KANTAR PUBLIC

Centre for Longitudinal Studies

Early Life Cohort Feasibility Study - Public Dialogue Report

May 2022

Amy Busby, Bethany Dokal, Lucy Williams, Danny Price

4032240

Acknowledgement

The Kantar Public team would like to thank Lisa Calderwood, Alissa Goodman and Karen Dennison for their input to and oversight of this project.

We would also like to thank and acknowledge the stakeholders who participated in the dialogue workshops and those who helped review the materials.

We also extend our thanks to the 59 parents who gave up their time to participate in the dialogue and generously shared their views and ideas.

Contents

1. Executive Summary

1.1 Background to the Research

Funded by the Economic and Social Research Council (ESRC) and led by the Centre for Longitudinal Studies (CLS), the Early Life Cohort Feasibility Study (ELC-FS) is a two-year project, that will test the feasibility of a new UK-wide birth cohort study.

Administrative data is intended to be used in a number of ways to help improve the representativeness and inclusivity in the ELC-FS – for sampling, recruitment and retention of participants and to enhance the study with data linkages to administrative records, including embedding the ELC-FS in administrative data to a greater degree than previous birth cohort studies. The study team has undertaken an extensive consultation and development phase to inform the design of the ELC-FS and as part of this Kantar Public was commissioned in 2021 to undertake stakeholder interviews and a public dialogue. This report outlines the findings from the public dialogue workshops, which explored the attitudes of **parents of young children** to the proposed uses of identifiable and de-identified administrative data in the ELC-FS and their acceptability.

1.2 Research Design

1.2.1 Public dialogue

Public dialogue is a specialist qualitative research approach that involves citizens in complex and sensitive decision making and enables the public to shape processes and regulations that impact their lives. Dialogue involves extended discussions, to allow citizens to reach informed, considered and insightful views about complex policy processes by bringing together a diverse mix of the public with relevant policy makers, experts and information.

This dialogue comprised of two waves of workshops held online with the same participants from Northern Ireland, Scotland, Wales, North of England and the Midlands, and South and East England.

The first wave workshops were held on the 18th and 25th September 2021 and the second wave workshops were held on the 30th October and the 6th November 2021, and were conducted online via Zoom. In total, 59 parents of young children completed the dialogue process. Workshops were held on Saturday mornings and lasted three hours. Each workshop involved three break-out groups and each was accompanied by a Kantar Public moderator and an expert stakeholder, who were mainly administrative data users or administrative data controllers, and included members of the CLS-led ELC-FS project team.

1.2.2 Sampling

The dialogue was carried out with parents with young children aged six months to three years old. These parents were able to reflect on their recent experiences of parenthood.

A purposive sampling approach facilitated a demographically diverse sample that also included lesser-heard voices in longitudinal research. The sample was intended to reflect the general composition of the UK, in order to allow a diverse range of voices to be heard. Minimum quotas were set to ensure a range of demographic characteristics including for gender, socio-economic group, ethnicity, household configuration, whether this was the parent’s first child, and attitudes towards data sharing.

1.3 Key Findings

Some key themes emerged which contextualise the dialogue findings including the insight that **data was widely seen as something that is never fully secure.** There were **mixed levels of trust in public bodies to handle and share data,** although few parents were very trusting. **Data was also understood by parents to be commercially valuable**, in particular to private companies. A further important theme that emerged was that parents felt that **data that is about them belongs to them**. This meant parents saw their involvement in birth cohort studies as **'partnerships'** between the study team and participants and they wanted to be treated accordingly. The final key theme that emerged which is important to contextualise the dialogue findings was that **trust is key to parents' signing up to be part of a birth cohort study** and their on-going participation in it.

Overall, parents generally found most of the proposed uses of administrative data acceptable, as long as certain conditions were in place. Across the workshops, **seven factors emerged as key drivers of acceptable use of administrative data** in the ELC-FS:

- **Level of transparency:** Across all of the proposed uses of administrative data, where parents were able to understand who was using their data, why it would be used, the value of the use, and how it would impact on them and their child, they found this more acceptable.
- **Level of risk of potential harm to the individual**: Where parents thought there was a lower level of potential risk of harm to the individual, they also tended to view the use as more acceptable. The three types of harm parents were most concerned about were: (i) theft and hacking, (ii) the fear of social stigma, and (iii) lost freedom of choice.
- **Degree of control**: Parents feel they own data that is about them and therefore, where they feel they have a greater degree of choice over how their administrative data will be handled, they found the use more acceptable. Conversely, where parents felt pressurised or coerced into sharing their data, parents found this less acceptable.
- **Level of clarity:** Where information and communications are provided about exactly how data was being used and where the processes involved were clear, parents tended to find the proposed uses of administrative data more acceptable as it helped parents feel more informed and increased their trust in the study.
- **Level of data security in place**: Reassurances around data security also increased the acceptability of the proposed uses of administrative data.
- **Level of necessity:** Where the proposed use of administrative data was understood to be essential to the implementation of the study or a particular aim, then this was more acceptable to parents.
- **Degree of normalisation**: The proposed uses of administrative data were seen as more acceptable when parents knew data was used **routinely in this way**. This gave reassurance to parents that the proposed uses of administrative data were well established and the security measures involved robust.

The following sections of this Executive Summary discuss key findings from this qualitative research for each specific proposed use of administrative data in the ELC-FS.

1.3.1 Views about the proposed approach to sampling

Parents were introduced to the proposed sampling approach for England and Wales, Scotland and Northern Ireland.

Parents widely found the proposed sampling approach, using birth registration and maternity records data, acceptable as it was seen as necessary to access a sample through which the study team could find and contact parents to invite them to participate in the research. It was also seen as important to ensure the sample was representative and to help improve the inclusivity of the research, which was widely seen as important to parents.

1.3.2 Views about potential recruitment approaches

Parents were also presented with information detailing three different recruitment approaches (the One Step, Two Step and Alternative Two Step approach) which the study could take when accessing administrative data to sample families for the study and for contacting parents for recruitment to the study.

There was a strong preference for the Two Step recruitment approach, in which families would be able to opt-out before their details are passed on to the study team by the data holder. It felt intuitively more suitable than the One Step recruitment approach, which does not have a prior opt-out, to parents and met their expectations of what the research team should do. This was because the Two Step approach **increases the control** offered to participants over their administrative data and **improves transparency.**

Although most parents did not support the One Step recruitment approach, in which there is no prior opt-out before their details are passed on to the study team by the data holder, some parents saw benefits to this, including simplicity, being quicker and less resource intensive and likely to deliver a more inclusive sample. Some parents who had lower concerns about the risk of personal harm found the One Step approach acceptable.

1.3.3 Views about use for targeted recruitment and boosting

Parents were introduced to two methods of targeted recruitment that could help improve inclusivity of under-represented groups in the study sample:

1. *Tailored recruitment: This could involve having specific recruitment letters and leaflets that aim to be more appealing to certain groups or emphasises why it is important for them to take part or asking interviewers to make extra efforts to interview these groups.*
2. *Boosting: This is when the study recruits additional participants from groups that form smaller parts of the population; e.g. those from minority ethnic backgrounds. This process allows more meaningful findings to be drawn about these groups and uses sampling frame variables to identify specific individuals to recruit.*

Data use for targeted recruitment was generally acceptable to parents as long as it is done sensitively to avoid generating feelings of stigmatisation. Boosting also made sense to parents and was seen as important to improving the social value of the findings of the study by further ensuring diverse experiences are represented in research findings.

1.3.4 Views about use of administrative data for retention purposes

It was explained to parents that as study participants are randomly sampled in birth cohort studies, they cannot be replaced. This means it is important to keep participants in the study. Parents were further told how identifiable administrative data can be used to trace participants the study team has lost contact with because they have moved and not informed the study team of their new address.

Parents found tracing acceptable if participants were clearly informed about it during recruitment as this would ensure they are offered control and transparency over administrative data use. However, most parents thought tracing using administrative data should only be undertaken once other contact methods provided by study participants had been used which do not use administrative data, because they were considered more suitable and simpler.

1.3.5 Responses to proposed uses of de-identified administrative data in the ELC-FS: for linkage to survey responses

Parents were introduced to this proposed use of de-identified administrative data and shown two videos (from ADRUK and Next Steps) outlining what data linkage is, how it is carried out and the measures in place to protect the security of the data.[1][2] *Parents were also shown information highlighting the benefits and risks of data linkage and discussed which were the most important to them. After this, parents were shown further information about what types of data the ELC-FS study team wish to link with study participants' survey responses and parents discussed in a sorting exercise which types of data they were the most and least comfortable being linked, including linking to health, social services, income and benefits, education and criminal records data.*

Overall, parents were supportive of data linkage to survey responses. This was because the use of de-identified administrative data was perceived to have low levels of risk of potential harm to the individual and high levels of data security were understood to be in place. However, parents had greater reservations around the study team accessing more sensitive types of data including mental health, social services and

---

[1] Administrative Data Research UK – ADR UK, https://www.youtube.com/watch?v=ij4mNEFBdyg

[2] Next Steps, longitudinal study of young people in England born 1989-90. https://www.youtube.com/watch?v=W6ZuK3lYW6Y

criminal records information. Parents tended to have fewer concerns over sharing their own data compared to their child's.

1.3.6 Views about potential consent models for data linkage

Parents were presented with five potential consent models to collect participant agreement to the proposed data linkages.

①

Participants are not told their survey responses will be linked to their admin data at any point

It **may not be legally required** for participants to be told or to ask for their permission.

②

Participants are told their admin data will be linked to their survey responses and by signing up to the study they are agreeing to this

The initial **invitation letter and booklet** would explain their admin data will be collected and analysed if they participate in the study.

③

Participants are told their admin data will be linked to their survey responses, unless they say they do not want this to happen

During the survey, when told this information, they would need to spontaneously say 'no' to opt-out of this. This information would also be in the invitation letter and booklet

④

Participants are told the types of admin data that will be linked and are asked to select if they:

4a.

Agree or disagree to all of them

During the survey, they would be asked **one** question and they can select if they agree or disagree to this. Information about data linkage would also be in the invitation letter and booklet

4b.

Agree or disagree for each type

During the survey, they would be shown a list of different admin data the study would like to link to, and asked to select if they agree for each type. Information about data linkage would also be in the invitation letter and booklet

Overall, parents had a good understanding of the key issues surrounding consent models for both the participant and the study team. This included the competing demands of the study team's need to collect as many linkages as possible from as many study participants as possible to improve the inclusivity of the study findings, to build a more detailed picture and to complete further analysis, balanced against the study participant's desire to select which linkages to agree to, which is likely to diminish the number of people consenting and consents collected.

There were mixed preferences regarding the consent models; although **Model 4B (Participants are told the types of administrative data that will be linked and are given the option to agree or disagree for each type)** was overall their preferred model and **Model 1 (Participants are not told their survey responses will be linked to their administrative at any point)** was seen as the least acceptable as it offered no choice to study participants and they were not informed of the linkages. For some, **Model 2 (Participants are told their administrative data will be linked to their survey responses and agree to this by signing up to the study),** whilst not their preference, was seen as sufficiently acceptable.

Despite parents having mixed views around their preferences for consent models there was consensus around principles that should be used regardless of approach, and these included ensuring **transparency, minimising participant burden, and providing reassurances to build trust.**

1.3.7 Views about use of de-identified administrative data to check representativeness of the study sample

Parents were introduced to the concept of non-response bias which explained that certain types of families are less likely to take part in studies like the ELC-FS, leaving them underrepresented in social research. Parents were shown stimulus materials that identified the different groups that would make up the selected sample and the study sample, to explain how de-identified data could be used to check representativeness. These groups were:

- *The overall selected sample*
- *Those in the selected sample who had opted-out when receiving the letter from data controller*
- *Those in the selected sample who opted-out when receiving an invite letter from the study team*
- *Those in the selected sample who were invited by the study team, but did not respond to their invite letter or later recruitment attempts*
- *Study sample, which is those who have taken part in the study*

Initially parents generally found the concept of use of de-identified administrative data to check the representativeness of the study sample acceptable due to the social value of having a representative sample. However, when exploring the processes of checking representativeness in greater detail, parents' views evolved and became more nuanced. This was primarily due to concerns about this use of data for those who opted out of their data being transferred from the data controller to the study team. A range of views emerged as to whether it was acceptable to use the data of those who had not responded to the study team's recruitment attempts, but parents widely felt that using the administrative data of those who had explicitly opted-out of the study either with the data controller or with the study team was unacceptable.

Some parents were comfortable with the data controller using data for those who didn't take part in the study to check the sample representativeness as the data would remain with the data controller and not be passed to the study team.

1.3.8 Using weighting to improve representativeness

Parents also discussed weighting as a potential approach the study team could use to improve the study sample's representativeness which uses de-identified administrative data.

Parents widely found the use of de-identified data, including on the selected sample from the sampling frame, for weighting acceptable but struggled to understand this highly technical concept.

1.4 Conclusions and recommendations

The following key drivers of acceptable use of administrative data have emerged and can used to inform decisions to help ensure that use of administrative data is acceptable for potential study participants.

- **Be transparent: inform and explain.** Transparency around the use of their administrative data was a key driver of acceptability for parents. Across all the proposed uses of administrative data, parents wanted to know who was using it, why it is being used, the value of the use, and how it will impact on them and their child, and that it is not being used for commercial gain.
- **Paint a picture for participants** of the social value of the study, including inclusivity – for society and individuals. Where parents could see the value of the ELC-FS and the impact it could deliver, this tended to inform how acceptable they found the particular uses of administrative data were.
- **Give control by giving choice** about whether and how data is used, as parents see themselves as the owners of their data and want to be treated as partners when it is used.
- **Set the scene** of research norms. Explaining when and how data is already used routinely in research can help to reassure participants that this is not a new approach and that established practices and processes are in place to help reassure them about the potential for data loss or misuse to occur.
- **Reassure about data security.** Provide information about robust processes and procedures that are in place to protect data as this increases the acceptability of the proposed uses of administrative data.

2. Research Design

2.1 Background to the Research

Funded by the Economic and Social Research Council (ESRC) and led by the Centre for Longitudinal Studies (CLS), the Early Life Cohort Feasibility Study (ELC-FS) is a two-year project which started in April 2021, that will test the feasibility of a new UK-wide birth cohort study.

The study aims to test the feasibility of a major new national study that will paint a nationally representative picture of the circumstances and lives of a new cohort of babies born in the 2020s. The primary scientific aim is to understand how inequalities in early child development are changing over time, and to learn whether the social and biological factors driving these trajectories are evolving and in particular to look at the following themes: cognitive, social, and emotional development of infants, infant-parent relationships, and the early home environment; infant health, including growth, nutrition and sleep; mental health of parents and the developing child; social, environmental and neighbourhood influences on infant and family; inequality, disadvantage, and social mobility; genomics, early adversity and biological embedding of stress.[3]

The feasibility study is planning to recruit around 3,000 families with new babies in their first year of life, collecting information about their families and their development through face-to-face interviews with both mothers and fathers, combined with innovative measures to capture babies' development and interactions with parents. These may include direct observations and recordings by trained fieldworkers in the home, and data collection via a smartphone app and wearable sleep and activity devices. Bio-samples for DNA extraction are also planned for a sub-sample of families.

The feasibility study will be evaluated, and if judged to be successful, the commissioning of a new, larger main study is anticipated.

A representative sample is key to the robustness of national birth cohort studies such as ELC-FS. Inclusivity will be key to the ELC-FS, which will work to include the voices of groups who are often under-represented in these studies such as minority ethnic groups and own-household fathers (OHFs) who live apart from their children. In order to help improve the representativeness and the inclusivity of the study, the ELC-FS is intended to be embedded in administrative data to a greater degree than previous birth cohort studies. Administrative data is planned to be used in three ways to help improve inclusivity in the ELC-FS:

- Firstly, it will be used as the **sample frame**. Ideally this will involve using linked birth registrations and NHS maternity records to sample participants and build a more comprehensive and inclusive sample. Securing access to a high-quality sample frame will be crucial to the feasibility of the study. The approach to sampling will vary across the four nations.
- Secondly, for **operational purposes** including the recruitment and retention of participants. This will include using administrative data for targeted recruitment approaches and to assess representativeness and for non-response analysis and adjustment (e.g. weighting). It will also be used to trace those who drop out of the study.
- Lastly, for **substantive research and analytical purposes**, the study data will be enhanced with data linkage to administrative records – for the child, parents and potentially other family members (including siblings). This could include health, education and financial records as well as criminal and social services data. A range of consents will be required, and maximising consent rates will be important to the success of the feasibility study.

The study team has undertaken an extensive consultation and development phase to inform the design of the ELC-FS. This phase has included consultation on draft questionnaire and non-questionnaire measures with academic, policy and practice communities, qualitative research with own household fathers and low-income families, focus groups with parents and young people, and interviews with stakeholders and a public dialogue with parents. As part of the development phase, the study team also sought and were granted ethics approval from the Institute of Education's (IOE) Research Ethics Committee.

---

[3] https://cls.ucl.ac.uk/cls-studies/early-life-cohort-feasibility-study/

Kantar Public was commissioned in 2021 to undertake the stakeholder interviews and the public dialogue. This report outlines the findings from the public dialogue workshops. The findings from the stakeholder interviews with data controllers and data holders are available in a separate report.

The stakeholder interviews provided insight into the feasibility and ethical considerations of the proposed uses of administrative data in the ELC-FS from a specialist perspective, whilst the public dialogue provided insight from a public perspective into the acceptability of these uses. Due to timing constraints, the stakeholder phase and the public dialogue ran concurrently, which meant the extent to which the findings from the stakeholder interviews fed into the development of the dialogue materials was limited.

In summary, the stakeholder phase consisted of **14 60-minute depth interviews**. Recruitment of the stakeholders drew on CLS' existing networks, and these fell into the following broad categories across academic, policy and practice communities: **data holders** (e.g. of birth registration records and NHS maternity records) and **data users** (e.g. academic, government and industry researchers). The interviews were conducted remotely either by telephone or Zoom.

2.1.1 Research objectives

The objectives of these public dialogue workshops were to explore the attitudes of **parents of young children** (aged 6 months to 3 years) to the following proposed uses of identifiable and de-identified administrative data in the ELC-FS and their acceptability:

- **Sampling:** Exploring views about the use of linked birth registration data and NHS maternity records as the sample frame for this project, and preferences between a one and two-step recruitment process
- **Operational:** Exploring views about the use of administrative data to:
    a. Assess the representativeness of the sample (to assess non-response bias and assist weighting);
    b. For improving representation of minority ethnic groups and other groups (through boosting)
    c. For targeted recruitment approaches (e.g., engaging teenage mothers)
    d. To help with retention in the study (e.g., to trace people when they move)
- **Substantive:** Exploring views about:
    a. Proposed data linkages - for the child and parents (including potentially health, education, HM Revenue and Customs (HMRC), Department for Work and Pensions (DWP), and social services data)
    b. Proposed consent models (including implicit/ explicit, and opt-in/ opt-out models)
    c. Substantive use of administrative data to assess and counter non-response bias, including over time to assess non-response bias by comparing characteristics of cohort study families to the general population.

2.2 Research Design

2.2.1 Public dialogue

Public dialogue is a specialist qualitative research approach that involves citizens in complex and sensitive decision making and enables the public to shape processes and regulations that impact their lives. Dialogue involves extended discussions, to allow citizens to reach informed, considered and insightful views about complex policy processes by bringing together a diverse mix of the public with relevant policy makers, experts and information.

This dialogue brought together parents of young children (aged six months to three years) to explore their views about and the acceptability of the proposed uses of administrative data in the ELC-FS. It comprised of two waves of workshops held online with the same participants from Northern Ireland, Scotland, Wales, North of England and the Midlands, and South and East England.

The first wave workshops were held on the 18th and 25th September 2021 and the second wave workshops were held on the 30th October and the 6th November 2021, and were conducted online via Zoom. In total, 59 parents of young children, reflecting a range of demographics and attitudes to data privacy, completed the dialogue process. Workshops were held on Saturday mornings and lasted three hours. The moderators of each workshop followed a topic guide to ensure consistency across all workshop locations. Each workshop involved three break-out groups and each was accompanied by a Kantar Public moderator and an expert stakeholder, who were mainly administrative data users or administrative data controllers, and included members of the CLS-led ELC-FS project team.

Across both waves, parents were introduced to the potential uses of administrative data. The first wave focused on the use of administrative data for sampling and operational purposes, which involved the use of identifiable data. The second wave focused on analytical and substantive uses of administrative data including linkage to survey responses and checking non-response bias, which would be using de-identified data. Parents were presented with a range of stimulus to introduce them to the uses of administrative data and completed tasks which helped to facilitate discussion and engagement with the proposed uses. The full stimulus materials used can be found in the Workshop Materials Appendix.

2.2.2 Stakeholder involvement in the dialogue

The involvement of expert stakeholders with a range of perspectives is key to the success and credibility of public dialogues. Successful public dialogue relies on genuine two-way engagement between public participants and experts to ensure citizens can impact on the policy process. Involving stakeholders in public dialogues, and harnessing their expertise, also ensures that the information presented to participants is accurate, up to date, and balanced. In this dialogue stakeholders:

- Reviewed and inputted into workshop materials to ensure that they were accurate and balanced
- Attended workshops with the public to observe and provide expert knowledge
- Offered insight during the workshop debrief sessions to enrich analysis and build trust in the findings of the research.

