KANTAR PUBLIC

Centre for Longitudinal Studies

4032240 Early Life Cohort Feasibility Study – Stakeholder Research Report

June 2022

Amy Busby, Bethany Dokal, Lucy Williams

For internal use only

Contents

Executive Summary

Background to the research

Funded by the Economic and Social Research Council (ESRC) and being 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.

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 of the study 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.[1]

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.

---

[1] https://cls.ucl.ac.uk/cls-studies/early-life-cohort-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.

Kantar Public was commission in August 2021 by CLS to undertake both the stakeholder interviews **(data controllers and data users)** and the public dialogue. This report outlines the findings from the stakeholder interviews to the proposed uses of administrative data in the ELC-FS. The stakeholder phase consisted of 14 60-minute depth interviews with 15 stakeholders. 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.

The findings from the public dialogue are published in a separate report.

Research Objectives

The objectives of the stakeholder interviews were to explore the **attitudes of administrative data controllers and users** to the following proposed uses of administrative data in the ELC-FS and their acceptability. The specific research objectives included:

- **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.

Methodology

In total, **15 stakeholders** from across the four nations took part in one-hour online depth interviews, which included 9 data controllers and 6 data users. Interviews took place between 2nd September to 22nd October 2021.

The interviews were conducted remotely either by telephone or Zoom.

Main findings by topic

Stakeholders widely agreed that now is a good time for a new UK birth cohort study and were excited by the prospect and thought it would provide invaluable insight into the lives of a new generation.

The use of administrative data for sampling purposes

Overall, stakeholders thought the **sampling approach proposed of using linked birth registration data and NHS maternity records seemed feasible**, although having different approaches for each of the four nations was seen as **cumbersome**. The approach was also seen as beneficial as it would allow for **almost**

complete sample frame and this would help to improve the representativeness of the sample. However, stakeholders also thought the proposed sampling approach had limitations: some **groups are likely to be missing**, potentially there are **missing health characteristics,** and **not all information** which the study team would like to access may be **available**. Further practical concerns were raised which included the **burden providing this information may place on busy government data controllers** and challenges accessing this data due to **delays caused by the Covid-19** pandemic

Stakeholders thought either the **one step or two step recruitment** options proposed were possible.[2] Although, due to the reputational risk for and potential ethical concerns associated with the one step approach, thought it was unlikely data controllers would agree to this approach.

The use of administrative data for operational purposes

Stakeholders reported that four proposed uses of administrative data from the sampling information for operational purposes were fairly standard and in line with what the expected.

These included using administrative data to **assess the representativeness of the sample**, which stakeholders again thought was important to ensure the representativeness of the sample. Although stakeholders did question whether the inclusion of the data of those who had opted out was ethical and recommended this was undertaken in a secure environment.

Stakeholders thought the **use of administrative data to improve the representation of particular groups including those from minority ethnic groups** was necessary to help improve the inclusivity of the sample. Although, they flagged even this approach may still be missing those who are homeless or living in insecure accommodation.

Stakeholders saw the use of administrative data for **targeted recruitment approaches to maximize engagement and inclusivity of under-represented groups as sensible** as it would further help to improve the inclusivity of the research and the likelihood that smaller groups are included. However, stakeholders did warn this would need to be administered in a sensitive manner to avoid participants feeling stigmatised from the communications (e.g. teenage mothers).

Stakeholders generally said the **use of administrative data to help with retention** in the long term of the study would help to combat participant attrition but should be undertaken to locate participants after alternative communication channels are explored first.

The use of de-identified administrative data for substantive and analytical purposes

Overall, stakeholders thought **data linkage was an acceptable use of administrative data as it uses de-identified data** and has been completed on other birth cohort studies. However, stakeholders did raise several practical and ethical concerns, **namely the length of time needed to collect the consent**s and that in some of the **nations administrative data may be 'patchy'** and the extent to which it is possible to collect **'informed consent'** on this complex topic.

Stakeholders agreed **linkages to the child are necessary** in a birth cohort study as you are following participants across their life course, although any data linkage comes with potential risks, including the risk of re-identification and potential data leaks if not undertaken in a secure environment. Stakeholders were aware of the **value of linking to data on the parents** and to understand the outcomes of the birth cohort, researchers often need to look at the generation before. Stakeholders also thought linkages to other family members would provide further **analytical benefits**. However, stakeholders recommended the study team need to demonstrate how they are factoring in that **family relations may be changing over time**.

