Childcare and Early Years Survey of Parents: Sampling frame investigation

Research report

January 2017

Tom Huskinson and Kevin Pickering – Ipsos MORI Social Research Institute

Contents

Abbreviations

This report makes use of abbreviations throughout in the interests of readability. For reference, a list of abbreviations is presented here.

BME	Black and Minority Ethnic
CAG	Confidentiality Advisory Group (of the NHS Health Research Authority)
CAPI	Computer-Assisted Personal Interviewing
CBR	Child Benefit Register
CEYSP	Childcare and Early Years Survey of Parents
DCMS	Department for Culture, Media and Sport
DfE	Department for Education
DH	Department for Health
DWP	Department for Work and Pensions
FRS	Family Resources Survey
HICBC	High Income Child Benefit Charge
HMRC	Her Majesty's Revenue and Customs
HSCIC	Health and Social Care Information Centre
LSYPE	Longitudinal Study of Young People in England
MIDAS	Medical Research Information Service Integrated Database & Administration System
NatCen	NatCen Social Research (formerly National Centre for Social Research)
NFER	National Foundation for Educational Research
NPD	National Pupil Database
ONS	Office for National Statistics
PAF	Postcode Address File
PDS	Personal Demographics Service
PSU	Primary Sampling Units
SEND	Special Educational Needs and Disabilities

1 Executive Summary

The Childcare and Early Years Survey of Parents (CEYSP) is commissioned by the Department for Education (DfE), and measures behaviours and attitudes relating to childcare and early years education in England. The survey samples children (aged 0 to 14) from the Child Benefit Register (CBR), and each survey wave (the survey is currently carried out every two years) over 6,000 face-to-face interviews with the parents of these children are carried out.

Until 2013, the CBR enjoyed almost universal coverage of children in England. Since the introduction of the High Income Child Benefit Charge (HICBC) in 2013, however, high income families no longer derive a financial gain from Child Benefit, and so may not register their child(ren) for Child Benefit. The present study sought to investigate the impact of the HICBC on the CBR's coverage of children in England, and the associated risk of bias to survey estimates. The study also sought to investigate the suitability of alternative sampling frames, and to make a recommendation as to the most suitable sample design for future waves of the CEYSP.

The investigation found that the HICBC has led to a tendency for parents with high incomes not to register their children for Child Benefit. Specifically, the HICBC is resulting in around eight per cent of children born in England each year not appearing on the CBR. As a consequence, continuing to sample children exclusively from the CBR will mean CEYSP survey estimates will become increasingly biased, with each successive wave, away from children in high income households, and it will also be difficult, if not impossible, to attribute changes across survey waves to real changes in the population.

The suitability of the following alternative sampling frames was considered for future waves of the CEYSP: the Postcode Address File (PAF), the Medical Research Information Service Integrated Database & Administration System (MIDAS), the National Pupil Database (NPD), databases of schools, following-up respondents to another survey such as the Family Resource Survey (FRS), and commercial databases. The investigation found that MIDAS, and following-up respondents to another survey, would not be suitable due to difficulties gaining access to these sampling frames; the NPD, schools databases, and commercial databases would not be suitable due to undercoverage of the CEYSP survey population; and the PAF (on its own) would not be suitable given the extensive and costly in-field screening efforts that would be required to identify addresses with eligible children.

The investigation recommends that future waves of the CEYSP employ a split-sample approach, sampling from the CBR in those postcode sectors in England that contain a negligible proportion of high income families (where the CBR's coverage will be largely unaffected by the HICBC), and sampling from the PAF in those postcode sectors in England that contain a non-negligible proportion of high income families (where the

CBR's coverage will be affected by the HICBC). Using this approach, high coverage of the survey population can be maintained, while fieldwork inefficiencies arising from in-field screening can be minimised.

2 Introduction

2.1 Background

The Childcare and Early Years Survey of Parents (CEYSP) is a high quality cross-sectional government survey, funded by the Department for Education (DfE), which aims to track parents' use of, and experiences relating to, childcare and early years education, and to provide evidence to assist with the assessment of policies that relate to childcare and early education.

With a history stretching back over ten years and a large sample size, the CEYSP provides an invaluable resource for researchers from government, academia, the charity sector, and other backgrounds to understand in detail the changing landscape of childcare and early years education in England.

To date, the CEYSP has used the Child Benefit Register (CBR), which is held and maintained by Her Majesty's Revenue and Customs (HMRC) as the sampling frame. Children aged 0 to 14 in England are sampled using a probability sampling methodology. This methodology means that each eligible child on the sampling frame has a known chance (greater than zero) of being selected for the survey. This makes it possible to produce unbiased estimates of population parameters by weighting sampled children by their probability of selection. For each sampled child, a face-to-face interview is sought with a parent or guardian with main or shared responsibility for childcare decisions for the child. For the most recent wave of the CEYSP, undertaken in 2014-15, 6,198 interviews were conducted, resulting in a response rate of 57 per cent.

The CBR has historically enjoyed very high coverage of the CEYSP survey population; however, this coverage is now deteriorating due to the introduction in 2013 of the High Income Child Benefit Charge (HICBC). The HICBC is a tax which leads high income families (those where one or both partners earn £60,000 or more per year) to derive no financial gain from Child Benefit payments. As a result, there has been a tendency for high income families to not register their child(ren) for Child Benefit, leading to the absence of these children from the CBR, and consequently, a risk of survey results being biased away from the behaviours and experiences of parents and children in these high income households.

DfE has commissioned the present investigation to assess which sampling frame will be the most appropriate choice for future waves of the CEYSP, given the recent changes to the CBR. The present investigation draws on a previous (unpublished) report, authored by Ipsos MORI for DfE in 2014, that provided an early assessment of the implications of the HICBC for the CEYSP sampling frame.

2.2 Aims and objectives

The aims of the present investigations were to:

- investigate the suitability of the CBR for future waves of the CEYSP, given the change in coverage of the CBR;

- investigate the suitability of alternative sampling frames for the CEYSP; and

- make a recommendation as to which sampling frame is most appropriate for future waves of the CEYSP.

In order to achieve these aims, the objectives were to:

- conduct desk research, analyses, and interview staff at HMRC to assess the CBR's qualities as a sampling frame for future waves of the CEYSP;

- identify potential alternative sampling frames and conduct desk research, analyses, and interview relevant staff at organisations holding the sampling frames, in order to assess their qualities as sampling frames for future waves of the CEYSP; and

- carry out a review of surveys similar to the CEYSP in order to gain insights from existing research.

2.3 Assumptions

This investigation assumes that future waves of the CEYSP will continue to:

- use a face-to-face in-home CAPI methodology, with an average questionnaire length of around 45 minutes;

- deliver in the region of 6,350 interviews with a representative sample of parents or guardians of children aged 0 to 14 living in England (allowing for the possibility that DfE may wish to narrow the age range to children aged 0 to 12, and allowing for the fact that DfE may wish to boost children of pre-school age relative to their proportion in the population); and

- require comparisons to be made across survey waves in order to measure changes in behaviours and attitudes over time.

2.4 Structure of this report

We would like to thank Naeem Abdulhussein and Hannah Rhodes at HMRC, Clare

We start this report by describing the current sampling methodology of the CEYSP (Chapter 3). We then describe the characteristics that are desirable in the sampling

frame that is to be used for the CEYSP (Chapter 4); we do this in order to provide the relevant methodological background and context for the remainder of the report.

In Chapter 5, we discuss the suitability of each sampling frame we have identified as being worthy of consideration for future waves of the CEYSP, with reference to the desired characteristics of the sampling frame described in Chapter 4. We describe potential sample design options for future waves of the CEYSP, along with their respective advantages and disadvantages. For each sample design option, we present our estimate of the number of interviews the design would deliver for the fieldwork budget required to deliver 6,350 interviews under the existing CEYSP sample design[1]. We also describe how each sampling frame has been used on similar surveys, and we draw upon these experiences where relevant.

In Chapter 6, we present our conclusions and recommendations for future waves of the CEYSP.

2.5 Acknowledgements

We would like to thank Naeem Abdulhussein and Hannah Rhodes at HMRC, Clare Watson at the Office for National Statistics, James Gray at the Health and Social Care Information Centre, and Thomas Leach at NatCen Social Research. They have all been extremely helpful and provided valuable information for the investigation.

A number of Ipsos MORI colleagues have also contributed to the investigation and we would like to extend our thanks to Patten Smith, Sarah Tipping, Nicholas Gilby, Sam Clemens, and Julia Pye.

We are also grateful to Max Stanford at the Department for Education for his support throughout the investigation.

---

[1] It should be noted that the estimates provided are ballpark estimates, and as such are subject to change. Furthermore, these estimates are Ipsos MORI's own estimates, and may not reflect the estimates that would be arrived at by other survey contractors.

3 Current CEYSP sampling methodology

In this chapter we describe the sample design that has been implemented in the four most recent waves of the CEYSP (the 2010-11, 2011-12, 2012-13, and 2014-15 surveys)[2].

3.1 Survey population

The survey population for the CEYSP is defined as children aged 0 to 14 living in England[3].

Although children are the sampling unit, interviews are conducted with an appropriate adult, defined as an adult within the child's household with 'main or shared responsibility for making decisions about the child's childcare'. Where a sampled child is found to have moved address during fieldwork, the child is, wherever possible, traced to his or her new address, and an interview is carried out with an appropriate adult at that address.

The selection of children is important because the questionnaire collects detailed information about the sampled child (to avoid long and onerous interviews for parents with multiple children), and this information is used to derive child-level survey estimates (for example, the proportion of children in England attending a particular type of childcare provider). The questionnaire also collects information about the responding parent, and about the household. These questions are used to derive family-level survey estimates (for example, how often parents engage in home learning activities with their children, and the proportion of families that find it difficult to cover their childcare costs given their household income).

3.2 Sampling frame

The sampling frame for the CEYSP is the Child Benefit Register (CBR), held by HMRC. We describe the characteristics of the CBR, both prior to and after the introduction of the High Income Child Benefit Charge (HICBC), in section 5.1.2.

---

[2] Prior to the 2010-11 survey, the CEYSP sample design was more complicated and, while interviews still represented the population of children aged 0 to 14, a greater degree of corrective weighting was required, leading to a loss of precision and to less reliable survey estimates. Further details can be found in Section B.3 of the Childcare and Early Years Survey of Parents 2014 to 2015 report: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/516924/SFR09-2016_Childcare_and_Early_Years_Parents_Survey_2014-15_report.pdf.pdf

[3] Children living in communal establishments, for instance in a children's home or a care establishment, are excluded.

3.3 Sample design

Children are sampled using a two-stage approach: i) selection of Primary Sampling Units (PSUs), and ii) selection of individual children within each PSU.

At the first stage (selection of PSUs), HMRC creates a list of all postcode sectors in England with counts for each of the number of children on the CBR aged 0 to 14, and the number of children on the CBR aged 2 to 4[4]. To reduce clustering, postcode sectors containing fewer than 250 children are grouped with neighbouring postcode sectors. The list of grouped postcode sectors is then stratified by region, population density, proportion of households in managerial professional and intermediate occupations, and the proportion of the population that is unemployed. PSUs are then selected with probability proportionate to number of children they contain. For the 2014-15 CEYSP, 431 PSUs were selected.

At the second stage (selection of individual children within each PSU), for each selected PSU, HMRC lists all eligible children aged 0 to 14 on the CBR (sorted by postcode and Child Benefit number to help to avoid children from the same household being selected), and selects 26 children from this list. In selecting these children, HMRC uses a weighted design to boost (by a factor of 1.728[5]) the number of children aged 2 to 4 that are selected.

The benefit recipient for each sampled child is then sent an advance letter introducing the survey, and giving them at least two weeks to opt out of the research should they so wish. Those who do not opt out are visited by an interviewer, who attempts to carry out a face-to-face Computer-Assisted Personal Interview (CAPI) with an appropriate adult at the household.

To ensure children born into these households after the sample is drawn have a chance of selection, the interviewing program randomly re-selects the child that is to be the focus of the child-specific questions from among all children (including the new-born child) in these households.

---

[4] Separate figures for the number of children aged 2 to 4 is required because this age group is boosted relative to their prevalence in the population. This is done to obtain sufficient numbers of children attending early years provision to allow for separate analysis of this group.

[5] This boost factor was used in the first survey in the CEYSP series (in 2004) to boost the number of interviews achieved with parents of children aged 2 to 4 by around 900. Since 2004, the target sample size has been reduced from 7,200 to 6,350, meaning the boost results in approximately 800 additional interviews with parents of children aged 2 to 4.

3.4 Survey response

The response rate for the 2014-15 CEYSP was 57 per cent. This figure reflects the proportion of productive interviews (6,198) across all eligible addresses (10,898). Across the four most recent waves of the CEYSP, the response rate has ranged from 57 per cent (in 2010-11 and 2014-15) to 59 per cent (in 2012-13). This response rate is in line with other major government-commissioned cross-sectional social surveys. For instance, the response rate to the most recently published Family Resources Survey (2013-14), the Department for Work and Pensions' flagship survey, was 57 per cent, and the response rate to the most recently published Taking Part survey (2014-15), the Department for Culture, Media and Sport's flagship survey, was also 57 per cent.

4 Desirable characteristics of the sampling frame for the CEYSP

In this chapter we describe the characteristics the sampling frame for the CEYSP would, ideally, possess. In Chapter 5 we refer back to these characteristics in appraising the suitability of each sampling frame we consider.

4.1 Accessibility

Of fundamental importance, the sampling frame must be accessible for use on a long-term basis. Inaccessibility is most likely to arise due to one (or more) of the following issues: administrative burden or related delays, legal or data security restrictions, and cost.

4.2 Format and fields

The sampling frame should be held in electronic format and should be easy to manipulate. For instance, a sampling frame that is stored in a database that does not allow the sampling frame holder to perform the various calculations required for the selection of survey elements (for instance, calculating the total number of children of a certain age on the frame), will not be appropriate. This characteristic is of particular importance for the CEYSP because data protection legislation means it will not be permissible for an entire sampling frame containing personal details to be provided to the survey contractor, from which the contractor will select the children whose parents should be approached for interview. Rather (as occurs currently) the sampling frame holder would be required to follow a precise set of calculations and steps to select the individuals to be approached, and to send details of only these individuals to the survey contractor so that they survey can be carried out. A sampling frame that is not easily manipulated electronically would make this task difficult and error-prone, if not impossible.

