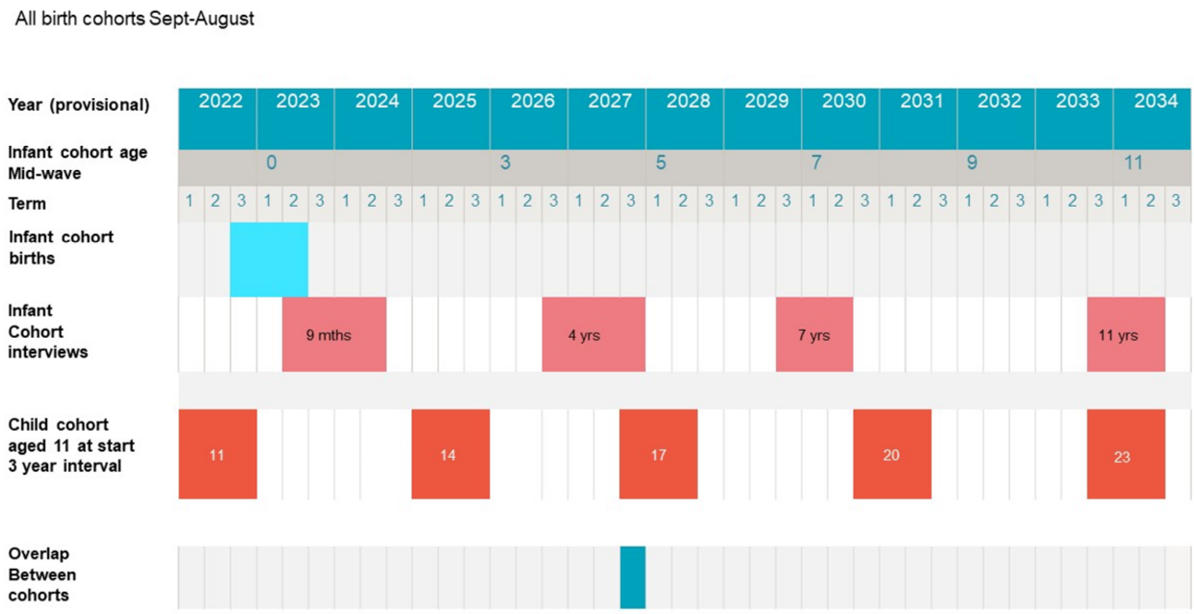


Figure 2.2 substitutes a 4 year interval for the child cohort, which reduces the number of surveys but hardly cuts the periods when both surveys would be in the field simultaneously. Of course this could be handled given adequate resources, but the two year intervals between 3, 5, 7 in MCS were challenging, and limited the scope for learning and other activities between waves. Adequate resourcing across relevant teams would be essential if such frequent surveys were under consideration, but we also offer a more parsimonious scenario.

Figure 2.3 Time line for fieldwork in a dual cohort : parsimonious version



The parsimonious version in Figure 2.3 proposes a reduction in the number of visits to the infant cohort also, and, for the sake of argument, replaces the surveys at 18m, 3yr and 5yr with one at 4yr. 7 and 11 corresponds with MCS and NCDS This sequence avoids almost all fieldwork overlaps, but comes at the cost of scientific considerations. We note that some of the difficulty with fieldwork overlaps in MCS could be reduced by keeping the birth cohort strictly to 12 months of birth rather than 16.5 months. We recommend the options are scoped (in terms of feasibility, cohort size, and survey mode). They should also be scoped in conjunction with any plans to collect data on the offspring of MCS and Next Steps.

If there were **not** to be a child cohort, the greater frequency of the infant cohort as in Figures 2.1 and 2.2 would be desirable.

Options for the co-ordination of a dual cohort (continued)

Infant cohort	Child cohort
5. Dates/ ages of follow-ups	**Dates/ ages of follow-ups**
To replicate MCS, ages would be 9mos 3, 5,7,11, 14, 17. However, there are many developmental milestones between infancy and 3 which a new cohort would ideally capture. Subject to the feasibility of scheduling 3 years (pre-school) allows one to record lots of detail on childcare, maternal return to work (albeit retrospectively) etc.) and 5 years the transition age to formal school for most children) allows collection of info on child's/parent's preparation for formal school, issues around school choice; school preparedness. Consistency with BCS and, especially, MCS would allow comparisons over time, especially on changes over time in areas like attendance at pre-school and school readiness.	The ages in the two cohorts should be chosen jointly. Figures 2.1-3 show some variants. Figure 3 replicates MCS ages, 11, 14, 17 etc. If the study started in 2022, they would have been born in '2011'. Bynner and Elias in their paper for the ESRC Longitudinal Studies review (2017) used the illustration of a child cohort starting at age 9 in 2021, at the same time as an infant cohort. Such a cohort would have been born in 2011-2.
6. Size of Infant Cohort 20,000 would be in line with MCS and earlier cohorts, and provide sufficient power for general coverage of minority groups. The need to maintain effective sample size in the infant cohort when they reach the cross over age (9-11) and beyond must be taken into account. It will need to be decided whether the sample is to be boosted in Scotland Wales and NI, within or beyond the total. See Chapter 4.	**Size of child cohort** Up to 20,000 in line with MCS and earlier cohorts. This may be limited by budget to closer to 10,000 if two cohorts are being followed. A child cohort sample size of 25% smaller than the infant cohort can be justified as early years attrition will not be a factor. Tentatively a boost to the child cohort in Scotland might be linked with the revival of the Second Birth cohort of Growing Up in Scotland, born 2010. This would follow ca 4,000 children, and is being proposed by the Centre for Applied Developmental Psychology, University of Edinburgh. It would focus on adolescent mental health. Any link to the UK child cohort depends on a nearby birth year being chosen for the UK cohort, both studies being funded, and agreement

<table>
<tr><td></td><td>being reached with all stakeholders about the basis of collaboration.</td></tr>
<tr><td>7. Source of sample frame (More details in Chapter 3.)

Ideally a combination of birth registration and NHS notifications, but there may be administrative and legal barriers, including obtaining consent. See Chapter 3</td><td>Sample frame

The sample frame for the child survey is likely to be based on the (state) school census. This should be augmented by information on independent schools, and potentially also the home-schooled. Consideration must be given to the question of how to include looked after children. GP registrants data is one potential avenue here.</td></tr>
<tr><td colspan="2">8. Simple or Complex Design

See Chapter 4 for more details.

The following options for either cohort can be summarized as follows
1) Simple Random Sample (SRS), n = 20,000: Sample representative of population, but not powerful enough for some subgroup analyses
2) Complex sample, n = 20,000 (or more), oversampling of one/two subgroups (as yet unspecified). Design effect (DEFF) an issue and more complex analysis with weights and corrected standard errors required
3) SRS, n = 40,000 (or more): Sample representative of population, AND powerful enough for subgroup analysis (not all subgroups, but some)
4) Clustered sample, n = 20000 for fieldwork logistics reasons. DEFF and complex analysis an issue for analysis, fieldwork easier.
(3) would be advantageous in terms of statistical power, but clearly has cost implications. This number may be beyond the sample size which can be considered.

A larger sample would increase the ability to analyse smaller population groups, including both those who can be identified at birth (e.g. ethnic minorities) and those who may be identified later on (e.g. children with disabilities or those who enter the care system). Population representativeness should not be conflated with subgroup analyses. Specific subgroup surveys could be commissioned to do this appropriately.
See Chapter 4 for further discussion of stratification, clustering and weighting, an Annex 1 for a discussion about the impact of design elements of MCS
The critical thing is that the sample should be representative of the total relevant population from which it was sampled and can be longitudinally reweighted and interpreted by users with relative ease. Complexity is a barrier to maximising the analytical use of the data by the widest range of disciplines and academic and other user groups. Heavy clustering or over-sampling of sub-groups can complicate re-weighting and interpretation hugely, in an area which is sufficiently complicated, given the need for longitudinal attrition weights.</td></tr>
<tr><td>9. Informants

The direct informants in past cohorts have included the cohort member, mother and father figures, teachers and head teachers. Absent parents may also be considered as potential respondents, possibly on a postal basis.

Consideration could be given to asking for ongoing consent from both parents at wave</td><td>Informants

The same set of respondents would be desirable for the child cohort as for the infant cohort. At primary level, a teacher questionnaire would provide a valuable additional perspective on the child’s development, as well as providing information about the school and class. At secondary level, this should also be considered, although there is no one ‘class teacher’ for each child as in primary school. The mode of a survey</td></tr>
</table>

1 to improve the chances of follow up if the parents split up, since it is difficult to ask resident parents for help locating absent parents.	of teachers does not need to be face-to-face. There are various options to consider.

10. Record Linkage

The potential for linkage to health, educational, economic and geographic data to enhance the surveys is very strong. It is strongly advocated both for enriching analysis and for supporting tracing rather than replacing survey data. Any administrative data used in constructing the sampling frame could be used for the analysis of non-response as well as helping to avert subsequent attrition. However the use of such information has its own costs and imperfections. There are many types of data that cannot be collected administratively and others where the routine date is incomplete or inaccurate. The sources can fruitfully complement each other, but there may also be barriers to academic access. The legal basis for linkage and onward sharing will need to be established.

11. Pilots

Mainstage fieldwork should be preceded by two pilots, one of which would test consent to the sampling procedure at Wave 1, and at all subsequent waves one to test instruments (with enough time to analyse results, and one to test fieldwork procedure). At least at wave 1 need to allow at least 12 months to develop procedures of this complexity. Time also needs to be allowed for a development phase for consultation, questionnaire design, ethical approval etc. The substantive pilot for school age sweeps in both cohorts might have to be with children in the preceding academic cohort. It may also be advisable to collect some school-based information in the previous year. We would recommend two longitudinal pilot samples for the study (one for each cohort). These would be involved in piloting each wave of the study as it moves forward.