Fifteen expert stakeholders took part in the dialogue, from a range of academic and professional backgrounds and represented a range of views on the topic. Seven of these stakeholders were part of the ELC-FS study team. Stakeholders attended an online briefing session once recruited, received a detailed stakeholder briefing pack outlining their role, attended a short online briefing session before each wave, and took part in a de-brief session after each wave. The stakeholders took part in the workshops, allowing participants to talk directly with experts and ask them questions. This involved attending the plenary sessions and then the break-out groups, where they could provide expert input and answer technical questions.

2.2.3 Content of the workshops, and pre and interim tasks

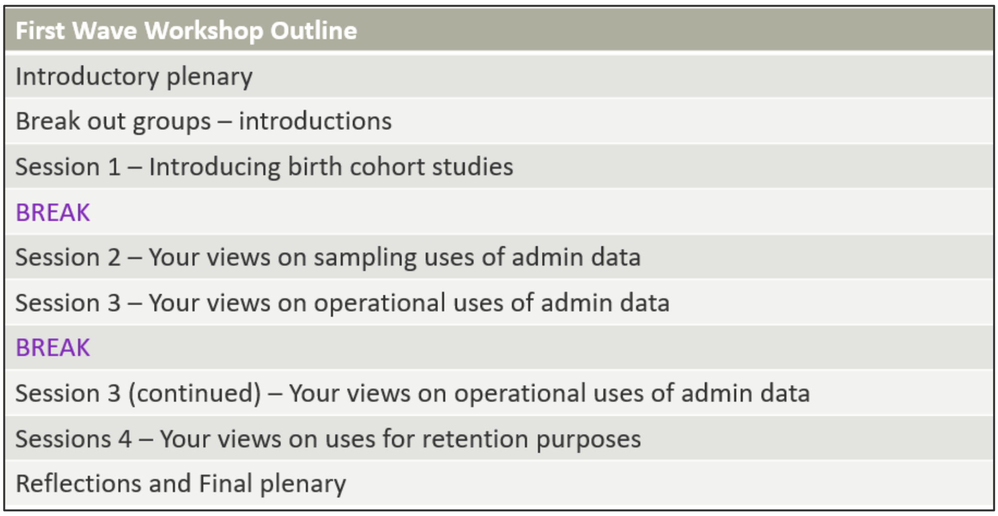
Prior to taking part in the first workshop, participants were sent a task to complete and return to the Kantar Public study team. They were asked to watch a short video about a previous birth cohort study, 'The National Child Development Study', providing context and preparing participants for the topic of discussion at the upcoming workshop.[4] Parents were then asked to answer four questions which gauged their familiarity with birth cohort studies, their views on their potential benefits, their concerns, and their initial interest in taking part in similar studies.

The first wave of the dialogue focused on the uses of administrative data for sampling and operational purposes. The groups began with discussion about participants' views on administrative data and birth cohort studies in general. Parents were then introduced to the ELC-FS specifically, with moderators explaining how administrative data would be used in the study. These uses included using administrative data to create the sample frame, to recruit participants and to trace participants if the study team lost contact with them. Moderators then led discussions and tasks to explore parents' views about these proposed uses. The outline for the first wave workshop is shown below in Figure 1.

---

[4] The National Child Development Study, https://ncds.info/home/about/ncds-at-60/

Figure 1 Outline for the first wave workshop



First Wave Workshop Outline
Introductory plenary
Break out groups – introductions
Session 1 – Introducing birth cohort studies
BREAK
Session 2 – Your views on sampling uses of admin data
Session 3 – Your views on operational uses of admin data
BREAK
Session 3 (continued) – Your views on operational uses of admin data
Sessions 4 – Your views on uses for retention purposes
Reflections and Final plenary

Between the first and second wave of workshops, parents were asked to complete a “Homework task”. Parents were sent a fictional information sheet and invitation to take part in the ELC-FS. They were asked five questions concerning their attitudes to taking part in the study, following the first dialogue session and to explore the views of their families after having discussed the study at home.

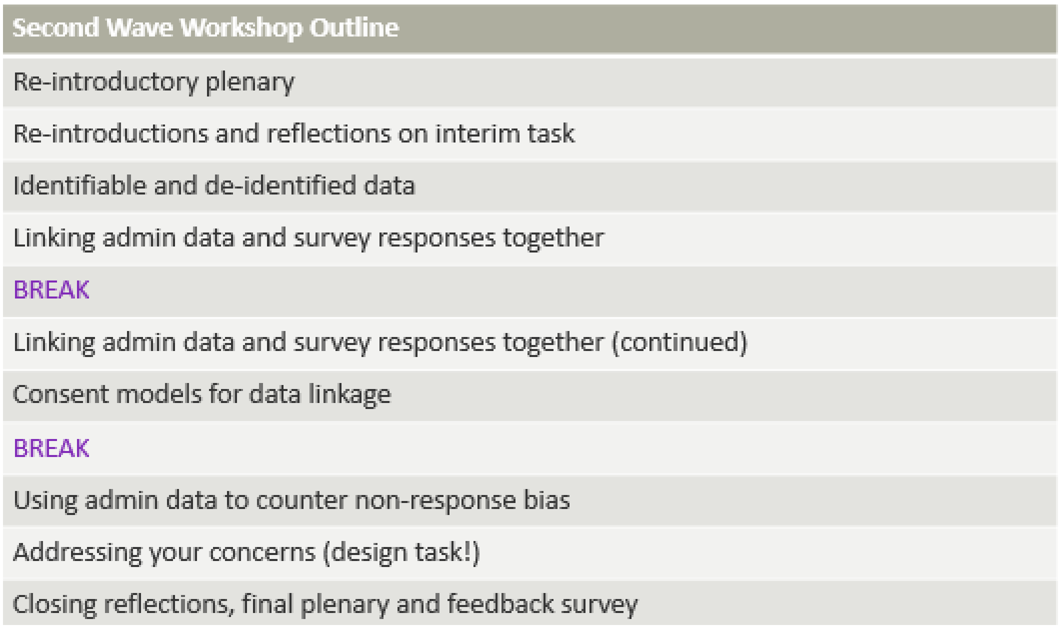
The focus of the second wave of the dialogue was on the substantive uses of administrative data. Parents were introduced to two types of administrative data, identifiable and de-identified data. They were then shown two videos outlining how de-identified administrative data from different sources can be linked together and how it can be linked to survey responses, to help answer questions about society.[5] Moderators then discussed with parents which administrative data linkages they would be comfortable with and how participants should be informed about the use of their administrative data in the study. Moderators also explained how the study team could use administrative data of those involved in the study and compare it with the selected sample to check representativeness. The final task of dialogue asked participants to rank the study team’s proposed uses of administrative data from most to least acceptable. The outline for the second wave workshop is shown below in Figure 2.

---

[5] Administrative Data Research UK video, https://www.youtube.com/watch?v=ij4mNEFBdyg

Next Steps video, https://nextstepsstudy.org.uk/home/faqs/adding-other-information/

Figure 2 Outline for the second wave workshop



Second Wave Workshop Outline
Re-introductory plenary
Re-introductions and reflections on interim task
Identifiable and de-identified data
Linking admin data and survey responses together
BREAK
Linking admin data and survey responses together (continued)
Consent models for data linkage
BREAK
Using admin data to counter non-response bias
Addressing your concerns (design task!)
Closing reflections, final plenary and feedback survey

For a more detailed account of what was covered in the dialogue workshops, please see the topic guides and stimulus materials used for both waves, included in the Workshop Materials Appendix to this report.

2.3 Sampling

The dialogue was carried out with parents with young children aged six months to three years old. These parents were able to reflect on their recent experiences of parenthood and have children at a similar age to when families are likely to be recruited to the ELC-FS.

The sample included parents from across the four nations of the UK and the workshops were structured by country and region to reflect the different legal and ethical frameworks in place across nations. There were two workshops for England, and one for each of Scotland, Wales, and Northern Ireland.

Beyond this, a purposive sampling approach facilitated a demographically diverse sample that also included lesser-heard voices in longitudinal research (e.g. minority ethnic groups). The sample was intended to reflect the general composition of the UK, in order to allow a diverse range of voices to be heard, and minimum quotas were set to ensure a range of demographic characteristics. The key demographic variables for this dialogue were gender, socio-economic group, ethnicity, household configuration, whether this was the parent’s first child, and attitudes towards data sharing. These variables underpinned the project’s recruitment approach.

The dialogue involved five workshops with a total sample of **63 participants at the first wave (comprised of 31 fathers and 32 mothers) and 59 at the second wave (comprised of 27 fathers and 32 mothers)**. To be eligible, parents needed to have access to Zoom, appropriate technology to participate fully, and a reliable internet connection. Those who work in research, are or have been part of a research panel, or have been involved in relevant campaign groups / activities were excluded. Parents were recruited using free find methods and received £140 to thank them for their time.

For a complete account of the achieved sample please see the Workshop Materials Appendix.

2.3.1 Qualitative analysis approach

The following types of data were created from the workshops: video recordings of each breakout group, moderator notes, and stakeholder notes/feedback. These raw data were all used in the analysis process.

The analysis process was comprised of two elements: process-driven and interpretative. The process driven element used a matrix mapping framework technique, in which data was coded and systematically summarised into an analytical framework organised by issue and theme. The framework was developed to reflect the research objectives, the workshop topic guides and the themes which emerged from brainstorm sessions following each wave and data coding. Through the framework, the video recordings were systematically recorded (including verbatim quotes) according to emerging hypotheses, which enabled some comparison across areas and key sub-groups of interest (e.g. regions, attitudes to data sharing, social grades, and ethnic groups). Video recordings of the workshop session were analysed, as well as moderator notes from after the workshops and notes from stakeholder feedback.

The second analysis element, which was interpretative, focused on identifying features and patterns within the data, mapping the range and nature of data, finding associations, defining concepts, creating types, and undertaking sub-group analysis. This process created descriptive accounts and explanatory data, which came not only from aggregating patterns but by weighing up the salience and dynamics of issues and searching for structures within the data that have explanatory power. Researcher analysis sessions were used to support interpretation of the data, during which the research team came together to discuss and test emerging themes and insights.

2.3.2 Limitations of the dialogue research

There are three key limitations to keep in mind when reading this report:

- **Non-generalisable findings**: The study used a qualitative sample with purposively selected parents. Findings are not representative and generalisable in a statistical sense, but are reflective of views of the population for the dialogue in a broader sense.
- **The parents became better informed about the topic over the course of the dialogue**: Having attended two three-hour dialogue sessions the parents in this study were provided with more detailed information about the use of administrative data in the study than those who will receive recruitment information for the ELC-FS. It is important to note this context when reading the report.
- **This report is limited to what was explored directly in relation to the research objectives for the dialogue and not all aspects of the ELC-FS**. Parents were provided with limited information about the design of the study and the study team's planned data collection approaches (e.g. biological samples). This should be considered when interpreting findings regarding parents' attitudes towards the proposed study and its elements.

2.3.3 Reading this report

The chapters of this report are presented by topic and not necessarily in the order they were discussed in the workshops. Chapter 2 provides key contextual factors which should be considered when reading this report and outlines the key drivers of acceptable use of administrative data and four types which help to explain variation in views across the sample throughout the report. Chapter 3 reports on parents' views about the proposed uses of identifiable administrative data in the ELC-FS and Chapters 4 and 5 report on their responses to the proposed uses of de-identified administrative data. Chapter 4 reports on the discussions about proposed linkage of administrative data to survey responses while Chapter 5 focuses on views about using administrative data to check the representativeness of the study sample. Whether the administrative data was identifiable or de-identified was dictated by its research purpose and informed how acceptable parents felt the use was and therefore report findings are structured accordingly for ease of reading.

Verbatim quotes are used throughout the report to illuminate and bring to life key findings and are attributed as follows: *"Quote."* (Location, Wave).

Due to the purposive sampling approach, we are unable to break down findings by all quota categories, but key subgroup differences have been pulled out where possible.

3. Context and key themes: drivers of acceptable use of administrative data in the ELC-FS

This chapter outlines some key contextual factors which should be kept in mind when reading the rest of this report on parents' responses to the specific proposed uses of administrative data in the ELC-FS.

It first briefly outlines some relevant findings about parents' wider attitudes to use of data which shaped their responses throughout the dialogue, it then introduces seven key drivers of acceptable use of administrative data in research which emerged from across the dialogue discussions and are applicable across the specific proposed uses. This chapter introduces four types of attitudes which emerged from across the sample and which help us to understand the variation in responses to the specific proposed uses of administrative data.

At the start of the dialogue, parents were introduced to the concept of birth cohort studies and then the plans for the ELC-FS specifically. Materials were shared to explain the importance of a representative and inclusive sample for the robustness of the results produced. We note (see Chapter 1 on research design) that the dialogue participants were therefore substantially more informed about these topics than participants who would be recruited the ELC-FS itself. More information about the discussions and the materials used can be found in the Workshop Materials Appendix.

Key findings:

- ***Views towards data:*** *Parents understood data as something which is never fully secure but is also valuable. Parents commonly felt data that is about them belongs to them. For parents to sign up to the study, they need to trust the study team and feel they are 'partners' in the research study.*
- ***Drivers of acceptability****: seven key drivers emerged which informed how acceptable parents found the proposed uses of administrative data: level of transparency, level of potential risk/harm to the individual, degree of control, level of clarity, data security, level of value, and degree of normalisation.*
- *There was some* ***variation in views across the sample*** *with four broad 'types' of attitudes emerging: Engaged Enthusiasts, Amenable Accepters, Engaged Sceptics, and Entrenched Cynics. Across these types there were varying levels of trust in sharing their data and varied views on how valuable they thought the birth cohort study was. Across the types, there was variation in the amount of information about the ELC-FS they required.*
- *Overall, there was a widely positive response to the ELC-FS and parents widely thought now is a* ***good time for a new birth cohort study*** *to understand more about children's' development, especially in light of Covid-19 and Brexit. However, parents were concerned by the* ***level of commitment*** *required, (although parents often assumed the time commitment would be greater than it would actually be) the* ***potential for harm,*** *and how they may feel '****judged****' as a parent if they participate in the study.*

3.1 Views about use of administrative data in social research

At the beginning of the first wave, parents were introduced to the term 'administrative data' and groups shared views around how comfortable they were with public organisations holding and sharing this data. Initially, parents were commonly unfamiliar with the term 'administrative data' but despite this initial lack of familiarity, they had good comprehension and were able to share views. These views were important and framed how parents responded to proposed uses of administrative data in the rest of the dialogue. From these initial discussions, several themes emerged which contextualise the rest of the dialogue findings.

The first of these themes was that **data was widely seen as something that is never fully secure.** Data was understood by parents as something which is inherently vulnerable in both the public and private sector. Some parents felt data held in the public sector was even more vulnerable due to the computer systems being more likely to be out of date, which made them more vulnerable to potential hacks.

Secondly, across the sample, there were **mixed levels of trust in public bodies to handle and share data,** although few parents were very trusting. Parents repeatedly cited similar incidents which shaped this view. This included intentional misuse or hacking as seen in the **NHS Data hack** (2017) **and the Facebook - Cambridge Analytica Data 'scandal'** (2018) or examples of human error as seen when a **government laptop was left on a train** (2008).[6][7][8]. Anecdotally, parents also cited examples where they thought government agencies held inaccurate data on them such as inaccurate HMRC records, benefit information or the information held by their GP surgery. This undermined their trust in public agencies to hold (and potentially share) accurate information.

Thirdly, **data was also understood by parents to be commercially valuable**, in particular to private companies. Parents have gained increased awareness and become savvier around companies collecting their data and selling it on to third parties. Parents were concerned by this use especially if they thought it could lead to repercussions for them, such as more expensive insurance. The collection and use of data without the individual's awareness or permission was seen as unacceptable. The example of the **Facebook - Cambridge Analytica** Data 'scandal' was referenced repeatedly across the workshops. This example undermined the trust of the general public in sharing their data, including with researchers. In the wake of this, parents wanted reassurances their data will be used exclusively for research purposes as advertised, and not for other purposes or commercial gain. This is because parents associated data being accessed for commercial gain as not being in their interest, and felt it is unfair that companies can profit from their data without them being aware. It should be noted that how data could be used for commercial purposes was not fully explored with parents in the dialogue and these associations stemmed from pre-existing ideas of commercial uses of their data.

Despite these concerns, parents were generally willing to share their personal data in their everyday lives, even when there are privacy concerns, where they feel they have a choice and there is an immediate service/benefit for them (e.g. on social media and through their use of other online services). This shows that people are willing to make trade-offs regarding the use of their personal data when the benefits for them are clear and are tangible.

A further important theme that emerged was that parents felt that **data that is about them belongs to them**. Conceptually, parents did not think that their data belonged to the data holder such as the ONS or NHS Digital.[9] Therefore, they think the study team should ask their permission to use and access their data. This meant parents saw their involvement in birth cohort studies as **'partnerships'** between the study team and participants and wanted to be treated accordingly. This partnership would be characterised by being **consulted and updated with the use of their data and emerging study findings to ensure they feel valued and respected**.

The final key theme that emerged which is important to contextualise the rest of the dialogue findings was that **trust is key to parents' signing up to be part of a birth cohort study** and their on-going participation. In particular, parents need to trust researchers to collect and store their data in accordance with ethical and legal procedures, to ask and consult them about how they use their data, and to not use their data for commercial gain but only the research purposes, as agreed. This trust is something that needs to be built over time with participants. Without this trust, parents are less likely to find the proposed uses of administrative data acceptable as reassurances around how their data will be handled seem less meaningful. This trust can be quickly and easily eroded if parents find that the study has used their data without their knowledge and without giving them a choice.

3.2 Drivers of acceptability

Overall, parents generally found most of the proposed uses of administrative data acceptable, as long as certain conditions were in place. However, there was variation in views across the sample and some parents retained higher levels of concern than others throughout the dialogue.

---

[6] https://www.bbc.co.uk/news/health-39899646, https://www.independent.co.uk/news/uk/politics/exclusive-new-batch-of-terror-files-left-on-train-847451.html

[7] https://www.bbc.co.uk/news/topics/c81zyn0888lt/facebook-cambridge-analytica-scandal. Link accessed 9 December 2021.

[8] http://news.bbc.co.uk/1/hi/uk/7449927.stm

[9] Note: This is parents' perception that data on them belongs to them and not the legally correct as this data belongs to the public bodies who hold it. However, under the legal frameworks governing the use of data, individuals usually have rights regarding the processing and use of data about them.

Across the workshops, **seven factors emerged as key drivers of acceptable use of administrative data** in the ELC-FS:

- **Level of transparency:** Transparency was a key driver of acceptability. Across all of the proposed uses of administrative data, where parents were able to understand who was using their data, why it would be used, the value of the use, and how it would impact on them and their child, they found this more acceptable. Underpinning this driver was that parents wanted to assess for themselves the level of risk the use of administrative data poses to themselves and their child. Practices which were viewed as opaque or underhand were felt to be suspicious, and reflective that a proposed use could be harmful to the participant.

- **Level of risk of potential harm to the individual**: Where parents thought there was a lower level of potential risk of harm to the individual, they also tended to view the use as more acceptable. The three types of harm parents were most concerned about were: (i) theft and hacking, (ii) the fear of social stigma, and (iii) lost freedom of choice. There was more concern about the use of sensitive data which was strongly associated with social stigma where they feared judgement as an individual or parent (e.g. mental health data).

- **Degree of control**: As mentioned, parents feel they own data that is about them and therefore, where they feel they have a greater degree of choice over how their administrative data will be handled, they found the use more acceptable. Conversely, where parents felt pressurised or coerced into sharing their data, parents found this less acceptable (for example, if an interviewer arrives at their door unexpectedly asking them to share their data without them being aware beforehand)**.**

- **Level of clarity:** Where information and communications are provided about exactly how data was being used and where the processes involved were clear, parents tended to find the proposed uses of administrative data more acceptable as it helped parents feel more informed and increased their trust in the study.

- **Level of data security in place**: Reassurances around data security also increased the acceptability of the proposed uses of administrative data. In the workshops parents were particularly reassured by the data security features highlighted in the Administrative Data Research (ADR) UK and Next Steps data linkage videos, which included using encryption, serial numbers, only reputable researchers having access to the data and deleting all identifiable records after use.[10][11] These measures seemed to reassure them that their data was being held securely and efforts were being made to reduce the likelihood of breaches or data mishandling, which in turn would reduce the potential for harm to an individual from a data hack, breach or human error.

- **Level of necessity:** Where the proposed use of administrative data was understood to be essential to the implementation of the study or a particular aim, then this was more acceptable to parents. For example, in the sampling process, most parents thought it was acceptable as they valued the inclusivity of the approach, but also, they could not think of another way the study team would be able to build a complete and inclusive sample frame. Meanwhile, when it came to recontacting participants, some parents felt the use of administrative data in tracing was less necessary, as other approaches are available, and therefore it was less acceptable.

- **Degree of normalisation**: The proposed uses of administrative data were seen as more acceptable when parents knew data was used **routinely in this way**. This gave reassurance to parents that the proposed uses of administrative data were well established and the security measures involved robust. Parents found the absence of any significant data incidents involving linking administrative data in birth cohort studies reassuring.