The sampling frame should preferably include each sampled element (child) as a separate entry, rather than grouped within a hierarchy such as households, addresses, or families. Such grouped entries can be problematic as reformatting the sampling frame from a grouped format to an individual level format may be time-consuming, expensive, and error-prone, while selecting grouped elements may result in an inefficient and more complicated sampling strategy.

With respect to the fields contained on the sampling frame, these should include the child-level demographic data required for sample selection and the calculation of weights. Ideally, for a sample design in which a child is the selected element for a face-to-face

survey with a parent or guardian of the child, the full name of the child, the child's date of birth, and the child's gender should be available. The full address of the household (to allow for a face-to-face methodology) should also be available, and the name of the child's parent(s) or guardian(s) (to allow an advance letter to be addressed to a named individual) is preferable, but not essential.

Other fields should be included for specific purposes. If the sample design is to boost particular groups of children, relevant fields should be included on the sampling frame. For instance, if the survey design intends to boost children with special educational needs and disability (SEND), then the sampling frame should identify whether or not each child has a SEND. It should also be possible to link the elements on the sampling frame to area-level geodemographic data (such as population density, and the proportion of the population that is unemployed). This can be readily achieved by matching against postcodes. This enables sampled elements to be clustered, which is necessary for efficient face-to-face fieldwork. It also allows a range of stratifiers to be used, which increases the precision of survey estimates by ensuring that the sample drawn is representative of the population (in terms of the stratifiers used), thereby minimising the risk of drawing an atypical sample.

4.3 Coverage

Coverage describes the extent to which a sampling frame accurately includes (or covers) each eligible element of the survey population. An ideal sampling frame from a coverage perspective will list every member of the survey population once only. Problems of coverage can arise in the following ways:

- Missing elements (elements in the survey population are not present on the sampling frame). Missing elements result in undercoverage, which will bias survey estimates to the extent that those missing from the survey frame differ from those present in ways related to the topics of interest. Missing elements can arise either because they have not been added to the frame, or because they are present but incorrect (or absent) contact details means they cannot be contacted.

- Foreign elements (the frame includes elements that are not in the survey population). Foreign elements result in overcoverage, and require screening exercises in the field. This reduces the efficiency of fieldwork, increasing survey costs. The inclusion of foreign elements also has the potential to cause annoyance or alarm to those (wrongly) selected and approached for interview.

- Duplicate elements (an element in the population is recorded more than once on the sampling frame). Duplicate elements are a problem because some elements on the sampling frame will have more than one chance of being selected, biasing survey estimates towards these elements. While attempts can be made to remove

duplicate elements prior to selection, these attempts can be time-consuming, costly, and not wholly effective; and while attempts can be made to address this problem via the calculation of survey weights, this reduces the effective sample size, damaging the precision of survey estimates.

4.4 Stability

In the case of repeated cross-sectional surveys such as the CEYSP, consideration must also be given to the stability of the characteristics of prospective sampling frames over time. For instance, if a sampling frame's coverage of the survey population is expected to change over time, it will be difficult, if not impossible, to attribute changes in survey estimates across waves to real changes in the population, rather than to an artefact of the falling coverage of the sampling frame.

If a survey's primary objective is to obtain the most accurate possible population estimates for any given survey wave, the sampling frame that currently possesses the 'best' characteristics in terms of format, fields, and coverage should be chosen, regardless of whether it can be used, or will have changed, when the sampling is carried out for subsequent survey waves. If, on the other hand, a survey's primary objective is to assess whether attitudes and behaviours in a population are changing over time, then the stability of the sampling frame is of far greater consequence, and it may be sensible to prefer, for instance, a stable sampling frame with lower coverage of the survey population, over an unstable sampling frame with higher coverage of the survey population.

5 Suitability of sampling frames

In this chapter we discuss the suitability of those sampling frames from which samples of children (or parents) can be obtained, and which may therefore have potential for use on the CEYSP. These sampling frames were identified in the following ways:

> 1.   By conducting a review of the sampling frames that have been used on major cross-sectional and longitudinal UK surveys of the public (a list can be found on the UK Data Service website[6]);

> 2.   By conducting a review of the sampling frames that have been used over the past two decades on surveys that have sampled children or parents specifically. These sampling frames were identified by conducting searches of the government's www.gov.uk website on combinations of relevant keywords (including 'survey', 'study', 'evaluation', 'parent', 'parents', 'child', 'children', 'family', 'families', 'young people', 'early years' and 'early education'), and by conducting further separate internet searches using these keywords; and

> 3.   By consulting with relevant colleagues at Ipsos MORI with experience of working on large-scale social surveys, surveys that have sampled children or parents specifically, and on other sampling frame investigations.

The sampling frames we have identified are: the Child Benefit Register (CBR), the Postcode Address File (PAF), the Medical Research Information Service Integrated Database & Administration System (MIDAS), the National Pupil Database (NPD), databases of schools, respondents to another survey, and commercial databases.

For each of these sampling frames, we describe its characteristics (with reference to the desirable characteristics of the sampling frame for the CEYSP set out in Chapter 4) and the implications for the use of the sampling frame for future waves of the CEYSP. We also describe how each sampling frame has been used for other similar surveys (those identified at stage 2 above), and draw on these experiences wherever relevant.

In Table 7.1 (in the Appendix) we present a summary of the characteristics of the sampling frames we have considered.

---

[6] Available at: https://www.ukdataservice.ac.uk/get-data/key-data/

5.1 Child Benefit Register (CBR)

5.1.1. Introduction

The Child Benefit Register (CBR) is held by HMRC, and is used to make Child Benefit payments to eligible parents or guardians. Child Benefit is paid at the rate of £20.70 per week (as of 2016/17) for the eldest or only child, and at the rate of £13.70 per week for each additional child. Only one person can receive Child Benefit for a given child.

An important change was made to Child Benefit in 2013. Prior to 7 January 2013, any parent or guardian responsible for a child under 16 (or under 20 if in approved education or training) was eligible to receive Child Benefit. Child Benefit was universal and non-taxable; all families with a child or children stood to make a net financial gain if they registered to receive the benefit.

On 7 January 2013, the government introduced a tax charge called the High Income Child Benefit Charge (HICBC) which requires a Child Benefit recipient to repay some, or all, of their Child Benefit if they, or their partner, has an adjusted net income greater than £50,000 per year[7]. For each additional £100 over the £50,000 threshold that an individual earns, the tax charge due increases by one per cent, such that any recipient whose income (or partner's income) exceeds £60,000 will be liable to repay their entire Child Benefit entitlement. Claimants affected by the HICBC can elect to opt out of receiving Child Benefit, ceasing their receipt of payments altogether.

Prior to the introduction of the HICBC, around 7.9 million families were paid Child Benefit at a total cost of £12 billion[8]. The HICBC is estimated to affect around 1.2 million of these families, of whom 70 per cent are subject to the full charge, with savings estimated at around £1.5 billion a year[9].

5.1.2. Characteristics

Accessibility

Use of the CBR as a sampling frame requires permission from HMRC. A wide variety of surveys have used the CBR as a sampling frame, including every wave of the CEYSP, and the CEYSP's predecessor surveys dating back to 1997. We do not envisage any

---

[7] Adjusted net income is total taxable income before any personal allowances, and less certain tax reliefs such as trading losses, donations made to charities through Gift Aid, pension contributions paid gross (before tax relief), and pension contributions where the pension provider has already provided tax relief at the basic rate.

[8] Available at: http://www.parliament.uk/briefing-papers/SN06299.pdf

[9] *Ibid*

barriers to accessing the CBR for use as the sampling frame for future waves of the CEYSP.

Format and fields

The CBR is held in electronic format, can be readily manipulated, and has a long history of successful use as the sampling frame for surveys with complex probability-based sample designs.

The CBR contains one entry per child, and includes the child's full name, date of birth, gender, resident address, the benefit recipient's name, and details of other children in the household for whom Child Benefit is also claimed. Geodemographic information can be linked to the CBR via postcodes, allowing for stratification prior to selection. This information is sufficient for the CEYSP's two-stage random probability sample design[10]. The CBR does not identify other child-level characteristics (such as whether a child has a SEND, or the child's ethnicity) precluding boosts of such groups at the sampling stage.

Coverage

Prior to the introduction of the HICBC in January 2013, Child Benefit was a universal, non-taxable benefit, and published HMRC estimates show it enjoyed a high take-up rate of around 96 per cent of eligible children and young people across the UK[11].

This official take-up rate includes children in Wales, Scotland and Northern Ireland, as well as young people aged 16 to 19 in approved education or training, and may not, therefore, be an accurate estimate for take-up among the CEYSP survey population: children aged 0 to 14 in England.

To derive an estimate of the CBR's coverage of the CEYSP survey population prior to the introduction of the HICBC, we compared the number of children aged 0 to 14 in England for whom child benefit was being claimed in both 2011 and 2012 (from published HMRC

---

[10] A two-stage random probability design refers to a sample design where primary sampling units (PSUs) are selected at the first stage, and survey elements are selected at the second stage. This is usually done in order to limit fieldwork costs. For the CEYSP, the first stage refers to the selection of postcode sectors (with probability of selection proportionate to the number of resident children), and the second stage refers to the random selection of 26 children within each postcode sector.

[11] Available at:
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/502601/cwtcchb-take-up2011-12-Corrected-1602-V1.pdf

data[12]), against the relevant mid-year population estimates of the number of children aged 0 to 14 in England (from published ONS data[13]).

This analysis shows that the CBR's coverage of the CEYSP survey population prior to the introduction of the HICBC was 98 per cent (98.0 per cent in 2012, and 98.3 per cent in 2011).

In practical terms, the CBR's coverage of the CEYSP survey population is ultimately lower than 98 per cent, because of the following missing elements:

1.   Children who are their parents' first child, and who are born between the date the sample is drawn from the CBR and the start of fieldwork (a duration of approximately five months[14]). These children have no chance of selection. We estimate that these children account for around one and a half per cent of the survey population. It should be noted that children born after the sample is drawn into households already containing a child or children do have a chance of selection. Specifically, the CEYSP questionnaire begins with a 'household grid' section, via which the interviewer records details of each household member (name, date of birth, gender, and relationship to other household members). Where a household member has been born after the cut-off date for the version of the CBR used as the sampling frame, the interviewing program randomly re-selects the child that is to be the focus of the child-specific questions from among all children (including the new-born child) in these households.

2.   Children excluded by HMRC prior to drawing the sample from the CBR. These children fall into the following groups: i) those taken into care or put up for adoption, ii) those not living at the same address as the claimant, and iii) those who are subject of correspondence between the benefit recipient and the Child Benefit Centre (because the reason for correspondence cannot be ascertained and may be sensitive). These exclusions constitute approximately one half of one per cent of children on the CBR. To account for such exclusions (as well as for errors on the CBR), where an interviewer finds that the number of children in the household is greater than the number of children recorded on the CBR (excluding new births), and Child Benefit is found to be received for some but not all children in the household, the interviewing programme re-selects the child that is to be the

---

[12] Available at: https://www.gov.uk/government/collections/child-benefit-geographical-statistics#history

[13] Available at:
https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates

[14] Updates to the CBR are released quarterly. Typically, fieldwork commencing in October will be drawn from the May release of the CBR, meaning that the youngest sampled child on commencement of fieldwork will be around five months old.

focus of the child-specific questions from all children in the household. This re-selection is likely to account for some, but not all, of the HMRC exclusions.

3. Children who have moved address, and whose new address cannot be traced by the interviewer. Recent waves of the CEYSP show that these children account for around 11 per cent of children on the CBR.

As a result of these missing elements, the CBR's de facto coverage of the CEYSP survey population, prior to the introduction of the HICBC, is likely to have been around 86 per cent.

Despite this imperfect coverage, the CBR has been the preferred sampling frame for the CEYSP due to the efficiency with which fieldwork can be conducted given that it enables direct contact with parents. Prior to the introduction of the HICBC the undercoverage of the CEYSP has been of a nature that has not raised any serious concerns about the introduction of systematic bias into survey estimates[15].

We now describe how the introduction of the HICBC has reduced the CBR's coverage of the CEYSP survey population, and how we might expect the nature of this reduced coverage to change over the coming years. Coverage has decreased, and can be expected to decrease further over time, via two processes: i) existing Child Benefit recipients electing to stop receiving Child Benefit (opt outs), and ii) families with new children not claiming Child Benefit (non-claimants).
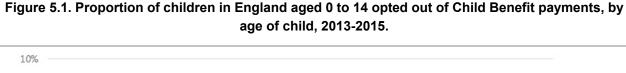
Opt outs

In 2013, HMRC introduced the option for families already in receipt of Child Benefit to elect to stop receiving Child Benefit altogether, by contacting the Child Benefit Office directly (referred to as 'opting out'). Opting out is attractive to families that are subject to the full HICBC as it spares them the task of declaring the Child Benefit they have received (and must pay back in full) via a self-assessment tax return.

---

[15] The CEYSP does not collect data from children and families that have moved and who cannot be traced (movers), so it is not possible to compare the characteristics of movers to those of non-movers in terms of survey estimates. One can, however, compare the initial sample profiles of movers and non-movers, both in terms of address-level information recorded on the CBR (for instance, the number of children in the household for whom Child Benefit is claimed), and in terms of postcode sector-level geodemographic information from the 2011 Census (for instance, the local unemployment rate). This analysis, conducted for the present investigation, finds a similar profile between mover and non-mover addresses for the 2014-15 CEYSP sample in terms of: the number of children in the household for whom Child Benefit is claimed (2.0 vs 2.1 respectively); the population density of the local area (36.8 vs 33.2 persons per hectare); the proportion of the local resident population in managerial professional and intermediate occupations (45.7% vs 46.1%); and the local unemployment rate (3.5% vs 3.4%).

Families opting out remain on the CBR and can continue to be sampled for surveys; there is precedent for this approach from the CANparent trial evaluation (2014), which we describe in section 5.1.3. However, opt outs are likely to lead to a decrease in coverage because families opting out are (in the opinion of the HMRC staff member to whom we spoke) less likely to inform HMRC of a change of address when they move[16].

HMRC publish detailed annual estimates of the number of children for whom Child Benefit is claimed (including, since 2013, the numbers that have been opted out)[17]. Combining these data with the relevant mid-year population estimates for England by age, published by ONS, indicates that in 2013, 5.1 per cent of children aged 0 to 14 had been opted-out, rising to 6.0 per cent in 2014, and to 6.2 per cent in 2015.

Figure 5.1 illustrates how the proportion of children that have been opted out varies by the age of the child, for each of the three years for which data is available (2013 to 2015).