12. Tracing

Ideally the administrative record used for drawing the original sample could help tracing and the minimization of subsequent attrition, but this also would depend on (perceived) public acceptability.
We suggest that signed parental/guardian consent to track through named administrative records would be secured at each round of interview. We need legal advice about recording and holding alternative contact addresses under GDPR and general DP regulations. Also, there are a battery of well-established practices for a cohort retention strategy besides those adapted to new digital technology. The details are beyond the scope of this report, but it is nevertheless of paramount importance that they be effectively in place.

2.2 Data collection modes and measurement

A new dual cohort will combine face to face data collection with record linkages and will exploit new technologies for innovative forms of measurement, alongside content which allows comparability to MCS and previous cohorts in key domains. A fuller discussion of measures is set out in an accompanying report on following the offspring of the Millennium Cohort Study and Next Steps by Fitzsimons and Bornstein. In that report we set out an approach to measurement which is multi-domain, multivariate, multimodal and multi-informant, providing the most complete and robust insights into infant and child development. The collection of biological specimens is an option whose content and timing should be considered especially carefully.

Addressing the issues of cohort engagement and attrition will be vital to the success of the project. Factors to consider include minimising respondent burden, using engaging instruments and modes, involving cohort members in the study, and maximising the perceived social value of the study.

We note that content will focus on material that is not possible to collect retrospectively or from administrative sources, such as aspirations, mental health of both the child and parents, cognitive attainment, and social networks.

2.3 Enrolment of immigrants

The child cohort would automatically include children who had not be born in the UK, who should be identified as such in analysis. Provisions to deal with international migration e.g. through immigrant boosts - could be envisaged but are outside the scope of this report.

2.4 Ethics and governance

It will be vital to develop a structure for governance and ethics, addressing child protection issues, data access, and consultation with funders and the research community. In addition the channels for international collaboration and comparison should be kept open.

2.5 Multiple respondents

As well as mothers and cohort members, it will be important to engage fathers with the study. The vast majority of fathers are resident at birth, and therefore there is an opportunity to gain paternal consent for independent follow-up, which will be particularly important in case of family break-up. The study team will be able to build on the work of the Nuffield Foundation carried out for the Life Study in considering how best to engage fathers (Kiernan 2016).

Setting aside concerns of respondent burden, adding grandparents (for those living) would have some advantages: potential genetic analyses, long range social mobility, childcare, capital transfers (though some of this can be done by asking parents about grandparents). If grandparents are involved at the outset they could also be a resource for tracing later on, However, contact with grandparents would be challenging, many will not live near the CM or even in the same country. Administrative data linkage consent for grandparents could be an option.

Siblings could be incorporated, but including all siblings could be overly burdensome. In MCS, up to two older siblings were included. Younger siblings could potentially be picked up in pregnancy. Sibling data would be used to support sibling analyses, e.g. sibling difference models to control for unobservable characteristics, but they would not constitute a convenience sample of older/younger age groups, as these would only be representative of children with a cohort child in their family. Administrative linkage consent for all siblings might be considered.

Table 2.1 Summary of design issues

<table>
<tr><th>Infant cohort</th><th>Child cohort</th></tr>
<tr><td>1. Timing of Recruitment
Sep – Aug birthdays front runner for both cohorts</td><td>How many years before infants?
8-12</td></tr>
<tr><td>2. Period of births covered by infant cohort
2022/3 – 2024/5</td><td>Period of births covered by child cohort
2021/2022 -2024/5</td></tr>
<tr><td>3. Age at first survey
6mth-11mth</td><td>Age and date at first survey
7, 9, 10, or 11, with some preference for 11</td></tr>
<tr><td>4. Synchronisation
Avoid fieldwork overlaps</td><td>Start child cohort first
To recover missing cohort and allow time to pilot new approach to sampling births</td></tr>
<tr><td>5. Dates/ ages of follow-ups
18m,3, 5, 7, 9/11 if no child cohort, else 3 or 5, 7/9, 11,
in either case post 11: 14, 17, 22</td><td>Dates/ ages of follow-ups
(9) 11, 14, 17, 22</td></tr>
<tr><td>6. Size of infant cohort
20,000 (up to 40,000 but budget constraints make this unlikely)</td><td>Size of child cohort
15,000-20,000</td></tr>
<tr><td>7.Source of sample frame
Registration +NHS notification if new procedures produce adequate response</td><td>Source of sample frame
NPD + independent schools; GP registrants</td></tr>
<tr><td>8a. Clustering of sample
Some geographical sampling, on area wider than an average ward, may save initial fieldwork costs, however adds analytical complexity</td><td>Clustering by address or school?
Primary schools quite a small cluster as far as field work economies</td></tr>
<tr><td colspan="2">8b. Stratification
Both cohorts PPS by Region and size of PSU. Simple structure for both has attraction for users and implementation of sampling.</td></tr>
<tr><td colspan="2">8c. Disproportionate stratification, complicates analysis, but some over-representation may be of interest.
Strongest case for groups expected have low response rates and high attrition</td></tr>
<tr><td colspan="2">9. Informants
Resident parents or caregivers, and (potentially) non-resident parent and teacher/ head teacher, cohort member.</td></tr>
<tr><td colspan="2">10. Record linkage
Should play a valuable role in sample selection and maintenance, and then enhance but not replace survey data</td></tr>
</table>

Chapter 3. Drawing a nationally representative sample

This section should be taken as preliminary, and subject to further investigation with NHS-Digital, and the Office for National Statistics (ONS). It should be informed by legal advice and may be affected by developments in the use of government data for research. Securing an opt-out sampling frame will be central to the success of this endeavour, given the lower response rates associated with opt-out sampling (Dezateaux and Elias 2016). We have no doubt that this will be possible with the political will behind it.

Collaboration with official statistical authorities is essential to ensure joint activities in linking data, particularly for setting the sample frame. The venture needs to be understood as an investment in a national data resource, not just a research project.

We have assumed that the practical obstacles to drawing a sample and obtaining consent in (or before) pregnancy, as experienced in Life Study and the US National Children's Study, rule out starting a new birth cohort study before birth.

We suggest that the younger cohort will be sampled from surviving live births, with data on these births collected some time during their first year of life..

As concerns infants, NHS-Digital hold data from ONS birth registration, NHS Birth Notifications and Maternity HES, for England and Wales. Combined, these datasets include the address at registration, postcode, name of mother and of child, the NHS number of both, and whether the birth was live. There are also other items of information which might be useful to consider in designing (or validating) the sample, including gestational age, ethnic category of mother (from HES) and of baby (from NHS birth notification), country of birth of each parent, among others.

One potential option for drawing a new birth cohort sample of infants would be to use ONS data from birth registration without further linkage to NHS data. However this approach would have severe limitations because it does not have any updating of addresses after registration. Registration must take place no more than six weeks after the birth, and is at a time when families frequently move home; it does not have information on ethnicity, and crucially, it is not clear how ONS might permit an approach to potential sample members on an opt-out basis or some other effective form of encouraging consent to participation, yet to be developed .

One example of a linked dataset which combines Birth Notification and NHS data has been constructed by Professor Alison MacFarlane for 2005-2014. Another one was set to be used in drawing the sample for the national probability component for the Life Study, abandoned in 2015. Linking has to be done with care, and is not currently done on a routine basis. Currently ONS has published a table on ethnicity and gestation in its Birth Characteristics Tables which features information from NHS-Digital as well as the ONS birth register. Although the combination of these datasets is not routine, NHS-Digital have informally indicated that it could be feasible for a new project. NHS Digital refreshes the infant data base frequently and would be able to update addresses of families who had notified a move to their GP, and remove any cases where the child died in the early months.

Such a linked dataset would have many of the properties needed to draw a nationally representative sample which could be stratified by region and by clustered within regions to facilitate field work efficiency and even to over-represent localities with characteristics taken from small area statistics such as the Index of Multiple Deprivation or Census Small area statistics. It would also in theory be possible to over sample on the basis of information included in the individual birth register, such as ethnicity, maternal age or parent's country of birth.

Although this potential sampling frame has many attractive properties, it may need to be created specifically at the time needed. The legal and practical possibilities to exploit it also need to be established. Setting aside the important question of whether the statistical agencies have, or could be funded to have, the necessary practical resources, is the even more key question of whether they could legally divulge the identities and whereabouts of the members of the public thus chosen without their consent. ONS is governed by Statistical legislation which obliges them to obtain consent, an 'opt-in', to release details to an external body such as a fieldwork agency. NHS Digital operates under different legislation, the NHS Act of 2006. This provides conditions, in Section 251, under which permission may be granted to approach people who have been drawn in a sample on the basis of 'opt-out'. A researcher can use Section 251 to enable the common law duty of confidentiality to be temporarily lifted so that confidential patient information can be transferred to an applicant without the discloser being in breach of that duty of confidentiality. Subject to confirmation, selected participants can be sent a letter saying their details will be passed on if they don't object (as was done for the Millennium Cohort Study by the holders of Child Benefit records) and/ or visited in the field by an interviewer to discuss and obtain informed consent in person. There is, of course, no guarantee that families selected for a survey will agree to participate at the end of this process. They may not feel inclined to participate even when asked in person, but co-operation is less likely, and more likely to be biased, if the first approach is an impersonal invitation to send back written consent. Indeed the evidence from the Life Study pilot was that fewer than 20% did so. A comparison of participants with all invitees indicated a significant bias towards participation by more socially advantaged groups. A second pilot was to be conducted for Life Study using an opt-out, whereby potential participants would be approached, having had the opportunity to refuse to participate. With the cancellation of Life Study in October 2015, all plans for the second pilot were terminated, so this approach was not tested, but for the MCS, including an opt-out, the response rate was around 70%.

Since then the Digital Economy Act of 2017 has introduced new rules for statistical and academic research, and new procedures whereby ONS invites people to join a study, putting them in direct contact with the fieldwork agency. These procedures have yet to be piloted and it is not yet known what sort of response rate might be expected. ONS have kindly agreed to investigate the legality and practicality of this route.