---

[10] Administrative Data Research UK, *What is administrative data and why is it so useful for research?,* https://www.youtube.com/watch?v=ij4mNEFBdyg

[11] Centre for Longitudinal studies, *How does adding data help your generation*?, https://www.youtube.com/watch?v=W6ZuK3lYW6Y

3.2.1 Variation in views across the sample

Most of the proposed uses of administrative data in the ELC-FS were generally seen as acceptable as long as certain conditions were in place. However, there was variation in views across the sample and some parents retained higher levels of concern than others throughout the dialogue. Across the sample, how acceptable parents found the proposed uses of administrative data was driven by two key factors:

(1) The extent to which they valued birth cohort studies and,
(2) The level of trust they had in the study team to handle their data safely.

Those who had higher levels of trust in the study team and those who saw birth cohort studies as more valuable tended to find the proposed uses of administrative data more acceptable, whilst those who had less trust and perceived the study to be less valuable were more reticent towards the proposed uses of administrative data.

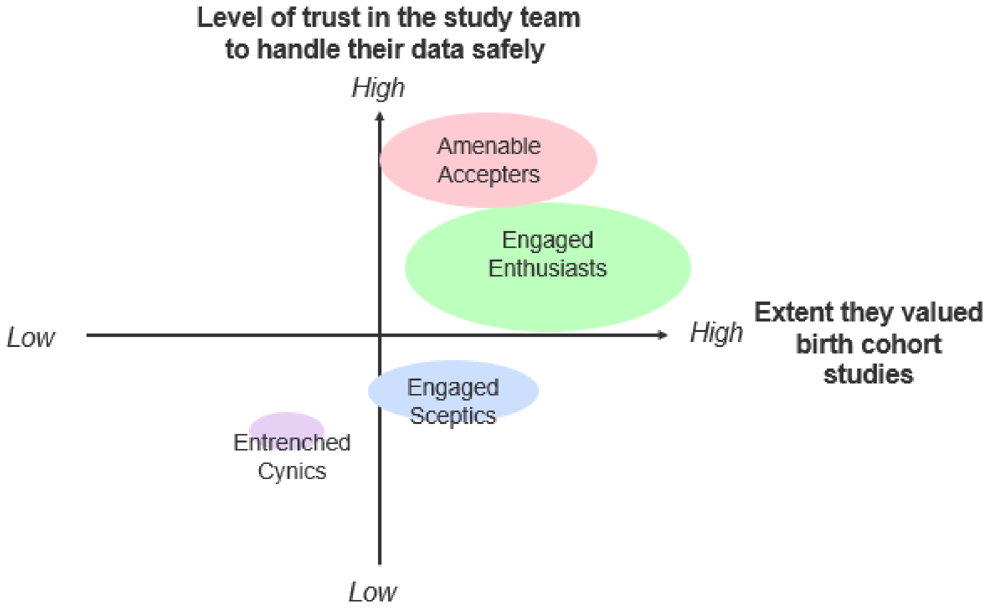
Four types of attitudes emerged from the sample across these two axes: **the Engaged Enthusiasts, Amenable Accepters, Engaged Sceptics,** and the **Entrenched Cynics**. These are illustrated below in Figure 3, which also sets out the relative size of these types across the qualitative dialogue sample.

These types emerged in the analysis process where the value parents placed in birth cohort studies and the level of trust they had in the study team emerged as key drivers which shaped parents' views to the proposed uses of administrative data. This hypothesis was then tested with moderators from the workshops, who were then asked to code parents from their workshops into a type. Analysis for how each type responded to the proposed uses of administrative data was completed using the analysis framework. The types were also analysed for differences by key characteristics including gender, ethnicity, region, profession and level of qualification but due to the small sample sizes differences by all these characteristics could not be established. However, it is worth noting that within this small qualitative sample, proportionately more mothers were classified as Amenable Accepters and proportionately fewer as Entrenched Cynics, than fathers.

Across the qualitative sample (for both mothers and fathers), Engaged Enthusiasts were the most common type, followed by Amenable Accepters, then Engaged Sceptics, and Entrenched Cynics were the least common type in the sample. Please note these types are reflective of the dialogue's qualitative sample, the size of these types cannot be extrapolated to the general public.

Parents views about the acceptability of the use of administrative data in the study evolved throughout the dialogue, and in general, the more the parents knew about the uses the more acceptable they tended to find them.

Figure 3 Parent 'Types'.



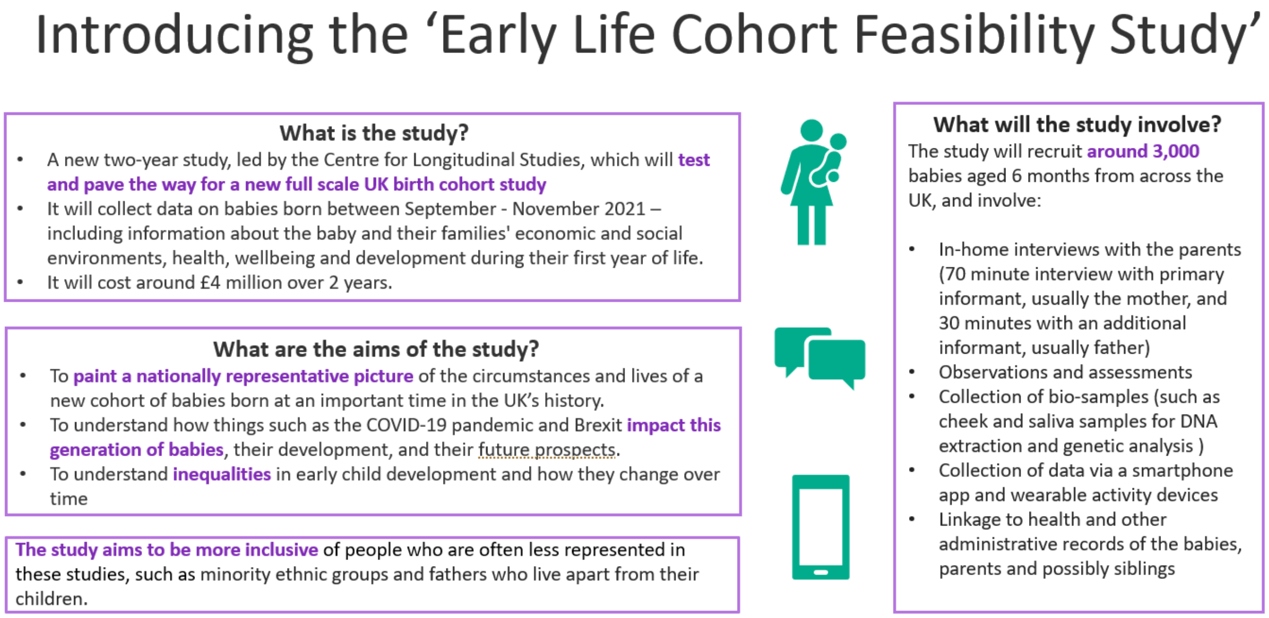
These four types can be understood as follows:

- **Engaged Enthusiasts:** This type believed most strongly in the value to society of birth cohort studies, and prioritised the inclusivity of the study, which if compromised they saw as compromising the overall value of the study. However, they also had some concerns around the extent they trusted the study team to handle their data securely. They wanted detailed information that reassured them around the processes and safeguards involved to ensure data is handled safely. This could include videos on the proposes uses their purpose, how they will be administered, any potential risks and any measures that can be used to offset these. When reassurances were provided, they found the proposed uses of administrative data more acceptable. This was the most common type, especially in Scotland and North England and the Midlands.
- **Amenable Accepters:** This type saw the value of birth cohort studies, although they tended to be less familiar with the benefits of the studies than Engaged Enthusiasts. This group tended to be more trusting of the study team and needed fewer reassurances their data would be handled securely. They tended to assume that safeguards and protocols were in place and did not need the specific uses listed in order to feel secure. This type had less of a need for more detailed information than the previous group.
- **Engaged Sceptics:** This type could see the merits of a birth cohort study but valued them less than the previous two groups, and had lower trust in sharing their data generally, which included the handling of their data by the study team. This type was also more worried about the cost of participating in a birth cohort study at an individual level. This includes it being too time consuming or the sharing of personal data, which may make them more susceptible to potential harm. This meant they were likely to think it was good that others participated in these types of studies but were less inclined to do so themselves. Similarly, to the Engaged Enthusiasts, they wanted more detailed information that informs them about how their data will be handled and the safeguards in place to ensure it is protected and there is no commercial use of their data. Parents from South and East England more commonly fell into this type.
- **Entrenched Cynics:** This type were less bought in to the value of birth cohort studies and had lower levels of trust in the study team handling their data securely. This type was much less convinced by reassurances of the protocols of how data will be handled and thought data will always be at high risk of being hacked or mishandled. This type were more fearful of potential commercial use of their personal data. This type were less motivated by participating for the ‘social good’ and commonly perceived participation as posing too high a cost to the individual. This was the least common type in the dialogue.

3.2.2 Perceptions of the ELC-FS

Parents were presented with stimulus which introduced them to the ELC-FS, which is shown in Figure 4. This stimulus was to provide key context on the aims of the study, which would further aid understanding to the uses of administrative data proposed throughout the study. The stimulus outlined the study would involve in-home interviews with the parents, observations and assessments, the collection of bio-samples (such as cheek and saliva samples for DNA extraction and genetic analysis), the collection of data via a smartphone app and wearable activity devices and the linkage to health and other administrative records of the babies, parents and possibly siblings. It did not to provide detailed information about these data collection elements and what they would involve, which should be kept in mind when interpreting parents’ perceptions of the ELC-FS.

Figure 4 Stimulus introducing the ELC-FS



Introducing the 'Early Life Cohort Feasibility Study'

What is the study?

- A new two-year study, led by the Centre for Longitudinal Studies, which will **test and pave the way for a new full scale UK birth cohort study**
- It will collect data on babies born between September - November 2021 – including information about the baby and their families' economic and social environments, health, wellbeing and development during their first year of life.
- It will cost around £4 million over 2 years.

What are the aims of the study?

- To **paint a nationally representative picture** of the circumstances and lives of a new cohort of babies born at an important time in the UK's history.
- To understand how things such as the COVID-19 pandemic and Brexit **impact this generation of babies**, their development, and their future prospects.
- To understand **inequalities** in early child development and how they change over time

The study aims to be more inclusive of people who are often less represented in these studies, such as minority ethnic groups and fathers who live apart from their children.

What will the study involve?

The study will recruit **around 3,000** babies aged 6 months from across the UK, and involve:

- In-home interviews with the parents (70 minute interview with primary informant, usually the mother, and 30 minutes with an additional informant, usually father)
- Observations and assessments
- Collection of bio-samples (such as cheek and saliva samples for DNA extraction and genetic analysis)
- Collection of data via a smartphone app and wearable activity devices
- Linkage to health and other administrative records of the babies, parents and possibly siblings

Overall, parents were generally very positive about the ELC-FS and agreed it was a useful time to launch a new birth cohort study to more fully understand the development of children born in the 2020s, whose childhood is likely to be different from those born in 2000 due to advancements in technology and experiencing the COVID-19 pandemic.[12] However, parents widely felt uncertain about whether they would like to participate themselves due to concerns around what participating in the study would involve and how burdensome this would be for them and the nature of the long-term commitment.

> "*From a technology point of view, we live in a completely different world now to compared to 20 years ago. Our lives are so much more dependent on technology now. It's really important we can reflect on these changes in society*." (Wave 1, North England and the Midlands)

Key benefits

The following key benefits of the study were identified by the dialogue participants after they were introduced to the study:

- Parents thought that the 2020s is a **timely opportunity to launch a new birth cohort study**. Parents picked out the references in the stimulus which included 'Brexit' and 'Covid-19' and thought it is important to understand how these events have impacted on society and in particular, families with young children. For example, parents from Northern Ireland thought that children in the pandemic were less socialised with the world outside their home than children born pre-Covid-19 and thought it would be valuable to learn more about how these types of experiences impact on child development. Parents also referenced that children born now grow up with exposure to far more technology than ever before, which they assumed would also have an impact of children's development.
- **Parents** with children who had **a health condition or disability**, could see how potential findings could benefit their family. They anticipated that the study could shed light on how their child's development is impacted by their health condition or disability.
- Parents also thought the ELC-FS would help to **uncover useful findings around children's development**. For example, in Northern Ireland parents referenced studies that had highlighted mixed

---

[12] The last national birth cohort was born in in 2000-02, the Millennium Cohort Study, https://cls.ucl.ac.uk/cls-studies/millennium-cohort-study/

race children had a poorer performance at school - they thought birth cohort studies could enable researchers to find these participants and follow them to determine why this is occurring.

"*Obviously being more well informed is going to be better than anything so I couldn't see what the harm would be...I can't see any concerns*." (Wave 1, Northern Ireland)

"*For me actually, the greater good of the community is the greater good of my community and people that I know, so I think that I'm willing to give that information to help my community."* (Wave 1, Northern Ireland)

Key concerns

Parents had five key concerns initially about the ELC-FS, these are listed below.

- Across types, parents widely had concerns about the **level of commitment** being part of the study would entail, leading parents to think the study was a good idea but not necessarily something they would be able to commit to participating in themselves. This concern arose partly as specific details on the running of the study such as how long initial visits would last, the number and frequency of follow-up interviews were not provided as part of the dialogue. Clarification from stakeholders, for example that participants would likely be contacted every several years and that interviews may typically last around 90 minutes, reassured parents that participation was less burdensome than they had initially imagined. Some parents who belonged more with the Engaged Sceptics or Entrenched Cynics were also more reticent about signing up to a study on their child's behalf, who in the future may wish that they had not.
    - In South and East England parents were concerned that many of the benefits of participating in a birth cohort study would not be apparent until the future– whereas all the costs of participation happen in the short-term. These parents felt they would have to determine whether this short-term cost is worth waiting for the long-term benefits.
- Concerns around **data security** emerged here too, with parents being particularly concerned how their data would be stored and whether it would be vulnerable to hacking, which may result in individual harm being caused to them.
- Parents were also concerned about what their data will be used for and did not want it shared with third parties, who may use it for their own **commercial gain**. Parents objected to their data being used for the profit of others and feared the selling on of their data could have potential negative repercussions for them in the future or for their children. For example, if we have a more privatised health care system in the future, being able to access health data on an individual could lead to more expensive health insurance for them was a common example. We note that the ELC-FS is not proposing to share participant data to third parties for commercial gain.
- Parents also raised worries about some of the **data collection elements involved. As these elements were not related to administrative data and were not the core focus of the dialogue, parents were provided with limited information on what these involved, therefore the concerns of parents about these should be read with these caveats in mind. These concerns included:**
    - Parents widely felt the collection of biological samples for DNA extraction could be particularly sensitive and were cautious about the potential impact participation in this element may have on their child, such as making them feel like a 'research specimen.' Some parents felt this would be mitigated if they were informed about the reasons for DNA collection and how this data will be used, accessed and protected. For example, they wanted reassurances this data would not be sold to or shared with private companies (including insurance companies) that may penalise participants in the future.

    "*Personally wouldn't feel comfortable someone taking DNA from my child.*" (Wave 1, South and East England, Wave 1)

    - Those who were more concerned about data privacy disliked the idea of the app, which collects data on the child's development, and were concerned their phone activity may be tracked (e.g. websites they visit and spending habits). Parents associated apps with companies recording and tracking their behaviours for commercial use. However, other parents responded positively to a channel where they may receive more information about their child's development, and some likened it to using pregnancy and baby milestone apps.

> "*I first thought this is invasive, that would be the word I'd use. It kind of scares me to think that in 20 years my child would ask what has happened to all the information, but at the same time it is understandable that you need for this type of study more than a few random questions.*" (Wave 1, Northern Ireland)

- Parents were cautious about **how participating in the ELC-FS may reflect on them as a parent** and could **make them feel 'judged'.** This included fearing they would be misrepresented in the administrative data if it was out of context, or may cover previous behaviour, such as criminal history or detail prior drug or alcohol use, which they feel does not reflect them as a person anymore. Part of this fear stemmed from some parents thinking the study team would analyse them as individuals. Parents were also concerned they would feel responsible if their children were predicted to have more negative life outcomes and if their children were aware of this it may give their children a more negative outlook on life as a result.
    - In Scotland, parents worried their children would be unhappy they had been put into this study once they were older and wanted reassurance their child could opt-out in the future if they chose to.

> "*Am I being judged, I'm giving certain answers about my home life. I'm a single parent, is the person interviewing thinking 'oh we know what this means'. I know I shouldn't really care but I would. Also who are the researchers - what qualifications do they have. We don't know anything about them and I'm not talking about safeguarding but having a stranger in your house.*" (Wave 1, North England and the Midlands, Wave1)

3.3 Views on participation in the ELC-FS

There was a mix of views on whether parents would consider participating in the ELC-FS across the dialogue sample. Overall, parents did not tend to shift their own view on whether they would be interested in participating in a birth cohort study over the course of the dialogue. This was because practical concerns including the potential time burden and how intrusive the data collection elements would potentially be were key and remained in place by the end of the process.

Views of other members of the household, predominantly from partners, also shaped whether parents thought they would participate. This is reflective of parents seeing birth cohort studies as a whole household commitment, and where family members had greater concerns, this diminished parents' certainty over whether the household would be willing to participate. Although, in some instances, for Engaged Enthusiasts and the Amenable Accepters where their household had concerns, they took on an advocate role for the ELC-FS, to help alleviate their concerns.

We note that views towards the use of administrative data were not generally reported to be a key driver in determining whether parents would want to participate in the ELC-FS or not, with concerns around the potential time commitment involved and how intrusion the data collection elements involved would potentially be still paramount for some. Whilst for others, the potential benefit to society was a key driver to participating.

> "*My husband is a very private person, unless he could see the value for him, he couldn't see beyond the intrusion,*"(Wave 2, South and East England)
>
> "*We were all agreed in our house it was a good thing to take part in, it will be better for society overall."(*Wave 2, North England and the Midlands)
>
> "*My partner was initially quite suspicious about it. But then a lot of people are much more protective about their data. But once I explained more to her about it, she understood and could see the benefit in taking part. She works with young people as well as me, so she can anticipate the positive impact it will have*." (Wave 2, North England and the Midlands).
>
> "*My husband felt the same, he is laid back and had not issues with his data being shared or people contacting him through maternity records but it is something else putting on parents to make a long-term decision on behalf of their child(ren)."* (Wave 2, Scotland)

4. Responses to proposed uses of identifiable administrative data in the ELC-FS: for sampling, recruitment and tracing

This chapter explores parents' responses to the proposed uses of identifiable administrative data in the ELC-FS. It first describes parents' general concerns around use of identifiable data in research, before outlining views towards the specific ELC-FS proposed uses of identifiable administrative data. It covers responses to the proposed approach to sampling, recruitment, targeted recruitment (tailored recruitment materials and boosting) and retention of the sample through tracing.

In the workshops parents were introduced to the proposed sampling approach for England and Wales, Scotland and Northern Ireland. Stimulus materials showed the different types of identifiable administrative data and the sources (data holders) from which the study team hopes to access this data.

Parents were also presented with information detailing three different recruitment approaches which the study could take when accessing administrative data to create the sampling frame and contacting parents and discussed the acceptability of each approach. Further, parents were introduced to two methods of targeted recruitment (tailored recruitment materials and boosting) that use identifiable data that could help improve inclusivity of under-represented groups in the study sample.

Finally, it was explained to parents that as study participants are randomly sampled in birth cohort studies, they cannot be replaced. This means it is important to keep participants in the study as they cannot be replaced by other participants. Parents were further told how identifiable administrative data can be used to trace participants the study team has lost contact with because they have moved and not informed the study team of their new address.

Key findings:

- *Parents were widely concerned about* ***data security*** *and harm that can result from identifiable administrative data misuse and mishandling, particularly for types of data they felt could more easily lead to individual identification. Despite parents not raising as many of these concerns when discussing the specific proposed uses of identifiable data for ELC-FS, these underlying concerns were raised during earlier general discussions about administrative data use and remained relevant in the background of the discussions.*
- *Parents widely found the proposed sampling approach, using birth registration and maternity records data, acceptable as it was seen as necessary to access a sample through which the study team could find and contact parents to invite to participate in the research. It was also seen as important to ensure the sample was representative and to help improve the inclusivity of the research, which was widely important to parents.*
- *There was a strong preference for the Two Step recruitment approach, in which families were able to opt-out before their details were passed on to the study team by the data holder. It felt intuitively more suitable than the One Step recruitment approach, which does not have a prior opt-out, to parents and met their expectations of what the research team should do. This was because the Two Step approach* ***increases the control*** *offered to participants over their administrative data and* ***improves transparency.***
- *Although most parents did not support the One Step recruitment approach, in which there is no prior opt-out before their details are passed on to the study team by the data holder, some parents saw benefits to this, including simplicity, being quicker and less resource intensive and likely to deliver a more inclusive sample. Some parents who had lower concerns about the risk of personal harm found the One Step approach acceptable.*
- *Data use for targeted recruitment was generally acceptable to parents as long as it is done sensitively to avoid generating feelings of stigmatisation. Boosting also made sense to parents and was seen as* ***important*** *to improving the* ***social value*** *of the findings of the study by further ensuring diverse experiences are represented in research findings.*

- *Overall, parents found tracing acceptable if they were clearly informed about it during recruitment as this would ensure they are offered* ***control*** *and* ***transparency*** *over administrative data use. However, most parents thought tracing using administrative data should only be undertaken once other contact methods provided by study participants had been used which do not use administrative data, because they were considered more suitable and simpler.*

4.1 Views about identifiable data

As described in Section 2.1, parents held some general overarching concerns about administrative data use in research, with particular concern being expressed towards uses of data that could more easily lead to individuals being identified. However, they understood that the use of identifiable data was important and necessary to the study and were generally happy with its use if certain conditions were in place and they were reassured about data security and data handling processes.