Figure 5.1. Proportion of children in England aged 0 to 14 opted out of Child Benefit payments, by age of child, 2013-2015.

[16] It should also be noted that new children born into these families are unlikely to have a claim for Child Benefit made on their behalf, as there is no financial incentive for doing so; however, these missing children can be identified and given a chance of selection in-field.

[17] Available at: https://www.gov.uk/government/collections/child-benefit-geographical-statistics#history

The data in Figure 5.1 show that, among children born after the introduction of the HICBC (those aged 0 in 2013, those aged 0 to 1 in 2014, and those aged 0 to 2 in 2015), only around one to two per cent are being opted out of Child Benefit. This is far lower than the total proportion of children aged 0 to 14 that have been opted out of child benefit, which stands at 6.2 per cent in 2015.

This difference is consistent with a tendency for children in high income households born after the introduction of the HICBC to not have a claim for Child Benefit made on their behalf in the first place (non-registration), compared to a tendency for children in high income households born before the introduction of the HICBC to already be in receipt of Child Benefit, and to be opted out by their parents. Extrapolating this trend (and assuming no further changes are made to the administration of Child Benefit), we would expect the approximate proportion of children aged 0 to 14 that are opted out of Child Benefit to be a uniform one per cent across all children aged 0 to 14 by 2027, when children born in 2013 will be aged 14.

We now turn to a direct analysis of the proportion of children that are not being registered for Child Benefit.

Non-registration

High income households with a new-born child who are not already in receipt of Child Benefit, and who therefore do not appear on the CBR, may decide not to complete a Child Benefit form for their new-born child, as they do not stand to make a net financial gain from doing so. This is in spite of explicit advice from HMRC that new-born children should be registered for Child Benefit regardless of the family's income, and regardless of whether or not the family intends to opt out of receiving Child Benefit payments[18].

To estimate directly the proportion of new-born children that are not being registered due to the HICBC, and who will therefore not appear on the CBR, we compared HMRC's annual figures of the number of children for whom Child Benefit is claimed for 2012 to 2015 (including those children that have been opted out) against the relevant mid-year population estimates for England by age, published by ONS. These data are presented in Table 5.1.

---

[18] The Child Benefit claim form (CH2) states 'If you or your partner have an individual income of more than £60,000, the tax charge will be equal to the total amount of Child Benefit so you might want to stop getting payments and not have to pay a tax charge. But it is really important to fill in the Child Benefit form if you have a new child in your family. Filling in the Child Benefit claim form ensures you are registered to receive National Insurance credits which can help to protect your State Pension and help your child get their National Insurance number.' Available at:
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/501129/CH2___CH3_combined_for_web_English.pdf

Table 5.1. Proportion of children in England present on the CBR (including opt outs), by age of child, 2012-2015

Age of child	2012	2013	2014	2015
	(%)	(%)	(%)	(%)
0	92.9	88.0	84.7	84.6
1	97.1	94.0	89.5	87.3
2	97.9	97.4	94.0	89.9
3	98.0	98.0	97.2	93.9
4	98.4	98.4	98.1	97.3
5	98.4	98.2	97.9	97.6

Turning first to children aged 0 (the first row in Table 5.1), coverage has fallen from 92.9 per cent in 2012 (before the introduction of the HICBC), to 84.6 per cent in 2015 (after the introduction of the HICBC), a fall of 8.3 percentage points. It should be noted that the coverage of children under 1 has historically been lower than the coverage for older children, due to the lag between a child's birth and a claim for their Child Benefit being processed.

Turning to children aged 1, coverage has fallen from 97.1 per cent in 2012 (before the introduction of the HICBC; children aged 1 in 2012 will have been born in 2011), to 87.3 per cent in 2015 (after the introduction of the HICBC; children aged 1 in 2015 will have been born in 2014), a fall of 9.8 percentage points.

Turning to children aged 2, coverage has fallen from 97.9 per cent in 2012 (before the introduction of the HICBC; children aged 2 in 2012 will have been born in 2010) to 89.9 per cent in 2015 (after the introduction of the HICBC; children aged 2 in 2015 will have been born in 2013), a fall of 7.9 percentage points.

The coverage estimates for children aged 3 to 5 are less instructive in understanding the impact of the HICBC, as all of these children will have been born before the introduction of the HICBC. It should be noted, however, that there has been a four percentage point fall in coverage between 2012 and 2015 for children aged 3. This may reflect children born in 2012 (who would have been aged 3 in 2015) not being registered for Child Benefit due to parents' knowledge of the imminent introduction of the HICBC. Coverage for children aged 4 and 5 is very high, reflecting the historical high coverage of the CBR.

Taken together, these data suggest that the HICBC is resulting in around eight per cent of children born in England each year not appearing on the CBR.

Conclusions about the impact of the HICBC on the CBR's coverage of the CEYSP survey population

Prior to the introduction of the HICBC in January 2013 the CBR had very high coverage (around 98 per cent) of the CEYSP survey population, and even though some children were ultimately not covered in the field (primarily those who had moved address and who could not be traced) the nature of this undercoverage did not raise any serious concerns about the introduction of systematic bias into survey estimates.

The introduction of the HICBC has led to a tendency, since 2013, for parents with higher incomes not to register their children for Child Benefit. We estimate that this has increased the CBR's undercoverage of the CEYSP survey population by around two percentage points overall, from 98 per cent to 96 per cent. What is of concern, however, is not this absolute fall in coverage, but that:

1.   children in high income households can be expected to differ from those in lower income households in ways relevant to survey estimates. For instance, the 2014-15 CEYSP found that children in households with higher incomes were more likely to: use both formal and informal childcare; use a greater number of hours of formal childcare (among those using formal childcare); have a greater amount spent on their childcare per week; have parents who found it easy to cover childcare costs; and receive the entitlement to government funded early education[19].

2.   the extent of the undercoverage is not uniform, but varies by the age of the child. Around 98 per cent of children born before January 2013 will appear on the CBR, compared to around 90 per cent of children born after January 2013. As a result, the sample profile of children is becoming progressively skewed away from children in high income households with each successive year. This trend will persist until 2027 (assuming no further changes are made to the administration of Child Benefit), when children born in 2013 will be aged 14, and the coverage will be a uniform 90 per cent across children aged 0 to 14.

The combination of these two factors means that if the current CEYSP sample design is retained, survey estimates will become increasingly biased with each successive year away from children in high income households. More consequentially, it will be very difficult, if not impossible, for researchers to be confident that changes in survey estimates across waves are due to real changes in the population, rather than to an artefact of the falling coverage of the sampling frame. These difficulties will be particularly acute when making comparisons between age groups that differentiate between children born before and after the introduction of the HICBC.

---

[19] Available at: https://www.gov.uk/government/statistics/childcare-and-early-years-survey-of-parents-2014-to-2015

Stability

We are not aware of any changes to the eligibility criteria of Child Benefit in the foreseeable future, and the analyst to whom we spoke at HMRC confirmed that there are no current plans to change the eligibility criteria. It is of course open to this and future governments to vary the eligibility criteria for Child Benefit, and this could either restore the efficacy of the Child Benefit Register as a sampling frame, or could accelerate the decline in its suitability.

5.1.3. Use of the CBR on similar surveys

We describe in this section recent large-scale social surveys (other than the CEYSP) in which parents or children have been sampled from the CBR. These studies demonstrate that the CBR has been used successfully on many previous surveys; however, with one exception, all of these surveys were carried out before the introduction of the HICBC, when the CBR enjoyed almost universal coverage of children in the UK. The exception is the CANParent trial evaluation (2014), for which fieldwork for the second survey of the evaluation took place soon after the introduction of the HICBC (between August and November 2013). This evaluation sampled parents who had opted out of Child Benefit in order to maximise coverage, and the extent of the undercoverage arising from non-registration would have been small and limited to children aged 0 given how soon after the introduction of the HICBC the survey took place. As detailed in section 5.1.2, non-registration poses a more serious threat to the CEYSP given the changing nature of the undercoverage, and the survey's need to track changes in population parameters over time.

The Millennium Cohort Study

The Millennium Cohort Study, run by the Centre for Longitudinal Studies, is the most recent of Britain's cohort studies. It is multidisciplinary survey that follows the lives of around 19,000 children born in the UK in 2000-01. Six waves have been conducted to date, the earliest in 2001-02 when cohort members were aged nine months, and the most recent in 2015 when cohort members were aged around 14.

Children were sampled from Child Benefit records[20]. In 2000, Child Benefit was a universal and tax-free benefit, with take-up among those eligible nearing 100 per cent, meaning Child Benefit records provided an almost complete sampling frame for the survey population. While it would have been possible to use the Office for National

---

[20] Those eligible were those born between 1st September 2000 and 31st August 2001 in specified wards in England and Wales, and between 23rd November 2000 and 11th January 2002 in Scotland and Northern Ireland.

Statistics (ONS) birth registration records to sample children, this method was rejected because, while these records are essentially complete (they contain the baby's address at the time of registration, an event that must take place by six weeks after birth), ONS required that families opt in in writing to the study, rather than be approached with the ability to opt out, and this would have damaged response rates.

CANparent trial evaluation (2014)

The CANparent trial evaluation was an evaluation of a government trial of high quality universal parenting classes to support the parenting skills of mothers and fathers. The study was commissioned by DfE, and the survey was carried out by TNS BMRB. An initial survey was conducted, entailing face-to-face interviews with c.3,000 parents of children aged 0 to 5, split across trial and comparison areas. A second survey entailed c.3,000 face-to-face interviews with parents of children aged 0 to 7, again split across trial and comparison areas.

For both the initial and second surveys, parents were sampled from the CBR. Fieldwork for the second survey was carried out between 19th August and 21st November 2013, around six months after the introduction of the HICBC in January 2013. The authors note in their final report on the evaluation that 'Due to a change in eligibility for Child Benefit after the first survey was conducted, at Wave 2 the sample was drawn from both Child Benefit records and from an HMRC listing of parents who had chosen to opt out of receiving Child Benefit, thus ensuring that an almost universal coverage of parents was maintained across both waves'[21]. Given the brief period between the introduction of the HICBC and the sampling for the second survey, the extent of undercoverage would have been small, confined to new-born children of only a few months old.

Relationship Support Trials for New Parents: Evaluation (2013)

The Relationship Support Trials for New Parents: Evaluation was commissioned by DfE, and carried out in 2013 by a consortium of TNS BMRB, BPSR, London Economics and OnePlusOne. It was an evaluation of services set up in five areas in England designed to help new parents (via either face-to-face sessions or online workshops) to keep their relationship healthy at a stressful time. The evaluation included a baseline postal survey, of new parents and expectant parents, in trial and comparison areas, before the start of the trials (in Summer 2012). Around 1,800 questionnaires were returned. While a follow-up survey was initially intended to take place once the trials had been running for around a year and a half (in early 2014), take-up was far lower than anticipated and consequently the trials were ended in June 2013.

---

[21] Available at: https://www.gov.uk/government/publications/canparent-trial-evaluation-final-report

For the baseline survey, the CBR was used to select parents of children aged four months or older (new parents), and Bounty was used to select parents who were expecting their first child within three months. Bounty is a commercial database consisting of expectant mothers who have registered to be sent freebies and offers relevant to them (see section 0). Of the 5,000 households selected for the survey, 1,162 took part, equating to a household response rate of 23 per cent.

Diet and Nutrition Survey of Infants and Young Children (2011)

The Diet and Nutrition Survey of Infants and Young Children was commissioned by the Department for Health and the Food Standards Agency, and carried out by NatCen in 2011. The survey aimed to collect detailed information about the food consumption, nutrient intake and nutritional status of infants and young children in the UK aged between 4 and 18. Face-to-face interviews were conducted with c.2,500 parents in the UK, with the sample drawn from the CBR.

Families and Children Study: Waves 1-10 (1999-2008)

The Families and Children Study was a series of annual surveys between 1999 and 2008, commissioned by DWP and carried out by NatCen. The basic design was of a panel study, with sample boosts to ensure cross-sectional representativeness in each year The survey explored topics including: the effects of work incentive measures; the effects of policy on families' living standards; changes in family circumstances over time; the impact of benefits and tax credits in supporting families with young children; and barriers to work and measures to overcome such barriers. Each wave comprised face-to-face interviews with between c.5,000 and c.8,000 parents with children aged 16 years or under, or 18 years or under and in full-time education, with the sample drawn from the CBR.

Mental health of children and young people in Great Britain (1999, and 2004)

The Mental health of children and young people in Great Britain survey was commissioned by the Department of Health and the Scottish Executive, and carried out by the Office for National Statistics in 1999, and again in 2004. The survey collected information about the prevalence of mental disorders among children, using both structured and open-ended questions, to inform policy about the need for child and adolescent mental health services. Families in Great Britain containing a child or children aged 5 to 16 (or 5 to 15 in the 1999 survey) were sampled, and information was collected from up to three sources: parents, children, and teachers. Around 10,000 interviews in total were conducted in 1999, and c.8,000 interviews in total were conducted in 2004.

The sample was drawn from the CBR. The Office for National Statistics notes, in their Summary Report of the 2004 survey, that: 'The use of centralised records as a sampling frame was preferred to the alternative designs of carrying out a large scale postal sift of

the general population or sampling through schools. The design used enabled direct access to parents, which would not have been possible with a school-based sample, and it was more efficient than a sift.[22]'

Parents' Demand for Childcare: Waves 1-2 (1999, 2001)

The Parents' Demand for Childcare Survey is one of the two predecessor survey series to the CEYSP (the other survey being the Survey of Parents of Three and Four Year Old Children and Their Use of Early Years Services). It was commissioned by the Department for Education and Skills (DfES), and was carried out by NatCen in 1999 and again in 2001. The questionnaire collected comprehensive information on the current use of childcare, both formal and informal, as well as information about parents' views and experiences of childcare. Each wave comprised face-to-face interviews with c.5,000 parents of children aged 0 to 14 in England, with the sample drawn from the CBR.

Survey of Parents of Three and Four Year Old Children and Their Use of Early Years Services: Waves 1 to 6 (1997 to 2002)

The Survey of Parents of Three and Four Year Old Children and Their Use of Early Years Services was one of the two predecessor survey series to the CEYSP (the other survey being the Parents' Demand for Childcare survey, described above). It was commissioned by DfES, and was carried out by NatCen annually between 1997 and 2002. The questionnaire collected information about children's use of nursery education and childcare, as well as information about the quality of this provision. Each wave comprised face-to-face interviews with between c.4,000 and c.7,000 parents of children aged 3 to 4 in England, with the sample drawn from the CBR.