---

[2] Recruitment approaches tested with stakeholders:

The one step approach: Parents details are passed directly to the study team from the organisation that had collected and held their data. The study team contacts the parents via post and invites 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.

The two step approach: Contact from data holder - the organisation that holds the parents details would first contact them via post, giving them the opportunity to opt out of being invited to the study before their details are passed to the study team. Contact from the study - the details of parents who do not opt out are then passed to the study team. The team then contacts the 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 can then decide if they want to take part in the research or not.

The use of administrative data to understand non-response bias

Stakeholders noted **using identifiable administrative data to find participants who are no longer actively participating in the study** was a good approach to combat survey attrition and is a standard approach in longitudinal research, although queried how long this consent would be valid for and noted the validity of the consent gathered is contingent on the consent wording used.

Stakeholders generally thought **analysing de-identified administrative data for those who do not take part** was also acceptable**, although noted that the proposed blending of different sets of data** as linkage between multiple sets could make **participants more identifiable**. However, this they felt could be mitigated through functional separation in the systems used or through using an independent researcher to carry out the analysis who is not part of the study team.

Key principles emerging from the stakeholder interviews

On the whole, stakeholders thought the proposed uses of administrative data in the ELC-FS were acceptable as long as the study team could attempt to demonstrate and adhere to the following principles.

- **Communicate the value of the research:** where participants know what is being done and why, they tend to be more accepting.
- **Collect informed consent:** to collect this the study team need to explain to participants in plain English what they intend to do with the data.
- **Re-establish consent at regular intervals**: Regardless of whether an implicit or explicit consent model is used for data linkage, the type of consent needs to be collected every few years to reflect that participants can change their minds and in birth cohort studies children also come of age to make their own decisions about how their data is used.
- **Ensure the necessary agreements and legal frameworks are in place with data controllers:** this was seen as essential and would help to ensure that the study team are handling and storing administrative data securely.

1. Background and objectives

1.1 Introduction and Background

Funded by the Economic and Social Research Council (ESRC) and being 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 of the study 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

---

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

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.

Kantar Public was commissioned in August 2021 by CLS to undertake both the stakeholder interviews **with key data controllers and data users** and the public dialogue. This report outlines the findings from the stakeholder interviews to the proposed uses of administrative data in the ELC-FS.

The stakeholder phase consisted of 14 60-minute depth interviews with 15 stakeholders. 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.

The interviews with the stakeholders and the public dialogue were undertaken concurrently, which limited the extent findings from the stakeholder interviews influenced the development of workshop materials for the public dialogue. The findings from the public dialogue are published in a separate report.

1.2 Research objectives

The objectives of the stakeholder interviews were to explore the attitudes of data controllers and users to the following proposed uses of 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.

1.3 Methodology

In total, **15 stakeholders** from across the four nations took part in one-hour online depth interviews, which included 9 data controllers and 6 data users. Interviews took place between 2nd September to 22nd October 2021.

Recruitment of the stakeholders drew on CLS' existing networks and these fell into the following broad categories across academia, 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). Stakeholders interviewed were not members of the ELC-FS study team. Data controllers from all four nations of the UK were invited to participate (with participants from three of the nations taking part) to ensure views were collected on the proposed sampling approaches for the ELC-FS across the nations.

The interviews were around one hour long and conducted remotely either by telephone or Zoom. No incentives were provided. The topics covered in the interview are show in the figure below and the full topic guide used in the interviews can be found in the Appendix B.

Figure 1 Topics covered in Stakeholder interviews

Topic Area
Introduction to their role
Views on using admin data for social research
Perceptions of the ELC-FS
Views about the ELC-FS proposed sampling approach
Views about proposed uses of admin data for operational purposes
Views about proposed data linkages and potential consent models
Using admin data to understand non-response bias
Reflections and Close

1.4 Reading this chapter

In order to ensure stakeholders are not identifiable, the chapter does not break down findings by stakeholder type but does include references to differences by nation.