The route to gaining permission to operate an opt-out is via an application to CAG, the Confidentiality Advisory Group of the Health Research Authority, which covers England and Wales, and the equivalent authorities in Scotland and Northern Ireland. In Scotland, the body to approach is the Public Benefit and Privacy Panel of ISD. In Northern Ireland, the first port of call would be NPIH

The application needs to show that data is needed for a medical research purpose, that the purpose cannot be served on an opt-in basis, demonstrate 'patient engagement' and that suitable data security provisions are in place. Applications to CAG are made in tandem with a preliminary application to the Research Ethics Committee (REC) and for Data Security Clearance. CAG makes a conditional recommendation to the Secretary of State, taking a number of points into account. The legal grounds for processing such data would be that ‘processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller’ (GDPR Article 6).

The Committee meets every 2 months, and provides for some pre-application consultation. Clearance would take at least three months, especially if the Committee requests amendments to proposed plans. Before this there should be discussions with the NHS-Digital Data Access Request Service, but plans cannot be taken very far until CAG Clearance is granted.

3.1 Administration of Sampling

It is arguable that ONS has (or has had) greater experience and capacity in practice of drawing and administering samples than NHS - Digital, but we need to check that they still have the capacity and are not bound by the need to get a written opt-in. This does not seem to have prevented them acting as agents to drawing the sample for the Mental Health of Children and Young People in England Survey.

3.2 Lessons learned from Life Study

Set against the success of the Millennium Cohort and its predecessors, the unfortunate experience of the Life Study need not be taken as demonstrating the impossibility of ever having another successful birth cohort study. However, there are important lessons to be learned from Life Study.

Scientific ambition must be balanced against practical feasibility and respondent burden. Keeping the study instruments relatively minimal at the outset – with the possibility of adding more complexity later on – represents a strategy to engage mothers and families and increase the likelihood that they will commit to the study.

More generally, cohort engagement and a focus on methods to increase uptake and retention will be vital to the success of the study.

Sampling pregnant women presents two difficulties – the absence of a national listing of pregnant women, and the hesitation mothers may have when they don’t know what and when the outcome of pregnancy will be. Given the absence of a national listing, studies need the co-operation of ante-natal service providers. Attempts to sample pregnant women have had low response rates (e.g., US National Children’s Study). Although the Pregnancy component of Life Study was confined, in the end, to just 4 localities, the response rate in the pilot was low (Dezateaux, Colson, Brocklehurst, Elias 2016) . Pregnancy samples have succeeded in local areas, with strong buy-in from local communities, and Health Service staff (Born in Bradford, ALSPAC). This does not mean that it could be scaled up to national level. All recent successful national cohorts that we are aware of have sampled infants, post-natally (e.g. GUI, NEPS).

Here are some constructive conclusions about recruiting participants offered by Professor Peter Elias (private communication) who was involved in Life Study at the time:

“*Section 251 of the NHS Act 2006 provides for an ‘opt-out’ approach to sampling for a new birth cohort, provided there are sufficient important medical issues to be addressed, that the records to be sampled are medical records, that consultation has indicated the acceptability of the approach and that evidence has been provided to indicate that no alternative approach yields a workable solution.*

Linked NHS notifications with birth registration records provide an up-to-date population frame for a birth cohort. The richness of information on these linked records affords opportunities for targeted samples, as well as providing vital information on non-respondents.

Given the time required to sample these records, contact with mothers/fathers when the child is 6 months old is possibly the earliest time for contact to be made.

Samples can be issued with adjustments for the term of the pregnancy – a very important issue that has not hitherto been addressed in the UK birth cohorts. There are some differences in linkage procedures in Scotland compared with England and Wales, but these are well recognised and surmountable. Northern Ireland does not have such good access to linked NHS/birth registration records. This is an issue that requires further investigation".

These processes cannot be actually initiated until it has been decided if there will be a new cohort study, and who is responsible for it, but no such venture is likely to get off the ground without having mapped out the steps outlined in this note. It is also more likely to succeed if the official agencies and statisticians are actively participating with academics and fieldwork agencies. We strongly recommend enhancing the links between academia and ONS and/or NHS digital via secondment of an academic.

3.3 Other possible sources of a sampling frame for infants

Child Benefit records might be suitable for drawing a sample of families claiming Child Benefit, but this has not been pursued since a growing number of families have too much earning to qualify, and it is not at all clear that DWP would permit the use of their records, even if they were co-funders of the study. The Labour Force Survey with ca 500,000 people sampled every year would cover ca 5,000 infants or children of a given single year of age, which is inadequate for these purposes.

3.4 Sampling Frames for the Child Cohort

NHS-Digital's Population Demographics, based on GP registration, is a possible source of a sampling frame for school aged children. NHS-D has agreed this would be feasible, although the cost of this (and indeed other routes) has not yet been estimated. NHS may not have developed the expertise or resources to offer such a service (in all 4 countries). It was out-sourced to ONS for sampling the Child and Young Person Mental Health Survey in England in 2017. The data holds the possibility of stratifying by ethnic group (subject to data quality and the considerable 'unknown ethnic group'). Whether or not permission would be granted has not been established at this stage. This approach might be better at finding the minority of children who are home-schooled than an approach relying on school registers.

The NPD School census would allow sample of individuals or a two-stage sampling schools then pupils. Primary schools mainly have one class in a given year averaging around 27 pupils, A few have two or three classes in a year, in which case the design could take a random class within a school. Whether 1 or two or three stages, it would need a separate exercise in 4 countries and for independent schools (very few outside England) and resources might be needed to obtain information from or on schools.

LSYPE1 took a two stage sample of (secondary) schools, with independent schools a separate stratum. It oversampled schools with high numbers on Free School Meals and boosted the samples of ethnic minorities within state schools based on individual level data in NPD. Users are provided with weights reflecting this design. There is also a two stage design to sample 15 year olds for PISA.

Growing up in Ireland (GUI) had a two-stage clustered design, built around the primary School system, which included a stratum for a small number of private schools. The clustering in primary schools, with up to 40 pupils taken per school, worked very effectively and allowed the multi-level modelling, but clustering would have been lost in secondary schools at age 13 wave, and was never established for the infant cohort when they reached school age.

For both school and NHS sample in UK, identifiers would enable routine data to be linked to survey - notably birth and health records - if sampling yielded an NHS number.

Chapter 4. Sample design, simple or complex? Stratification, clustering and weighting

A new cohort study requires a sample design with a structure aiming to create a representative sample from which generalisations can be drawn about a population, with a known probability of having been selected. Two features of a sample design affect both the practicality of administering a survey and the certainty with which inferences can be drawn: clustering and stratification. Proportionate stratification, with probability proportional to size within strata ensures that the groups of interest within the population –e.g. regions- have presence in the sample which reflects their relative size, and helps minimise sampling error. Disproportionate stratification magnifies the size of sub-samples of groups of interest and also reduces sampling error around estimates of what would otherwise be small samples. A clustered design affects the efficiency of inference depending on how much of the variability in the population is contained within rather than between clusters.

Unlike in a cross sectional survey, when drawing a stratified or clustered sample for a longitudinal survey it is necessary to consider the longevity of factors used in constructing the design. The concerns of future researchers may focus on variables not recognised, or at least reliably identifiable, at the initial stage. Eligibility for particular policy interventions, for example, may be transitory, so may parental employment or marital status. Particular health conditions or traits may not be apparent at the time the sample is drawn. Area of residence is likely to change, which will dilute any informative properties of the original clustering. Belonging to an ethnic minority is a strong candidate to consider for a sample boost, as it changes little over time and there are many questions for analysis which may go unaddressed if the sample size is inadequate. Policymakers may be particularly interested in vulnerable groups who are 'hard to reach' and often lost to follow-up, including 'looked after' children, children of prisoners and children of refugees. Specific efforts would be needed to reach and retain these groups which cannot be identified and oversampled at birth.

We assess the success of over-sampling specific areas for high ethnic minority populations in MCS (Chapter 4, and annex) and conclude its effectiveness was limited, for the following reasons:

a) any fieldwork efficiencies that were gained early on in the study by sampling larger numbers of ethnic minorities in particular areas were probably very limited because irrespective of their geographical concentration, absolute numbers were small
b) beyond the first data collection, such fieldwork efficiencies were further dissipated by the fact that geographic mobility is typically very high among families with young children, so these minority children no longer remained highly geographically clustered for long
c) ethnic minority families clustered within specific geographical areas are not typically representative of the broader national profile of ethnic minorities, and so clustering may have come at the cost of national representation of a broader set of minority groups.

These suggestions, together with the higher population prevalence now of many ethnic minorities compared to 18 years ago would suggest unclustered national sampling is likely to be a stronger strategy. Some of these considerations are very likely to apply to other types of over-sampling local areas for families at risk. However the evidence would need to be carefully considered on a case- by –case basis (depending on what type of at-risk-group is being considered).

For example, some of the groups that policymakers are most keen to reach, e.g. ‘looked after’ children, will not be reached by geographic-area oversampling. Indeed the majority will not be identifiable at birth. Reaching such groups is likely to require focussed sub-studies, which could with sufficient budget, potentially be built into the cohort design. Whether or not these would be geographically clustered would have to be carefully considered. As above, it is likely that geographic-clustering may come at the cost of national representation, and the efficiencies of doing so may be both limited and short-lived

The following section illustrates, in a simplified way how the practice of disproportionate sampling may serve to increase the power of a survey to make estimates about minority groups. Clustering, while potentially reducing fieldwork costs has the statistical cost of a increasing standard errors in a sample of given size. The Annex discusses the lessons learned from the specific design implemented in MCS.

4.1 Power Calculations

The degree of certainty about the predictive power of a sample depends on the size of the sample and the size of the differences we wish to detect. The smaller the sample and the smaller the difference, the lower the ‘power’: the probability of correctly rejecting the null hypothesis (for example, of no mean difference between groups).