5.1.4. Implications

We now consider how the CEYSP might be conducted were the CBR to be retained as the sampling frame.

Option A: CBR, status quo

Under this option, the CEYSP would continue with its current sample design, sampling children present on the CBR (including those who have been opted out, as these children remain on the CBR).

This option would not affect survey costs or administration, excepting that in several years' time fieldwork efficiency may decline - due to address moves among those opting out – to the extent that fieldwork costs would need to increase by a small amount.

---

[22] Available at: http://www.hscic.gov.uk/catalogue/PUB06116/ment-heal-chil-youn-peop-gb-2004-rep2.pdf

As described above, the introduction of the HICBC means that if the CEYSP continues to sample from the CBR using the current sample design, survey estimates will become increasingly biased with each successive survey wave away from children in high income households. This will lead to difficulties in attributing changes in survey estimates over time to real changes in the population, rather than to changes in the characteristics of the sampling frame, and these difficulties will be most acute when comparing children born before and after the introduction of the HICBC (who will be aged under 4, and 4 and over, respectively, for the 2017 CEYSP).

In practical terms, this would have the following implications for forthcoming waves of the CEYSP:

- A survey conducted in 2017 would still have good coverage of the survey population (the CBR could be expected to cover around 95 per cent of children aged 0 to 14 in England, with a de facto coverage in field of around 85 per cent), and could be expected to produce survey estimates that are accurate for children aged four and over, and to produce survey estimates that are largely accurate but which have a bias away from children in high income households for children aged three and under. For a survey conducted in 2018 (should the survey move to an annual timetable) survey estimates would be accurate for those children aged five and over, and would be largely accurate but with a bias away from children in high income households for children aged four and under. This pattern would continue with each successive survey year (with the proviso that there would be a slight and gradual deterioration in the accuracy of survey estimates derived from the older age group with each successive wave, reflecting address details becoming progressively more out of date among parents opting out of Child Benefit).

- In reporting data from the CEYSP, it would be highly advisable for DfE to include caveats describing the coverage issues resulting from the HICBC, to help data users interpret the results. These caveats would be particularly important in the case of comparisons across survey waves, especially where comparisons are made between children born before and after the introduction of the HICBC. Unfortunately, given the complex nature of the undercoverage resulting from the HICBC, many data users are likely to find these caveats difficult to interpret from the perspective of making sense of the data in which they are interested. This stands in contrast to surveys that have used sampling frames for which the degree of undercoverage is greater, but for which the nature of the undercoverage is simpler. For example, the Parental Experience of Services for Disabled Children survey (see section 5.4.3 for details) used the National Pupil Database (NPD) as the sampling frame. The NPD does not cover children in independent schools (around seven per cent of the school-aged population). The technical report to the survey states: 'How far this level of non-coverage biases the survey estimates is unknown', but 'so long as (a) the conditions for being listed on the NPD remain the

29

same, and (b) the broad distribution of household types remains stable, the level of non-coverage bias should hold steady across future waves of this survey. Consequently, any significant changes in performance indicator scores ought to reflect real change in performance rather than issues with the sample frame.' The same principle applies to the CEYSP: from the perspective of accurately tracking and interpreting change over time, it would be preferable for the sampling frame to have relatively lower but stable coverage, as opposed to relatively higher but varying coverage.

Given that a key assumption for the CEYSP is that it should enable comparisons to be made across survey waves (see section 2.3), because of these difficulties, we do not recommend that the CEYSP continues with its current sample design.

Option B: CBR, but redefine the survey population

Perhaps the simplest approach to dealing with the undercoverage resulting from the HICBC is to redefine the CEYSP survey population to avoid the coverage issue altogether. This would mean changing the survey population from 'children aged 0 to 14 living in England' (as currently) to 'children aged 0 to 14 living in England in families that are not subject to the full HICBC', or more fully 'children aged 0 to 14 living in England in families where neither parent has a net adjusted income exceeding £60,000 per year'. This redefined survey population is clearly arbitrary and idiosyncratic and does not relate to an easily conceptualised societal group. As such, data users are likely to experience difficulties in interpreting the data.

The sample would be drawn as per the current CEYSP design, with the exception that children in families that have opted out of receiving Child Benefit would be excluded from selection (as they will almost exclusively be in families subject to the full HICBC). In-field screening would then be required to exclude children living in families subject to the full HICBC that receive Child Benefit and pay the full amount back via the tax system.

The advantages of this approach are that:

- coverage will not change over time, meaning that changes in survey estimates across waves can be attributed to real changes in the population, rather than to an artefact of the falling coverage of the sampling frame; and

- only around eight per cent of children aged 0 to 14 in England will be excluded (broadly in line with the proportion that are excluded from surveys that use the NPD, given the NPD's exclusion of pupils in independent schools). This means that survey estimates would provide a reasonable guide to the characteristics of the original CEYSP survey population. Survey estimates that are highly correlated with income would have the highest risk of bias.

These advantages must be weighed against the following disadvantages:

- As discussed, the redefined survey population will be arbitrary and idiosyncratic, will not relate to an easily conceptualised societal group, and will therefore cause difficulties for data users when interpreting the data. This is in contrast to, for example, the NPD's undercoverage, which is relatively easily interpreted because it relates to a discrete societal group (pupils at independent schools). The redefined survey population will also not accord to other policy-relevant criteria. For instance, the income cap for eligibility for Tax-Free Childcare is £100,000 per year for either parent, higher than the cap for the HICBC, meaning that evidence gathered from the CEYSP to assess the impact of Tax-Free Childcare would be incomplete.

- Individuals approached to participate in the survey would need to answer screening questions to determine whether or not they are subject to the full HICBC. It is possible that some individuals will not know this, will be unwilling to provide this information, or will answer incorrectly. For instance, some individuals may think they are subject to the full HICBC, when they only pay it in part. Such errors have the potential to introduce error into survey estimates. Furthermore, the administration of the screening questions will increase the cost of administering the survey by a small amount, given the additional time these questions will take.

- While this approach would cost roughly the same as the status quo approach, the additional screening questions, and the reduced eligibility of the issued sample, would increase the overall survey cost by a small amount. We would expect the fieldwork budget required to deliver 6,350 interviews using the status quo approach to deliver approximately 6,000 interviews using this approach[23].

Because of the arbitrary and idiosyncratic population definition, and the resulting difficulties data users will experience in attempting to make sense of the data, we do not recommend that future waves of the CEYSP redefine the survey population such that children eligible for the full HICBC are excluded. Such a move would also violate a key assumption for the CEYSP, in that it should deliver interviews with a representative sample of parents and guardians of children aged 0 to 14 in England (see section 2.3).

---

[23] It should be noted that the estimates of the number of interviews that could be delivered under the various options explored in this report are ballpark estimates, and as such are subject to change. Furthermore, these estimates are Ipsos MORI's own estimates, and may not reflect the estimates that would be arrived at by other survey contractors.

5.2 Postcode Address File (PAF)

5.2.1. Introduction

The postcode address file (PAF) is a list of all known addresses in the United Kingdom (almost 30 million) that is maintained by the Royal Mail, and which is used to facilitate the efficient delivery of mail. Due to its good coverage of UK addresses, the ease with which it can be accessed, and because of the lack of alternatives, it has been the main sampling frame for random probability face-to-face surveys since the 1980s.

5.2.2. Characteristics

Accessibility

The PAF is available via a subscription service with Royal Mail. The PAF is a widely used sampling frame for high quality social surveys in the UK and there are no restrictions on its use.

Format and fields

The PAF is held in electronic format, and can be easily manipulated.

The PAF contains address details only. Because the PAF does not contain any information about residents, its use as the sampling frame for the CEYSP would be less efficient, and therefore more expensive, than the current design. In order to identify children aged 0 to 14 using the PAF, it would be necessary for interviewers to screen households for the presence of children in this age range. Such screening involves interviewers visiting a large number of addresses in order to identify a sufficient number of eligible households. It would not be possible to tailor advance communications and contact attempts to named individuals, as per the current CEYSP design.

The PAF can be readily linked to Census and other geodemographic data via postcodes, allowing for stratification by a range of variables. The lack of any information about residents, however, means that boosts based on child-level characteristics could only be achieved via in-field sampling. To boost children aged 2 to 4, for instance, it would be necessary for an interviewer to enumerate all children in the household, and for one child to be selected at random with a higher probability of selection given to children aged 2 to 4.

Coverage

The PAF covers nearly all addresses in the UK, and as such, has almost complete coverage of the survey population (children aged 0 to 14 living in England). Lound (2014) recently estimated that the PAF covers 98 per cent of the resident population of England and Wales[24].

The PAF is updated quarterly, leading to a small degree of undercoverage where new addresses have not yet been added, and a small degree of overcoverage where addresses are found to be 'deadwood' and must be screened out in the field (for instance, vacant, non-residential, or demolished addresses; typically, face-to-face surveys sampling from the PAF find between that between five and ten per cent of sampled addresses are deadwoodd). The nature and extent of the undercoverage does not pose a serious risk to the accuracy of survey estimates, and the nature and extent of the overcoverage increases survey costs by a relatively small amount.

The PAF's use for the CEYSP poses a more serious problem given that, as mentioned, households with children aged 0 to 14 are not identified in the file, and therefore expensive in-field screening is required to address this overcoverage and identify eligible households.

Stability

The PAF has been the preferred sampling frame for random probability face-to-face surveys since the 1980s, and estimates of its coverage have remained high and stable over the last three decades; for instance, Foster (1993) estimated that the PAF covers around 97 per cent of UK households, which is similar to the 98 per cent coverage of the resident population of England and Wales quoted above (Lound, 2014)[25]. We are not aware of any proposed changes to the PAF, or forthcoming restrictions on its use.

5.2.3. Use of the PAF on similar surveys

While most surveys of parents use a sampling frame in which eligible parents (or children) can be identified, we are aware of two relatively recent surveys of parents that have used the PAF as the sampling frame: the Parental Opinion Survey (2010), and the National Survey of Parents and Children (2008). These surveys have employed an in-field screening approach to ascertain eligibility and identify target respondents.

---

[24] Available at: https://gss.civilservice.gov.uk/wp-content/uploads/2015/09/QIF-Projects-Quality-Assessment.pdf

[25] Foster, K, (1993) The coverage of the Postcode Address File as a Sampling Frame, Survey Methodology Bulletin, no. 34, January 1994, OPCS, pp 9-18

The PAF was used for these surveys because permission to use the CBR (which would have enabled the identification of eligible parents prior to fieldwork) was not granted, and because the eligibility rate of parents (30 per cent of households) was high enough for an in-field screening approach to be feasible. These studies provide a precedent demonstrating that the CEYSP could be successfully carried out were it to use, or incorporate, a sample drawn from the PAF.

Parental Opinion Survey (2010)

The Parental Opinion Survey was commissioned by DfE, and carried out by TNS BMRB in 2010. The survey aimed to assess parents' opinions on a range of issues focusing on their role as parents, and in particular their confidence as parents and their views about the services that they or their children used. Around 2,500 face-to-face interviews were conducted with parents in England.

PAF was chosen as the sampling frame, with a full screening approach employed. The sampling and eligibility criteria were the same as the earlier National Survey of Parents and Children (2008), described below; the only difference was that non-resident parents were also deemed eligible.

National Survey of Parents and Children (2008)

The National Survey of Parents and Children was commissioned by the Department for Children, Schools and Families (DCSF), and carried out by BMRB in 2008. The survey aimed to understand better parents and carers of children and young people aged 0 to 19, gaining insights into family attitudes and dynamics, and how these support the well-being, behaviour and learning of children and young people. Around 2,500 face-to-face interviews were conducted with parents in England.

The technical report of the survey notes that several sampling frames were considered for the survey: the NPD, the CBR, and the PAF[26]. The NPD was rejected because it excludes households which contain only pre-school children, those in independent education, and those containing only 16 to 19 year olds living at home and no longer at school. Permission to use the CBR was not granted by HMRC, precluding its use. The technical report also notes that the CBR presented two disadvantages: it excluded children aged 16 to 19 who were neither in full-time education nor registered for work, and around 10 per cent of those listed on the CBR were expected to have moved from their sampled address at the point the interviewer called.

---

[26] Available at:
http://webarchive.nationalarchives.gov.uk/20100202100446/dcsf.gov.uk/research/data/general/dcsf-rr059-tr.pdf

The PAF was chosen as the sampling frame as it provided the most comprehensive listing of residential addresses in England, and a full screening approach was considered feasible given that Census data showed around 30 per cent of households in England would contain an eligible child. To increase fieldwork efficiency and control survey costs, before the sample was drawn those postcode sectors where the proportion of households with eligible children was less than 20 per cent were excluded. This led to the exclusion of around six per cent of all postcode sectors, equating to around three per cent of all households with eligible children. An analysis of the socio-economic profile of all postcode sectors against this subset of postcode sectors did not show great differences.

In order to preserve the random probability sampling approach, at each sampled address a dwelling unit was randomly selected where there was more than one at the address. In selecting a parent for interview, where necessary the interviewer randomly selected a 'parenting unit' in the household, where a parenting unit was a set of parents or a single parent of a child (households could contain more than one parenting unit if, for example, there were three generations at an address, such as a parent with a teenage child who had a child of their own). This ensured that single parents living in multi-generational households were not under-represented. Where a parenting unit comprised two parents, one parent was randomly selected for interview[27].

The sample design anticipated that eight per cent of the issued PAF sample would be deadwood, 30 per cent of remaining households would contain an eligible child, and that at 65 per cent of eligible households an interview would be achieved. Fieldwork outcomes showed that six per cent of the issued PAF sample was deadwood, 27 per cent of remaining households were found to contain an eligible child (interviewers asked neighbours in instances where contact at the main address could not be made, and the authors believe this proportion may be somewhat deflated due to 'a proportion of households falsely claiming they had no children living with them in order to avoid taking part in the survey'), and at 57 per cent of eligible households an interview was achieved. The response rate (the eligibility rate among those addresses where eligibility could not be determined was assumed to be the same as among those where eligibility could be determined) was 55 per cent.

5.2.4. Implications

The PAF's characteristics make it a suitable sampling frame for the CEYSP. It is accessible, has good coverage of the survey population, has been used before on similar

---

[27] Because the CEYSP samples children rather than parents, defining and selecting among parenting units would not be applicable.

surveys, and is a widely used and well regarded sampling frame for face-to-face probability surveys of the public.