Verbatim quotes are used throughout the report to illuminate and bring to life key findings and are attributed as follows: “Quote.” (Nation, Stakeholder).

2. Main findings by topic

This chapter outlines the findings by topic which includes views about the use of administrative data in social research, and linking together administrative data to survey responses, and stakeholder views about the proposed sampling, operational, and substantive uses of administrative data.

Overall stakeholders generally thought the proposed uses of administrative data were feasible and acceptable, although some concerns around the ethical and practical use of the administrative data were flagged.

Key Findings:

- *Stakeholders were widely excited by the prospect of a new birth cohort study for the UK as they thought it would provide invaluable insight into the lives of a new generation.*
- *The proposed uses of administrative data for sampling, operational, and substantive research purposes were as expected for many stakeholders, especially for data controllers who tended to already have some familiarity with the ELC-FS due to earlier discussions with the study team, and were generally acceptable.*
- *Concerns and queries were mostly raised around how these uses will take place in practice, which will be contingent on agreements with data controllers being in place. Concerns and queries typically revolved around how the study team will ensure the use of data is ethically robust and, through a practical lens, how the study will account for limitations in the administrative data where it is 'patchy' or missing key groups or variables.*

2.1 Views about the use of administrative data in social research

There was some variation in how familiar stakeholders were with the use of administrative data in social and health research across the nations. Stakeholders from England, Wales and Scotland, where the use of administrative data was said to be more common and more established, had greater familiarity with processes and legislation in place to facilitate research access. These stakeholders were more confident the study team would be able to achieve the sample frame planned. Although they had queries about exactly what the arrangement between data controller and the study team would look like, and in particular, at what point data would be transferred between them and the study team. They noted the exact arrangement would be determined by the data sharing agreements between the data controller and the study team.

In comparison, stakeholders from Northern Ireland had less familiarity with the sharing of administrative data for research purposes and the practice was seen as less established. In turn this made stakeholders from this country more cautious of potential security risks or misinterpreting legislation on the topic. In addition, stakeholders thought researchers made fewer administrative data requests in their nation as they were less aware that they could or were uncertain of how to go about this process. Stakeholders from this country were not sure data controllers would necessarily have the (current) infrastructure and capacity to deal with too many of these requests from researchers.

The use of administrative data in social research was widely seen by stakeholders as important and useful in health and social research as it offers a number of significant benefits to researchers and to the general public. These include making the research **process more efficient** as researchers can access data sets on education, health and social care on a large number of people without individually contacting each one. This means important insights can be gathered from pre-existing data both efficiently and economically too. This is especially useful when researchers cannot fit all of the questions, they would like into survey data collection or have research questions which cannot be addressed using survey data.

Secondly, the use of administrative data can **reduce the burden on the general public** to 'actively' participate in health and social research as data subjects do not need to be contacted to collect the necessary information. The value of this approach was further highlighted by one stakeholder who cited the example of Finland, where researchers must first determine whether they can answer research questions through the use of pre-existing data before asking participants, which also saves public money spent on survey research.

> "*It diminishes the burden particularly on the participants in terms of data collection and it increases the scope of research you can look forward in ways that participants cannot look forward, in 10-20 years' time or time or look back to things they may have not remembered.*" (England, Stakeholder)

Stakeholders further noted that linkage has enabled researchers to answer research questions through offering **alternative approaches to research**. For example, in cancer research where patient experience of care was linked to routinely collected health care data, this allowed researchers to look at variation by treatment and variation by outcome.

The use of administrative data can also provide **insight into non-response bias** where cohorts are linked to administrative data researchers can see where gaps are; e.g. understudied populations.

Lastly, the use of administrative data in social research was seen **to enable researchers to understand prevalence and how generalisable their findings are**. This is because they are able to carry out whole population analysis using de-identified records of the population, with wide coverage.

However, stakeholders also raised some concerns around the use of administrative data in health and social research. Notably this included that **administrative data can be limited when used in isolation** as it does not provide rich understandings of behavioural, attitudinal, and social drivers, which help to explain associations researchers find. Administrative data should be used in tandem with other types of primary data collection to fully explore these.