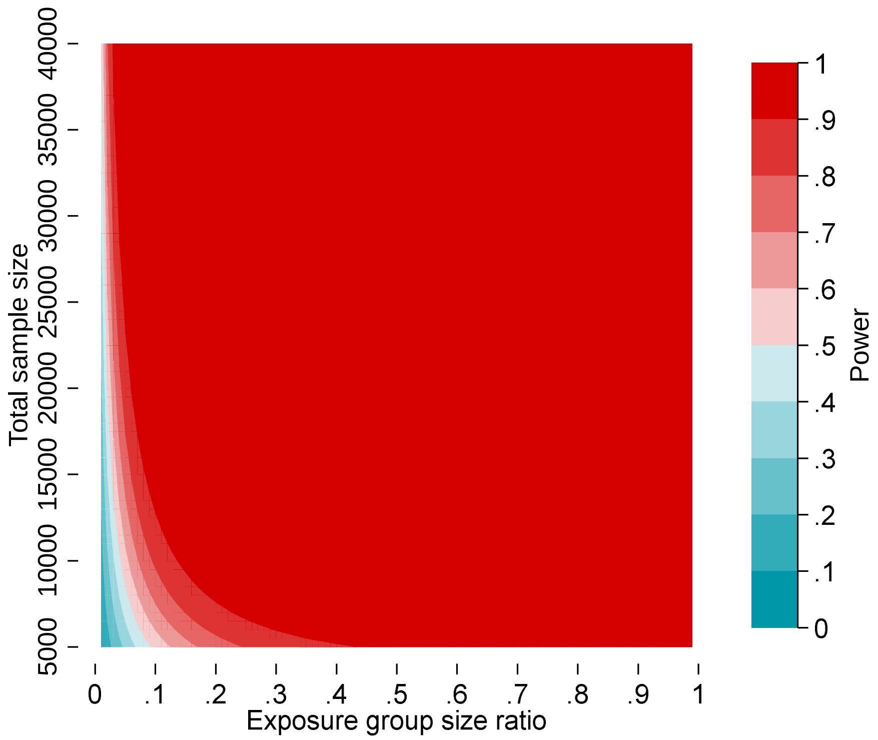
Analyses across the whole sample

We examine the potential power of analyses of samples of different sizes in two separate ways: i) by varying the exposure prevalence (for a given effect size) and ii) by varying the effect size (for a given exposure prevalence).

The impact of exposure prevalence. We examine this analytically, assuming simple random sampling and interest in estimating a mean difference between two groups, where the outcome has mean 0 in one group and mean 0.1 in the other, with common standard deviation of 1. We consider total sample sizes between 5,000 and 40,000 and exposure group size ratios between 0.01 (i.e. the smaller group is 1% the size of the larger group) and 1 (i.e. the smaller group is the same size as the larger group). Calculated powers are shown in Fig. 4.1. Power increases as the exposure group size ratio increases (i.e. as the exposure groups become more similar in size), for a given sample size, and as the sample size increases, for a given exposure group size ratio.

With sample sizes of 10,000, 20,000 and 40,000 the power would be greater than 90% for an exposure group size ratio of at least 0.14, 0.06 and 0.03, respectively. If the sampling departed from simple random sampling, the design factor, discussed below, would reduce the effective sample size, thereby reducing the power of the sample.

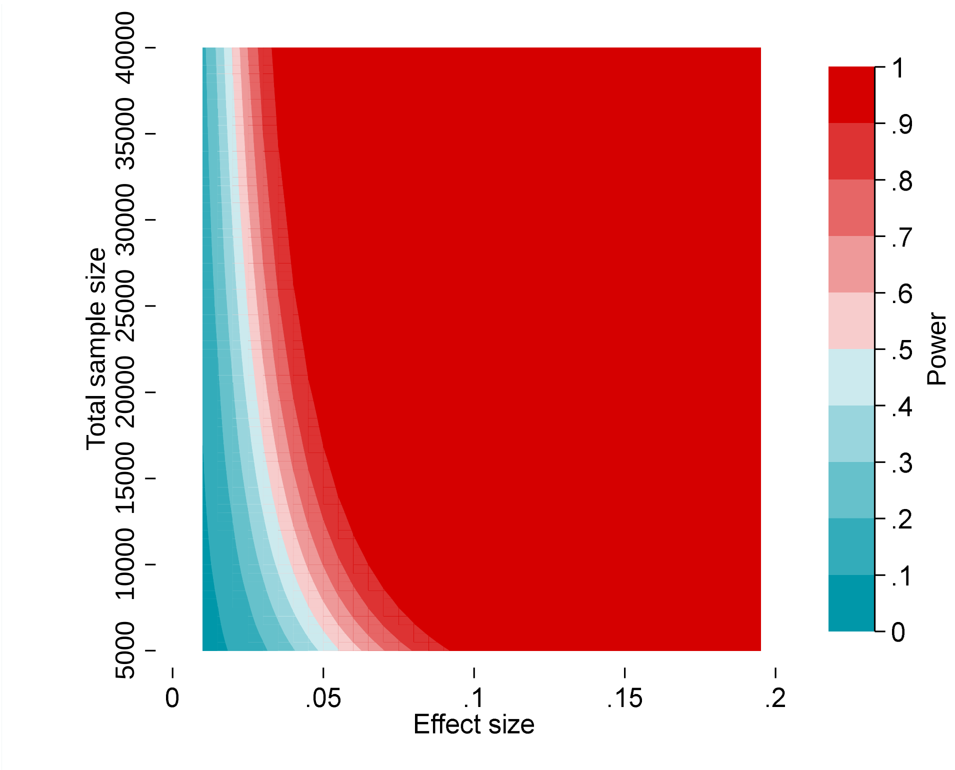
Fig. 4.1. Power by exposure group size ratio and sample size



The impact of effect size. We examine this analytically, assuming simple random sampling and interest in estimating a mean difference between two evenly distributed groups, where the outcome has a common standard deviation of 1 in the two groups. We consider total sample sizes between 5,000 and 40,000 and mean differences of between 0.01 and 0.2 between the two groups. Calculated powers are shown in Fig. 4.2. Power increases as the effect size increases, for a given sample size, and as the sample size increases, for a given effect size.

With sample sizes of 10,000, 20,000 and 40,000 the power would be greater than 90% for an effect size of at least 0.07, 0.05 and 0.04, respectively. If the sampling departed from simple random sampling, the design factor, discussed below, would reduce the effective sample size, thereby reducing the power of the sample.

Fig. 4.2. Power by effect size and sample size

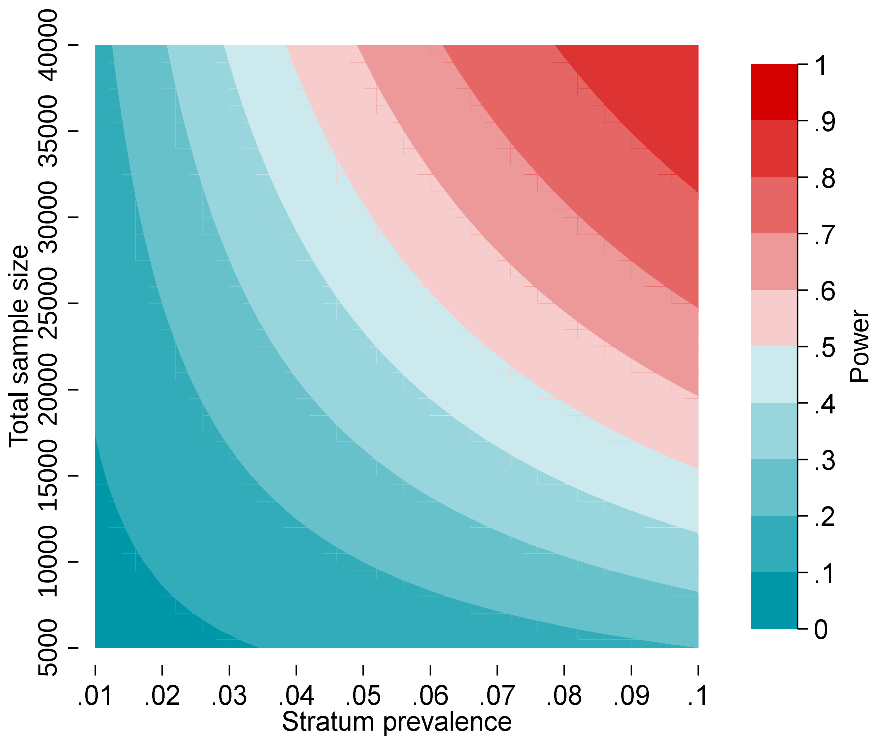


Analyses within specific strata

If we wish to estimate something, for example an association, within a specific stratum or subgroup of a population, then our power to do so will be determined by the size of the stratum of interest within our sample – or equivalently by the combination of the total sample size and the sample prevalence of the stratum of interest. We examine this analytically, assuming simple random sampling and interest in estimating a mean difference between two evenly distributed groups, where the outcome has mean 0 in one group and mean 0.1 in the other, with common standard deviation of 1. We consider total sample sizes between 5,000 and 40,000 and sample prevalences of the stratum of interest between 1% and 10%, meaning sizes of the stratum of interest within the sample between 50 and 4000. Calculated powers are shown in Fig. 4.3. Power increases as the stratum prevalence increases (i.e. as the variable of interest becomes less uncommon), for a given sample size, and as the sample size increases (for a given stratum prevalence).

With a sample size of 20,000, at 10% stratum prevalence the power would be 61% and at 5% stratum prevalence it would be 35%. With a sample size of 10,000 the power would be no more than 35% at any of the stratum prevalences considered. With a hypothetical sample size of 40,000 there would be at least 80% power to detect this difference at stratum prevalences of at least 8%, which could be achieved within a context of a dual cohort if the two cohorts, each 20,000, were pooled for some purposes. If the sampling departed from simple random sampling, the design factor, discussed below, would reduce the effective sample size, thereby reducing the power of the sample.

Figure 4.3 Power by stratum prevalence and sample size for within-strata analyses



4.2 Complex sampling strategies

The design effect – calculated as the ratio of the variance under complex sampling to the variance under simple random sampling – can be used to assess the extent to which complex sampling reduces the effective sample size of a survey. For a given complex sample size, dividing by the design effect gives the effective sample size had simple random sampling instead been used. Here we employ an analytical approach to explore the effects of weighting (necessitated by unequal selection probabilities) and clustered selection on the design effect.