The PAF's primary drawback is that it does not directly identify the survey population; eligible children, or parents, are not identified on the PAF, and consequently a relatively expensive in-field screening exercise is required to ascertain eligibility at each household. This overcoverage also makes the accurate calculation of response rates difficult given that eligibility among those addresses not screened must be estimated; however, this would not pose a serious risk to the quality of the survey.

We now consider how the CEYSP might be conducted were the PAF to be used as the sampling frame. We describe two possible sample designs: the PAF with full screening (option C); and a split-sample design using the PAF in areas affected by the HICBC, and the CBR in areas unaffected by the HICBC (option D). We also discuss why we do not consider two other potential sample designs using the PAF (the PAF with a focused enumeration approach, and the PAF using a push-to-web approach) to be suitable.

Option C: PAF with full screening

A sample design under which addresses are sampled from the PAF, with full in-field screening carried out to identify eligible households, would be similar to the approach used for the Parental Opinion Survey (2010) and the National Survey of Parents and Children (2008), which we describe in section 5.2.3.

Under a PAF with full screening design, addresses would first be selected from the PAF using a random probability sampling approach. Data from the 2011 Census, and from the Health Survey for England, indicate that approximately 25 per cent of households in England contain a child aged 0 to 14 (and about 22 per cent contain a child aged 0 to 12). Advance letters could be sent to selected addresses introducing the survey, and explaining that an interviewer will visit with a view to interviewing a parent of a child aged 0 to 14 (or 0 to 12) in the household[28]. Interviewers would then visit sampled addresses

---

[28] An argument could be made for not sending advance letters given that most households sampled will be ineligible. Rather, interviewers would call at addresses 'cold' and would introduce the survey on the doorstep, with an explanatory letter to hand over where appropriate. However, given the sensitivities involved with interviewers calling at addresses without prior warning and asking about the presence of children, we would guard against this method. Indeed, the technical report of the National Survey of Parents and Children (2008), for which advance letters were not sent, states that 'feedback and calls to DCSF and BMRB contact lines suggested that it may have been preferable to post an advance letter, due to the sensitivities of asking about presence of children without pre-notifying households. This should be borne in mind if conducting a similar study in the future'. Furthermore, the printing and postage expense of mailing out advance letters would largely, if not wholly, be offset by increased fieldwork efficiencies arising from interviewers calling at addresses where residents are aware of the survey.

and screen for the presence of an eligible child at each household, and where one is identified, would attempt to interview an adult with main or shared responsibility for childcare decisions for that child.

Fieldwork efficiency could be increased by: i) excluding those postcode sectors containing less than a certain proportion of eligible children (as per the approach taken for National Survey of Parents and Children (2008)), and ii) by introducing an element of 'micro-clustering,' whereby within each sampled postcode sector, groups of neighbouring addresses are selected, rather than all addresses being randomly distributed across the sector (a third approach by which fieldwork efficiency could be increased is 'focused enumeration', which we discuss separately below).

Both of these measures would reduce the amount of time interviewers spend travelling between addresses, relative to the amount of time they spend carrying out interviews, and with a suitable sample design, would increase the number of interviews achieved for a given cost. However, this would be at the expense of a slight but, in our view, acceptable loss of accuracy in survey estimates. For excluding postcode sectors containing less than a certain proportion of eligible children this would be because, those postcode sectors containing relatively few children may differ systematically from those containing a relatively larger proportion of children. For 'mirco-clustering' this would be because households in a given postcode sector that are adjacent to each other may be more similar than those households which are further away from each other.

The advantage of an approach using the PAF with full screening is that:

- coverage of the survey population (children aged 0 to 14 in England) would be very high, and most importantly, would not vary across survey waves. This means that changes in survey estimates across waves of the CEYSP could be attributed with confidence to real changes in the population, rather than to an artefact of the changing characteristics of the sampling frame.

The disadvantages of an approach using the PAF with full screening are that:

- the approach would be costlier than the current CEYSP design, given that most addresses issued to interviewers will prove to be ineligible and will be screened out (that is, they will not contain a resident child aged 0 to 14). We estimate that the fieldwork budget required to deliver 6,350 interviews using the status quo approach would deliver under half this number (c.2,900) were a PAF with full screening

sample design to be used[29]. The number of interviews delivered would fall further should certain groups of children (for instance, pre-school children) need to be boosted relative to their proportion in the population, because this would require a greater amount of screening in-field. In addition, should the eligible age range of children be narrowed from 0 to 14, to 0 to 12, the number of interviews delivered would also fall, again because more screening would be required. Either of these measures could be expected to reduce the total number of interviews delivered for the current CEYSP fieldwork budget to in the region of 2,500 interviews, and lower if implemented in combination. This reduction in the number of interviews achieved is at odds with a key assumption for the CEYSP, namely that it should deliver in the region of 6,350 interviews per year (see section 2.3).

- advance letters (and interviewers' doorstep introductions) could not be addressed to named individuals, reducing the impact of the survey invitation, and potentially reducing the response rate.

While this option is therefore attractive in terms obtaining high coverage of the survey population, the large-scale in-field screening exercise required does not make it a cost-effective solution.

Option D: PAF in areas affected by the HICBC, CBR in areas unaffected by the HICBC

Those families eligible for the full HICBC (whether they receive Child Benefit and pay it all back via the HICBC, have opted out of Child Benefit, or have never registered for Child Benefit) will not be spread evenly throughout the country; rather, they will be disproportionately resident in more affluent postcode sectors. Indeed, there are likely to be many postcode sectors across England in which few, if any, families are eligible for the full HICBC. In these postcode sectors, the CBR's coverage will be unaffected by the HICBC, and the CBR could be used as the sampling frame without fear of biasing survey estimates. On the other hand, for those postcode sectors in which a non-negligible proportion of families are eligible for the HICBC, the use of the CBR would bias survey estimates (because higher income families would not be covered), and the PAF could be used as the sampling frame instead.

This split-sample design requires us to estimate, with a reasonable degree of confidence, the proportion of families in each postcode sector that are not on the CBR as a result of the HICBC. Using these proportions, postcode sectors can be allocated either to the CBR sample, or to the PAF sample. These proportions can be estimated by reference to the

---

[29] Exclusion of postcode sectors containing less than a certain proportion of eligible children, and micro-clustering, could be expected to result in a small proportionate increase in the number of interviews delivered (in the region of five per cent, equivalent to around 150 interviews).

proportion of families on the CBR that have opted out of receiving Child Benefit. Specifically, in those postcodes sectors in which few, if any, families are eligible for the full HICBC, one would expect few, if any, families to have opted out of receiving Child Benefit. On the other hand, in those postcodes in which a comparatively large proportion of families are eligible for the full HICBC, one would expect a comparatively large proportion of families to have opted out of receiving Child Benefit[30].

To explore the feasibility of this approach, HMRC provided us with counts, for each postcode sector in England, of the number of families with children aged 0 to 14 that: are paid Child Benefit; and have opted out of Child Benefit.

This enabled us to calculate, for each postcode sector in England, the proportion of families that have opted out of receiving Child Benefit. Using these proportions, we then considered a range of cut-off points below which postcode sectors would be drawn from the CBR, and above which postcode sectors would be drawn from the PAF. A cut-off point of five per cent would mean, for example, that postcode sectors in which more than 1 in 20 families have opted out of receiving Child Benefit would be drawn from the PAF, and postcode sectors in which 1 in 20 or fewer families have opted out of receiving Child Benefit would be drawn from the CBR.

Table 5.2 shows (in column 2) the proportion of postcode sectors that would be drawn from the PAF for cut-off points between 1 and 10 per cent. Column 3 shows our estimate of the relative coverage of children born after the introduction of the HICBC, compared to before the introduction of the HICBC, for each of these cut-off points. And the final column shows, for each cut-off point, our estimate of the relative coverage of children born after the introduction of the HICBC, compared to before the introduction of the HICBC, for those children living in households eligible to pay the HICBC.

---

[30] We note that it is not possible to estimate the proportion of families eligible for the full HICBC via data from tax records, because many of those eligible for the full HICBC will have opted out of receiving Child Benefit, or will not have registered for Child Benefit it in the first place, meaning their tax records will not contain any information about the HICBC. Furthermore, the staff member to whom we spoke at HMRC confirmed that there were likely to be legal barriers in accessing and analysing data from tax records.

Table 5.2. Allocations of postcode sectors to PAF versus CBR sample, and associated coverage of children

Cut-off: postcode sector allocated to PAF if proportion of opt outs greater than…	Postcode sectors allocated to PAF	Relative coverage of children born post-HICBC vs pre-HICBC	Relative coverage of children born post-HICBC vs pre-HICBC, for children in high income households
(%)	(%)	(%)	(%)
1	78.4	99.9	98.6
2	62.8	99.7	95.4
3	52.4	99.4	91.8
4	44.8	99.2	88.1
5	39.0	98.9	84.4
6	34.2	98.6	80.6
7	30.8	98.4	77.5
8	27.2	98.1	73.7
9	24.4	97.9	70.4
10	21.8	97.6	66.8

The data in Table 5.2 show that relative coverage of children born after the introduction of the HICBC would be very high for any cut-off point up to and including 10 per cent. For instance, using a cut-off point of 10 per cent, the coverage of children born after the introduction of the HICBC would be around 98 per cent of the coverage of children born before the introduction of the HICBC.

This relative coverage relates to all children born since the introduction of the HICBC, regardless of their household's income. Instead, to properly assess the impact of the cut-off point, we should consider the relative coverage of those children in high income households, whose parents are eligible to pay the HICBC. Using a cut-off point of five percent would mean that: i) 39 per cent of postcode sectors would be drawn from PAF, with the remaining 61 per cent drawn from the CBR; and ii) the coverage of children living in high income households and born after the introduction of the HICBC would be around 84 per cent of the coverage of children living in high income households and born before the introduction of the HICBC.

Using a higher cut-off point of 10 per cent would reduce the number of postcode sectors drawn from PAF to 22 per cent, but at the cost of reducing the relative coverage of children in high income households to around 67 per cent.

So, using a higher cut-off point lowers fieldwork costs (because fewer addresses are drawn from the PAF meaning less screening is required), but raises the risk of bias to survey estimates (because there is lower coverage of children in high income households).

Were a five per cent cut-off point to be used, we estimate that around 4,300 interviews could be delivered for the fieldwork budget required to deliver 6,350 interviews using the status quo approach[31]. This is significantly higher than the c.2,900 interviews that could be delivered using a full PAF screening approach, and would cover the great majority (84 per cent) of children in high income households born after the introduction of the HICBC; however, it is still lower than the 6,350 interviews that it is assumed that future waves of the CEYSP will deliver (see section 2.3).

We consider this approach to be promising, because although it does not deliver the full 6,350 interviews desired, it maximises the number of interviews that can be achieved, with only a small degree of compromise on coverage. Furthermore, the coverage can be expected to be stable over time, meaning that changes in survey estimates across waves could be attributed with confidence to real changes in the population.

Two shortcomings of this approach should be noted. The first is that it will become increasingly difficult with each successive year to estimate with accuracy the proportion of families missing from the CBR due to the HICBC. This is because many families with new children do not register for Child Benefit in the first place, so the numbers of families opting out of Child Benefit will become progressively smaller[32]. This will not present a problem in the short term however, and should the numbers of families opting out become so low as to make accurate estimation impossible, one could defend using data from previous years to calculate the estimates.

---

[31] As per the PAF with full screening approach, the number of interviews delivered would fall were a boost of children of certain ages to be incorporated into the sample design, or were the eligible age range of children to be narrowed to 0 to 12. Both of these approaches would require a greater amount of screening among the PAF sample, but would not affect the CBR sample (for which children's ages are known in advance). Either of these measures could be expected to reduce the total number of interviews delivered for the current CEYSP fieldwork budget to in the region of 4,000 interviews, and lower if implemented in combination.

[32] Specifically, very few children born after the HICBC was introduced in 2013 are being opted out of Child Benefit, because parents in high income households tend not to register these children for Child Benefit in the first place (see Fig 4.1). Over time, this means the proportion of children opted out of Child Benefit will fall, reaching a uniform 1 per cent (approximately) by 2027, when those children born in 2013 will be aged 14. Because of this, using the proportion of children opted out of Child Benefit as the proxy measure to determine whether a postcode sector contains sufficient children in high income households to warrant allocating it to the PAF sample will become less robust.

The second shortcoming is that the split-sample design introduces complications both in terms of fieldwork administration (interviewers would need to be trained both on approaching addresses drawn from the CBR, and on approaching addresses drawn from the PAF), and in terms of calculating the survey weights; however, neither of these complications pose serious risks to this approach.

Other PAF-based approaches

We now turn to a brief discussion of two PAF-based sample designs which we have considered, but discounted: PAF using a focused enumeration approach, and PAF using a push-to-web approach.

Turning first to PAF using a focused enumeration approach, focused enumeration enables one to reduce the cost of screening with a PAF sample. The method was first developed about 35 years ago (Brown and Ritchie, 1981)[33] and has evolved over the years. It has been used mainly to identify black and minority ethnic (BME) groups. It generally involves selecting addresses in neighbouring 'micro-clusters,' and allowing households to screen out neighbouring addresses within their micro-cluster if they are certain that they do not contain someone who is eligible for the study. Interviewers do not visit addresses that are screened out, meaning focused enumeration is more efficient, and therefore less costly, than an approach requiring full screening.

Concerns have been raised in recent years about the use of focused enumeration (e.g. Smith et al., 2010)[34], as analyses of its use for samples of BME groups have indicated that the method fails to identify a substantial proportion (around a third) of eligible households. Given that unidentified households may differ systematically from identified ones in survey-relevant ways, using focused enumeration risks introducing bias to survey estimates.

Furthermore, a focused enumeration approach for the CEYSP would require members of the public to divulge information to a third party about whether or not their immediate neighbours have very young children. This approach raises far greater sensitivities than traditional focused enumeration approaches which have attempted to identify adults (for instance, those from BME backgrounds). Some individuals may even feel it is inappropriate for DfE to commission research that asks them to inform on their

---

[33] Brown C, Ritchie J (1981) Focussed Enumeration: The Development of a Method for Sampling Ethnic Minority Groups, Policy Studies Institute/SCPR, London.

[34] Smith P, Pickering K, Williams J and Hay R (2010) The Efficacy of Focused Enumeration. Royal Statistical Society presentation. Available at: https://www.ipsos-mori.com/Assets/Docs/Publications/The_Efficacy_of_Focused_Enumeration.PDF

neighbours in this manner, and this may lead to complaints to both DfE and the survey contractor.