Additionally, there are **data protection risks** (including data can be leaked or participants details could be incorrectly assigned to another), which means how this data is managed and stored is extremely important. However, stakeholders agreed that these risks can be mitigated through adherence to standard and widely used procedures to ensure data is handled securely. Mitigation strategies include research accreditation panels which scrutinise applications for use of data and adherence to the 5 safes model.[4]

The last concern raised around the use of administrative data in social research was **whether the value of the research findings justified access to an individual's personal records**, and that in some cases it might not. To access this data, stakeholders felt researchers need to demonstrate its value sufficiently, which includes explaining why the research is being conducted and how it is valuable to the general public or to the groups affected by the research.

2.2 Views about linking administrative data to survey responses

A further topic which was explored with stakeholders was the linking of administrative data to survey responses. This is usually undertaken through collecting participant consent to the linkages. Overall, stakeholders were widely familiar with this approach and thought it provided a number of key benefits. This included **further enhancing researchers' understanding of people's experiences or opinions** through supplementing self-reported responses in the survey, with administrative data.

> *"Admin[istrative] data is large N and small P and survey data is small N and large P. N the number of cases and P being the number of attributes you have about each case." Survey data less cases but you know much more about each one."* (England, Stakeholder)

Linkage can also provide additional value to longitudinal studies through **offering interpolation of data between waves**. For example, in longitudinal research where change over time is measured. However, due to costs researchers may survey people relatively infrequently. Through linking administrative data to the survey data, it provides a way of **adding in data between survey waves**.

Additionally, linking together administrative data with survey responses can make the data more **comprehensive and inclusive** through **continuing to link to administrative data based on consent, after people have dropped out/lost touch with the study**. For example, administrative data can be used to fill in the gaps for survey participants where they may have dropped out or lost touch with the study team. For example, adding in the educational attainment of participants even if they are no longer participating in the surveys. This helps the research to be **more inclusive as certain groups are more likely to drop out/ lose contac**t with the study team over time.

---

[4] More information on the 5 safes model can be found at: https://ukdataservice.ac.uk/help/secure-lab/what-is-the-five-safes-framework/

Lastly, administrative data can be used to **verify self-reported information**. This is because as surveys rely on participants' recall, information can be inaccurate. In these cases, administrative data can be used to provide objective information such as how many times a person has visited the GP in the last 12 months or information about less socially desirable behaviours or whether someone has a criminal record or had previous interactions with social services.

However, similar to stakeholder views on the use of administrative data in social and health research, stakeholders raised some concerns around linking together administrative data and survey responses. Namely, this included **ethical concerns** around the practice. This is because some stakeholders felt researchers could not **collect fully informed consent** from participants through a survey framework to complete these linkages. Some were concerned about how valid these consents are as you cannot be sure of participant comprehension of the linkages and the collection of them may put too much burden on the participant. Moreover, some stakeholders queried whether a tick box at the end of a survey asking if participants are happy to link with their health records is sufficient for researchers to then access confidential patient information. We note that most surveys which collect consent take steps to ensure it is informed and more than a tick box.

> "*You need to communicate benefits to participants, so they become the advocates. It stops it from being a siloed industry that their research is being used for but not having any direct input from them, they are creators, and they are not given the product of it. So, need to be verified as a proper stakeholder and involved in the process*." (Northern Ireland, Stakeholder)

> "*We absolutely need to take the public with us. We've embedded public engagement into everything we do... if you don't actually explain to people how the data is going to be accessed and used, and what those people think about it you've lost the social contract to use it.... Within a study, it's a conversation with study participants about what other data might be linked to that and are they ok with that. I know there are legal routes where you don't have to ask for consent but personally, I think we have to come up with constructive ways of asking the question and explaining to people the benefits upfront*." (England, Stakeholder)

Some stakeholders also raised the issue that administrative **data has gaps in its coverage of the general population** in its sample. This is because it systematically misses some important marginalised groups, including those who are homeless or those who are from the travelling community.