Parents felt data was never fully secure, and open to mishandling and misuse leading to personal harm. Parents described these concerns during general workshop discussions about administrative data storage and handling by public bodies. They were worried administrative data can risk personal harm through:

- Loss of data through theft and hacking, leading to administrative data being used for financial fraud.

  "There are always hackers who may access the data." (Wave 1, Scotland)

  "It's not the organisations who hold our data that I don't trust, but the cyber criminals out there who could attack their systems. This happened not so long ago in one of the health departments in the Republic. It was all over the news." (Wave 1, Northern Ireland)

- Social stigma, where data is leaked leading to judgment or embarrassment, or risk of actual harm by something from their past; e.g. participants felt access to information about previous births could be sensitive if the mother had experienced miscarriages.
- Loss of freedom of choice by administrative data being used without permission; e.g. participants described Cambridge Analytica's use of administrative data as being seen to influence election results and *"big brother"* uses of administrative data to monitor and influence.

4.2 Views about the proposed approach to sampling

Parents were shown the ELC-FS sampling plans for England, Wales, Scotland and Northern Ireland that detailed the various administrative data sets that may be accessed by the study team to create a sample see Workshop Materials Appendix). Moderators explained how this process could be used to create a sample for the ELC-FS by identifying families with a baby born within the chosen timeframe. The information held would also help ensure a more representative and inclusive study sample. For example, accessing further information about the mother (such as age and country of birth) can help the ELC-FS ensure the sample is representative.

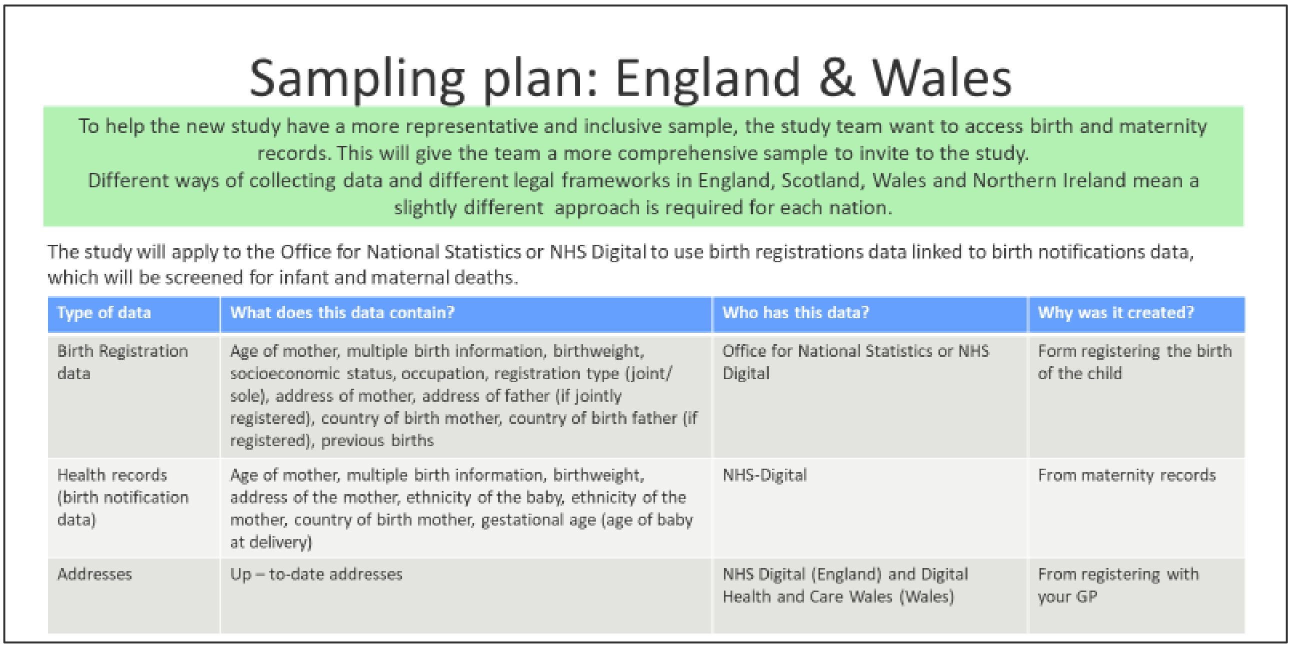
In all nations the study team want to use linked birth registration and/or health records, along with address information.

Figure 5 Summary of sampling plan

Type of data	**Examples of what this data contains**	**Data holders: England and Wales**	**Data holders: Scotland**	**Data holders: Northern Ireland**
Birth Registration data	Age of mother, multiple birth information, birthweight, occupation, address of mother, country of birth mother, country of birth father (if registered), previous births	Office for National Statistics or NHS Digital	National Records Scotland (NRS)	Northern Ireland Maternity System

Health records (birth notification data)	Age of mother, multiple birth information	NHS-Digital	Public Health Scotland	Northern Ireland Statistics and Research Agency
Addresses	Up – to-date addresses	NHS Digital (England) and Digital Health and Care Wales (Wales)	Community Health index	National Health Application and Infrastructure Services data

Figures 6 Example of stimulus material used to show the study team sampling plans



Sampling plan: England & Wales

To help the new study have a more representative and inclusive sample, the study team want to access birth and maternity records. This will give the team a more comprehensive sample to invite to the study.
Different ways of collecting data and different legal frameworks in England, Scotland, Wales and Northern Ireland mean a slightly different approach is required for each nation.

The study will apply to the Office for National Statistics or NHS Digital to use birth registrations data linked to birth notifications data, which will be screened for infant and maternal deaths.

Type of data	What does this data contain?	Who has this data?	Why was it created?
Birth Registration data	Age of mother, multiple birth information, birthweight, socioeconomic status, occupation, registration type (joint/ sole), address of mother, address of father (if jointly registered), country of birth mother, country of birth father (if registered), previous births	Office for National Statistics or NHS Digital	Form registering the birth of the child
Health records (birth notification data)	Age of mother, multiple birth information, birthweight, address of the mother, ethnicity of the baby, ethnicity of the mother, country of birth mother, gestational age (age of baby at delivery)	NHS-Digital	From maternity records
Addresses	Up – to-date addresses	NHS Digital (England) and Digital Health and Care Wales (Wales)	From registering with your GP

Parents were generally supportive of this proposed use of administrative data, and the different types of data obtained from different data holders in the sampling plan, because they easily **saw the necessity** of it **to be able to draw a sample and to improve the social value of the study** through accessing a comprehensive and accurate sample frame. Parents quickly recognised the need for the ELC-FS to have access to an accurate sample frame including identifiable data to be able to invite parents to the study. They also mentioned the social benefits of having an inclusive sample that represents the experiences of a range of people. Parents easily saw the proposed usage as a logical and sensible way to achieve this. A number of parents even commented they could not see another way to achieve this outcome. The clear way that parents valued and understood the need for using administrative data, and the clear information they received in the workshop about the importance of comprehensive and accurate sample frames, led to wide acceptability of this use.

> *"It makes sense to use this initial basic data like X [Name of Stakeholder] said, you need to reach these people. Also, it means that you are not wasting time. If you sent a letter through everyone's door, a lot of those people won't be people that you are looking for – so if you are looking for certain ethnicities or disabilities, at least you know you are contacting the right people. It whittles it down and saves time and money."* (Wave 1, North England and the Midlands)
>
> *"To be fair, to get efficient and effective data to start the process of finding the right people, I don't see any other way. So I think it makes sense, and then getting in contact. There's no alternative, there's nothing I can think of."* (Wave 1, North England and the Midlands)

Where parents recognised that the different proposed types of administrative data were already routinely accessed and shared, it **normalised the wider use of this administrative data** and led to the greater acceptability of its use for the ELC-FS. A number of parents stated that their data was *"already out there"* and held, accessed and used by many individuals. For example, some parents described how midwives routinely access general health records including NHS and telephone numbers. Similarly, a number of parents explained how they regularly and freely share much of the information the study wants to access with others; e.g. the NHS and other government organisations routinely ask for ethnicity, gender and address. Further, after learning this from dialogue stimulus, parents described how other social research projects already use their linked administrative data without prior permission. These understandings normalised proposed types of administrative data being accessed, leading to greater acceptability of this proposed use.

> *“You've already disclosed that information anyway, and other people have access to it too. Like my midwife for instance, she's already got access to that and lots of other information about me. Why would I really care [if the study team has access to this also]?”* (Wave 1, Northern Ireland)

Further, acceptability towards the proposed use of administrative data for sampling was gained because the data had been or is already routinely collected, which **reduced parental levels of concern about risk of harm**. A number of parents, particularly Amenable Accepters, felt more comfortable with the use of proposed types of administrative data because the study team was not collecting and storing new kinds of information. Parents appear to have seen this process as less worrying because it was not adding to the range of information stored by others about them. There was even a sense among some that since the data had already been collected it might as well be used for a positive purpose, which they understood the research to be. These attitudes may be informed by general views that data is vulnerable and can be used to harm, and as a result it’s best for less types of information to be stored about individuals.

> *"I think it sounds fine, it's all data that exists already on national databases.”* (Wave 1, Scotland)

> *"Already stored so might be good to make use of it."* (Wave 1, South and East England)

Despite aspects of the proposed sampling approach normalising the sharing of identifiable administrative data and reassuring about individual harm, **concerns about risk** were still expressed. As previously mentioned in Section 2.1, parents were concerned about risk of harm from use of administrative data, especially when it is identifiable. During conversations about sampling parents raised these concerns, including about the rigour of data security. For example, parents commented on a case where civil servants lost laptops on a train[13] and the importance of protecting individual identities. These concerns about harms led to the views that certain kinds of administrative data could be more sensitive and so less acceptable to share. Concerns about data sensitivity appear informed by fear of judgment and social stigma that might result from the loss or leaking of data. For example, sharing maternity or birth records was seen to be more sensitive (particularly by mothers) depending on the personal circumstances of participants. Parents imagined that this information may feel more sensitive for those who had been through miscarriage, rape or still births, in comparison to those whose pregnancy and birth had been straightforward. Further, certain groups were recognised to be more concerned about sharing addresses, such as women experiencing domestic violence, those in witness protection, immigrants, and refugees, and parents wondered how the study could account for this. Additionally, one Northern Irish participant felt that occupation information might be sensitive for those who work in more vulnerable employment, such as those in security, the army or the police. It’s important to note here that parents were primarily reflecting on how they thought others who are part of these groups might feel, rather than themselves.

> "*I think some are more sensitive issues, like previous miscarriage, or some kind of trauma, like people who've had to deal with rape or anything like that. We have to be mindful they may not want to disclose any information regarding that*." (Wave 1, North England and the Midlands)

> "*I think in Northern Ireland we are in more of a unique position in term of security, so I read the slide and the one thing that popped out to me was occupation*.” (Wave 1, Northern Ireland)

Finally, a few participants raised worries about the accuracy of the administrative data and the **reduced social value** this could lead to. For example, parents mentioned how GP address records may not be up to date as not all people reregister or may not have had time to re-register if recently moved. Similarly, others thought birth records might not be complete. Parents were worried these outdated and incomplete records would impact the quality of the sample frame. Additionally, one group felt mothers were primarily identified within the sample frame which might exclude the equal participation of the fathers in the research.

---

[13] http://news.bbc.co.uk/1/hi/uk/7449927.stm

> *"It's a good source for the sample, but some birth records aren't complete."* (Wave 1, South and East England)

4.2.1 Northern Ireland Registration of Birth

In Northern Ireland, the proposed sampling plan, as presented in the dialogue, uses health records only. This is because unlike in other countries, birth registrations are not routinely linked to health records in Northern Ireland. Therefore, as well as discussing the sampling plan, parents in Northern Ireland also were asked about birth registration and how they would feel about being informed when registering their child's birth about the potential for being contacted about the ELC-FS and their birth registration data being used for the study. As the sample of 11 parents in Northern Ireland is small and due to workshop time constraints, these findings should be taken as indicative.

Overall, parents from Northern Ireland wanted to be informed that they might be asked to take part in the study and there was a range of views about how best to do this. Parents liked the idea of being notified before being invited and felt it was suitable for this to take place at the birth registration. One parent said that receiving information before the initial invitation would mean study participants were less likely to think the invite was a scam and that getting this information at the registration of a birth would offer high levels of reassurance as it is during a legal process. Another said being notified at the registration would feel normal and unsurprising since parents of new-borns are provided with numerous information about their child by many different sources; e.g. doctors, midwives and health visitors. However, two parents expressed feeling very busy and hassled around the time of registering the birth, trying to manage a new-born and other children. They felt being informed at registration was not appropriate and suggested the research team inform parents later on when they are less busy, such as when registering the child with the GP for vaccinations.

> *"When my child was born there was health visitors coming in and out asking the same kind of questions anyway."* (Wave 1, Northern Ireland)
>
> *"I remember feeling quite hassled when I registered my last baby. I had two other little children with me and my husband was stressing about parking. You just want to go in and get it done."* (Wave 1, Northern Ireland)

4.3 Views about potential recruitment approaches

Stimulus materials (see Workshop Materials Appendix 6.14, 6.15) were presented to parents detailing three different recruitment approaches the study could take when accessing administrative data to create the sample and contact parents.

Figure 7 Summary of potential recruitment approaches

One Step approach

Parents details would be passed directly to the study team from the data holder. The study team would contact parents via post and invite them to take part in the study, which is followed up with an interviewer in person to confirm recruitment for those who have not opted-out. Parents are told which data holder provided their data to the study and they can confirm this with the data holder if they wish to. Parents then decide if they want to take part in the research or not

Two Step approach

The data holder would first contact parents via post, giving them the opportunity to opt-out of being invited to the study before their details are passed to the study team. The details of parents who do not opt-out are then passed to the study team. The team would then contact the parents via post and invite them to take part in the study, which is followed up with an interviewer in person to confirm recruitment. Parents can then decide if they want to take part in the research or not.

Alternative Two Step approach

The data holder would contact parents via post giving them the opportunity to opt-out of being invited to the study or to opt-in. The data holder sends an interviewer to the parents' homes who have not responded to confirm they are happy for their details to be passed on to the study team to be invited to the study. The details of parents who have opted-in are then passed to the study team. The team then contacts parents via post and invites them to take part in the study, which is followed up with an interviewer in person to confirm recruitment. Parents then decide if they wish to take part or not.

4.3.1 One Step approach

Overall, most parents did not support the One Step approach, seeing it as **un-transparent** and **offering less participant control**. The preference for transparency and control was strongly and intuitively felt, and quickly expressed by most parents. Parents felt ownership over their administrative data and as a result felt it was appropriate and ethical for the study team to communicate with them in certain ways that respected this; e.g., to be clearly informed of how administrative data would be used before it is and to feel in control of its use. For many this was not because they would want to opt-out of administrative data being shared, but because they felt it was the appropriate and fair thing to do. The One Step approach was not seen to provide sufficient transparency and control by most parents because by administrative data being shared with the study team before participants are informed, participants are not offered information about and a say over how their data is being used. Some parents suggested that trust would be lost as a result.

Despite this, some parents saw benefits to this approach as they felt the other approaches were overly complex and **unnecessary,** and recognised this approach being quicker and less resource intensive and likely to deliver a more inclusive sample. Parents understood the One Step approach would be quicker and less resource intensive (as less rounds of letters) and likely deliver a better sample as fewer participants would opt-out. For a small number of parents, this led to a preference of this approach.

> "*In my mind the one step process seems an easier process for the participant and the researcher…less paperwork*". (Wave 1, Scotland)
>
> "*It will probably cost less money and be easier for the researchers*." (Wave 1, North England and the Midlands)

Additionally, a small number of parents, where they had low concerns about **the risk of personal harms from administrative data**, preferred this approach over others. Some parents felt unconcerned with their data being shared without prior knowledge. They recognised few risks and generally had little to say about the different approaches. These participants had had lower concerns about data security throughout the dialogue and primarily were the Amenable Accepters.

> *"I prefer that One Step one. It just seems a bit forceful telling me you need to remember to opt-out. I'd just go for the one step, it's less hassle overall."* (Wave 1, Northern Ireland)

4.3.2 Two Step approach

There was an overwhelming preference for the Two Step approach as the **increased control** it offered over administrative data and **improved transparency** felt intuitively more suitable and met parent expectations. Parents felt a sense of ownership over administrative data that was about them. As a result, they wanted the study team to treat study participants in a transparent way and offer them a say over administrative data use before it was shared. The Two Step approach met these expectations and was therefore preferred by most parents over the One Step approach.

> *"I prefer the Two Step. The process gives me more information as an individual, and it's up to me to then decide to stay in or opt-out for any further things to do with the study. It gives me the choice."* (Wave 1, Northern Ireland)
>
> *"Ethically it should be done, so you have the opportunity to opt-out of if you don't want to be contacted."* (Wave 1, Northern Ireland)
>
> *"Control in the hands of the participant with this approach."* (Wave 1, South and East England)
>
> *"Going directly to the [participant] would undermine the legitimacy of the study." (Wave 1, South and East England)*

4.3.3 Alternative Two Step approach

Like the Two Step, the Alternative Two Step approach was seen to give fair and appropriate levels of transparency and control to study participants over the use of administrative data. However, parents expressed several strong concerns about the role of interviewers when discussing the Alternative Two Step approach, specifically around the perceived inappropriateness of a data controller interviewer visiting homes after the first letter and the purpose of this visit being only to give permission for the data transfer.

It should be noted, stimulus materials about the one and two step recruitment approach were adapted after the first weekend of the first wave workshops to make clearer that an interviewer would visit homes in these two approaches, and to help to ensure that findings around the role of interviewers could accurately distinguish between the different purposes of their visits in the One Step, Two Step and Alternative Two Step.

Parents felt that study participants should be communicated with several times before an interviewer visits in the Alternative Two Step approach where the purpose of the visit is to discuss whether or not they are happy for their details to be passed to the study team for them to be invited to the study. This was for a number of reasons. Firstly, the first letter from the data holders could have been lost, missed or ignored by residents. Across the sample, parents raised a concern that if this happened and parents were not pre-informed about the interviewer visit, the visit, could feel invasive, inappropriate and coercive; they thought it was *"too much and too intrusive."* Some felt it could feel like a "*cold caller*" and others (particularly women) expressed how they might feel unsafe. Further, because study participants would likely know little information about the study after receiving one letter, it would inhibit their ability to make informed decisions if they were rushed into this by the interviewer. One parent described it as likely feeling "*forceful*", while others saw it as "*badgering*" and "*pressurising*" them to hand over their data. Some parents described how they would feel more comfortable expressing their opinions about whether or not to agree to their details being passed on over the phone and would prefer this type of communication instead of an interviewer. Additionally parents felt the presence of a data controller interviewer would force them into making certain decisions i.e. agreeing to the transfer of their administrative data to the study team.

> *"It is already a bit intrusive, I haven't even given thought to it and they are already showing up at my house."* (Wave 1, South and East England)
>
> *"That's really weird. If someone knocked on my door, even if they showed ID, I think a lot of mothers would feel quite vulnerable."* (Wave 1,Wales)
>
> *"My suggestion would be an initial letter opt- in or opt-out followed by a phone call and you can say no and then hang up. Coming to your front door is maybe a step too far."* (Wave 1, Northern Ireland)
>
> *"Difficult to say no in the interviewers face...I would be much more comfortable getting a call."* (Wave 1, Scotland)

In comparison, parents saw the Two Step approach as more acceptable because they believed the interviewer would visit at the end of the recruitment process. This would mean more communications received by parents beforehand, better informing them about the study and the potential for an interviewer to visit. Overall, other forms of communications were felt to be less invasive than the presence of an interviewer, such as emails, texts and telephone calls; parents thought these communications should be used before an interviewer visit[14]. There was also a sense that the nature of the interviewer visit in the Two Step approach, asking for participation in the study, was a less daunting request than one about the transfer of administrative data in the Alternative Two Step approach.

Despite these commonly held concerns, a small number of parents expressed a preference for the Alternative Two Step approach, as they felt it **increases participant control** over administrative data. Some parents thought an opt-out approach puts the onus onto individuals to act to prevent things from happening they do not agree to, and since families have busy lives they can fail to take action. For example, in the Two Step approach, failing to opt-out leads to the transfer of administrative data to the study team. Further, as letters can be easily lost in the post, sent to the wrong house, or ignored by residents, study participants might not get the opportunity to opt-out. As a result, the Alternative Two Step approach was deemed more appropriate by some as it actively confirms agreement for data to be passed on through sending interviewers to homes of parents that have not responded to the first letter. A small number of Engaged Enthusiasts and Engaged Sceptics therefore had a preference for this approach which reflects their generally higher levels of concern about data security.