For these reasons, we do not recommend a PAF based focused enumeration approach for future waves of the CEYSP.

Turning now to a PAF using a push-to-web approach, this would entail sending letters to sampled addresses requesting recipients to follow a web-link and complete a survey online. While such approach is less costly than a face-to-face methodology, DfE wish to retain the CEYSP as a face-to-face survey, because adapting the survey to one suitable for self-completion online would require fundamental changes to the survey that would seriously compromise the amount and the quality of the data that could be collected. A push-to-web approach could also be expected to deliver a far lower response rate than that enjoyed by the CEYSP's current face-to-face design. For these reasons, we do not recommend a PAF based push-to-web approach.

5.3 Medical Research Information Service Integrated Database & Administration System (MIDAS)

5.3.1. Introduction

The Health and Social Care Information Centre (HSCIC) maintains a database called the Medical Research Information Service Integrated Database & Administration System (MIDAS). MIDAS allows HSCIC to provide 'cohort management services'. For instance, using MIDAS a researcher can define a cohort, which may be a group of patients fulfilling various demographic and clinical criteria, and can track outcomes, such as mortality rates, for these patients over time.

MIDAS is populated directly from the Personal Demographics Service (PDS), which is the national electronic database of NHS patient demographic details, and has almost complete coverage of the population of England and Wales. MIDAS therefore has very good coverage of the CEYSP population, and would in theory be an ideal sampling frame for the CEYSP; however, as it contains sensitive personal data (patients' names and contact details) its use is restricted, and it is highly unlikely that permission for its use would be granted for the CEYSP.

5.3.2. Characteristics

Accessibility

Access to identifiable patient details held on MIDAS is highly restricted, and is governed by a number of legal restrictions set out within the Data Protection Act 1998, the Health and Social Care Acts 2012 and 2014, HSCIC's Code of Practice on Confidential Information, and which arise from the common law duty of confidence.

Reflecting these restrictions, HSCIC state that they can only provide access to identifiable patient information if the following conditions are satisfied[35]:

> 1.   The individuals have given permission for their information to be shared, or the law allows it to be shared;
>
> 2.   The information is used to promote healthcare or support the delivery of care services in England; and

---

[35] Available at: http://www.hscic.gov.uk/article/4963/What-we-collect

3.   The organisation requesting the information has demonstrated to assurance bodies that it will be looked after according to the law and good Information Governance practice.

While both the DfE and the commissioned survey contractor for the forthcoming CEYSP can be expected to have the appropriate data security measures in place to satisfy condition 3, conditions 1 and 2 are problematic.

With respect to condition 1, the individuals one would wish to sample will not have given their permission for their NHS records to be shared, and in order for permission to be obtained, it is likely that the CEYSP's survey methodology would need to change from the current opt out approach to an opt in approach, whereby HSCIC would contact parents or guardians of sampled children requesting them to provide explicit consent to be approached to participate in the CEYSP. The survey contractor would then attempt to contact and interview only those who had opted in. This change would lead to an unacceptable fall in the survey's response rate given that many individuals would not opt in. Furthermore, there are reasons to believe that those opting in will differ systematically from those not opting in (for instance, those opting in may have a greater interest in childcare, and in turn, this may be because they use more childcare, or certain types of childcare), and these systematic differences will result in bias to survey estimates.

With respect to the second part of condition 1 (that the law must allow identifiable patient information to be shared in cases where consent has not been obtained) the legal bases permitting identifiable patient records to be shared with other organisations include statutory bases for disclosure, a court order, or a public interest justification, none of which are of relevance to the CEYSP. A further legal basis for the disclosure of identifiable patient records is via support from the Secretary of State for Health under Section 251 of the National Health Service Act 2006. This section enables the common law duty of confidentiality to be temporarily lifted so that confidential patient information can be transferred to an applicant without the discloser acting illegally. However, reflecting condition 2 above, clause 12 of Section 251 requires that the activities for which the confidential patient information are being shared 'must fall within a medical purpose to be considered, so, for example, requests to access patient information to inform road traffic management planning could not be approved as the primary purpose would not support health service improvements.'[36] Because the CEYSP's primary purpose is not medical in nature, this would preclude it from Section 251 approval.

Applications to access patient information without consent via Section 251 approval must be made to the Confidentiality Advisory Group (CAG) of the NHS Health Research

---

[36] Available at: http://www.hra.nhs.uk/documents/2014/02/cag-frequently-asked-questions-1.pdf

Authority[37]. The application must include detail on the following: a description how the proposed use of patient information will improve patient care and serve the wider public interest; a description of how patient and user organisations/representatives have been involved in the development of the activity for which support is sought; and detail about why it is impracticable to conduct the activity via an alternative method.

Format and fields

MIDAS is an electronic database and can be readily manipulated. We spoke to a staff member at HSCIC who confirmed that, in theory, MIDAS could be used to draw a two-stage clustered probability sample equivalent to the current CEYSP design. Indeed, there are precedents: MIDAS has previously been used to draw a probability sample on the What about YOUth? Survey: Trial Study (2014), and is currently being used on the Survey of the Mental Health of Children and Young People (2016) (see section 5.3.3). Both of these surveys were considered to 'fall within a medical purpose' from the perspective of gaining approval to use MIDAS.

MIDAS contains variables which would accommodate the current CEYSP sample design, namely: full name, gender, date of birth, and full address. Because MIDAS entries are based on NHS numbers (which are automatically allocated to UK born individuals at birth, and to immigrants at their first contact with the NHS) MIDAS is structured as an individual level database, rather than a household level database. The staff member we interviewed at HSCIC confirmed that, in practical terms, advance letters would need to be addressed to, and interviewers would need to seek contact with 'The parent or guardian of [child's name]', rather than with a named parent directly.

MIDAS entries can be linked to relevant geodemographic data (via postcodes) to allow for clustering, and stratification. MIDAS does not contain variables flagging demographic details such as ethnicity, disability, or special educational needs, precluding boosts of these groups of children.

Coverage

The PDS - and by extension MIDAS - has almost complete coverage of the population of England and Wales. It includes all those who have had contact with the NHS in England, Wales or the Isle of Man, and additionally all babies born in England, Wales or the Isle of Man since October 2002 will have been automatically added to the PDS via birth registration records, providing, in theory, complete coverage of children aged 0 to 14. There is a lag of only a couple of days between an addition or change to PDS, and the respective addition or change to MIDAS.

---

[37] Available at: http://www.hra.nhs.uk/resources/confidentiality-advisory-group/

Undercoverage on MIDAS can arise in a number of ways:

- Out of date address details. Addresses will be out of date where an individual has moved but has not notified their GP. The HSCIC staff member whom we interviewed was unable to provide information about the proportion of MIDAS entries that may be affected in this way; however, the proportion is likely to be lower than the proportion of out of date address details on the CBR (around 11 per cent), given that one can receive Child Benefit without notifying HMRC of a change of address (Child Benefit is paid directly into the recipient's bank account), but cannot continue to receive NHS services without confirming one's current address with a GP or hospital.

- Those who use private health care exclusively, and who have not registered with their local GP or had contact with NHS services. The HSCIC staff member whom we interviewed was unable to provide information about the proportion of the population that may be affected in this way, however, he believed the proportion would be very small. Furthermore, given that all babies born in England since October 2002 will have been automatically added to PDS via birth registration records, this undercoverage should not affect children aged 0 to 14, although it is likely the proportion of out of date address details will be higher among those children who have used private health care exclusively.

- New-born children. New-born children are added to PDS once they have been registered with the General Register Office, and will appear on MIDAS a couple of days later. In England and Wales, children must be registered within 42 days (six weeks) of birth, and while many children will be registered well before this period has elapsed, this does allow for a degree of undercoverage of babies of only a few weeks of age. Nevertheless, this undercoverage can be accounted for by in-field sampling of children born after the sampling cut-off date, and is lower than the respective degree of undercoverage on the CBR (for which in-field sampling is required for babies aged under five months).

- Duplicates and confusions. Some individuals are incorrectly allocated two NHS numbers (duplicates), which would double their chance of selection. Others are incorrectly allocated an existing NHS number (confusions), which reduces their chance of selection. HSCIC expend considerable efforts on resolving such instances, and the HSCIC staff member to whom we spoke confirmed that such instances affect only a tiny proportion of entries, and are thought to occur at random, rather than being more common among particular demographic groups.

Stability

We are not aware of any proposed changes to the PDS, and by extension to MIDAS. Given the importance of the PDS to the day-to-day running of the NHS, and given that

the NHS, for the foreseeable future, can be expected to remain free at the point of use to every legal resident in England, we see no reason to be concerned about the stability of MIDAS as a sampling frame.

5.3.3. Use of MIDAS on similar surveys

To our knowledge, MIDAS has been used as the sampling frame on two surveys, which we now described. Each of these surveys are health-related surveys, satisfying the condition that identifiable patient information can only be shared where doing so promotes healthcare or supports the delivery of care services. As such, neither of these surveys sets a precedent that leads us to believe that access to MIDAS would be granted for future waves of the CEYSP.

Survey of the Mental Health of Children and Young People (2016)

The Survey of the Mental Health of Children and Young People (2016) was commissioned by HSCIC and is currently being carried out by NatCen and ONS. It is designed to assess the prevalence of mental illness among children and young people aged 2 to 19 in England and Scotland, and will support the planning, commissioning and improvement of mental health services. The survey involves face-to-face interviews with young people and parents (the survey is in progress at the time of writing so further details are unavailable).

The survey fulfils HSCIC's condition that any patient details that are transferred are 'used to promote healthcare or support the delivery of care services in England'.

What about YOUth? Survey: Trial Study (2014)

The What about YOUth? Survey: Trial Study was commissioned by HSCIC on behalf of the Department for Health (DH), and was carried out by Ipsos MORI in 2014. It was a study to explore the feasibility of conducting the main What about YOUth Survey, a survey of 15 year olds in England about health and wellbeing relating to public health outcomes. A postal self-completion methodology, with the option to complete online, was used, and approximately 3,000 questionnaires were completed.

The MIDAS database and the NPD were considered as possible sampling frames. MIDAS was chosen because it had better coverage given that the NPD excludes those in independent schools. An application to use MIDAS was made to HSCIC, and was granted. It should be noted that the application for the use of MIDAS had to satisfy a number of requirements before access was granted; for instance, it had to demonstrate how 'the proposed use of patient information will improve patient care and serve the wider public interest', and how 'patient and user organisations/representatives [had been involved] in the development' of the survey.

The main What about YOUth? survey was carried out in 2015, and ultimately used the NPD as the sampling frame (the main survey is discussed in section 5.4.3) because permission to use MIDAS was not granted by HSCIC due to heightened concerns around patient confidentiality at the time of commissioning.

5.3.4. Implications

Although MIDAS's characteristics mean it is, in theory, the ideal sampling frame for the CEYSP, strict legal restrictions around sharing identifiable patient data mean that a formal application by DfE for MIDAS to be used as the sampling frame for the CEYSP would have no realistic chance of success. Given the significant resources required to submit a formal application, we therefore do not recommend that an application is made.

5.4 National Pupil Database (NPD)

5.4.1. Introduction

The National Pupil Database (NPD) is a pupil level database held by DfE that contains detailed information about pupils in England's educational system. It is primarily compiled from the School Census and the Early Years Census. These are statutory censuses that require schools and early years providers to collect information about their pupils, as well as information about the schools and early years providers themselves. The NPD covers all pupils in England attending state schools or funded early years education on a defined Census day, and includes a wide range of school level and individual level educational and demographic variables.

The School Census is carried out each term (Autumn, Spring and Summer), and the Early Years Census is carried out annually. The datasets are available soon after Census day (for example, the dataset relating to the Spring 2014/15 School Census was released on 11th June 2015).

5.4.2. Characteristics

Accessibility

To obtain a sample of pupils from the NPD, an 'NPD data request application form' must be completed and submitted to DfE. Applicants must demonstrate compliance with the Data Protection Act 1998 and registration with the Information Commissioner's Office, and must also provide a 'basic disclosure' certificate for each user of the data.

Permission to use the NPD as a sampling frame has been granted for use on surveys similar to the CEYSP (we describe these surveys in section 5.4.3), and we are not aware of any reasonable requests that have been denied.

Format and fields

The NPD is an electronic database and includes a range of data about individual pupils, including name, gender, date of birth, and address. The NPD also includes additional demographic details, including ethnicity, whether the child has a SEND, and educational attainment, enabling boosts of these groups.

The NPD does not hold contact information about pupils' parents or guardians, meaning that in the case of face-to-face surveys of parents, advance letters need to be sent to pupils' home addresses, addressed to 'The parent or guardian of [pupil's name]'.

Coverage

The NPD covers school-age children, and those attending funded early years education. The NPD has good coverage of children aged 5 to 16 (because education is compulsory for this age group), but its coverage of children aged 2 to 4 is much poorer (given that early education is not compulsory for this age group), and the NPD does not cover children aged under 2. The NPD could therefore only be used for the CEYSP if there was a viable alternative sampling frame for children under school-age. To the best of our knowledge, no such sampling frame exists.

In addition to not covering the full age range of the CEYSP survey population, using the NPD would result in further undercoverage. Most importantly, it is currently optional for independent schools to supply the relevant data for inclusion on the NPD. Even for those independent schools that do supply this information, no contact information is recorded for their pupils, meaning that the NPD contains no contact information for any pupils attending an independent school. Around seven per cent of the school-aged population attend an independent school, and the characteristics of these pupils can be expected to differ systematically from those attending state schools in a number of ways; for instance, they are likely to have parents with a high income. It is therefore likely that the NPD would generate a sample of school-age children that would be subject to similar biases as a sample selected from the CBR. The NPD also does not cover pupils that are home-educated or excluded from school on Census day, leading to a small amount of additional undercoverage.

Stability

We are not aware of any proposed changes to the NPD.

5.4.3. Use of the NPD on similar surveys

The NPD has been used as the sampling frame on a range of surveys, three of which we now describe. These surveys have only covered children of school age that attend a state school, given the NPD's undercoverage of pre-school children, and those at independent schools. None of these surveys provides a precedent to suggest that the NPD could be used successfully as the sampling frame for future waves of the CEYSP.

Taking Part Child Boost (2015)

The Taking Part Child Boost was commissioned jointly by the Department for Culture, Media and Sport (DCMS) and DfE to boost children aged 11 to 15 as part of the Taking Part Survey. Taking part is DCMS's flagship survey of participation in arts, culture, sport and heritage, and is designated by the UK Statistics Authority as a National Statistic. Almost 2,000 children were interviewed face-to-face, in home.