2.2.1 Views about the ELC – FS

Stakeholder perceptions of the ELC-FS were also collected. There was variation amongst stakeholders of how familiar they were with the ELC-FS, with data controllers who were already in discussions with CLS around accessing administrative data tending to be the more familiar, whilst data users were less so.[5]

Stakeholders widely agreed that now is a good time for a new UK birth cohort study and were excited by the prospect as they thought it would provide invaluable insight into the lives of a new generation. This was because there has not been one since 2000 with the Millennium Cohort Study and they felt that there were **significant differences in childhood now for those born in the 2020s,** caused by the impacts of new technologies and the Covid-19 pandemic, which has had a significant impact on young families both financially and socially. A new birth cohort study would offer an opportunity to access self-reported data on these differences and their impact on family life experiences.

A further aspect which stakeholders found exciting about the prospect of a new birth cohort study was birth cohort samples are typically larger for particular age groups than other longitudinal research panels, which means **researchers have the opportunity to report findings more precisely.** As each 'sweep' relates

---

[5] For those who were less familiar with the ELC-FS, they were given the following information during the interview Funded by the ESRC and being led by the 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 will recruit several thousand new babies, collecting information about their families and their development. The feasibility study will be evaluated in early 2023, and if judged to be successful, commissioning of a new, larger main study is anticipated in 2023. The Early Life Cohort Feasibility Study will collect rich data on a new generation of babies born across the UK between September – November 2021, capturing information about their economic and social environments, and their health, wellbeing and development during their first year of life. Recruitment will take place either at 6 months or at 9 months. The study aims to paint a nationally representative picture of the circumstances and lives of a new cohort of babies born at a critical time in the UK's history. The primary scientific aim of the study 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."

specifically to a stage of development, unlike other longitudinal studies where participants may be asked the same questions every wave.

> "*There are insights into people's life course which you uniquely get from the birth cohort studies compared to more general panel studies like Understanding Society. You have got a much bigger sample of people in a particular age cohort so you can say things much more precisely about them and also because you are only surveying a single cohort of people. In each sweep you have a questionnaire absolutely tailored towards that stage of development. Whereas with Understanding Society the same questions are asked of everybody*." (England, Stakeholder)

Stakeholders also thought the potential greater use of administrative data in the ELC-FS, compared with previous birth cohort studies, could be highly beneficial, as it can provide the opportunity **to gather more detailed data when combined with survey responses which is useful for analytical purposes.**

Accessing the administrative data was seen as a **valuable way to combat participant attrition** too. For example, identifiable administrative data could be used to locate participants if the study team lose track of them over time, as well as deidentified administrative data can be used to fill in outcomes for participants where they had dropped out or lost touch with the study team for research purposes.

It was also felt that there are **now new ways to collect data** that could also be utilised in the ELC-FS. For example, devices that capture eye tracking or measures cognitive function, which can be used in researching child development. Although these should be used without causing too much burden or being too invasive to participants.

Stakeholders also **commented on the use of a feasibility stage** prior to the main study and felt this stage was not always undertaken in research but would help to deliver a more successful and robust main study. Meanwhile stakeholders from Northern Ireland were **pleased with the inclusion of Northern Ireland** which was not always included as studies are often limited to Great Britain or England

2.3 Views about the proposed sampling approach for the ELC-FS

Stakeholders were introduced to the sampling frame plans for how the study team would like to access routinely linked birth and maternity records to achieve a representative and inclusive sample frame. Due to differing bodies in each nation holding the birth and maternity records, a different approach by nation is proposed as part of the ELC-FS and was explored with stakeholders.

Overall, stakeholders thought the sampling approach proposed seemed feasible. Although stakeholders thought having different approaches for each of the four nations is necessary but cumbersome.

Stakeholders **from England** thought the approach of the ELC-FS seemed feasible but was contingent on collecting necessary approvals from the Office for National Statistics (ONS) and/or NHS Digital to do so. From a practical lens, for birth registration information in the UK there is a legal requirement for parents to register the birth within 42 days and this data takes around two months to be processed. This could pose a potential delay to the study team receiving this data but as the study is looking to recruit infants aged between six to nine months this should have minimal impact. Data users were concerned this approach did not provide participants the opportunity to opt out of their details being shared with a third party, which seemed less ethical. However, directly sharing identifiable details with the study team without first undertaking an opt out was less of a concern for other stakeholders who noted ONS, and NHS Digital, were likely to use a third-party organisation to send out initial communications to participants anyhow, which already involves third-party sharing.