> *"You shouldn't just assume consent just because no one's got in touch with you."* (Wave 1, Scotland)

Further, in Northern Ireland there were some concerns for interviewer safety, applicable to all approaches. For example, a parent said he was unsure if all areas of Belfast would be safe for an interviewer, particularly at certain times of the year, due to sectarian issues. Another parent expressed worries about interviewers walking in certain areas.

---

[14] Study team would not have access to email addresses and telephone numbers so would not be able to contact parents through these channels

> *"And obviously you need to take into consideration the history of Northern Ireland. If people walk into certain areas at the wrong time it's not always safe."* (Wave 1, Northern Ireland)

4.3.4 Communications during recruitment

Across the recruitment approaches, a number of parents expressed the importance of communications coming from recognisable organisations to build trust and legitimacy of the study. A number of parents were unfamiliar with University College London (UCL) and said this could impact the likelihood of their participation. In contrast the NHS, health visitors, and healthcare staff felt more familiar and were suggested as alternatives.

4.4 Views about use for targeted recruitment

Parents were introduced to two methods of targeted recruitment that could help improve inclusivity of under-represented groups in the study sample:

3. Tailored recruitment: This could involve having specific recruitment letters and leaflets that aim to be more appealing to certain groups or emphasises why it is important for them to take part or asking interviewers to make extra efforts to interview these groups.
4. Boosting: This is when the study recruits additional participants from groups that form small parts of the population e.g. those from minority ethnic backgrounds. This process allows meaningful findings to be drawn about these groups and uses sampling frame variables to identify specific individuals to recruit.

These processes use administrative data to identify groups to target. Parents were given explanations of general approaches to tailored recruitment and boosting, rather than specific information about the plans for the study.

4.4.1 Tailored recruitment materials

Overall, parents understood the **social value** and **necessity** of using administrative data for targeted recruitment as it would improve the representativeness of the study. This understanding drove widespread acceptability of its use for this purpose. As previously described, parents understood the social value of having a representative sample that includes a range of people from different backgrounds. They saw targeted recruitment as an effective means to ensure this through sending communications that positively explain the importance of individual participation of people from these groups and how it aids research findings to be relevant to those from similar backgrounds. Parents listed a range of groups they thought would be good to target e.g. same sex couples, those with adopted children, those who used IVF and disabled parents.

> *"I also think people will be more inclined to engage if they feel needed in some way."* (Wave 1, Scotland)
>
> *"For example, say if it is a teenage mother, if you send her a specific letter and explain why this research group needs teenage mothers it becomes personal to them then. I'm thinking you would have a better chance of people taking it up and agreeing."* (Wave 1, Northern Ireland)
>
> *"I would be even more likely to respond if there was an explanation why I was contacted...if the information in mail is vague I would be unlikely to take any notice. Would be more likely to take part if it was targeted at me."* (Wave 1, South and East England)

However, there was a common concern that unless targeted recruitment is done sensitively, it could be inappropriate and generate feelings of stigmatisation. Parents were concerned single mums and those from low socio-economic backgrounds may feel uncomfortable and singled out, and this could generate concerns they may be judged through participating in the study. Further, some parents felt that if the letter had detailed information about their background, it could make these groups feel particularly unnerved or concerned; this was seen as especially inappropriate for minority ethnic groups.

Parents felt targeted recruitment needs to be done sensitively for example, by stating the importance of a range of people taking part including certain groups, and not specifically stating that someone was contacted because of their own specific characteristics. It is important to note here, parents were primarily reflecting on how they thought others who are part of these groups might feel, rather than themselves.

> *"Why would you think you were being contacted for any good purpose when your situation as a migrant, or a young mum, a single mum - whatever it is - why would think someone independently*

> *contacting would be a great reason? They must have so much contact that is negative towards them."* (Wave 1, North England and the Midlands)
>
> *"I know that if I received a letter in the post that said 'we're contacting you because you're on benefits, your wife is disabled and you're her carer' I wouldn't be comfortable with that and I wouldn't want to be part of it. If they said something like 'we want to take people of all backgrounds in to account and we value everyone's opinion. Something more blanket rather than 'you there as a single mum'.* (Wave 1, Wales)

4.4.2 Boosting

Boosting made sense to parents and was seen as **necessary** to improving the **social value** of the findings of the study by further ensuring diverse experiences are represented in research findings. Parents understood boosting would allow for smaller groups, and for differences within these groups, to be understood.

Parents were unconcerned about boosting, apart from similar concerns raised about tailored recruitment materials surrounding communications not making study participants feel targeted or judged.

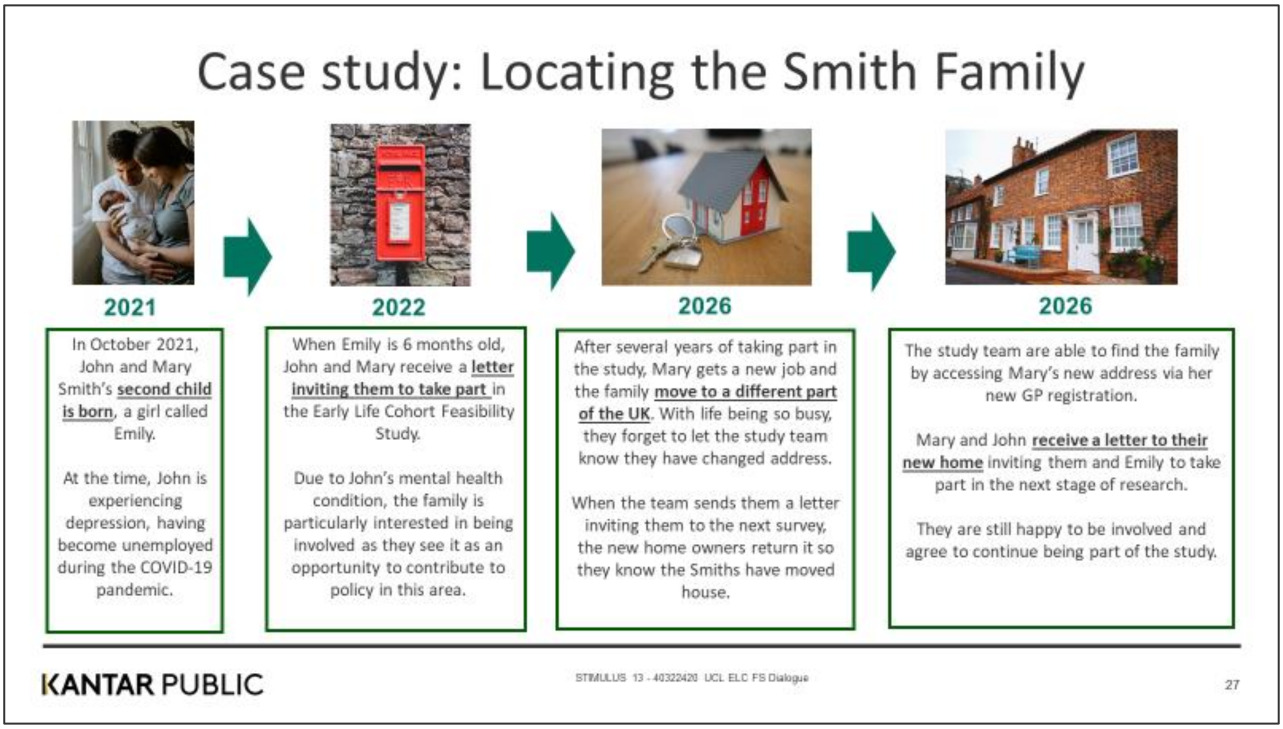
> *"The booster was something I was thinking about myself. If you increased the number of ethnic minorities than you needed, if there were previous studies to show that was the group that dropped off more quickly, then it would make sense to add more than required because you would rather have to take people out than have to find people."* (Wave 1, North England and the Midlands)

4.5 Views about use of administrative data for retention purposes

It was explained to parents that as study participants are randomly sampled in birth cohort studies, they cannot be replaced. This means it is important to keep participants in the study as they cannot be replaced by other participants.

Parents were further told how administrative data can be used to trace participants the study team has lost contact with because they have moved and not informed the study team of their new address. Stimulus materials were used to show an example of how someone might drop out of a study and be traced.

Figure 8: Tracing participants case study



Case study: Locating the Smith Family

2021

In October 2021, John and Mary Smith's **second child is born**, a girl called Emily.

At the time, John is experiencing depression, having become unemployed during the COVID-19 pandemic.

2022

When Emily is 6 months old, John and Mary receive a **letter inviting them to take part** in the Early Life Cohort Feasibility Study.

Due to John's mental health condition, the family is particularly interested in being involved as they see it as an opportunity to contribute to policy in this area.

2026

After several years of taking part in the study, Mary gets a new job and the family **move to a different part of the UK**. With life being so busy, they forget to let the study team know they have changed address.

When the team sends them a letter inviting them to the next survey, the new home owners return it so they know the Smiths have moved house.

2026

The study team are able to find the family by accessing Mary's new address via her new GP registration.

Mary and John **receive a letter to their new home** inviting them and Emily to take part in the next stage of research.

They are still happy to be involved and agree to continue being part of the study.

KANTAR PUBLIC

STIMULUS 13 - 40322420 UCL ELC FS Dialogue

27

This use of administrative data was viewed as acceptable as parents understood the usefulness of tracing to keeping the research **socially valuable** and saw it as **not impacting their control** over participation in the study. Parents understood the importance of maintaining the sample and how tracing would assist this. For example, parents described how tracing ensures study participants are not lost if they forget to update the study of new addresses when they move house. Additionally, some recognised its ability to allow those who have dropped out of the research for several years the opportunity to re-start participation. After learning about the importance of maintaining a sample, a number of Engaged Enthusiasts and Amenable Accepters even suggested that they would expect the study to trace participants because of the value of keeping study participants. Parents gaining an understanding about the purpose and value of tracing led to this acceptability.

> *"I do feel you are within your right [the studies right] to try and regain contact and if I was to sign up, I suppose that would be an expectation. I would expect that level of communication."* (Wave 1, North England and the Midlands)

> *"From a researcher's point of view, you've invested a lot of time and money into each family that you pick so you have right to find out if you're definitely opting out or not because otherwise you could be wasting more time and money. If that is clear from the start, there should be that expectation."* (Wave 1, North England and the Midlands)

As well as understanding the value of this process, parents also did not see it as affecting **control** over participation. When parents considered the impact this has on study participants' ability to choose whether or not they wanted to participate in the study, they concluded study participants are always able to withdraw from the study, even when traced. For example, parents described how even if a family have specifically chosen not to respond to a survey or not informed the study team of their new address as they didn't want to take part, once traced they can choose to formally withdraw. This understanding drove acceptability to the use of tracing.

> *"The only way you're going to know if something is ok, is to ask. It's ok to ask, you're not hounding me, you're giving me the option, you're not forcing the person to do it [when you trace them]. And if they then decide they no longer want to be involved, then that's ok too, you've given them the choice."* (Wave 1, North England and the Midlands)

Despite understanding the value and continued ability to opt-out, parents generally still wanted the study team to inform them that tracing might take place from the outset of the study, as this offers them **control** and **transparency** over the use of administrative data. As with the Two Step recruitment approach, parents felt it was right and fair for the study team to inform them about tracing before it takes place. Similarly, this was not because parents were unhappy with tracing, but because they thought it was the correct way to treat study participants. Parents wanted information about tracing when signing up to the study so that participants were opting into the study were well informed. If this was done, parents were generally happy with this proposed use.

> *"There needs to be some transparency at the start, where they say if we lose contact with you, we'll use these methods to try and make contact again."* (Wave 1, North England and the Midlands)

One concern parents did have was that tracing could be problematic for those wanting to protect their address information. For example, worries were expressed for any participants who had left abusive partners, as it was felt they may be concerned the partner may get access to their new address through tracing.

> *"Maybe someone's split up from a violent partner…then one party might be worried that the other might get access to their personal information."* (Wave 1, North England and the Midlands)

Furthermore, there was a sense among a number of parents that the use of administrative data for tracing was not the most **necessary** of the proposed uses discussed, as there are other contact methods available, which are more suitable, simpler and do not draw on administrative data. Some parents felt it would be simpler for the study to collect participant email addresses and telephone numbers and use these to make contact in the first instance. Similarly, collecting contact information for a secondary contact (such as a family member or friend) who could put the study back in contact with families was seen as more acceptable. Incentives were further felt to be a good means to prevent drop-outs and the need for tracing. Parents expected these options to be explored and actioned before administrative data was used to trace people.

> *"Yes it makes sense to have a second contact. If it's not the father of the baby, better another family member rather than a friend, as they are more likely to change over time."* (Wave 1, Northern Ireland)

5. Responses to proposed uses of de-identified administrative data in the ELC-FS: for linkage to survey responses

This chapter outlines the response to one of the proposed uses of de-identified administrative data in the ELC-FS. It explores parents' views about linking survey responses to administrative data for analytical purposes and potential consent models to gain participants' permission to do so.

Parents were introduced to this proposed use of de-identified administrative data at the second wave of the dialogue. Parents were shown two videos (from ADRUK and Next Steps) outlining what data linkage is, how it is carried out and the measures in place to protect the security of the data.[15][16]

Parents were also shown information highlighting the benefits and risks of data linkages and discussed which were the most important to them. After this, parents were shown further information about what types of data the study team wish to link with study participants' survey responses and parents discussed in a sorting exercise which types of data they were the most and least comfortable being linked.

Finally in this part of the workshops, parents were introduced to five potential consent models the study team could use to gain permission to link together survey responses with administrative data. Parents shared their initial views around the consent models before discussing for each type of data, which consent models were acceptable for the study team to use. The stimulus materials shown to parents can be found in the Workshop Materials, and were slightly different in each nation.

Key Findings

- *Overall, parents were supportive of data linkage to survey responses. This was because the use of de-identified administrative data was perceived to have low levels of risk of potential harm to the individual and high levels of data security were understood to be in place. However, parents had greater reservations around the study team accessing more sensitive types of data including mental health, social services and criminal records information. Parents tended to have fewer concerns over sharing their own data compared to their child's.*
- *There were mixed preferences regarding the consent models; although* ***Model 4B (Participants are told the types of administrative data that will be linked and are given the option to agree or disagree for each type)*** *was overall their preferred model and* ***Model 1 (Participants are not told their survey responses will be linked to their administrative at any point)*** *was seen as the least acceptable as it offered no choice to study participants and they were not informed of the linkages. For some,* ***Model 2 (Participants are told their administrative data will be linked to their survey responses and agree to this by signing up to the study),*** *whilst not their preference, was seen as sufficiently acceptable.*
- *Despite parents having mixed views around their preferences for consent models there was consensus around principles that should be used regardless of approach, these included to ensure* ***transparency, minimise participant burden and provide reassurances to build trust.***

5.1 Views about linking de-identified administrative data to survey responses

5.1.1 Overall responses to data linkage

Similarly to parents' initial awareness of administrative data, parents were not aware of data linkage as a process and the term was viewed as abstract. The stimulus consisted of two videos from ADRUK and Next Steps (another cohort study run by CLS) which explained what data linkage is, how it is completed and its

---

[15] Administrative Data Research UK – ADR UK, https://www.youtube.com/watch?v=ij4mNEFBdyg

[16] Next steps, longitudinal study of young people in England born 1989-90. https://www.youtube.com/watch?v=W6ZuK3lYW6Y

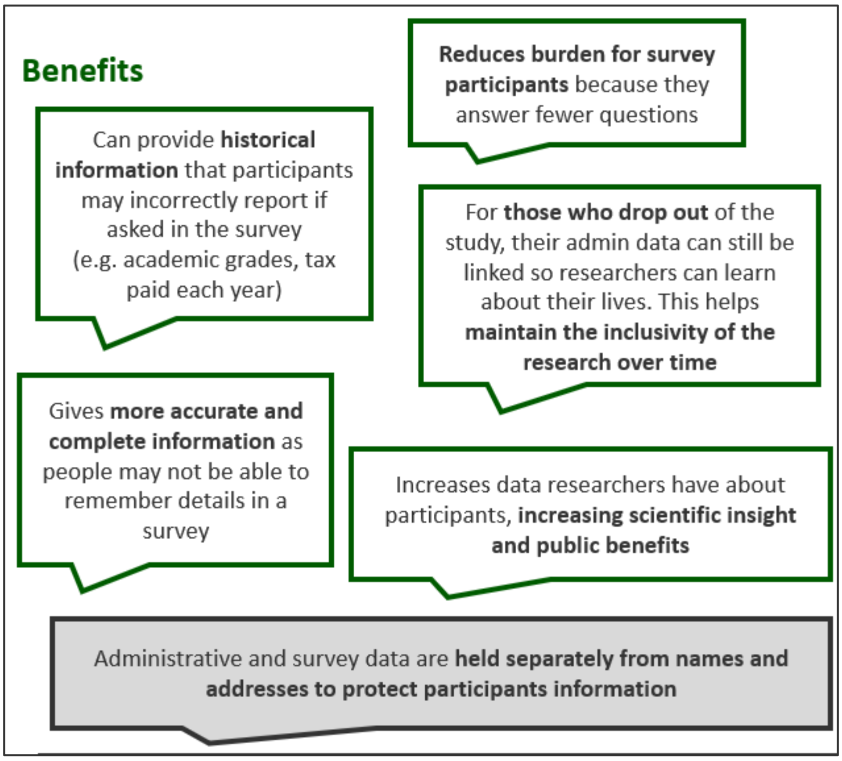
benefits and risks[17]. Parents found these videos accessible, and they helped them to build a clearer image of the data linkage process. Consequently, once introduced to the stimulus materials, parents had good comprehension of data linkage to survey responses, the benefits and the risks.

Overall, parents were supportive of data linkage to survey responses. This was because the use of de-identified administrative data was perceived to have low levels of risk of potential harm to the individual and high levels of data security were understood to be in place. Parents felt reassured the information accessed would not be used to the detriment to the participant. Additionally, the security measures mentioned in the videos about how the data is stored and transferred also provided reassurance to parents, these included the stripping the data of identifiers and the destruction of files with their identifiers once completed. Parents were reassured that they could withdraw their consent for data linkage to their survey responses at any point, which they thought was essential.

5.1.2 Perceived benefits of data linkage to survey responses

At the second wave, parents were introduced to a series of benefits which linking survey responses to administrative data can provide. From these benefits, parents selected three benefits as being particularly valuable to them, which are outlined below Figure 9 below shows the benefits parents were introduced to.

Figure 9 Data linkage benefits shown to parents



Benefits

Reduces burden for survey participants because they answer fewer questions

Can provide **historical information** that participants may incorrectly report if asked in the survey (e.g. academic grades, tax paid each year)

For **those who drop out** of the study, their admin data can still be linked so researchers can learn about their lives. This helps **maintain the inclusivity of the research over time**

Gives **more accurate and complete information** as people may not be able to remember details in a survey

Increases data researchers have about participants, **increasing scientific insight and public benefits**

Administrative and survey data are **held separately from names and addresses to protect participants information**

- Firstly, that linking survey responses to administrative data can **provide more information** about study participants which provides a richer picture which helps to enable greater analysis for researchers. This was more important to parents who were Engaged Enthusiasts or those who were Amenable Accepters.
- Secondly, the use of administrative data was seen to **reduce the burden** to study participants, and this partially mitigated the practical concern parents had about participating in a birth cohort study. In turn, parents thought the study having minimal burden for participants would also help keep study participants engaged over time too.
- Lastly, linking together administrative data and survey responses was seen as offering improved **accuracy,** as the study team are less reliant on participant recall in survey responses. This would be especially useful for gathering information on job and academic history, which parents thought would be

---

[17] Data Linkage videos sources: ADR UK – YT: What is administrative data and why is it so useful for research? https://www.youtube.com/watch?v=ij4mNEFBdyg. Next Steps – How does adding data help your generation? https://www.youtube.com/watch?v=W6ZuK3lYW6Y

harder to recall as participants get older. Parents also highlighted for more sensitive data, such as criminal records information, participants may be reticent to share this information due to fear of being judged. Parents thought using data linkage as a means to access these more sensitive types of data was a useful approach, which would avoid participants feeling uncomfortable or judged. Parents thought ensuring the ELC-FS has accurate data was highly important and without it, the credibility of the research was undermined. For parents who belonged to the Engaged Enthusiasts and the Amenable Accepters this was especially important.

> “*What’s the point in doing the study if you’re not getting correct data.*” (Wave 2, Scotland).

> *“The reality is…we leave a trail of data behind us publicly anyway, Linkedin, Facebook…in the modern world that’s the way things are*.” (Wave 2, South and East England)

> “*The benefits clearly outweigh the risks for using our data. I wouldn’t think twice about it personally.*” (Wave 2, Northern Ireland)

5.1.3 Concerns about data linkage to survey responses

Despite parents understanding and valuing many of the benefits of linking survey responses with administrative data, they also had concerns around the potential risks they may incur as a result. Parents considered five risks of data linkage to survey responses as particularly concerning (described below) and this influenced how acceptable they found linking survey responses to administrative data. Unlike the benefits of data linkage (described in the previous section), the concerns parents had about data linkage to survey responses, which are outlined below, were a combination of ones they were introduced to in the workshops and spontaneous ones too. Figure 10 below shows the risks parents were introduced to as part of the workshop.