The sample of children comprised children aged 11 to 15 in state-school secondary education, and was drawn from the NPD. These children were combined with respondents from the Taking Part child sample (which includes children in all types of secondary education, and originates from the PAF).

What about YOUth? Survey (2015)

The What About YOUth? Survey was commissioned by HSCIC on behalf of DH, and was carried out by Ipsos MORI. It was a survey of 15 year olds in England about health and wellbeing relating to public health outcomes, with topics including diet, physical activity, free-time, smoking, drinking, drugs, and bullying. A postal self-completion methodology with the option to complete online was used, and approximately 120,000 questionnaires were completed, representing a response rate of 40 per cent.

Two potential sampling frames were considered for the survey: MIDAS, and the NPD. An application was put forward to HSCIC to use MIDAS as the sampling frame, and while permission was granted for its use in the What about YOUth Survey Trial Study, permission was not extended for its use as the sampling frame for the main What About YOUth Survey (details of the use of MIDAS in the Trial Study are provided in section 5.4.3). The NPD was therefore used as the sampling frame.

Parental Experience of Services for Disabled Children (2010)

The Parental Experience of Services for Disabled Children Survey was commissioned by the Department of Children, schools and families (DCSF), and carried out by TNS BMRB. It aimed to measure the views and experiences of parents in England with a disabled child aged between 0 and 19 on a number of areas related to services, including information, assessment, participation, and feedback. A postal methodology was used, with around 31,000 questionnaires completed, representing a response rate of 56 per cent.

The originally proposed sample design intended to use three databases: the CBR, the element of the NPD containing children with Special Educational Needs, and the Disability Living Allowance register. By using the CBR as the 'backbone' sampling frame, and supplementing it with the element of the NPD containing children with Special Educational Needs and the Disability Living Allowance register (in which survey eligibility was expected to be substantially greater), it was expected that near full coverage of the survey population would be achieved. However, neither the CBR nor the Disability Living Allowance register were made available in time for the survey, and consequently the entire NPD database was used as the sampling frame, with screening carried out to identify children with disabilities. This led to undercoverage of certain groups, primarily children under the age of 5 or over the age of 16, and children at independent schools. The technical report for the survey states that 'How far this level of non-coverage biases the survey estimates is unknown', but that 'so long as (a) the conditions for being listed

on the NPD remain the same, and (b) the broad distribution of household types remains stable, the level of non-coverage bias should hold steady across future waves of this survey. Consequently, any significant changes in performance indicator scores ought to reflect real change in performance rather than issues with the sample frame.'

5.4.4. Implications

While the NPD has been used successfully as the sampling frame for a number of surveys of pupils and parents, we do not recommend its use for the CEYSP as it only includes children of school age that attend a state school, and those attending funded early education, resulting in an unacceptable degree of undercoverage of the CEYSP survey population. There is no efficient approach that could be used to include those children that are not covered.

5.5 Databases of schools

5.5.1. Introduction

An alternative approach to using the NPD would be to sample pupils via schools. DfE maintains a database called EduBase that includes all schools in England[38]. This database can be used to sample schools, from which pupils can be sampled directly.

5.5.2. Characteristics

Accessibility

A dataset of all schools in England is publicly available and can be downloaded from the EduBase public portal hosted by the DfE[39].

EduBase does not contain information about individual pupils at each school. In order to draw a sample of pupils from selected schools, schools must either supply lists of pupils from which a sample can be drawn, or must carry out the sampling themselves.

Format and fields

The EduBase database is held in electronic format and can be readily manipulated. EduBase includes (or can be matched against the School Census to derive) information about the number of pupils in each school in England by gender and age, as well as a range of variables that could be used for stratification when selecting a sample of schools.

Coverage

EduBase includes all schools in England and Wales, meaning that only a very small proportion of school-age children are not covered (for instance, those who are home-educated). However, databases of schools allow one to select samples of school-age children only, meaning that children under 4 (and most children aged 4) are not covered.

Also of note is the fact that, over recent years, response rates among schools have experienced a stark decline. For instance, for the long-established Smoking, Drinking and Drug Use Among Young People in England survey (see section 5.5.3), 74 per cent of eligible schools took part in the survey in 2003, compared to just 40 per cent in 2014.

---

[38] The National Foundation for Educational Research (NFER) also maintains a database of schools in England, called the 'Register of Schools'.

[39] Available at: http://www.education.gov.uk/edubase/home.xhtml

Additional non-response among sampled pupils resulted in a net response rate of 35 per cent in 2014.

Stability

We are not aware of any plans to change EduBase, and as it is used by government for administrative purposes, it is highly likely that it will continue to be maintained in its current or equivalent form.

5.5.3. Use of databases of schools on similar surveys

Schools databases have been used as the sampling frame on a range of surveys, four of which we now describe.

These surveys have not covered children of pre-school age, given schools databases' undercoverage of this group. There is also evidence that response rates from surveys that sample schools are low and declining. None of these surveys provides a precedent to suggest that schools databases could be used successfully as the sampling frame for future waves of the CEYSP.

Smoking, Drinking and Drug Use among Young People in England (1982 - ongoing)

The Smoking, Drinking and Drug Use Among Young People in England is an ongoing survey series which started in 1982. The survey is commissioned by HSCIC, and the most recent survey in the series was carried out by NatCen in 2014. The survey collects information about the experience of smoking, drinking and drug use, including consumption of cigarettes and alcoholic drinks in the previous week, and awareness of specific named drugs. In the most recent survey, approximately 6,000 paper self-completion interviews were completed by secondary school pupils, aged 11 to 15, in England, representing a net response rate of 35 per cent.

Children are sampled from schools in a two-stage process. At the first stage, schools (a total of 522 in 2014) are sampled from the National Foundation for Educational Research (NFER)'s Register of Schools, with the probability that a school is selected proportionate to the number of pupils at the school. At the second stage, 35 pupils are selected in each school from the school's register of pupils.

While the pupil response rate within schools has remained at a similar level in recent years of the survey, the response rate among schools has declined; in 2003 74 per cent of eligible schools took part in the survey, compared to just 40 per cent in 2014. The main reasons for schools declining to participation in 2014 were: no available time for pupils to complete the survey; the burden of participation on staff or pupils; and a reaction to the large number of school surveys currently being conducted.

Second Longitudinal Study of Young People in England (LSYPE2): Wave One (2013)

The Second Longitudinal Study of Young People in England (LSYPE2), also known as Our Futures, was commissioned by DfE, and Wave One was carried out by TNS BMRB and GfK NOP in 2013. LSYPE2 intends to follow a sample of young people through the final years of compulsory education to other forms of education, training, employment, and other activities, and to collect information about their career paths, as well as their lives and experiences. Seven annual waves are planned, up to and including 2019.

The first stage of sampling involved sampling schools: the Schools Census was used to sample maintained schools and pupils, and EduBase was used to sample independent schools, within which pupils were sampled directly from pupil registers. For each sampled pupil, a face-to-face interview was sought with the pupil, and with a nominated parent, at the pupil's home. Interviews were carried out around 13,000 households, equating to a household level response rate of 72 per cent.

First Longitudinal Study of Young People in England (LSYPE1): Waves One to Seven (2004-2010)

The First Longitudinal Study of Young People in England (LSYPE1), also known as Next Steps, was commissioned by DCMS, and was carried out variously by NatCen, BMRB, NOP, and Ipsos MORI between 2004 and 2010. The study aimed to gather evidence about the transitions young people make from education or training to economic roles in early adulthood, to evaluate the effects of policy, and to provide an information base for future policy development. The study followed 15,500 young people in Year 9 (or equivalent) over seven annual waves, with interviewing primarily conducted face-to-face (with the option of other modes).

The first stage of sampling involved sampling schools: the Schools Census was used to sample maintained schools and pupils, and EduBase was used to sample independent schools, within which pupils were sampled directly from pupil registers. At Wave One 73 per cent of sampled schools co-operated with the study. School level non-response was a specific problem in inner London (56 per cent co-operating) and in the independent schools (57 per cent co-operating). Among co-operating schools at Wave One, 74 per cent of sampled pupils participated, resulting in a net response rate of 53 per cent.

Extended Schools Survey of Schools, Pupils and Parents (2008)

The Extended Schools Survey of Schools, Pupils and Parents was commissioned by DCSF, and carried out by Ipsos MORI. Its aim was to assess current provision, usage and perceptions of extended schools services among schools, pupils (aged 11 to 18), and parents (of children aged 5 to 18) in England. A total of around 2,000 postal and telephone interviews were carried out with schools (representing a response rate of 53

per cent); around 1,000 face-to-face interviews were carried out with parents (representing a response rate of 75 per cent); and around 1,000 face-to-face interviews were carried out with pupils (representing a response rate of 71 per cent).

The samples were drawn using a two-stage approach. First, schools were sampled from EduBase, and were contacted for the purposes of conducting the schools interviews. Second, samples of parents and pupils were drawn, using the NPD, from a subset of schools selected at the first stage.

5.5.4. Implications

We would not recommend sampling via schools for the CEYSP as children under school-age would be not be covered, and there is no efficient method to include them. Furthermore, relatively high levels of non-response among sampled schools can be expected, which would result in a far lower net response rate than that achieved from the current CEYSP design.

5.6 Respondents to another survey

5.6.1. Introduction

One approach to obtaining a sample of parents of children aged 0 to 14 would be to follow up respondents to another survey who have consented to be recontacted for further research. We would not recommend that the entire CEYSP sample was drawn in this manner, because of likely difficulties accessing a sufficiently large recontact sample, and because the net response rate would be far lower than the response rate the CEYSP currently enjoys; rather, another survey could be used to 'fill in' the CBR's undercoverage arising from the introduction of the HICBC. As described in section 5.1.2, we estimate that the introduction of the HICBC has resulted in around eight per cent of children born after 2013 being absent from the CBR. A cost-effective and efficient approach to obtaining samples of families missing from the CBR due to the HICBC is to sample them directly from another high quality survey, and invite them to participate in the CEYSP.

The Family Resources Survey (FRS) would be the ideal survey from which to select this sample because it asks respondents whether there are any children in the household for whom Child Benefit is not claimed because of the HICBC. We therefore discuss this approach with reference to the FRS[40].

The FRS is a continuous survey that was launched in 1992 to collect information about the living conditions and resources of households in the United Kingdom. It is commissioned by the Department for Work and Pensions (DWP), is carried out by ONS and NatCen, and is designated by the UK Statistics Authority as a National Statistic. The survey allows DWP to develop, monitor and evaluate welfare policies. Around 20,000 UK households are interviewed annually.

5.6.2. Characteristics

Accessibility

Obtaining a follow-up sample from a large-scale government survey requires permission from the government department that commissions the survey. In the case of the FRS,

---

[40] If it is not possible to follow up FRS respondents, another large-scale government survey could be considered, although we are not aware of any that collect the required information to identify the eligible sample. As a minimum, information about the presence of children in the household would be required (which all large-scale government surveys collect), and either household income, or individual income for each parent in couple households. If household income alone is available, it would be necessary to select all households with a child aged 0 to 14 with a household income of £50,000 or more, and carry out in-field screening to identify those households containing children for whom Child Benefit is not claimed because of the HICBC.

this is DWP. As described in section 5.6.3, a number of surveys have successfully followed up FRS respondents, so there is precedent for permission being granted. It is also possible that DWP would look favourably on a request from DfE to follow up FRS respondents given DWP's interest in the CEYSP.

However, permission may be refused if DWP receives other requests to follow up respondents (to avoid overburdening respondents), and while DWP may grant permission to use the FRS for the 2017 CEYSP, there is a risk that they would not commit to its use on the CEYSP on an ongoing (annual or biennial) basis, meaning that were the FRS to be used, further discontinuity in the sampling frame for the CEYSP should be anticipated for future waves of the survey.

Format and fields

The FRS data are held in electronic format, can be readily manipulated, and contain the relevant fields by which eligible families can be identified and selected for follow up research. We discuss how this selection might be carried out in section 5.6.4.

Coverage

The FRS is sampled from the PAF, and as such, has good coverage of the CEYSP survey population (we discuss the PAF's coverage of the CEYSP survey population in Section 5.2.2).

Stability

We are not aware of any plans to cut or substantively change the FRS. The FRS has been carried out continuously since 1992, has National Statistic status and provides the data for a number of other DWP National Statistic publications (including the Households Below Average Income series, and the Pensioners Incomes series), and is DWP's flagship survey. As such it is likely to continue in its current (or similar) form for the foreseeable future.

5.6.3. Use of the FRS on similar surveys

The FRS has been used as the sampling frame for at least three surveys, which we now describe. While these surveys successfully used an FRS follow-up sample, each was a stand-alone survey, rather than part of a tracking survey series. As such, none of these surveys provides a precedent for an FRS follow-up sample being used on a continual basis, as would need to be the case for future waves of the CEYSP.

Poverty and Social Exclusion in the UK study (2011)

The Poverty and Social Exclusion in the UK study was commissioned by the Economic and Social Research Council (ESRC), and was carried out in 2011 by a collaboration of

the University of Bristol, Heriot-Watt University, The Open University, Queen's University Belfast, University of Glasgow and the University of York. The study aimed to improve the measurement of poverty, deprivation, social exclusion and standard of living in the UK.

The sample was selected from respondents to the 2010-11 FRS who had given permission to be re-contacted. One reason for following up FRS respondents was to make analytical use of the wealth of data already collected from these respondents in the FRS. In addition, the FRS enabled boosts of BME respondents and those from lower income households.

Relationship separation and child support study (2008)

The Relationship separation and child support study was commissioned by DWP, and carried out by NatCen in 2008. The study aimed to gather the views of separated parents on the Government's plans for the redesign of the child support system. Around 2,000 face-to-face interviews were carried out with separated parents.

The sampling was designed to obtain samples for four groups from combinations of: parents with care (PSC) and non-resident parents (NRP), that had and had not used the Child Support Agency (CSA). The sample for the NRPs that had not used the CSA were selected from two successive years of the FRS (2004-5 and 2005-6). A total of 14,826 respondents from the FRS were selected based on the following criteria: that they were male, aged between 18 and 55, and had consented to be re-contacted.