In **Scotland**, stakeholders thought the proposed process in which "*Public Health Scotland (PHS) send their maternity records data to National Records Scotland (NRS) to link with the birth registration data held by NRS. The NRS would prepare the sample for ELC-FS and check for infant and maternal deaths. The NHS Central Register (NHSCR) team would flag the sample on their system and would also conduct further checks for deaths before the sample goes into field. Up to date addresses would be collected from the Community Health Index*" took a standard approach and was in line with what they expected a new birth cohort study to do.

In **Wales,** as the Secure Anonymised Information Linkage (SAIL) databank has provided similar sampling information to other studies in the past, stakeholders thought this approach was feasible. Although one stakeholder did raise that SAIL does not have any personal identifiers (e.g. name, address) only the attribute data.[6] Therefore, they can find people with specific characteristics and demographics in their data bank but cannot link this information to who they are. In the case of ELC FS, the study team will also have to contact NHS Digital to reidentify these individuals.

Of all the nations, stakeholders were the least certain of the feasibility of the proposed sampling approach within **Northern Ireland**. Within the interviews, this was outlined as "*the ELC-FS will apply to the Northern Ireland Business Services Organisation (BSO) to link National Health Application and Infrastructure Services (NHAIS) data for the feasibility study. and hope this will pave the way to use birth registrations held by Northern Ireland Statistics and Research Agency (NISRA) for the main study."*

Some stakeholders were less familiar with research that has taken a similar approach and thought the approach would require a lot of discussion between the data controllers. Stakeholders were less certain of the legal pathways needed for the study team to access this information. One stakeholder commented that the Northern Ireland Statistics and Research Agency (NISRA) does not have the same legal coverage under the Statistics and Research Act and would need Information Commissioner's Office (ICO) guidance to clarify. Additionally, the study team would rely on a legal basis to support General Data Protection Regulation (GDPR) public task, but in this case NISRA does not have a legal basis to use the public task. Therefore, they would need to use GDPR legitimate interest, which would be new territory, and this could lead to other researchers requesting birth registration information for research.

However, other researchers were familiar with similar approaches adopted in Northern Ireland, for example for the Northern Ireland Cohort Longitudinal Study of Ageing (NICOLA) study, where the researchers approached the Honest Broker Service and pulled addresses without named personal information and wrote directly to addresses to invite participants to the study.

Despite concerns around the feasibility of this approach, stakeholders thought the proposed sampling approach would be beneficial to the study as it would allow for a **complete as possible sample frame** to be accessed in a **timely** manner. Stakeholders felt this would help to ensure the representativeness of the sample and would further enable the study team to 'boost' for certain groups to **support inclusivity in recruitment** as they will know who has not signed up to the study so they can construct statistics about their characteristics.

However, stakeholders did raise **some limitations** which undermined the completeness of the sample frame. This included **certain groups are more likely to be missing from the sampling data** such as those from the Gypsy Traveller community, illegal immigrants, as well as those who are homeless as they are less likely to be registered with a GP, and generally interact with public services less frequently. This reduces **the completeness and accuracy of the sample frames**, which ultimately limits the potential representativeness of the study and how inclusive the sample is. Ultimately, this will **limit how generalisable the findings are.** Stakeholders also felt that as these groups are marginalised in society, this increases the importance of understanding the outcomes and experiences of these groups, which could potentially lead to policy action to address less favourable outcomes for these groups.

Additionally, some stakeholders mentioned that there are **potentially missing data on health characteristics at birth** which may be useful to ensure these groups are in the study sample. For example, in England, the birth notifications and registrations data include weight and gestational age but miss conditions and important events in post-natal care for the child, and antecedent risks and pregnancy complications affecting the mother. This further detail is important as it **provides information on the health of the child and measures of health characteristics at birth** which are **highly predictive of long-term health, life morbidity, and education outcome predictors into adolescence**. These are important factors that services can respond to from a public health perspective.