Figure 10 Data linkages risks shown to parents

Risks

If data is not stored securely and handled carefully, there could be a greater **risk of data breaches**

Large quantities of information about individuals will be held by CLS

Data linkages may not always be accurate, and records of different people can be mistakenly attached together

Some groups may not have as much information to be linked as other groups, meaning less can be learnt about them (e.g. those recently moved to the UK, or who do not regularly engage with public services)

Participants may not fully understand how and why their data is being linked when they agree to this being done.

If data linkages are **not done for everyone** (if people don’t consent to linkages)**, the linked data may not be as inclusive or representative**

- Firstly, parents were concerned about their **administrative data being used without them being notified which was not something they were comfortable with.**
- **Secondly, they were concerned administrative data could potentially be used against the wishes of those who have actively withdrawn from the study**. For both the first and second concerns parents felt that if the study team accessed their data without their permission, or without them being notified, this would undermine the transparency of the ELC-FS and erode their trust in the study.
- Thirdly, **parents also feared being misrepresented as administrative data is ‘black and white’.** Parents were concerned where administrative data is sparse and does not provide a holistic picture, it could lead researchers to draw inaccurate findings, such as interpreting a participants’ situation from

their health records, although the participant does not feel this truly reflects them. Parents with this concern felt the study team would be able to access greater context if they used survey responses instead.

- A fourth concern parents had was **the study team would hold too much data on them.** Parents had concerns around the volume of data the study team would hold on them. The stimulus shown to parents gave only summary information about the different types of data, and not how much data would be collected and what specifically this data would include. This concern fed into parents' fear of potential harms to the individual through data linkage.
- A fifth concern parents had was that they would feel **responsible for their child's outcomes.** For example, if through the study linking together parents' mental health data with their child's outcomes, the study team found that those children whose parents had mental health issues were shown to have more negative life outcomes as a result. Parents felt if they could identify themselves as the parents with the mental health issues, this would leave them feeling responsible.
- Lastly, parents were concerned around **the potential individual scrutiny** they may incur as a result. Parents feared how they appeared on an individual level through their administrative data and had a sense of being 'judged'. This is reflective of parents' fear of **social stigma** as a form of harm from data linkage. Consequently, parents thought study participants would not want to share more sensitive types of data e.g. criminal records information. Parents also noted for those with a criminal record who had changed their lifestyle they may not want to be associated with this anymore. However, once stakeholders explained the way this data is analysed will be at a collective level this fear was mitigated. Some parents also felt this could be mitigated if they were able to see the administrative data the study had collected on them.

Some other concerns were also raised. Some parents' observed that the majority of the benefits of linking together administrative data and survey responses was for the researchers, whereas the potential risks would be shouldered by the participants in the study. They felt this was unfair and would have preferred that the risks were shared equally making the study seem more of a partnership.

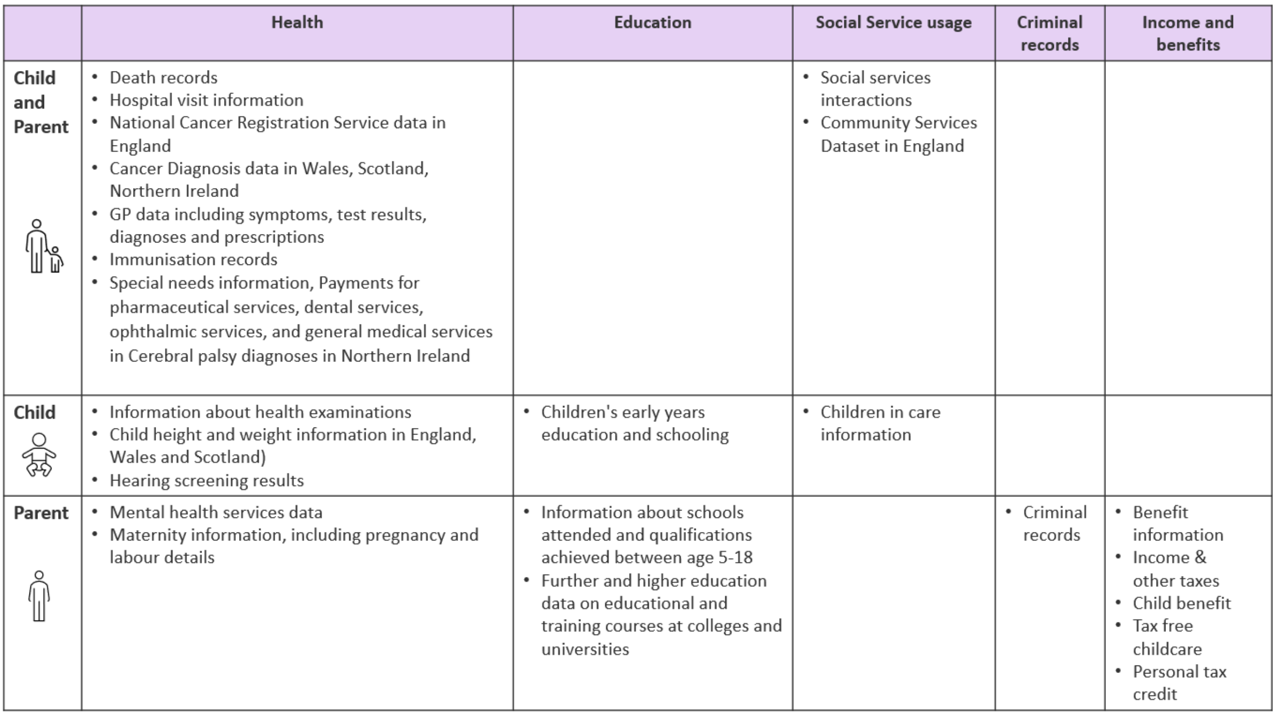
> *"Seems like the risks are more towards us, whereas the benefits are more towards the researchers. They're more to make it more efficient for the researchers*" (Northern Ireland, Wave 2)

Whilst in South and East England, parents also had concerns around whether these linkages could be 'future proofed' and worried in 20 years' time this data will be less secure. Parents in Wales also raised the point that the administrative data the study team intend to access may be limited. For example, the administrative data available may not contain groups who interact with public services less.

5.2 Types of linkages proposed (per nation)

In the second wave, parents were presented with the potential types of administrative data the study team would like to link with the survey responses. These are shown in Figure 11.

Figure 11 Data linkages across nations combined



	Health	Education	Social Service usage	Criminal records	Income and benefits
Child and Parent	- Death records - Hospital visit information - National Cancer Registration Service data in England - Cancer Diagnosis data in Wales, Scotland, Northern Ireland - GP data including symptoms, test results, diagnoses and prescriptions - Immunisation records - Special needs information, Payments for pharmaceutical services, dental services, ophthalmic services, and general medical services in Cerebral palsy diagnoses in Northern Ireland		- Social services interactions - Community Services Dataset in England		
Child	- Information about health examinations - Child height and weight information in England, Wales and Scotland) - Hearing screening results	- Children's early years education and schooling	- Children in care information		
Parent	- Mental health services data - Maternity information, including pregnancy and labour details	- Information about schools attended and qualifications achieved between age 5-18 - Further and higher education data on educational and training courses at colleges and universities		- Criminal records	- Benefit information - Income & other taxes - Child benefit - Tax free childcare - Personal tax credit

Overall, parents found the proposed data linkages for their nation acceptable and there was little variation across nations and regions in England. However, parents did identify several types of data as more sensitive than others, namely mental health, social services and criminal records information. Even though many parents did not have a record themselves for these types of data, they imagined for those that did, they would be reluctant to share this information. These data types were seen to be sensitive as they carried a certain amount of social stigma. Although parents could appreciate how this information would be useful to measure how these characterises impact on children's development and outcomes, even though this was likely to be in a negative way.

Health data

Across groups parents appreciated the value of health data being included in the study data and how it can be used to monitor children's development. However, they felt mental health data was likely to be more sensitive and study participants may feel uncomfortable sharing this personal data. Some parents raised concerns around the accuracy of this data, and some had experiences where their information held by their GP was out of date or inaccurate. GP data was also seen as potentially sensitive as it could include information such as if participants had reported experiencing depression or domestic violence. Parents were also curious how this data would be presented and specifically, the level of detail the study team would be able to access. This is reflective that administrative data is still an unfamiliar concept. Parents preferred information from administrative data to be provided with less specificity, for example either as counts or in bands.

In South and East England, some parents felt that prescription and immunisation records data was sensitive as they could have potential implications for the study participants in the future e.g., impact on insurance. Additionally, parents in South and East England thought mothers who had had difficult pregnancies and births may also feel the birth and maternity record data is sensitive and be less comfortable sharing it. On the other hand, some parents felt accessing this information through data linkages could provide a way for researchers to access this information without participant's having to relive traumatic memories, and therefore may be preferable.

> "*There is a lot under one title – you have mental health, cancer, immunisation and this all under one umbrella now. The certain things about my health data that I am more sensitive about than others. I don't care if you know about my Covid jab or Flu jab or whether I had chickenpox as a child, but less so whether you know I've been to the doctors to say I'm depressed or domestic abuse. There is more sensitive data held by my GP than other data.*" (Wave 2, North England and the Midlands).

> “*I’m comfortable with all of it if it is depersonalised – why does any of it matter. Why would anyone not be comfortable if depersonalised. The more information you can give the researchers that is depersonalised the better they can do their job.” (*Wave 2, Northern Ireland)

Social services data

Across groups, social services data was identified as one of the most sensitive types of data the study team were attempting to link. Social services data was seen to carry a considerable amount of social stigma, and for those who have interactions with social services, this may not be something that study participants would like to be linked.

> “*Could be a touchy subject…a point of shame if you’ve had involvement with social services.”* (Wave 2, North England and the Midlands).

Parents also raised (for those who have this type of data) they may worry about potential repercussions if this data is shared and linked with their survey responses. However, parents could also see the value in accessing this type of information, with those who worked with children, highlighting the importance of understanding these children’s outcomes.

In South and East England, parents raised concerns around the accuracy of the social services data given the social services sector is seen to be stretched, which may lead to out of date, incorrect or patchy data being held. Whilst, in North England and the Midlands, parents agreed this data was highly sensitive but essential for the study so researchers could understand, for those who interact with social services, how this relates to their life outcomes. Consequently, they did not think that study participants should be given a choice on whether these are linked to their survey responses.

Criminal record data

Criminal record data was felt to be highly sensitive, especially in Northern Ireland where some parents speculated that participants who have a criminal record associated with the Troubles would not want to share this information. However, other parents, across all locations, felt less concerned about this being linked. Some parents noted that whilst you may not have a record now, this could change in the future, and this may be worrying for potential participants.

> “*Northern Ireland is quite unique in that with the troubles there was a lot of people that have criminal records and under the Good Friday agreement having a criminal record linked to the troubles, that isn’t allowed to hold you back in education, jobs.”* (Wave 2, Northern Ireland)

> “*They may have turned a corner and want to move on...so may be less forthcoming in handing over police records.”* (Wave 2, Scotland).

> “*It would feel like something held over you, and if you’ve changed and got past that, you would not want that shadow over you*.” (Wave 2, North England and the Midlands)

Education data

There was agreement that education data was the type of data parents most expected to be shared in a birth cohort study, and this was widely viewed as acceptable. Additionally, parents felt this data was not as sensitive as health data and had less impact on their lives.

> “[Education data] *is not as personal…I’d be more concerned with the health data and the mental health aspect, because that’s quite a private thing to people*.” (Wave 2, North England and the Midlands)

A few parents thought study participants who had gained fewer qualifications or had a harder time at school may be less willing to share this information.

> “*I don’t mind who knows I went to college or university. I had a successful education and it is not a problem for me and I feel comfortable about but for others I could see why it might do*.” (Wave 2, North England and the Midlands).

However, for the majority of parents, their education was something they saw as more ‘neutral’ and was not something they felt defined by. This view was further reinforced as the data would be de-identified, so parents felt the study team would not know who it belonged to anyway.

> "*Your education doesn't necessarily define who you are now, obviously for some people it can do, but it is what it is - you went to college or uni or you didn't*” (Wave 2, North England and the Midlands).
>
> "*It doesn't matter because they don't know who I am so go ahead*." (Wave 2, Northern Ireland)

Income and Benefits data

Overall income data was not viewed as sensitive, although some parents were initially unsure of the value of linking this data to survey responses and were concerned this may include their bank details, which they worried could be used in a nefarious way. However, once explained that this can be used to measure socioeconomic impacts on children, parents appreciated it was useful and that their bank account details would not be used, this mitigated these concerns. Broadly, as government agencies already have access to this information parents were happy to share it. In Northern Ireland, parents raised concerns that HMRC records on income may be inaccurate for those who are self-employed as they may not honestly declare their income.

Benefit data was seen as slightly more sensitive than income data, but parents still thought it was acceptable to access and link.

> “*If benefits, already part of that big system anyway so it's not something I would really think about or be uncomfortable about because it's already out there and it wouldn't really bother me*" (Wave 2, North England and the Midlands).
>
> “*Again, I don't think that really bothers me - I earn what I earn. if I get more, I get more and if I could get more I would get more. I’d love to be in a job that paid loads of money but the reality is that I don't.*" (Wave 2, North England and the Midlands).

Across all types of linkages, parents also had questions around how the study would achieve these linkages from a practical perspective. These included how the study team would access and incorporate administrative data across nations and from outside of the UK. For example, parents in Scotland raised that historic healthcare data may be harder to link, and Scottish records may not necessarily link up with English records. Whilst parents who were born outside the UK queried how the study team intended to access their childhood health and educational records.

5.2.1 Child versus parent linkages

Parents tended to have fewer concerns over sharing their own data compared to their child’s. This view may also be because parents were more comfortable deciding the use of administrative data for themselves, than for others.

> “*Don't forget mental health and emotional wellbeing are at an all time high. This may seem like highly sensitive information to share. It will be important to explain this isn't personal about the individual, it's not about me but society*” (Wave 2, North England and the Midlands)
>
> “*If there was any risk I would want to put myself in front of it first and foremost, before my child. If I'm not willing to give up my information then I shouldn’t be willing to give up my child's information*” (Wave 2, South and East England)

Some parents felt children’s social services data was particularly sensitive and carried a significant amount of social stigma.

> ‘*If you have your child taken away from you that would be more of a stigma and the shame of that. It would be more in the person’s head than they are trying to hide it and about the stigma that comes with it., so I completely understand people not being comfortable with some things*" (Wave 2, Northern Ireland).

Although knowing clearly how this data will be stored, who had access to it and what it will be used for and that it is de-identified helped to alleviate concerns around how they would be perceived.

> "*It will be important to explain clearly why the research is being done, who will use the data, what benefit it adds, and that's it's anonymous. It's really important that it's made clear its anonymous*. (Wave 2, Northern Ireland)

5.2.2 Linkage to parents' records from before the baby's birth

Despite, parents being more comfortable sharing their data than their child's, some parents felt that those who had had more challenging upbringing may not want to be reminded about the experience, which may make them more reticent to agree to data linkage. Other parents also felt how the 'types of data' the study team wanted to link was framed made it seem negative and felt the study team were interested in how different aspects from childhood lead to worse life outcomes.

> "*This might be a part of their life they don't want anyone to know as it is locked away. It's gone and they've changed their life. Again, it all depends on the person"* (Wave 2, North England and the Midlands)

> "*A lot of the data from a personal point of view could be seen as quite negative - do you have criminal convictions, do you have poor education, are you on benefits, did you drink and smoke during your pregnancy*." (Wave 2, North England and the Midlands)

Parents were also concerned they may feel they were being scrutinised and judged for previous behaviours that led to poorer life outcomes for their children, especially for behaviours during pregnancy. This concern was higher for Engaged Sceptics and Entrenched Cynics. These included smoking during pregnancy or drug or alcohol use. These parents suggested placing a limit on how far back researchers went to access these records to ensure the records reflect study participants now.

> *"If you haven't had issues, you'd be more inclined to consent…like women who smoke when they're pregnant or drink may not want that link, because they don't want to feel like something they've done has affected their child*." (Wave 2, North England and the Midlands).

> "*I expected most of it but it is quite a lot, I don't know they would go into parent information but I don't mind that and I guess it is information that is needed - you need to go back, but what is really my concern is if it is kept safe."* (Wave 2, South and East England)

Parents also thought women who had been through difficult pregnancies and labours may not want researchers to access this information as it felt more sensitive One participant also raised that people who had been adopted may not wish this information to be accessed by themselves or others.

> "*My mum was adopted…she chose not to see data about her from when she was a child. .if there are people in the survey like this, should [CLS] access regardless?"* (Wave 2, Wales)

On the other hand, most parents, who saw the benefit of this approach, thought it was acceptable and especially useful to further understand child development. Parents' with medical conditions in particular were interested to know how chronic disorders are inherited or may further impact on their child's development. For example, studying mixed race babies and barriers they encounter in the identification of jaundice or the long-term impact on children whose mothers had gestational diabetes. These parents preferred an approach that went as far back as records would permit to build a fuller picture, which would enable greater analysis. This was a preferred approach amongst Engaged Enthusiasts and Amenable Accepters. These parents thought the value and importance of these findings was a greater priority than offering individual participants' choice on whether they are comfortable for this data to be linked, which is reflective of how they value the findings from birth cohort studies. Although it should be noted that in the ELC-FS only where consent has been collected will data linkage take place. Again, parents raised practical concerns or how they would access this data for those who were born outside of the UK.

> "*Logically it makes sense to use birth records, it feels like the right type of data to include*." (Wave 2, Northern Ireland).

5.2.3 Linking data for those who have stopped completing surveys

Parents had mixed views on whether continuing to link the administrative data for those who have stopped completing surveys was acceptable. Overall, parents tended to think this was acceptable (for those the study loses contact with) if study participants had agreed to this from the outset of the study when they first gave their permissions to link to administrative data. Additionally, some parents would like consents to link to administrative data to either be recollected during their time in the study or for them to be reminded periodically about the consents they have agreed to and the option to withdraw these consents. Although it is noted that due to time constraints in the workshops, this was not explored with all groups.[18] Parents views about this were also informed by the first wave of the dialogue, where it was seen as acceptable to use identifiable data to contact participants whom the study team had lost track with. They felt that identifiable administrative data could be used to recontact these participants, and then they can choose whether they are happy for their de-identified administrative data to continue to be used for analytical purposes.

In Northern Ireland, parents did not think this needed to be full reconsent collected each time – but could be simply a reminder that participants have agreed to these linkages to take place and can withdraw at any time. They felt this balanced the study team's need to maintain a higher number of consents, whilst offering sufficient choice to study participants, which does not place too much burden on them.

> **"***If there's any uncertainty I think it's better to check*."(Wave 2, Scotland)
>
> "*Unless they've actually said they don't want to take part anymore then it would be alright.*" (Wave 2, Wales)

5.2.4 Linking data for those who have actively withdrawn from the study

Parents were generally surprised that data linkages could continue for participants who had actively withdrawn from participating in the survey. Parents felt this was against the wishes of those who had actively withdrawn from the study . Instead, parents thought at the point of withdrawing from the study, to continue these linkages, consent needs to be clearly collected then for this so that participants are offered a choice. We note that due to time limitations that this topic was not covered in depth in all groups.

> "*I think if people drop out of the survey then it all has to end...clearly the survey team can try to persuade you otherwise...but if people decide to end, their wish has to be respected and you can't continue to track, that seems really sneaky*." (Wave 2, Scotland).

5.2.5 Use of sampling frame data for linkage to survey responses for research purposes

Overall parents generally thought the **use of the sampling frame data for linkage to survey responses for research purposes was acceptable**. This was because it is an efficient use of the information the study team already has about study participants. This use of data appealed to parents' common sense and seemed practical. However, parents agreed this use of the data **should be made clear to participants from the outset of the study in the recruitment and introduction materials** to confirm their acceptability of it. Whilst there were mixed views, many parents also felt that consent should be explicitly collected for this, along with the other data linkage consents. We note that parents widely thought it was unlikely that study participants would object to this use but to maintain participants' **trust in the study** they should still be at least informed. In Wales, certain characteristics were highlighted as particularly sensitive including the country of birth of the mother or father for asylum seekers. Parents thought it was especially important to inform these groups of this use to carry out data linkage to survey responses.

A small group, primarily consisting of Amenable Accepters, felt gaining specific explicit consent was not necessary as they assumed the study team would be using the sampling information for this purpose already – but they still wanted study participants to be informed that data was being used for this purpose.