The men identified from the FRS were contacted by telephone (using the telephone numbers collected in the FRS) and asked to take part in a short telephone screening interview. This interview established their eligibility for the survey by asking if they: were a separated parent; had at least one child under the age of 15 (or aged 16 to 19 who was in full-time education) who did not live with them; and had never used the CSA. The telephone screening identified 361 eligible individuals from the original 14,826 selected from the FRS, who were then issued for a follow-up face to face survey. Of these, 189 were interviewed.

Factors affecting the labour market participation of older workers study (2003)

The Factors affecting the labour market participation of older workers study was commissioned by DWP, and carried out by NatCen in 2003. The study aimed to identify the factors that encourage labour market participation and influence labour market withdrawal among older workers.

The sample was drawn from a combination of: i) respondents previously interviewed on the FRS, and ii) a boost sample of those receiving Incapacity Benefit and/or Income Support, from benefit records. With respect to the FRS, all 2001-02 FRS respondents

that had given permission to be re-contacted, and who were aged between 49 and 69 at the time of FRS interview, were selected for interview. The number of 2001-02 FRS respondents that satisfied these criteria was 3,219. The response rate among the FRS sample was 64 per cent.

5.6.4. Implications

We now describe a sample design in which the CBR is used as the core sampling frame, with the undercoverage of children in high income households addressed by following up respondents to the FRS survey. While this approach is promising, it would require DWP to commit to ongoing use of the FRS follow-up sample, prioritising the CEYSP over requests from other researchers (to avoid overburdening respondents). DfE would need to work closely with DWP to ensure this commitment for the foreseeable future. The sample design would also require the FRS sample to receive large weights, which would damage the statistical efficiency of the approach by a small amount.

Option E: Status quo, with follow-up of FRS respondents

Following up respondents to another survey, in particular the FRS, could mitigate the CBR's undercoverage resulting from the HICBC. Specifically, the FRS could be used to supplement a sample of children drawn from the CBR to give almost full coverage of children aged 0 to 14 in England.

The FRS currently asks direct questions about the HICBC. Those parents with a child or children who do not include any income from Child Benefit in their responses to the income questions in the FRS are asked the following 'check' question:

CBChk

Can I just check, you did not say you were getting Child Benefit, is this because of any of the reasons given on this card?

1.   You have chosen to stop receiving Child Benefit payments due to having a high income

2.   You have not applied for Child Benefit

3.   Your partner receives this benefit

4.   Other reason

This question explicitly identifies those parents that have not applied for Child Benefit, and who would therefore not be included on the CBR. The question also identifies those opting-out of receiving Child Benefit. While these parents are included on (and could be sampled from) the CBR, this provides an element of future-proofing should it ever be

decided that parents opting-out should be removed from the CBR altogether, or if the contact information of parents opting out becomes too inaccurate over time.

FRS respondents who do receive Child Benefit are asked a separate question, to confirm whether the HICBC applies to them:

CBTax

Can I check - do you pay a high income Child Benefit tax charge on the Child Benefit you receive?
[INTERVIEWER: Since January 2013 families where at least one parent earns £50,000 or more a year have had to pay a High Income Child Benefit charge on the Child Benefit they receive. This charge is paid through the tax system. Some families may have decided to stop receiving Child Benefit payments rather than repay the money through tax.]

1.   Yes [respondent] pays a high income tax charge

2.   Yes [respondent's partner] pays a high income tax charge

3.   No, a high income Child Benefit tax charge is not paid

Using these two questions, FRS respondents with children aged 0 to 14 can be classified into one of the following four categories:

1.   Not present on the CBR;

2.   Present on the CBR, receiving Child Benefit, and not subject to the HICBC;

3.   Present on the CBR, receiving Child Benefit, and subject to the HICBC (in part or in full); and

4.   Present on the CBR, but opted out of receiving Child Benefit.

Those respondents in category 1 (who are not present on the CBR, and whose child(ren) therefore cannot be sampled from the CBR), and in category 4 (who are present on the CBR but have opted out of receiving Child Benefit, and who may have a new child or children in the household that have not been registered for Child Benefit and who therefore cannot be sampled from the CBR) would be sampled from the FRS.

Analyses we have conducted on the FRS data show that, in the 2013-14 FRS, there were 3,780 participating households in England that contained a child aged 0 to 14. With around 76 per cent of FRS households consenting to be re-contacted (and if we assume this percentage is the same for households with children as other households) this gives a pool of 2,873 households with a child aged 0 to 14 consenting to be recontacted. This

figure is smaller than the target sample size for the CEYSP (6,350), meaning that a supplementary sample drawn from the FRS would comprise of fewer cases than are missing from the CEYSP. Weighting would therefore be required to make the FRS sample representative of children not covered by the CBR due to the HICBC.

If we assume a response rate of 70 per cent for the FRS follow-up sample, the FRS sample would have weights approximately 3.5 times larger on average than the sample drawn from the CBR. This would reduce the efficiency of the combined sample by a small amount, but not to the extent that the overall viability of the approach would be compromised.

This approach is therefore not an ideal solution as the limited sample size of the FRS means that those children not present on the CBR would be under-represented in the unweighted CEYSP sample. The approach would, however, with appropriate weighting, reduce the bias in survey estimates compared to using the CBR alone and therefore ignoring these children altogether.

Assuming that fieldwork for the 2017 CEYSP starts in January 2017, there is adequate time to incorporate a follow-up sample from the 2015-16 FRS. The approach would likely incur a small additional fieldwork cost to enable interviewers to travel to FRS addresses outside of their core postcode sectors. We estimate that the fieldwork budget required to deliver 6,350 interviews using the status quo approach would deliver approximately 6,000 interviews under this approach. It should also be noted that the FRS sample would undercover very young children in a similar fashion to the CBR sample. Just as those children born after the cut-off date for the quarterly CBR update that is used as the sampling frame do not have a chance of selection at the sampling stage, neither would those children born after the FRS interview date have a chance of selection at the sampling stage[41].

The timings would mean that very young children (born after April 2015 and not on the CBR) would be under-represented in the 2017 CEYSP, so there would still be a small coverage error with this approach.

As noted above however, the greatest shortcoming of this approach is that it requires a firm commitment from DWP that the CEYSP can use the FRS on an ongoing basis, and

---

[41] The impact of this undercoverage will depend on the combination of two factors: the precise timings for drawing the CEYSP sample, and how soon after FRS fieldwork is completed that DWP can share the FRS re-contact sample. As with the status quo approach taken for the CBR sample, it will be possible to correct for some of this undercoverage in-field, by re-selecting the child that is to be the focus of child-specific questions from all children in the household where a child has been born into the household after the date on which the FRS interview was conducted.

as such, DfE would need to work closely with DWP to ensure this commitment for the foreseeable future. For this reason, we do not recommend that future waves of the CEYSP attempt to incorporate a sample from the FRS.

5.7 Commercial databases

5.7.1. Introduction

Some surveys have sampled parents from databases held by private sector companies. These databases often consist of expectant parents, or those with very young children, who have provided their details to a private sector company so they can be sent promotional items, or 'freebies'. Examples of these companies include Bounty[42], and Emma's Diary[43].

5.7.2. Characteristics

Accessibility

Permission must be sought from the company holding the database for it to be used as a sampling frame. This decision is at the discretion of the company, and use of the database would be charged for.

Format and fields

Relevant commercial databases will be held in electronic format. The fields they contain will vary, and will generally be limited to those fields provided at registration. For instance, in the case of Emma's Diary, these fields include the subscriber's: name, postal address, email address, telephone number (optional), whether the subscriber is pregnant and if so their due date, and the number of children the subscriber currently has. We are not aware of any commercial databases that request information about whether or not a parent is subject to the HICBC, and given the business model of companies holding these databases, it is very unlikely this information would be requested.

Coverage

A commercial database will include only those parents who have voluntarily provided their details to the company holding the database for the purposes of receiving promotional offers, and as such, will undercover parents in the population to a large degree. To the extent that there are systematic and survey-relevant differences between those parents registering with such companies, and those not registering, one would anticipate bias to survey estimates.

It is also possible that some parents will have registered with a given company more than once, resulting in duplicate elements, and that some individuals without children will have

---

[42] See: www.bounty.com

[43] See: www.emmasdiary.co.uk

signed-up in order to receive promotional offers on behalf of someone else, resulting in foreign elements. Both of these issues are problematic for an effective probability based sample design.

Of greatest consequence, commercial databases of parents tend to be limited to those parents with very young children. For instance, Bounty and Emma's Diary both focus on promotional offers directed at parents with babies and toddlers. Such commercial databases could therefore only be used as the sampling frame for the CEYSP were they combined with another sampling frame or frames.

Stability

Commercial databases of parents are held by private companies, and as such, it is not possible to draw firm conclusions about their stability over time.

5.7.3. Use of commercial databases on similar surveys

We are aware of one published study that has used a commercial database to sample parents, the Relationship Support Trials for New Parents: Evaluation (2013). This evaluation used a commercial database to sample parents expecting their first child in the next three months, and as such does not set a precedent to suggest that a commercial database could be used successfully for the CEYSP given the wider survey population.

Relationship Support Trials for New Parents: Evaluation (2013)

The Relationship Support Trials for New Parents: Evaluation was commissioned by DfE, and carried out in 2013 by a consortium of TNS BMRB, Bryson Purdon Social Research (BPSR), London Economics and OnePlusOne. It used both the CBR, and Bounty, a commercial database, to sample parents. It is described in section 5.1.3.

5.7.4. Implications

Commercial databases of parents tend to cover only those parents with very young children (babies and toddlers), and even among these parents there will be a high degree of undercoverage (given the voluntary opt-in nature of registration to these databases) which can be reasonably expected to lead to bias in survey estimates. Because of these problems, we do not recommend that commercial databases are used as the sampling frame for future waves of the CEYSP.

6 Conclusions and recommendations

Before the introduction of the High Income Child Benefit Charge (HICBC) in January 2013, the Child Benefit Register (CBR) had very high coverage (around 98 per cent) of the CEYSP survey population (children aged 0 to 14 in England). The CBR has therefore been a suitable sampling frame for previous waves of the CEYSP.

The introduction of the HICBC in January 2013 has led to a tendency for parents with high incomes to not register their children for Child Benefit. This has led to around eight per cent of children born in England each year not appearing on the CBR. As a consequence, should the CEYSP continue to sample children exclusively from the CBR, one can expect survey estimates to become increasingly biased with each successive wave away from children in high income households, and it will be difficult, if not impossible, to attribute changes across waves to real changes in the population.

In Chapter 5, we described five potential sample design options for future waves of the CEYSP. We summarise these options in Table 6.1.

Table 6.1. Sample design options for future waves of the CEYSP

Option	Coverage of children aged 0 to 14 in England	Ability to attribute changes over time to real change	Main disadvantages	Int-erviews delivered [44]
A: CBR, Status quo	High (but falling)	Low	Falling coverage will lead to increasingly biased estimates, difficulties attributing changes over time to real change, and particular difficulties comparing children born before and after the introduction of the HICB.	6,350
B: CBR, but redefine the survey population	Medium	Very high	Idiosyncratic and arbitrary population definition will undermine survey estimates, and will prove difficult for data users to interpret.	c.6,000
C: PAF with full screening	Very high	Very high	Much costlier (than status quo approach) as large-scale in-field screening exercise required to assess eligibility of households.	c.2,900
D: PAF in areas affected by HICBC, CBR in areas unaffected by HICBC	High	High	Costlier (than status quo approach) as in-field screening exercise required to assess eligibility of households in PAF sample.	c.4,300
E: Status quo, with follow-up of FRS respondents	High	High	Access to FRS would need to be granted on an ongoing basis. Relatively large weights required for FRS sample. More complex sample design.	c.6,000

Of these options, we do not recommend that the CEYSP continues with its current approach (Option A: CBR, Status quo) because of the difficulties just described. We also do not recommend that the CEYSP implements Option B (CBR, but redefine the survey population) because of the problems this would cause analysts and policy makers in attempting to interpret and make use of the survey data. Option E (status quo, with follow-up of FRS respondents) would go a long way to solving the undercoverage of children in high income households; however, its success is entirely contingent on an ongoing commitment from DWP that access to the FRS follow-up sample would be granted, which presents too great a risk for us to recommend this option.

---

[44] This column shows the approximate number of interviews that could be expected to be delivered for a fieldwork budget that would deliver 6,350 interviews under the status quo CEYSP design. These figures assume a survey population of children aged 0 to 14 in England, with no boosts by age.

This leaves the two options that draw samples from the PAF. While Option C (PAF with full screening) is promising in terms of providing very high coverage of the survey population, the extensive in-field screening required would make fieldwork inefficient and far costlier than the status quo approach. Option D (PAF in areas affected by HICBC, CBR in areas unaffected by HICBC) provides a cost-effective approach through which high coverage of the survey population can be maintained, and fieldwork inefficiencies arising from in-field screening can be minimised. We therefore consider this approach to be the most appropriate design for the forthcoming waves of the CEYSP.

Appendix

Table 7.1. Summary of sampling frame characteristics[45]

Characteristic		CBR, before HICBC	CBR, after HICBC	PAF	MIDAS	NPD	Databases of schools	FRS follow-up	Commercial databases
Accessibility	Accessible	✓	✓	✓	✗	✓	✓	?	?
Format and fields	Held in electronic format	✓	✓	✓	✓	✓	✓	✓	✓
	Contains contact details	✓	✓	✓	✓	✓	✓	✓	✓
	Contains child-level details	✓	✓	✗	✓	✓	✗	✓	?
	Linkable to geodemographic data	✓	✓	✓	✓	✓	✓	✓	✓
Coverage	No missing elements (i.e no undercoverage)	✓	✗	✓	✓	✗	✗	✓	✗
	No foreign elements (i.e. no overcoverage)	✓	✓	✗	✓	✓	✓	✓	✓
	No duplicate elements	✓	✓	✓	✓	✓	✓	✓	✓
Stability	Stable characteristics over time	✓	✗	✓	✓	✓	✓	?	?

[45] Table 7.1 summarises the properties of the sampling frames considered in this investigation against the desired characteristics of the sampling frame for the CEYSP, as described in Chapter 4. A tick signifies that the frame possesses the characteristic, a cross signifies that it does not, and a question mark signifies that there is uncertainty in this regard. For full details of the characteristics of these sampling frames please refer to Chapter 5.

© Department for Education 2017

Reference: DFE-RR647

ISBN: 978-1-78105-704-9

The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Education.

Any enquiries regarding this publication should be sent to us at: max.stanford@education.gov.uk or www.education.gov.uk/contactus

This document is available for download at www.gov.uk/government/publications