Unequal selection probabilities: For the sake of illustration, we assume the presence of only two population strata: a "majority stratum" and a "minority stratum". The interest is in oversampling the minority stratum. We consider minority stratum population prevalences between 1% and 20% and minority stratum sample prevalences of between 2% and 40% (with over sampling). We do not consider cases where the minority stratum sample prevalence is less than the minority stratum population prevalence as this would imply under-sampling of the minority stratum. Calculated design effects are shown in Fig. 4.4. When the minority stratum population and sample prevalences are equal this implies simple random sampling, meaning a design effect of 1. The design effect increases as the minority stratum population prevalence decreases (for a given minority stratum sample prevalence) and as the minority stratum sample prevalence increases (for a given minority stratum population prevalence), both of which correspond to greater oversampling of the minority stratum. As a specific example, a four-fold oversampling of black children assuming a 4.2% population prevalence (most recent estimate in Figure 2) would give a sample prevalence of 16.8% and result in a design effect of 1.11. An achieved survey sample of 20,000 children would therefore have power equivalent to a sample of 17,960 had simple random sampling instead been used.

Clustered selection: We consider average cluster sizes between 1 and 100 and intra-class correlation coefficients (ICCs) between 0 and 0.3 (though note that survey ICCs above 0.2 are seldom observed). ICCs are likely to be higher if a given cluster size is achieved by sampling 100% of a small area population than a smaller fraction of a larger, less homogeneous population. Calculated design effects are shown in Fig. 4.5 (note the difference in scale from Fig. 4.4). When the ICC is 0 there is essentially no clustering, meaning a design effect of 1. The design effect increases markedly as the average cluster size increases (for a given ICC) and as the ICC increases (for a given average cluster size).

Figure 4.4 Design effect by minority stratum population and sample prevalences

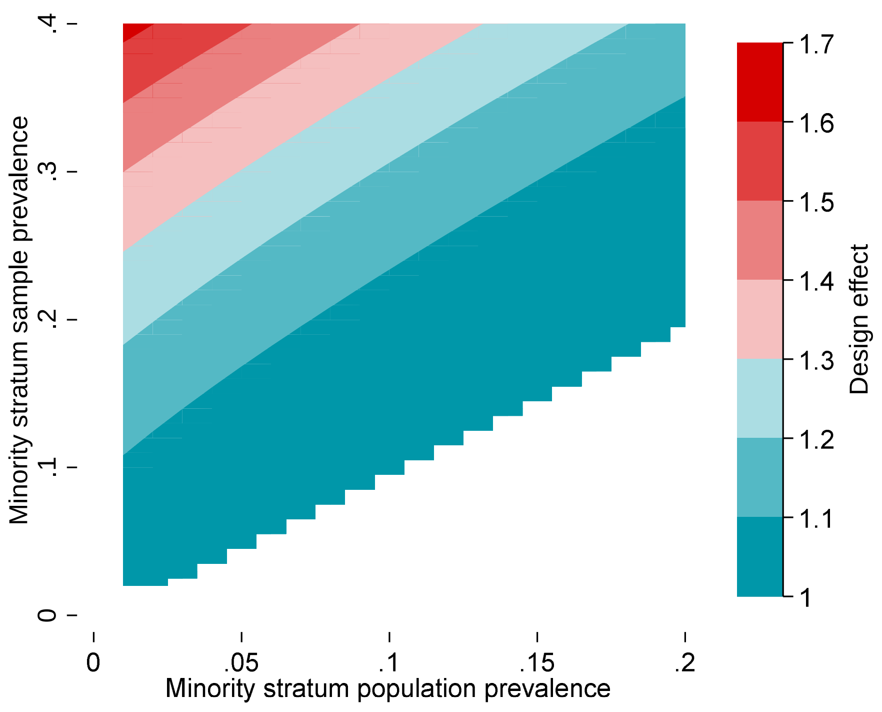
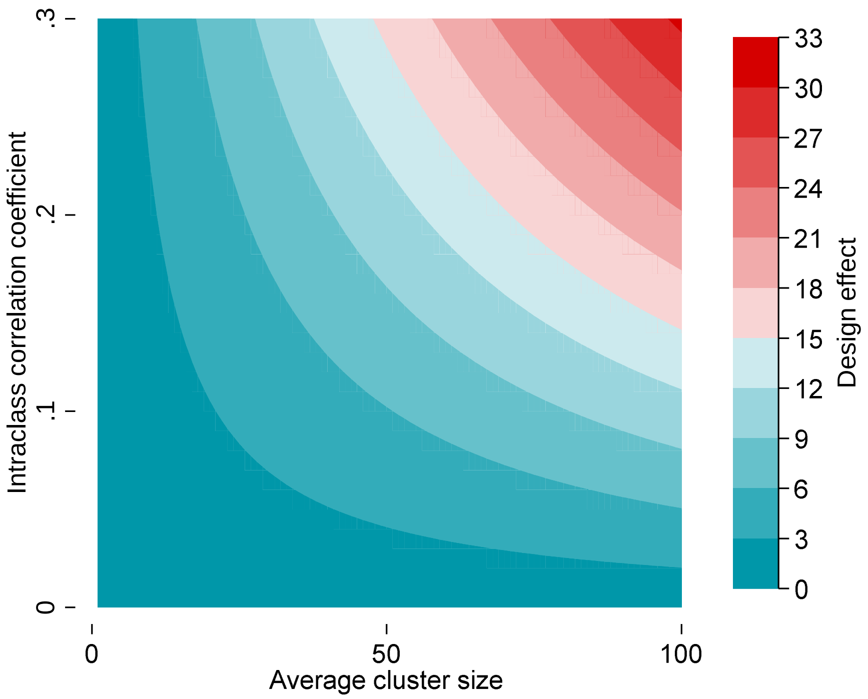


Figure 4.5 Design effect by average cluster size and intraclass correlation coefficient



While complex sample design can be handled by statistical adjustments including weighting by the inverse or the sample proportion (as illustrated in the Appendix on MCS), they do present some obstacle to less sophisticated users, particularly if they also wish to allow for differential attrition through weighting, or take sub-national samples. This is not illustrated here, but with complex patterns of longitudinal non-response, this additional source of weighting can get quite challenging**.**

4.3 Design for two Cohorts

One consideration might be that the two cohorts should follow the same general design in the interests of comparability. That is to say, if infants were to be clustered (disproportionately or not) by postcode district, the child cohort could follow the same principle, even the same districts. But it is not envisaged that the surveys would be in the field at the same time (or even necessarily have the same fieldwork agency) so the field work economies of having the same geography would not be great. However the important thing is to have a design based on a random sample for both cohorts which allows generalisation to a known population and which can be used relatively easily by analysts. Having the same general design for both is not a priority.

For children of school age there is the possibility of using school registers (as recorded in the National Pupil Database and equivalents) as a sampling frame, and taking a two-stage approach to sampling. The option remains of how small or large the sampling fraction of children of the required age within each school. A high fraction in a lower number of schools would open the possibility of estimating 'school ' or 'class' effects in multi-level models, but how useful this would be depends on how much variation there is between rather than within schools. It would also depend on how much mobility there was between schools, and for how many waves the survey was taking place at a primary school age. If child cohort did not start till the last year of primary school ('age 11'), there would only be one wave which could capture school effects. On the other hand, surveys starting at age 7 and 9 have not been entirely ruled out. If there were disproportionate sampling at either the school or individual level, provisions would have to be made in analysis for the complexity – weighting and corrections of sampling errors.

A new infant cohort does not have to start with the building block of the electoral ward. It, and also perhaps the child cohort, can start with a census-based geography, with less variable population size, and the boundaries derived from the 2011 census. This is due to be updated in 2022 to incorporate the results of the 2021 census. The geography should be used to ensure proportionate stratification across regions in England, at least. Geographical clustering within region may be useful to contain initial fieldwork costs, but it should have less granularity than the wards of MCS. Perhaps a Postcode Sector, or MSOA, with a sampling fraction well under the 100% used in MCS wards, to reduce the distortion of estimated variances, but even these might be rather sparse on infants. A postcode district, roughly 4 times the size of a postcode sector would yield more births from which to sample (ca 200 on average pa), but not necessarily enough to keep one interviewer fully occupied. Any disproportionate sampling that may be chosen could be done on the basis of smaller area indices, such as the IMD or its IDACI component at the LSOA Level. Census indicators of the ethnic composition of an area are not recommended as way of producing an 'ethnic boost' if the sampling frames contains information on individual ethnicity.

While there is no doubt that stratification by, say region would improve the precision of the sample, it is not so clear how much would be gained by *disproportionate* stratified sampling. It can focus on certain minorities of interest, but it will in perpetuity require the use of weights in analysis. It also need to be considered how closely clustered within the primary sampling units families need to be for initial field work economies, and how far strong design effects on estimates of standard errors can be avoided. Another consideration is that the process of selection a sample for a birth cohort is absolutely time critical. The children to be sampled get older by the day, and the sampling process cannot afford to make mistakes or run into delay. Piloting the process in advance (with the permissions in place) would be one precaution. A simple formula for the actual drawing of the sample would be another.

Chapter 5. Conclusions

We conclude with a set of conclusions based on the arguments in the chapters above.

1. A new nationally representative birth cohort study would address the need for longitudinal evidence on the development and life-chances of a new generation of children in the UK. To serve a wide set of scientific and policy questions, the study should be broad based.

2. Social scientific findings are contingent on time, place, and population (when, where, and who), and we cannot rely solely on data on previous cohorts to allow us to understand contemporary and future trends and processes and how these will affect new generations of children as they grow up.

3. We have provided detailed design options for an 'accelerated cohort' design featuring one cohort of infants, shortly after birth, and another cohort recruited in mid-childhood to fill a gap in knowledge on the generation already in mid-childhood in the UK. This would provide timely evidence to policymakers and others regarding school-age children. However if budget is constrained, a dual cohort limits the sample sizes available in each, and so depending on the overall budget envelope the trade-offs between a single and a dual cohort need to be carefully considered.