> *"If they've got the information [already] they might as well use it*." (Wave 2, North England and the Midlands)
>
> "*At one level the information is already there, so it seems fair to be able to use it. But I think it's important people are informed that's it being used and checking in to make sure it's ok. What harm is there in asking, if the person has agreed to be involved in the study, they're probably not going to say no to it*." (Wave 2, Northern Ireland)

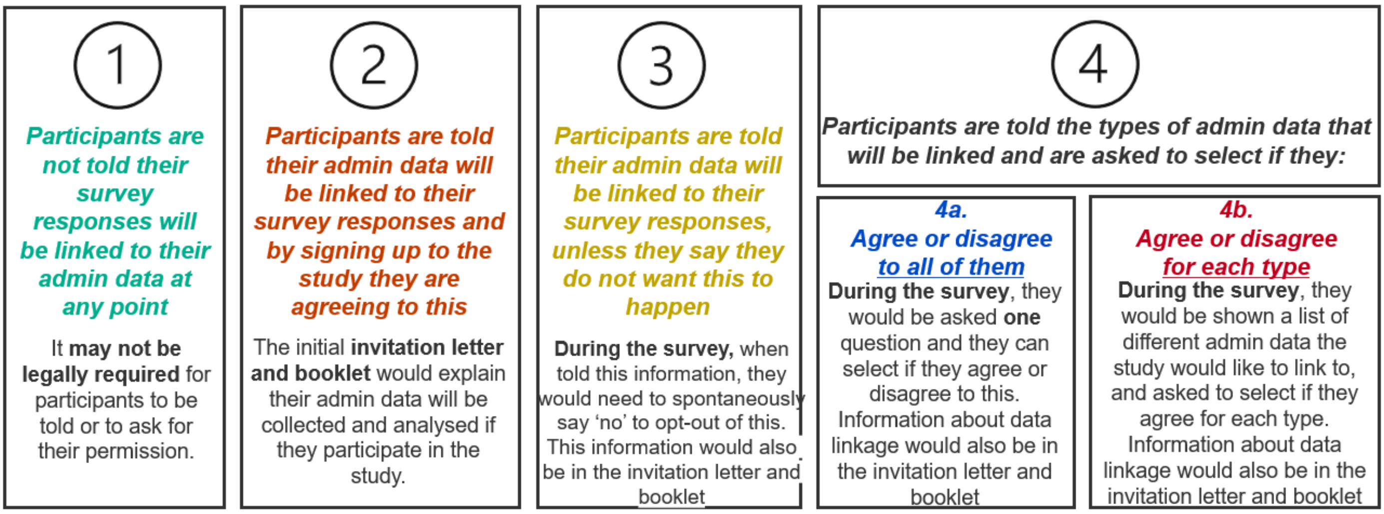
---

[18] To fully determine whether parents felt re-collecting consent periodically or whether reminding participants about the linkages they had already agreed to periodically are acceptable this would need to be explored further in more research.

5.3 Views about potential consent models for data linkage

At the second wave, parents were presented with five potential consent models to collect participant agreement to the proposed data linkages. These consent models are shown below in Figure 12.

Figure 12 Consent models for data linkage



1

Participants are not told their survey responses will be linked to their admin data at any point

It **may not be legally required** for participants to be told or to ask for their permission.

2

Participants are told their admin data will be linked to their survey responses and by signing up to the study they are agreeing to this

The initial **invitation letter and booklet** would explain their admin data will be collected and analysed if they participate in the study.

3

Participants are told their admin data will be linked to their survey responses, unless they say they do not want this to happen

During the survey, when told this information, they would need to spontaneously say 'no' to opt-out of this. This information would also be in the invitation letter and booklet

4

Participants are told the types of admin data that will be linked and are asked to select if they:

4a.

Agree or disagree <u>to all of them</u>

During the survey, they would be asked **one** question and they can select if they agree or disagree to this. Information about data linkage would also be in the invitation letter and booklet

4b.

Agree or disagree <u>for each type</u>

During the survey, they would be shown a list of different admin data the study would like to link to, and asked to select if they agree for each type. Information about data linkage would also be in the invitation letter and booklet

Overall, parents had a good understanding of the key issues surrounding consent models for both the participant and the study team. This included the competing demands of the study team's need to collect as many linkages as possible from as many study participants as possible to improve the inclusivity of the study findings, to build a more detailed picture and to complete further analysis, balanced against the study participant's desire to select which linkages to agree to, which is likely to diminish the number of people consenting and consents collected.

There were mixed preferences about the consent models, although **Model 4B (Participants are told the types of administrative data that will be linked and agree or disagree for each type)** was overall the most common preference. Whilst it may not be their first preference, for some, **Model 2 (Participants are told their administrative data will be linked to their survey responses and by signing up to the study they agree to this)** was sufficiently acceptable. **Model 1 (Participants are not told their survey responses will be linked to their administrative data)** was the least acceptable and widely felt not to be acceptable because it was not transparent and did not give study participants a choice.

In Table 1 below, a summary of parents' responses to the consent models is presented.

Table 1 Parents responses to proposed consent models

	Benefits	**Concerns**	**Overall acceptability**
Model 1 - Participants are not told their survey responses will be linked to their administrative data	• As research team are not asking permission, they can complete all data linkages for study participants. They build a more detailed picture of participants and are able to complete greater analysis. • Less burdensome on the study participants as they do not have to read through materials to	• Parents' felt this approach was the least ethical as it did not provide transparency and choice to parents. • Regardless of whether the study team are able to complete this approach, parents felt they would be upset to learn this had occurred and this would erode their trust in the study,	Unacceptable, although for a small number of Engaged Enthusiasts and Amendable Accepters it was acceptable, as they thought the benefit to the study took priority over individual choice.

	determine which linkages to consent to.	and potentially lead them to drop out.	
Model 2 - Participants are told their administrative data will be linked to their survey responses and by signing up to the study they agree to this	• Easier for the study team to administer. • Potentially a more representative approach than consent Model 3, since fewer study participants are likely to opt-out from data linkage as they may miss the information in the recruitment materials. • Preferable to Model 1, as there is a greater element of transparency.	• Although more ethical than Model 1 – parents still thought it was likely people may miss this information in the materials. • Does not fully inform study participants as it is unclear exactly what types of linkages will take place. • Study team may be able to complete fewer linkages than if using Model 1. • Study team may have fewer study participants willing to join the study as they do not view the linkages as acceptable.	A small number cited Model 2 as their preference. For some Engaged Enthusiasts, Amenable Accepters and Engaged Sceptics, it was sufficiently acceptable whilst not their overall preference.
Model 3 - Participants are told their administrative data will be linked to their survey responses, unless they say they do not want this to happen	• Viewed as more ethical than Model 1 and Model 2, as ensures study participant is made aware of linkages. • Offers participants agency to opt-out. • Likely to lead to greater numbers of consents gathered than Model 4a and Model 4b.	• Parents felt like the model was not especially clear, and that this approach was viewed as they study trying to reduce the potential number of opt-outs, which eroded their trust. • Model 3 still places considerable onus on the study participant to opt-out and it less clear how to opt-out. It may not be clear to participants they have the option to opt-out – especially for those who belong to more marginalised or vulnerable groups e.g. those where English is not their first language	Some parents felt Model 3 was acceptable and were from all four types.
Model 4a - Participants are told the types of administrative data that will be linked and agree or disagree to all	• Offers greater control for the study participant. • Less burden on the study participant to decide for all linkages as seen with Model 4b. • Potentially could lead to a higher number of consents for linkages compared to 4b as study participants may select all.	• If unhappy with one type of data linkage may reject all of them leading to researcher collecting fewer consents overall. • Less choice for study participants compared to Model 4b.	Widely seen as acceptable.
Model 4b - Participants are told the types of administrative data that will	• Offers the study participant the most choice on data linkage and therefore is the most ethical.	• Likely to result in a less inclusive, representative, and complete study, given the ease of opt-out for participants, especially around linking	Acceptable and most preferred model.

be linked and agree or disagree for each type	- Enables the study participant to be the most informed on what types of data linkage are taking place. - Model 4b also allows study participants to change their mind in the future.	data on mental health, criminal records and social services data. - Greater burden on the study team to administer. - Greater burden on study participants to read through materials and evaluate data types.	

> "[*On Model 2] It sounds like you're being sneaky to participants…going behind their backs. I think when you're running surveys like this it has to be completely open, honest and transparent*." (Wave 2, South and East England).

> "*I think Model 3 works better for the researchers. Otherwise, what's the point in doing the research if you don't have all the information*." (Wave 2, North England and the Midlands).

> [On Model 3] "*I think it's like 'I'm gunna take something of yours unless you say no, it's a sneaky way through the back door*" (Wave 2, Scotland)

5.4 Preferred models

Initially, when the models were introduced in the workshops, parents most commonly preferred **Model 4b (Participants are told the types of administrative data that will be linked and agree or disagree for each type)** as it offered the most transparency and choice to the study participant and therefore, was seen as the most ethical. **Model 1 (Participants are not told their survey responses will be linked to their administrative data**) was widely felt not to be acceptable as it was perceived to be unethical as it did not notify study participants or ask for their permission, which undermined the trust parents had in the study team.

However, as the models were explored in greater depth through the task of selecting which model was needed to link to the proposed data types, some parents shifted their position on which model should be used to collect consent. In particular, some Engaged Enthusiasts and Amenable Accepters, **who strongly believed in the value of the birth cohort study and its potential findings**, thought **Model 1 (Participants are not told their survey responses will be linked to their administrative data)** was acceptable to use to access sensitive data in certain instances. They felt that those with the social services information or a criminal record may be less inclined to agree for this data to be linked, but findings from this data were too important to be determined by the individual. Additionally, as the data was de-identified they found this reassuring and meant the potential for individual harm was limited.

> "[*On Model 1] If you don't tell them, it is unethical."* (Wave 2, South and East England)

> *"[On Model 1] If you found out further down the line that data was being shared and you hadn't been warned it would throw up some red flags and you'd probably bail*." (Wave 2, South and East England)

> *[On Model 1] "It's obviously not the most ethical way, but I think getting the right data is the most important thing. Because people are so fearful of having some things captured like criminal record, they may pull out. So, the only way to get all that information is model one, even though it's not necessarily ethical."* (Wave 2, Wales)

> *"[On Model 1] I would just say you need the data so just have it all, my comfortable level with different types is not really important -- it is all there, put it all together and de-identify me and use it"* (Wave 2, Scotland)

In contrast those who **valued their individual choice over the potential findings of the study** were unwilling to relinquish this control, which included the Entrenched Cynics and Engaged Sceptics.

> "[On model 4a] *This is a simple yes and no approach and it doesn't confuse people by bombarding them with too much choice and information*" (Wave 2, Northern Ireland).
>
> *"[On Model 4b] Everyone knows where they stand with this model, it clearly allows for choice*." (Wave 2, North England and the Midlands).
>
> "[*On Model 4b] Gives you the sense of control…from a user perspective. Also demonstrates a level of courtesy...considerate towards participants' thoughts"* (Wave 2, South and East England).

A small third group of parents also emerged who thought **Model 2 (Participants are told their administrative data will be linked to their survey responses and by signing up to the study they agree to this)** or **Model 3 (Participants are told their administrative data will be linked to their survey responses, unless they say they do not want this to happen)** was acceptable, which consisted of a mix of Engaged Enthusiasts, Amenable Accepters and Engaged Sceptics. They felt these options could provide **sufficient transparency and choice for the study participants.** Their initial preferences for **Models 4a (Participants are told the types of administrative data that will be linked and agree or disagree to all)** or **4b (Participants are told the types of admin data that will be linked and agree or disagree for each type)** were driven by being made aware of these options which provide greater choice to the study participant but potentially were not necessary. Underlying parents' views about the acceptability of **Model 2 (Participants are told their administrative data will be linked to their survey responses and by signing up to the study they agree to this)**, potentiality is that it offers a familiar set up through signing up to a service you agree to the terms and conditions, which creates a greater degree of **'normalisation'**. Additionally, these parents thought **Models 4a (Participants are told the types of administrative data that will be linked and agree or disagree to all)** and **4b (Participants are told the types of administrative data that will be linked and agree or disagree for each type)** put a **greater burden on the study participant** to consider each data type which, as it was focusing on using de-identified data, was not necessary, and this burden could also put study participants off from agreeing to the linkages.

> [*On Model 2, there is an] element of transparency…would hopefully instil confidence that your data wouldn't be shared without your consent.*" (Wave 2, South and East England)
>
> [*On Model 2] It's better to say what you're doing, as people will likely find out at some stage anyway. It's the right thing to do and help to build trust with the study participants*." (Wave 2, North England and the Midlands).
>
> "*If you presented me with Model 3 I'd probably still do it [agree to the linkages], but the fact you've shown me Model 4 means I'd be more likely to consider what data I'd agree to be linked." "What you don't know won't hurt you."* (Wave 2, North England and the Midlands).

5.5 Key principles:

Despite parents having mixed views around their preferences across the proposed consent models, there was consensus around key principles that should inform the approach that is used.

- **Transparency:** This means disclosing in the communication materials what the researchers are doing with administrative data and that participants are offered some level of choice and control over whether and how their data is used. Parents were also keen to know the purpose and value of linkage and this information could help to alleviate concerns and reticence.
- **Provide reassurances to build trust:** This should include emphasising data will be analysed at a collective level not at an individual level and is de-identified, which minimises potential harm to the individual and potential stigma experienced by the study participant. In South and East England, parents raised that since Cambridge Analytica, the public's trust in institutions and organisations accessing and using the public's data has been eroded. This suggests that this wider context is important, and that the approach used on the ELC-FS around accessing and using participant's data, is an opportunity to rebuild this trust.

- **Minimise participant burden:** The mechanism used to collect consent should put minimal burden on the study participant. This means communications should be easy to understand and accessible. This may mean making greater efforts to collect informed consent from those belonging to vulnerable groups, e.g. those with language barriers. These groups may lack confidence to opt-out from the linkages with a researcher present and therefore, greater efforts would need to be made to ensure they felt comfortable to do so and are aware it is a viable option. This approach removes the onus from the study participant on to the study team to ensure study participants are aware of the research approach and its ramifications.

6. Responses to proposed uses of de-identified administrative data in the ELC-FS: to understand the representativeness of the study sample

This chapter outlines parents' views towards the proposed use of de-identified administrative data to both assess and improve the representativeness of the study sample.

In the workshops parents were introduced to the concept of 'non-response bias (see Figure 12) which explained that certain groups are less likely to take part in studies like the ELC-FS, leaving them underrepresented in social research. Moderators explained how de-identified administrative data could be used to better understand non-response bias and the study representativeness. For example, this could be done through comparing the de-identified administrative data of the study sample (i.e.. those recruited to the study) to the selected sample (i.e. all those sampled for the study) to look for any variations. After describing the general approach to checking non-response bias, the workshop stimulus broke down the different groups that would make up the selected sample to gauge whether the study team's use of de-identified administrative data for each group was acceptable to parents (see Figure 13). It is important to note here that as parents learnt more about and discussed the processes of checking non-response bias using the selected sample de-identified data their views evolved. Parents also considered data holders checking the sample representativeness instead of the study team.

Parents also discussed weighting as a potential approach the study team could use to improve the sample's representativeness which uses de-identified administrative data.

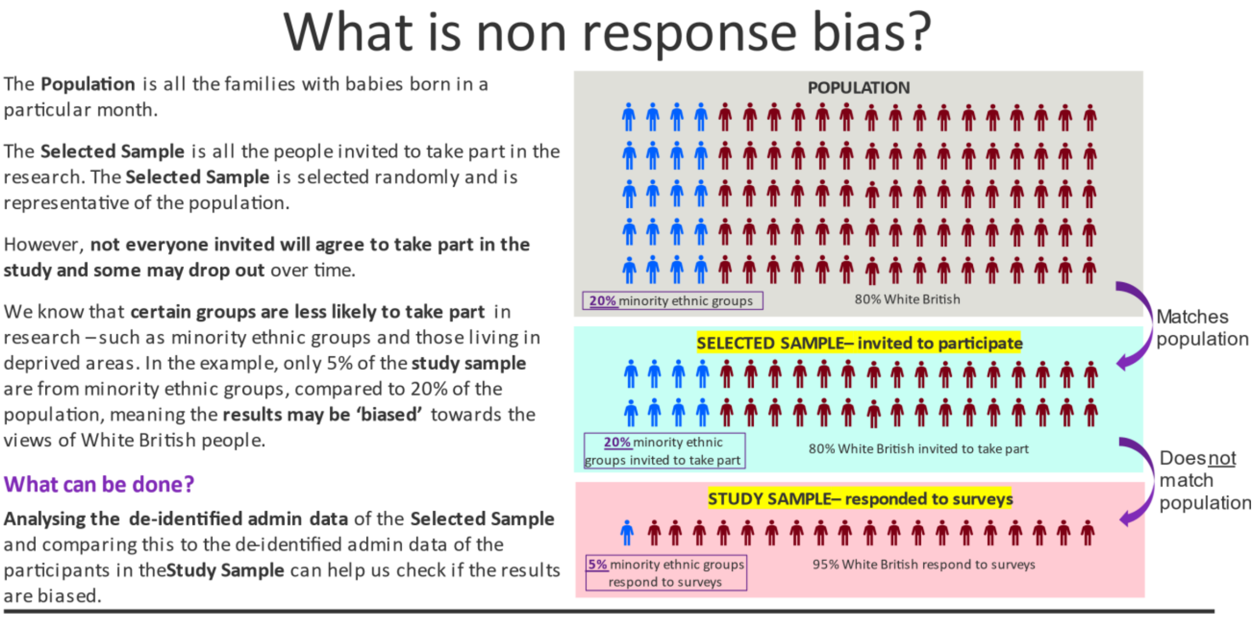
Key Findings

- *Initially parents generally found the concept of use of de-identified administrative data to check the representativeness of the study acceptable due to the social value of having a representative sample. However, when exploring the processes of checking representativeness in greater detail, parents' views evolved and became more nuanced. This was primarily due to concerns about this use of data for those who opted out of their data being transferred to the study team. A range of views emerged as to whether it was acceptable to use the data of those who had not responded to the study team's recruitment attempts, but parents widely felt that using the administrative data of those who had explicitly opted-out of the study either with the data controller or with the study team was unacceptable.*
- *Some parents were comfortable with the data controller using data for those who didn't take part in the study to check the sample representativeness as the data would remain with the data controller and not be passed to the study team.*
- *Parents widely found the use of de-identified data, including on the selected sample from the sampling frame, for weighting acceptable but struggled to understand this highly technical concept.*

6.1 Views on non-response bias and the use of administrative data to check representativeness of the study sample

During discussions about the use of de-identified data for understanding representativeness, parents were first introduced to Stimulus (Figure 13 – see below) which explained non-response bias and how it can be assessed.

Figure 13 What is non-response bias



What is non response bias?

The **Population** is all the families with babies born in a particular month.

The **Selected Sample** is all the people invited to take part in the research. The **Selected Sample** is selected randomly and is representative of the population.

However, **not everyone invited will agree to take part in the study and some may drop out** over time.

We know that **certain groups are less likely to take part** in research – such as minority ethnic groups and those living in deprived areas. In the example, only 5% of the **study sample** are from minority ethnic groups, compared to 20% of the population, meaning the **results may be 'biased'** towards the views of White British people.

What can be done?

Analysing the de-identified admin data of the **Selected Sample** and comparing this to the de-identified admin data of the participants in the **Study Sample** can help us check if the results are biased.

After exposure to the stimulus materials and explanation from the moderators and experts, there was generally good comprehension of the concept of 'non-response bias' and parents widely recognised the importance of achieving a representative sample. They understood without a representative and inclusive sample the research findings are less valid and accurate and therefore less useful, with one labelling it a potential "*waste of resources*." Parents understood that ethnic minorities are an example of groups that are more commonly underrepresented in social research and saw it as important to address this as there was wide appreciation of the importance of inclusion and diversity in our society.[19]

> *"[Representation] is a major concern…it should 100% reflect the population accurately."* (Wave 2, South and East England,)

At this point in the second wave dialogue parents generally found the use of de-identified administrative data by the study team to check representativeness as acceptable. It was seen as a low-risk process, given the data is de-identified and therefore it is unlikely to cause harm to the individual.

The degree to which parents valued social research was a key determinant to how acceptable they found the use of administrative data to check non-response bias**.** Parents all bought into the idea that the research aimed to produce positive and tangible outcomes at a societal level, which they valued to differing degrees. The degree of value they placed on the study's research aims was a key determinant in their level of acceptance of this use of administrative data – so it was more acceptable to Engaged Enthusiasts than Entrenched Sceptics.

The following factors drove the acceptability of this use of administrative data:

- **The use of de-identified data**: this made parents feel there was low risk of harm with this use of data as it was unlikely the data would be linked back to them as an individual.
- **The normalisation of processes**: where they understood this process had been routinely used in other social research studies, parents felt processes were robust and secure.

---

[19] Minority ethnic groups were used in the stimulus, as a tangible example of an underrepresented group to help with participant understanding of the issue of non response bias.

- **The necessity to use the administrative data**: where parents felt using administrative data was the only way the study team would be able to check the representativeness of the study sample and therefore found it more acceptable.

However, when exploring the processes of checking representativeness in greater detail, parents' views evolved and became more nuanced. This was primarily due to concerns about this use of data for those who opted out of their data being transferred to the study team. It should be noted that it was not explained to parents that the most value to the study would be to use this data for those who did not take part, and that this is central to the aim of improving representativeness.