4. Securing an opt-out sampling frame, as, will be central to the success of the project. Securing the political will to achieve this at a high level will be essential, and discussions with the ONS and NHS regarding administrative feasibility will be required.

5. We have considered a range of options for the sample design. Drawing lessons from the MCS, we recommend against tightly clustered disproportionate stratified sampling. The advantages of a relatively simple design include greater accessibility to the data for a range of analysts. If oversampling of ethnic minorities is required, we recommend this be done through the national sampling frame rather than geographical clustering of the sample.

6. We recommend a sample size of 20,000 for the infant cohort and 15,000-20,000 for the child cohort. There would be some scientific advantages to increasing these sample sizes (e.g. to 40,000 and 20,000), but this decision will of course be constrained by budgetary considerations. Should a total sample of only 20,000 be affordable, we recommend a single birth cohort (rather than a dual cohort design).

7. We have outlined a range of options for the timing of data collection in a dual cohort, taking into account both cross-cohort compatibility, key developmental stages and milestones, and practical considerations such as school terms. Our provisional recommendation is for data collection at the following ages: 9 months; 18 months; 3 years old; 5; 7; 9; 11; 14; 17; 20; 23, starting the child cohort at either 11 or 9. The intention would be to follow the cohort throughout their lives.

8. Consideration of both cohort engagement and wider public engagement will be vital – this is not addressed in this report as it is being considered in a parallel report by the Institute for Employment Research

Annex: Lessons from the sampling design of the Millennium Cohort Study

We take the design adopted by the Millennium Cohort (MCS) as a starting point to consider how any new cohort sample might be structured.

Section A1 presents an outline of its design. We then consider (in sections A2 to A5) the deviation of the unweighted survey numbers and percentages from what would have obtained, hypothetically in a simple random sample. Section A6 considers, how far the disproportionate sampling strategy served as an insurance policy against attrition, and section A7 explores how far the initial clustering of the sample was eroded over time.

A.1. The MCS Design

MCS was a disproportionately stratified sample. It was notionally a two-stage sample, first of geographical entities delineated by electoral geography, wards (or combinations of adjacent small wards). The second stage was only notionally a sample, as all infants resident in a ward at the time the sample was drawn were eligible. The wards were stratified by region, with selection proportional to size, but with disproportionate sampling probabilities across strata. The boosts (or oversample), were in the smaller countries of the UK, in each country from wards with high child poverty and in England from wards with high concentrations of ethnic minorities. This made nine strata. This complex design differs from a simple random sample in so far as children in each stratum had a different probability of being in the sample. Their distribution into clusters increased the estimated sampling error.

Target sizes were set for the total sample in each of the four countries. Within each, sampling was based on random (though disproportionate) selection of wards, the primary sampling units within which all families with a child in the target age were eligible, representative of all children living in the UK at age 9 months. All children in the national birth cohort had a known positive chance of selection, though this chance varied by stratum. Disadvantage was defined by a cut-off on an area-based index of children in families claiming means tested benefits. This cut-off represented approximately the poorest quarter of wards.

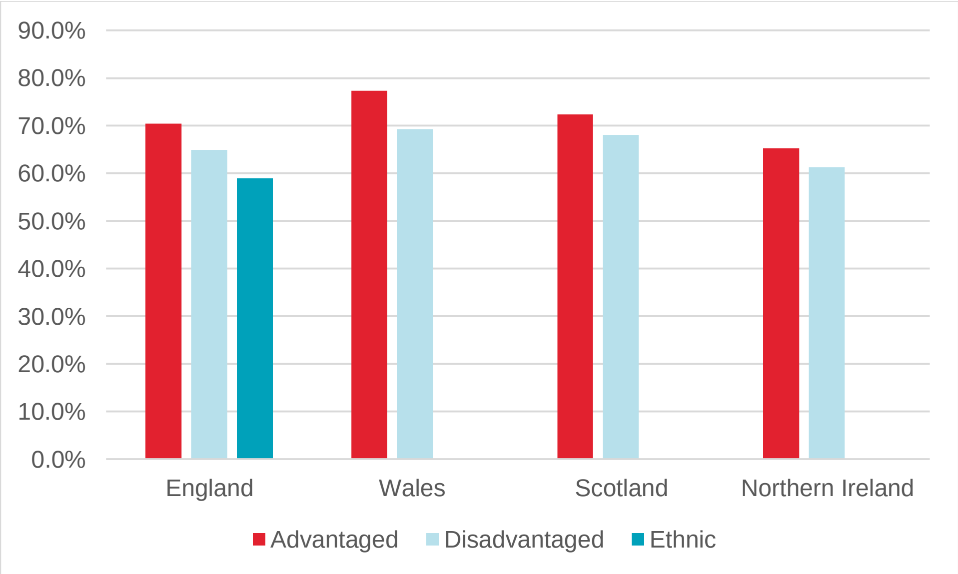
Beyond England, the prevalence of ethnic minorities was too low to attempt to over-sample them. Failing any listing of the infant population by ethnicity, ethnic minorities in England were detected from area-based information - the proportion of the ward population that was either black or Asian as far back as at the 1991 census. Such geographical concentrations of ethnic minorities was also regionally concentrated: 11 out of the 19 selected ‘ethnic’ wards in England were in London, 11 out of the 29 wards selected in London were in the ethnic stratum (Plewis 2007 table 5.2). Although the ‘ethnic’ wards were selected before the other wards in England, in practice those selected would also all have qualified as having high Child Poverty. .

.

A2. Initial response

Response from the initial sample in the 398 selected wards is shown in Figure A.1. The denominator is 28,926 families (including the New Families sample who had eligible addresses at MCS1 but were not interviewed until MCS2). Response counts as ever giving at least one interview in the first wav (MCS1) or as a New Family in England in MCS2), n=19,243. These are referred to below as the 'entry sample'. An obvious feature of Figure A.1 is that the 'ethnic' stratum yielded the lowest initial response rate. Disadvantaged wards had lower response than the non-disadvantaged, Northern Ireland had lower response, and Wales the highest.

Figure A.1 Percentage ever responding to MCS out of total initial sample, by stratum



Note to Figure A.1 Counts families with a cohort child only once even if they had twins or triplets. It includes some who turned out to be certainly or possibly ineligible after the names had been drawn, largely because of mobility , see MCS 3-7 Technical Report on Response

The nine strata were differentially sampled in proportion to the design weights shown in the first row of Table A., which reflect the ratio of the sampling fraction in a particular sample to the overall fraction of wards sampled. Only the two non-ethnic strata in England have weights over 1, effectively under-sampled. Among the others, the smaller the weight, the greater the degree of oversampling compared to a simple 'random' sample, with the most over-representation in the England Ethnic wards. The absolute numbers of cases by which the achieved (unweighted) sample exceeds and estimate of those that would have been interviewed with random sampling of wards is shown on the fourth row.

The positive entries can be thought of as the effective boost relative to simple random sampling. The England Ethnic sample is estimated to be 1632 more numerous than they would have been had there only been the average chance of their wards being selected. Northern Ireland and Wales (particularly Disadvantaged) also have a substantial over-representation, reflecting the intention to have more than proportionate sample numbers in these three countries, supported by the funding of additional sample boosts, referred to above. Table A1 is structured to show all 9 strata separately in the upper panel, with aggregations to the four constituent countries and a global total for UK in the lower panel.

Table A.1 Structure of the achieved sample

All families	England Advantaged	England Disadvantaged	England Ethnic	Wales Advantaged	Wales Disadvantaged	Scotland Advantaged	Scotland Disadvantaged	Northern Ireland Advantaged	Northern Ireland Disadvantaged
Sample design weight (UK)WEIGHT2	2	1.09	0.37	0.62	0.23	0.93	0.57	0.47	0.25
Unweighted ns families ever interviewed	4828	4805	2591	832	1928	1145	1191	723	1200
Reweighted by at MCS1/2 under equal UK probability by ward	9656	5237	959	516	443	1065	679	340	300
Over/ under sample, n (MCS1/2)	-4828	-432	1632	316	1485	80	512	383	900
As above, aggregated by country			**England**	**Wales**	**Scotland**	**Northern Ireland**		**UK**	
Unweighted ns families ever interviewed			12224	2760	2336	1923		19243	
Reweighted by at MCS1/2 under equal UK probability by ward			15852	959	1744	640		19195	
Over/ under sample, n (MCS1/2)			-3628	1801	592	1283		48	

Table A.2 Cohort members classified as non-white at Entry (MCS1/2)

All families	England Advantaged	England Disadvantaged	England Ethnic	Wales Advantaged	Wales Disadvantaged	Scotland Advantaged	Scotland Disadvantaged	Northern Ireland Advantaged	Northern Ireland Disadvantaged
Unweighted ns ever interviewed MCS 1/2)	343	807	2145	18	89	27	35	5	6
Reweighted under equal UK probability by ward	686	880	794	11	20	25	20	2	2
over/ under sample, n (MCS 1or2)	-343	-73	1351	7	69	2	15	3	5
As above, aggregated by country			**England**	**Wales**	**Scotland**	**Northern Ireland**	**England & Wales**	**UK**	
Unweighted ns families ever interviewed			3295	107	62	11		3475	
Reweighted by at MCS 1/2 under equal UK probability by ward			2359	32	45	4		2440	
Over/ under sample, n (MCS1/2)			936	75	17	7		1035	
% non white MCS 1/2			14.9 [12.5-17.7]	3.3 [2.3-4.8]	2.6 [1.9-3.5]	0.6 [0.3-1.3]	14.1 [11.9-16.9]	12.7 [10.7-15.1]	
External Estimates 2001			15.49*	3.8*	3.5**	1.3***	14.9*	13.3****	

Note: Ethnicity is based on the parent's report of the child's ethnic group. *Mid-year population estimates age zero, 2001; ** Census of Scotland 2001, dependent children in households age 0-2; .***Northern Ireland Census, 2001, Table T37, dependent children aged 0-2 in households; Based on n = 19218m known ethnicity.

A3. Sampling of Ethnic Minorities

The area-based approach was not ideal for finding concentrations of ethnic groups which had recently arrived, or individuals in any ethnic groups who did not live in concentrations of the major minorities. Furthermore it was not sensitive to the age structure of the presumed 'minority' population. Table A. displays the numbers of ethnic minority individuals (i.e. families with a non-white cohort child) achieved in the entry sample. There were 3,475 such cases observed, across the UK. 2,145 of whom (61.7%) were in the 'ethnic' stratum. They were 82.3% of the families recruited in that stratum, where they were over-represented by 1,351 cases (compared to a purely notional alternative of all the wards in UK having had the same chance of selection). However their overrepresentation in the whole UK sample, 1035, reflects an offset by the non-negligible numbers of ethnic minorities sampled in the other strata in England which were effectively under-sampled compared to the rest of UK. There were relatively few ethnic minority families sampled in the other three countries. MCS results were not compared against external data at the time, but we can see here, in the last row of Table A.2, that the estimated proportion of the population of infants belonging to a non-white minority was reasonably close to the best available external estimates. For the UK as whole the estimated percentage non-white was 12.7% compared to our estimate from official sources of 13.3%. For England and Wales, where there is an estimate of the population under age 1 in mid-2001, the MCS has 14.2% versus 14.9% in the mid-year estimates.

The concentration of ethnic minority individuals in the ethnic stratum and the unusually large populations of those wards have implications for the corrections needed to estimates of variance in a clustered sample and getting correct standard errors (see Plewis 2007, Appendix 2). These design factors are much greater for estimates involving ethnic group than most other variables in the study, which adds to the complexity facing the analyst.

Differential sampling by area of residence had limited success at locating non-white respondents. 62 % of the non-white children sampled lived in the target 'ethnic wards', which means 38% did not. When the sample is weighted to allow for differential chances of being selected, the estimate is that two thirds of non-white respondents did not live in 'ethnic' wards. According to reweighted estimates, 28% of the non-white children lived in Advantaged Wards of England, they were mainly Indian and mixed heritage, only 14% of the Pakistani, Bangladeshi or black children were estimated to live in such places. In practice the sample size for most disaggregated ethnic groups was small. Table A.3 3 shows that the sample sizes at entry, despite the over-sampling only exceeded 400 for Indian, Pakistani and black African groups. The 594 cases in the 'mixed' group are arguably not a meaningful group for most purposes. The 397 Bangladeshi families were heavily concentrated in the ethnic stratum as were the Pakistanis. The Black groups were spread more widely over the strata. The unweighted sample size for black Caribbean groups was less than the weighted estimate, indicating that those responding were not concentrated in the stratum intended to find them. It is also worth noting that there were many other ethnic minority groups in England and Wales at that time who do not show up in these tables – Non British/Irish whites such as immigrants from Eastern Europe, Arabs, Chinese for example. Some of these might have been picked in a less geographically based sampling scheme- had one been possible.

The weighted N's in the second row of Table A.3 Table A. show how many cases would have been encountered in the hypothetical simple random sample. Only Pakistani and Mixed groups would have attained more than 400 cases.

The MCS weighted estimates of the distribution of the MCS cohort across broad ethnic groups is compared with external evidence for England and Wales in the last two rows of Table A.. They are broadly similar. The national estimate of non-white infants, 14.9 %, is close to the 14.2% for MCS, well within its confidence band of 11.9- 16.9. It confirms Pakistanis as the most numerous ethnic group among infants in 2001, and also suggesting MCS may have a low response rate among them.

Table A.3 Major ethnic groups in the entry sample

	White	Mixed	Indian	Pakistani	Bangladesh	Black Caribbean	Black African	Other	All
UK									
MCS									
unweighted	15,743	594	497	954	396	259	419	356	19,218
weighted	16,730	607	343	562	185	187	287	270	19,171
weighted %	87.3 (84.9-89.3)	3.17 (2.7- 3.7)	1.8 (1.4-2.4)	2.93 (1.9- 4.6)	0.97 (0.6-1.6)	0.97 (0.6- 1.5)	1.5 (1.0-2.2)	1.4 (1.0-1.9)	100
England and Wales									
MCS									
unweighted	11,561	576	485	932	396	253	415	345	14,963
weighted	14,399	595	335	547	185	182	284	263	16,790
weighted %	85.8 (83.1-88.1)	3.6 (3.1-4.1)	2.0 (1.5- 2.7)	3.3 (2.1- 5.1)	1.1 (0.7-1.8)	1.1 (0.7- 1.7)	1.7 (1.1- 2.5)	1.6 (1.1-2.2)	100
External	**Mid- year pop estimates**								
	85.1%	4.0%	2.2%	4.1%	1.2%	1.0%	1.7%	1.8%	100

Figure A.2 Ethnic Group of Infants in England and Wales, 2001-2017

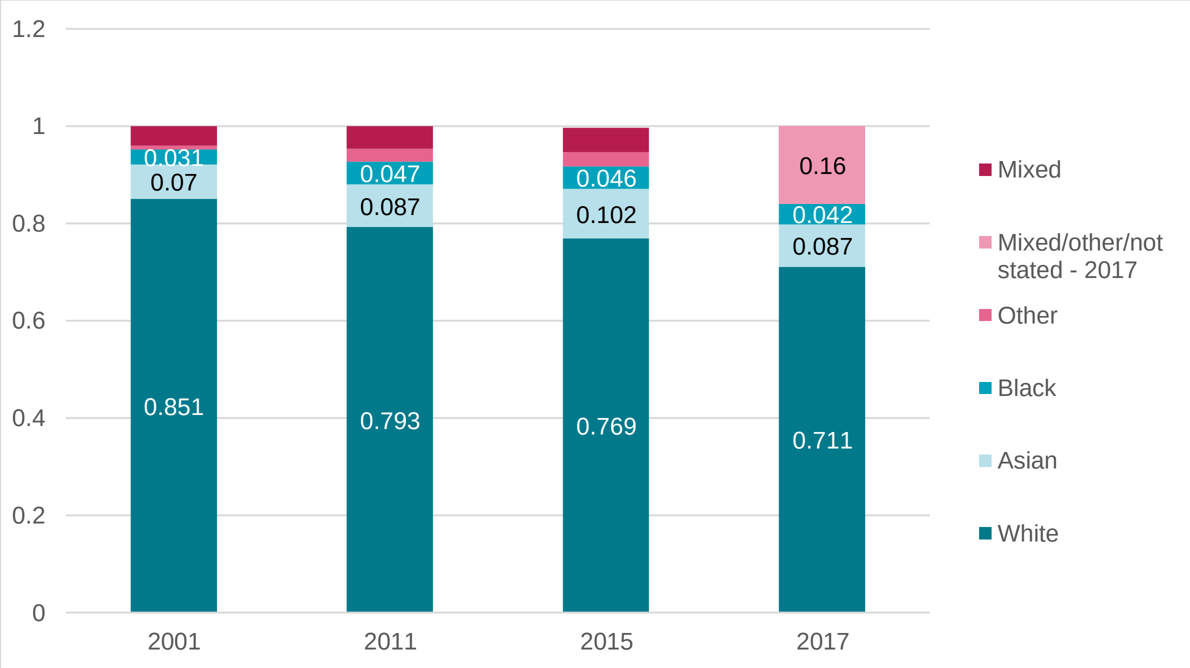


Figure A.2 shows the growing proportion of young children who belong to non-white ethnic groups (as grouped in recent statistics from the Current Population Survey). The available data on infants cover England and Wales only, the proportion white in UK would be slightly higher The data for 2001 are the mid-year population estimates for persons under the age of 1, ONS LAD tables. Those for 2011 and 2015 are based on the Current Population Survey. Those for 2017 from ONS Births Statistics are based on births in that year, according the ethnicity of the newborn reported by the midwife to NHS Birth Notifications (where some of the not stated cases may be white) England and Wales. Thus they are not completely comparable, nor up to date, but trend is clear. At face value the proportion of infants who were not classed as white appears to have nearly doubled from 15% in 2001 to 29% in 2017. By the time any new birth cohort is selected, the profile of particular minorities will have changed as will the options for selecting them, but taken as a whole, non-white groups will not be in as small a minority as they were in 2001. A cross section of primary school age children, born sometime between 2011 and 2015 would also have a smaller white majority among those born in UK than were born in 2001 and could also include recent immigrants.

A4. Over-sampling children in low income families

To assess the over-representation of low income families in the entry sample, Table A4 takes the numbers reporting net household income in the bottom quintile of the national distribution (at either MCS1 or MC2 for New Families), and comparing weighted and unweighted estimates. In UK as a whole there were 4877 low income families in the unweighted data, 1150 more than when re-weighted. There was over-representation of low income families in all four countries when their component strata are aggregated, but not in the two non-ethnic strata in England.

Given the definition of low income as the bottom fifth, around 20% of the cases were in the low income group in all countries when the sample is re-weighted. Approximately one tenth of cases in the Advantaged strata were in the low income group and three tenths in the Disadvantaged Strata, taken as a whole and including the Ethnic wards.

Taking the weighted estimates as population prevalence, just over half the 'poor' were in disadvantaged or ethnic areas, one third of the families in these areas had low income, compared to the expected 20% nationally. The oversampling of Disadvantaged and Ethnic wards 'yielded' 83 percent of respondents in those wards who were on low income, but the re-weighted estimates put one third of low income families outside such areas, and, along with their neighbours were not subject to boosted chances of inclusion. Thus targeting areas is only partially successfully at pinpointing poor families.