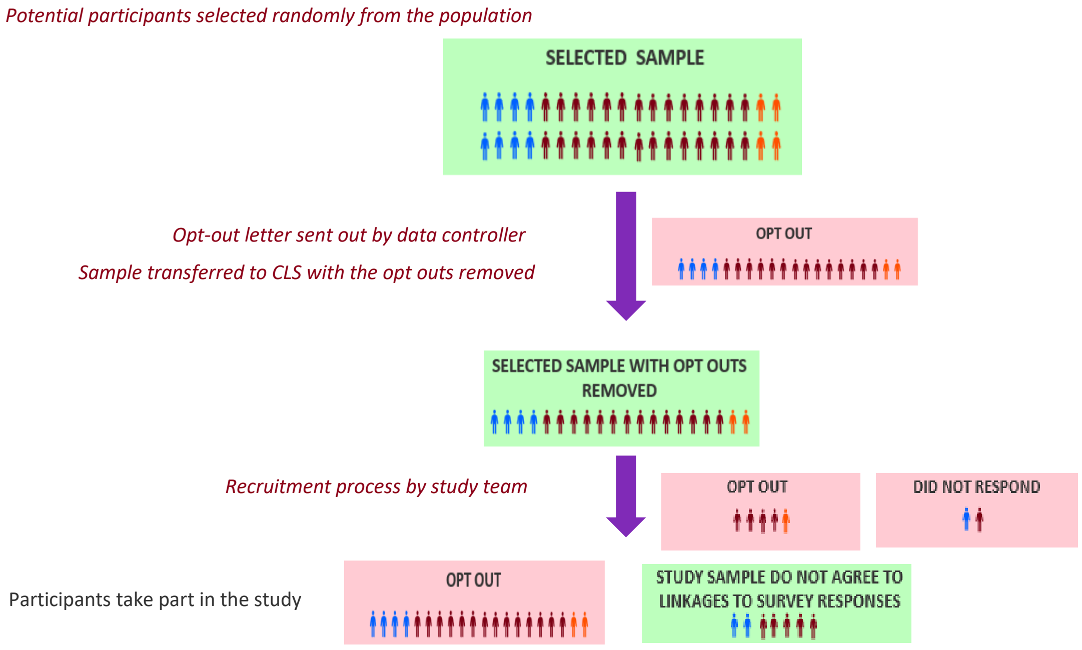
6.2 Views about use of different groups' data to check representativeness

To give parents a more detailed understanding of using the de-identified data of the selected and study sample to check representativeness, parents were shown stimulus materials that identified the different groups that made up the selected sample and the study sample (each group are outlined below in Figure 14), they included:

- The overall selected sample
- Those in the selected sample who had opted-out when receiving the letter from data controller – opt-outs are those who have actively let the data holder know they do not want administrative data about them passed to the study team
- Those in the selected sample who opted-out when receiving an invite letter from the study team – opt-outs are those who have actively let the study team know they don't want to be part of the study
- Those in the selected sample who were invited by the study team, but did not respond to their invite letter or later recruitment attempts – these are study participants who have not opted-in or opted-out of the research after receiving an invitation letter by the study team
- Study sample, which is those who have taken part in the study, including those who have and have not agreed to administrative data linkages to survey responses

Within the workshop the acceptability to using each groups' de-identified data to check the representativeness of the study sample was discussed in turn (and in the order presented in the list above). Parents' views towards acceptability of use changed over the course of discussions.

Figure 14 Checking representativeness of study groups



6.2.1 Selected Sample

Parents initially discussed views about the use of the selected sample's de-identified administrative data to assess the representativeness of the sample. There was initially widespread acceptance of the study team using the de-identified administrative data of the selected sample to check the study sample's representativeness because of the **perceived necessity of it to improve the value of the research**. The issue of representation resonated strongly with parents. They emphasised the importance of the study team building an inclusive sample to reflect an *'accurate picture'* of British society today and described that representation is vital to producing valid research findings.

> *"If there is not enough representation from a particular group of people then the results would be biased towards the majority… like white British people"* (Wave 2, Scotland).

The **necessity of this proposed use** was another key driver as parents recognised that the study team accessing the de-identified administrative data of the selected sample was the most effective and accurate way of checking the representativeness of the study altogether. Throughout the dialogue administrative data use was more acceptable when it was deemed essential to the study achieving its objectives.

> *"[Without a representative sample] you will not be able to dig out the issues that are inherent in society...you won't then be able to help to inform the solutions".* (Wave 2, South and East England,)

However, during these discussions about the use of the selected samples' de-identified data, parents became aware, or were informed by moderators or stakeholders, that the selected sample would include those who had opted-out of the study and those who had not responded to communications about the study. A handful of parents raised concerns at this initial point about this proposed use, particularly among the Engaged Sceptic and Entrenched Cynics who were more mindful **of participant's control of their personal data** throughout the dialogue. It is notable however, that at this initial point even these individuals tended to accept the proposed use of selected sample administrative data on balance for this purpose. (However, see section 5.2.2, 5.2.3 and 5.2.4 below for how parents views changed over the course of the workshops as they discussed the use of de-identified data for different groups of people in more detail).

> *"Maybe one of my concerns is that you're looking at linked data of a group who haven't consented…is that ethical?"* (Wave 2, Scotland).

At this stage in the dialogue, the variation in views for each of these groups suggests there is a **sliding scale of acceptability**, which is dependent on whether the person had opted-out and how and when in the study process this has occurred. Here, the deliberations of parents highlighted the recurring tension between the perceived social value of the research and concerns relating to control of personal administrative data - with those who placed more weight on the value of the research more willing to accept the study team's use of administrative data

6.2.2 Selected sample who opted-out from the data transfer from the data controller to study team

Next parents discussed in detail the use of de-identified data of those who, when contacted by the data controller, opted-outed of the transfer of administrative data from the data controller to the study team. There was a mixed response as to whether parents thought it was acceptable to use the administrative data of those who had opted-out of their data being passed to the study team from the data controller. However, the most common responses reflected caution because these participants had actively opted-out of sharing their identifiable administrative data with the study team which suggests that they might also have concerns about their de-identified data being shared with the study team. As a result, despite the understood value of this group's de-identified administrative data to the study **their opt-out was prioritised by parents,** who broadly felt that the study team should not use the de-identified administrative data of this group for checking representativeness. To use the administrative data of these study participants did not seem ethical or fair as the study team would be aware that they had opted-out of their identifiable data being passed to the study. The most unequivocal responses were voiced by the Entrenched Cynics and Engaged Sceptics who placed more value respecting the choice of the individual.

> *"If people have opted-out, whether identifiable or de-identified data it should not be used…because people have opted-out, that means they don't want their data to be used in any way…It would be unwise to go against their will."* (Wave 2, Northern Ireland).

Conversely, a minority of Amenable Accepters saw the loss of personal control over administrative data for this purpose as a worthwhile sacrifice for the good of the study. Throughout the dialogue, these parents were more trusting of the study team than the other types, believing in both the security of data in the study and that their interests aligned closely with wider society.

> *"The study is more important than upsetting someone who's not going to know...for the good of the study, what they don't know won't hurt them."* (Wave 2, North England and the Midlands).

While it was explained to parents that it may still be legal for the study team to use the de-identified administrative data of those who have opted-out of the transfer of administrative data to the study team for this purpose, and while parents understood the value of this to the study and for the research, most parents still felt it was unacceptable for the study to use de-identified administrative data of this group for non-response analysis. They saw the relationship between the study participants and study team as a “partnership” built **on trust and transparency** and were uncomfortable when it appeared to them that the study team were undermining this or were ‘changing the rules’ when it suited them.

> *"It ultimately comes back to transparency and ethics. You need to be upfront from the start about how you're going to use the data. Otherwise, you could really [annoy[20]] people if they found out another way later down the line."* (North England and the Midlands, Wave 2)

6.2.3 Selected sample who opted-out from data controller and study team invite letter

Parents broadly understood that there are two stages at which study participants could opt-out of the study. Most groups recognised that study participants could opt-out when sent the initial letter by the data controller, effectively refusing immediately to take part and for their data to be transferred to the study team (as discussed in section 5.2.2), or opt-out at a later stage when first contacted by letter from the study team. The difference between opting out at either stage was generally seen as unimportant to parents in discussions on this use of administrative data. This is because at both points study participants had actively opted-out from being involved with the study. As a result, the use of either of these group’s data for this purpose was not felt to be appropriate as it did **not respect their choice** to not be involved in the research.

> *"If I had chosen to opt-out that's me fully out…I'm fully uninvolved in any shape or form."* (Wave 2, Scotland)

6.2.4 Selected sample who were invited by the study team but did not respond to their invite letter or later recruitment attempts

There were more mixed views about the acceptability of using the data of those who did not respond to the study team’s invite letter and were not contacted in subsequent recruitment attempts. Some parents saw the use of these study participants’ administrative data as much more acceptable than those who had actively opted-out of the study. However, other parents viewed not responding to the study and opting-out as the same because in neither circumstance have these people actively agreed to participate in the study and therefore felt administrative data should not be used for this purpose.

For parents who accepted the use of this group’s administrative data, they argued the study team were not bound by ethical constraints preventing its use given these study participants had not actively opted-out of being in the study. This was particularly the case among Amenable Accepters and Engaged Enthusiasts. Further, some parents thought there were many reasons why a study participant may not have responded to recruitment attempts, for example missing an email through busyness. Therefore, they thought it would be a shame to exclude these study participants from this proposed use, which potentially may affect the quality of the research findings. In this case, parents understood that the study had worked to fulfil the principles of transparency and choice. Unlike with the administrative data of opted-out study participants, there was less of a sense that the study team was using information they did not have the right to.

> *"If you're not actively opting out you probably don’t care that much...I don’t see why not responding would be too much of a problem."* (Wave 2, Scotland)

Some parents, including some Entrenched Cynics and Engaged Sceptics, disagreed and felt non-responders’ data should be treated no differently to those who had opted-out of the study. In contrast, these parents placed higher value on **individual choice and personal control** over administrative data. They

---

[20] Expletive removed from quotation

believed that the study team should be actively given permission by study participants before accessing their administrative data.

> *“If they didn’t respond I think it’s safe to say they didn’t want to be a part of it.”* (Wave 2, South and East England, Wave 2).

Finally, some other parents, including Engaged Sceptics, viewed the data of non-responders as distinct from opt-outs but were less certain on how their de-identified administrative data should be used. These Engaged Sceptics understood why the study team would want to analyse this data and emphasised the importance of analysing representativeness but were unsure whether this was fully ethical. Ultimately, these parents could not come to a definite conclusion on acceptability of the use of this group’s administrative data for this purpose.

> *"If someone has dropped out, but they've told you they don't want you using any of their information anymore, then you shouldn't. But if they've not informed you and you know nothing about why they've [not responded], that's a different story. It's a bit of a grey area."* (Wave 2, North England and the Midlands).

6.2.5 Study Sample

Finally, parents discussed using the de-identified data of the study sample. There was unanimous acceptance across the groups of the study team using the de-identified data of the study sample to check for representativeness because parents recognised this group had actively agreed to take part in the study, unlike the opt-out and non-response groups. Parents were happy with the study team accessing the de-identified data of the study sample as this formed part of the research they had agreed to take part in. As a result, they felt using this groups de-identified administrative data **did not impact study participants choice and control** as the study team would be acting in line with study participants wishes to participate in the study.

Parents also reiterated **the necessity of the use of de-identified administrative data** to evaluate the representativeness of the study, which parents saw as important to the social value of the study. In addition, they recognised that without access to the study sample’s data, assessing representativeness would be altogether impossible and therefore access to this group’s data was deemed a necessity. Views were consistent across types as for those in the study sample parents did not need to weigh the study’s value against study participants loss of control.

> *"How else would you be able to tell us that ethnic minorities and low-income families don’t respond...otherwise you can’t account for them."* (Wave 2, Scotland)
>
> *“No research is worth anything if your research group is not representative. It will be weighted towards one community over another"* (Wave 2, North England and the Midlands).

Parents were further asked whether there was any difference in the acceptability of this use of de-identified administrative data for those who had and had not agreed to their survey responses being linked to administrative data (as discussed in the previous chapter) and overall there was felt to be little difference. Parents generally did not see study participants’ refusal to survey linkages as a barrier to acceptability of the use of de-identified administrative data for this purpose. They felt that despite study participants not agreeing to this substantive use, it did not imply they would be unhappy with their de-identified administrative data being used for this purpose.

> *"Yes, because they have agreed to take part in the study…it's just the linkage that they're not happy with."* (Wave 2, Scotland)
>
> *"As long as it is not being used for what they said no to and just used to check the representativeness then yeah, that is fine."* (Wave 2, North England and the Midlands)

There were several Engaged Sceptics who voiced concerns around **transparency**. For instance, a parent from North England and the Midlands warned the study team should always **be explicit with study participants** about how their data would be used. However, so long as this use was clear they saw it as ultimately acceptable.

6.3 Views on the data controller conducting non-response analysis on behalf of the study team

In some groups in the second wave, parents discussed the acceptability of data controllers conducting non-response analysis on behalf of the study team. This approach would mean that de-identified administrative data does not need to be passed to the study team for this task (as described in section 5.2), and could provide an alternative way of analysing the representativeness of the sample. This approach would involve data controllers comparing the study sample's de-identified data to the selected sample de-identified data.

Due to time constraints, there was limited coverage of the potential role of data controllers in analysing sample representativeness. The topic was only covered by two of the dialogue groups in the second wave (Scotland, Group 3 and Wales, Group 2), therefore these findings are indicative due to the small sample size. Some parents were receptive to data being analysed by data controllers for this purpose, especially if this was completed by **trustworthy organisations**, such as NHS Digital. We note it was not clear whether they were referring to the sample frame data due to lack of in-depth discussion on this topic.

Some parents found this approach acceptable as the de-identified administrative data of people who opt-out is not passed onto the study team. However, the participants' control of personal data remained a barrier to acceptability for a small number of parents as participants would not be asked if they are happy with data controllers using administrative data about them for this purpose (to assess the representativeness of the study sample for the study team).

> *"I think it's important to ask people, even if it's kept in house"* (Wave 2, Scotland)

While the lack of coverage limits the possibility of drawing any firm conclusions, initial discussions with the Scottish group suggest that participants may be comfortable with administrative data being used by trusted data controllers to check sample representativeness.

6.4 Using weighting to improve representativeness

In the first wave of the dialogue, parents were asked to discuss potential approaches the study team could use to improve the sample's representativeness, including weighting. Weighting is a process where researchers adjust the research data to correct bias and make it more accurate. Administrative data about the selected sample and the population would be used to guide this process. Although it was not explicitly explained to parents that this would use de-identified administrative data from the sampling frame, this discussion took place in the context of other uses of sampling frame data to improve representativeness and inclusivity of the study.

Parents initially struggled to understand the detail of the process of weighting within the first wave workshops, which resulted in parents questioning the robustness of the process. A common response of parents was to question the robustness of data that had undergone weighting, reflecting that statistical processes such as weighting were widely unfamiliar and abstract for the parents and that weighting was seen to amount to the study team manipulating the research findings. For this reason, several parents voiced a preference for boosting the sample with individuals from harder to reach groups, believing that this would negate needing to weight the data after the fact.

> *"That's just using statistics to get an answer, to get to the answer they want. Statistics can give you any answer you want if you use enough of them."* (Wave 1, Northern Ireland).
>
> *"Could the accuracy of what you are trying to be achieve be guaranteed if you are going to weight? That's my initial thought."* (Wave 1, South and East England).
>
> *"The other option, to recruit more people than you typically would, well if that gets you to the percentage…I think that sounds much more reliable for the data overall."* (Wave 1, Northern Ireland).

However, following explanations from stakeholders that provided further clarity, parents were generally reassured and felt this use of de-identified administrative data was acceptable. In this instance, the drivers of acceptability of trust in the study team, that weighting was a necessary process, and normalisation that weighting was a standard procedure were key.

> "*It is acceptable but it's about how it is explained and how it is done. People have to understand*." (Wave 1, North England and the Midlands).

7. Conclusions and recommendations

This chapter provides overall conclusions and general recommendations on the use of administrative data in the ELC-FS. It starts with key overarching themes that arose across the dialogue. and suggests some potential areas for future work arising from this dialogue research. It concludes with recommended principles to help support the acceptability of the proposed uses of administrative data in ELC-FS.

7.1 Key overarching themes

Across the dialogue several key overarching themes emerged, which are reflective of parents' understanding of data and how administrative data will be used in the study.

The first of these was a **tension between the research needs versus participant needs.** Across the dialogue, the proposed uses of administrative data were often highlighted and understood to be **beneficial for the study** as it would enable the study team to access more data and make the research process more efficient. Parents understood these uses would promote inclusivity and help to improve the scientific quality of the research. They were also accepting of the uses of administrative data where they could see they were necessary to complete the study, such as using maternity and birth records to access sampling information. Parents recognised findings from the ELC-FS would also benefit society, though they felt they were less likely to benefit them on an individual level. This suggests there is greater scope to emphasise how the use of administrative data can reduce the burden on the participant, but also how the benefits, such as the findings from birth cohort studies, are directly applicable to and can positively impact participants.

A further theme which emerged was the **need to build trust between participants and the study team.** Across the dialogue, parents frequently cited the importance of the study team building trust with the public. This was essential for parents to buy-in to the value of the study and to believe the security measures and protocols around handling administrative data would be adhered to. Parents wanted to feel like they were in a **'partnership'** with the study team – which reflects their sense that **data about them belongs to them**. Being in a partnership means consulting with study participants on the use of their administrative data not only to comply with the legal requirements but to go beyond this. Proposed uses where the study team may be able to access and use administrative data without notifying parents, using administrative data where participants have opted-out, or approaches which were seen to be intentionally opaque were viewed as suspicious with parents highlighting just because the study can use their data, does not mean they always should. Should these proposed uses be carried out, it was felt they are likely to erode parents' trust in the study.

7.2 Potential areas for further work

A **communications testing phase** for the initial opt-out letter from the data controllers and the recruitment letter from the study team (and any other recruitment materials that may be used) could further enhance the delivery of the ELC-FS. Throughout the dialogue parents did not generally shift from their initial views on participating in a birth cohort study, which highlights the importance of these initial communications. Parents did have concerns around the practicalities of participating in this research and how burdensome and invasive it would be Full information around this was not provided as part of the dialogue, as it will be in the actual study, and stakeholders were able to provide more information on these aspects during the dialogue which mitigated parents' concerns. However, the dialogue highlighted how important it is that initial communications anticipate parents' questions and concerns. These communications also need to be pitched appropriately to not deter too many participants from the outset, which would limit the initial sample and the representativeness and inclusivity of the study, whilst still being clear and transparent enough to accurately convey what the study will involve Further, parents had different types of information needs, with **Engaged Enthusiasts** and **Engaged Sceptics** wanting the most information. A communications testing phase would help the study team to develop participant information materials that are sufficient for those who want more information, whilst not being too burdensome for study participants who prefer less. This testing phase would also help to develop communications which use appropriate terminology and strike a balance between using specific and accurate terminology which means the study has been explicit and exhaustive in a legal sense and making the language accessible.

A further area that could also be explored is the development of participant-facing videos as part of the ELC-FS recruitment materials. The videos used as part of the dialogue, such as the ADRUK and the Next Steps videos, significantly improved parents’ comprehension and engagement with data linkage. Other videos could also be developed explaining the other proposed uses of administrative data for the ELC-FS to further enhance participants’ comprehension of these.

7.3 Key principles for increasing acceptability of use of administrative data in the ELC-FS:

From the discussions with parents as part of the dialogue, the following key drivers of acceptability have emerged and can used to inform decisions to help ensure that use of administrative data is acceptable for potential study participants.

- **Be transparent: inform and explain.** Transparency around the use of their administrative data was a key driver of acceptability for parents. Across all the proposed uses of administrative data, parents wanted to know who was using it, why it is being used, the value of the use, and how will it impact on them and their child, and that it is not being used for commercial gain. Ensuring participant communications answer these questions would help to increase the acceptability of the use of administrative data.
- **Paint a picture for participants** of the social value of the study, including inclusivity – for society and individuals. Where parents could see the value of the ELC-FS and the impact it could deliver, this tended to inform how acceptable they found the particular uses of administrative data were. Parents tended to think the ELC-FS was valuable either through:
    - **An individual lens:** If they had children who had a health condition or disability, they could see how potential findings could benefit their family. Conversely, if these parents felt participation in the birth cohort study posed too much of a cost to the individual, such as being too time consuming, encountering potential harm from sharing data or the incentive was not high enough, this diminished their buy-in to the study.
    - **The Social good:** Some parents were motivated by the potential findings from the ELC-FS being beneficial to the whole of society more widely, and in particular, findings will be used to influence policy to improve services for the whole of society. These parents prioritised the inclusivity of the study as it would ensure findings will be applicable to the general public.

Using these principles in combination should increase the value participants see in the study.

- **Give control by giving choice** about whether and how data is used, as parents see themselves as the owners of their data and want to be treated as partners when it is used. This does not necessarily mean giving them choice for every use, but as a minimum at least offering a mechanism through which they are informed about the use of administrative data and a way to opt-out of this use taking place.

- **Set the scene** of research norms. Explaining when and how data is already used routinely in research can help to reassure participants that this is not a new approach and that established practices and processes are in place to help reassure them about the potential for data loss or misuse to occur.
- **Reassure about data security.** Provide information about robust processes and procedures that are in place to protect data as this increases the acceptability of the proposed uses of administrative data. Being specific about the security measures in place can be particularly reassuring to those groups who are more concerned about data security and start from a point of being less trusting of public bodies.

Drawing on these principles can help to make the use of administrative data in the ELC-FS more acceptable to a broad range of parents across the UK.