Table A.4 Families with low income in the entry sample

	unweighted	Weighted
Percentage on Bottom Quintile Low income		
UK	25.5%	19.5%
England	25.4%	19.1%
Wales	28.5%	23.7%
Scotland	21.1%	18.8%
Northern Ireland	25.6%	22.0%
Disadvantaged and Ethnic Strata	34.6%	33.0%
Advantaged Strata	10.9%	10.5%
Proportion of low income cases in Disadvantaged or Ethnic wards	83.2%	67.4%
Proportion on non-low income in Advantaged wards	89.1%	81.3%
Proportion of all respondents in Disadvantaged or Ethnic wards	60.9%	39.7%
Number with income data reported or imputed	19107	19102

A.5 Oversampling Overall

Taking the two populations of interest – ethnic minorities and poor families, how much did the MCS sample design augment the cases available for analysis? On the argument that reweighted sample numbers reflect what might have been collected under, hypothetical simple random sampling, the results suggest that the strategy boosted each type of response by around one thousand cases (1,035 non-white cohort members, and 1,150 families on low income. On the hypothetical scenario of Simple Random Sampling, analysts would have had to make do with 2,440 ethnic minority cases and 3,727 low income cases, rather than 3,445 and 4,877 respectively.

In a future sample, if the proportion of ethnic minority births had doubled, a simple random sample yielding the same number of total births as MCS would anyway have double the number of ethnic minority sample members – more than was achieved by differential sampling at MCS. This might be 'adequate' for studies of non-whites taken as a whole, but still not big enough for many individual groups. How much a future sample would need to boost cases on low income needs a forecast of what level of poverty might need to be recognised. Clearly there will always be a bottom quintile, and they might always have a high propensity to drop out, as discussed in the next section.

A6. Attrition after Entry

One of the reasons efforts were made to over-represent these two sub populations was the concern that they might be especially prone to survey loss through attrition. We have already seen in Figure A.1 that residents of the 'ethnic wards' were less likely to enter the study in the first place. How far did the ethnic boost counteract differential attrition once a family had been recruited? From that point on we have evidence of actual ethnicity. Table A.5 takes a long-term overview, showing the percentages of the whole sample and of the subsamples of interest who were in contact with the fourth and sixth sweeps of the survey when the cohort was aged around 7 and 14. Overall, the numbers who were productive fell from 19,243 to 13, 857 at MCS4 and 11, 726, reflecting attrition of 28% and 39% respectively. Though referred to for shorthand as 'dropout', attrition was not entirely permanent. Some 'dropouts' dropped back in at later waves. Looking first across the early sweeps, up to Wave 4/age 7, the net loss from the ethnic stratum (35.4%) was, as expected, the highest of all strata. However by the time of the sixth, age 14, sweep the non-continuation rate in the ethnic stratum (39.7%) was no longer the highest, it was closer to the UK average of 39.1% The highest dropout rates to age 14 were in the Disadvantaged wards of Scotland, Northern Ireland and Wales (in that order.) This suggests that, at least in the long run, ethnic minorities were not particularly more prone to drop out of a longitudinal survey than other groups, on whom attention might also be fixed. At the age 14 survey the dropout rate by individual 'non-white' families in UK as a whole, 39.4 %, was very close to the overall rate of 39.1%.

Table A.5 Percentage of entry families not in productive contact at ages 7 and 14 (MCS4 and MCS6)

	England Advantaged	England Disadvantaged	England Ethnic	Wales Advantaged	Wales Disadvantaged	Scotland Advantaged	Scotland Disadvantaged	Northern Ireland Advantaged	Northern Ireland Disadvantaged
All MCS Families									
Age 7	21.5%	29.7%	35.4%	25.2%	27.6%	27.6%	32.9%	26.1%	30.2%
Age 14	32.9%	40.1%	39.7%	34.9%	41.5%	39.9%	51.7%	36.7%	45.2%
For cm's classified as non-white (at entry)									
Age 7	29.4%	34.4%	35.7%	50.0%	43.8%	40.7%	31.4%	0.0%	33.3%
Age 14	38.2%	41.0%	38.1%	44.4%	50.6%	63.0%	42.9%	20.0%	50.0%
For families classed in bottom quintile income at entry									
Age 7	29.4%	34.4%	35.7%	50.0%	43.8%	40.7%	31.4%	0.0%	33.3%
Age 14	48.8%	48.4%	41.0%	43.3%	51.4%	58.6%	64.2%	37.5%	49.4%

	England	Wales	Scotland	Northern Ireland	UK
All MCS Families					
Age 7	27.7%	26.9%	30.3%	28.7%	28.0%
Age 14	37.2%	39.5%	45.9%	42.0%	39.1%
For cm's classified as non-white (at entry)					
Age 7	34.7%	44.9%	35.5%	18.2%	35.0%
Age 14	38.8%	49.5%	51.6%	36.4%	39.4%
For families classed in bottom quintile income at entry					
Age 7	34.7%	44.9%	35.5%	18.2%	35.0%
Age 14	45.8%	50.1%	62.8%	47.5%	48.4%

Note: percentages (%) unweighted

If we consider individual families who are classified as having low income in the their first survey we see that this is a predictor of attrition which persists into the adolescent stage, by which time 48% had not responded compared with 39% overall (or 36% of ‘non-poor’). The excess attrition of the low income families was seen across all strata, but to a lesser extent in the England Ethnic stratum and Northern Ireland. Low income families in ‘Advantaged’ wards in England had as high rates of dropout as individual low income families in Disadvantaged wards. This suggests oversampling by area is not a completely satisfactory way of anticipating the dropout of low income individuals. Indeed it might be better to develop ways of retaining the participation of those who have been recruited (possibly including incentives) than undertaking a strategy of over –sampling to have ‘spare’ cases.

A7. Geographical Dispersion

By age 5 around half of the MCS families had moved home.

Table A. reports the distances between address of interview at ages 7, and 14 and that of entry interview, MCS1 (or MCS2 for New Families). It is defined in terms of kilometers by road from the original postcode, rather than any particular set of administrative or statistical boundaries. It gives some idea of the distances to be covered in fieldwork. This measure is defined only for families who took part in the later surveys. The shortest distance, within half a kilometre, is taken to be within the same 'neighbourhood', if not at the same address, which it also includes. Around half of these cases were thus still at, or very close to, the original point at age 7, while conversely around one half might not have had continued exposure to the original geographical and social cluster. This degree of proximity only applied to 41% of those responding at age 14, ie nearly 60% of those still in contact had moved further afield. The distances derived for age 11 (not shown) were more than half way between age 7 and age 14, but the time interval was also longer. These figures understate the numbers who dispersed away from their original addresses, in that they do not include non-respondents, among whom movers are known to be over-represented but they are also not available for other analyses.

At all these surveys, the distances covered were mainly moderate. Long distances, over 25km were not common, only reaching 10% at age 14 years. This suggests that a somewhat less tight clustering of follow up field work would have found many respondents within a somewhat wider localities than an electoral ward, and where there would have been more continuity of residence. For example by age 14, 72% of the participating families lived within five kilometers of their original address. Although the sample did not retain its original clustering most responding families remained within some sort of striking distance of their entry address.

Table A.6 Distance between entry address and productive interviews at MCS4 and MCS6

	Entry to MCS 4: age 7		Entry to MCS 6: age 14	
Distance by road	**N in UK, unweighted**	**percent with known distance**	**N in UK, unweighted**	**percent with known distance**
0 - 500m	7031	51.30	4688	40.90
0.5 - 2km	1926	14.05	1730	15.09
2 - 5km	1793	13.08	1822	15.90
5 - 10km	1099	8.02	1153	10.06
10 - 25km	768	5.60	904	7.89
25km +	1089	7.95	1165	10.16
Known distance	13706	100.00	11462	100.00
Distance not known	5537	percent of total not known	7781	percent of total not known
Total entry , all ever interviewed	19243	28.77	19243	40.44

Note: unknown distance if no interview or distance data missing in a few cases

The rates of dispersal were somewhat higher in (non-ethnic) disadvantaged wards of origin, particularly in England. This means that a weighted estimate of dispersal, not shown, would be somewhat lower. However, an estimate of dispersal that included the addresses issued to field but not productive would be higher.

With or without residential mobility the tight clustering did not ensure that cohort children were clustered in schools or even pre-schools. Even when they are not re-drawn, ward boundaries are permeable.

A8. Commentary

The MCS sample design had pros and cons which provide a learning opportunity in the planning of a new cohort. The tightly clustered and disproportionately stratified design offers users sufficient samples of some groups of interest, and in so doing also provides some mitigation of their higher propensity to drop out. It does this at the expense of requiring users to apply survey weights and corrections to standard errors for the clustering to all their analyses (and depending on the strategy taken for handling missing data, potentially in combination with non-response weights), which adds complexity. This may have impeded use by those not equipped to handle the complex design, such as some students, or government or third sector analysts.

The clustering into electoral wards could potentially have had fieldwork and scientific value. Given that the first round of the survey had to contact families month by month, an average of four births per ward per month would not necessarily be sufficient cases within any one ward to constitute a full interviewer load, unless the selected wards were reasonably closely located. The ideal clusters suitable to planning fieldwork would tend to have a larger radius. For example, with sampling points four times the size of the average ward, there would be 4 interviews to do per week rather than 4 per month. If there were to be a second stage of sampling rather than ‘all the births in a ward’, the geographical scope of the sampling unit would have to be even bigger. Even then, a longitudinal sample tends to lose its geographical clustering as people move home, particularly when they have young children as half of them did in MCS by age five. The fieldwork benefits of clustering by ward may thus have been short-lived.

The clustered design had another motivation, to provide evidence on the influence of neighbourhood conditions of the child's development. These ecological effects have proved elusive for several reasons and would not be investigated via sample clustering in a new survey, given advances in geo-coded supplementary data which could be linked to survey families' postcodes whether not they had moved. The survey design did not seek to observe interactions among cohort families, and for reasons of confidentiality could not have identified which neighbours were in the study. Once the children were old enough to attend childcare settings, let alone schools, cohort children were seldom grouped together, partly because of the mobility of families but also because there is not a unique school for all the children from a given set of addresses. Again, the consideration of preserving confidentiality and confidence therein are paramount for a longitudinal study. The analysis of local group and networks effects, may be better suited to a different type of less long-run study.

The findings in this Annex (particularly in respect of the clustering) strongly support the case made throughout the body of the report that it would be best to keep the design as simple as possible and as close to a SRS as possible, with an option for some boosted sampling of ethnic minorities. The anticipation of attrition should be handled in various efforts to maintain response and not rely on initial over-sampling.