A further issue is that **not all nations systematically collect all information that the study team would like to use for sampling**. For example, in Northern Ireland, it was raised that ethnicity is not systematically collected on birth and maternity records, which will limit the study team's ability to sample on ethnicity in this country. Possible alternatives included sampling of country of birth for the mother, but this would miss those who are born in Northern Ireland and are from a Minority Ethnic group. Stakeholders also raised ethical concerns, including uncertainties on the consent issues and what would be required by the data controllers. The study team will need be aware of the data controllers' sensitivities to new mothers and pregnant woman

---

[6] At the time of the stakeholder interviews, CLS were intending to use the SAIL data bank for sampling. This approach has changed to using ONS and/or NHS Digital, the same approach proposed for England. Therefore, in this chapter we report on stakeholder responses to using the SAIL databank, although this is no longer the approach the study intends to use.

being contacted and provide reassurances around this, for example making sure not to contact parents who have experienced the loss of a child.

Stakeholders also raised some **practical concerns with the sampling approach,** which included **the potential burden providing this information may put on the busy government data controllers**. Some stakeholders noted asking health agencies to assist in this process when busy, when the research is not their priority and expertise, is a challenge. Stakeholders also raised there may be **delays in the study team receiving this information**. This is due to a **backlog in birth registration services due to Covid-19**. Stakeholders were unsure on the extent of backlog and how long it will take to clear.

2.3.1 Views towards recruitment approaches

Within the interviews two possible recruitment approaches were tested: the one-step and the two step recruitment approaches, these are shown in the Figure below.[7]

Figure 2 Recruitment approaches explored with stakeholders

<table>
<tr><th>One step process</th><th>Two step process</th></tr>
<tr><td>Participants details would be passed from original data controller directly to ELC-FS team and then participants would be contacted directly by the data collection agency doing the survey</td><td>A postal opt out would first be conducted by the data controller before the details were shared with the ELC-FS and data collection agency</td></tr>
<tr><td colspan="2">The information that would be shared would be participants’ details and those of their baby (name, sex, date of birth, address etc), as well as other information on the sampling frame about the participants and baby i.e. parent’s country of birth, gestational age of baby at birth, ethnic group of baby and/or parents, age of parents, occupation of parents.</td></tr>
</table>

In principle, as long as a legal basis and data sharing agreement were in place, stakeholders generally thought either approach was viable. However, both data controllers and data holders widely thought it was unlikely data controllers would agree to the one-step approach as it posed considerable reputational risk for data controllers and raises potential ethical concerns.

One step approach: This approach was seen to be preferable for the study team as it enables them to access the most complete sample, which has a scientific value to the study. Some stakeholders felt that to be reassured by the one step process they would need to know how the study team intended to hold this information and whether the attribution information about the household would be kept with the identifiers. If these were kept separately following the data separation principle, stakeholders would have fewer concerns. A similar process had been undertaken for NHS national audits on cancer, so there is a precedent for this approach.

However, some data controllers and data holders commonly raised concerns about the one step approach and felt it carried greater reputational risk for data controllers and had the potential to alarm the general public. This is because it did not provide participants the opportunity to opt out of their details being shared with a third party before the initial opt out letter was sent, which seemed less ethical. However, this was less of a concern for other stakeholders who noted ONS, and NHS Digital, were likely to use a third party organisation to send out initial communications to participants and therefore, participant information was likely to be moved to a third party without participants aware regardless of approach. One stakeholder gave the example of the NICOLA study where although the study team were provided with addresses, they were

---

[7] Recruitment approaches tested with stakeholders:

The one step approach: Parents details are passed directly to the study team from the organisation that had collected and held their data. The study team contacts the parents via post and invites 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.

The two step approach: Contact from data holder - the organisation that holds the parents details would first contact them via post, giving them the opportunity to opt out of being invited to the study before their details are passed to the study team. Contact from the study - the details of parents who do not opt out are then passed to the study team. The team then contacts the 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 can then decide if they want to take part in the research or not.

not given any other personal identifiers, thereby limiting the level of personal data the study team had access to without the participants' consent. One stakeholder thought this concern could be further mitigated through the **public dialogue, and if the study team could demonstrate public support for this approach, this may mitigate data controllers' concerns.** From a practical perspective, stakeholders commented in England and Wales the study team would require section 251 permission, data sharing agreements and appropriate legal basis (GDPR) to do so.

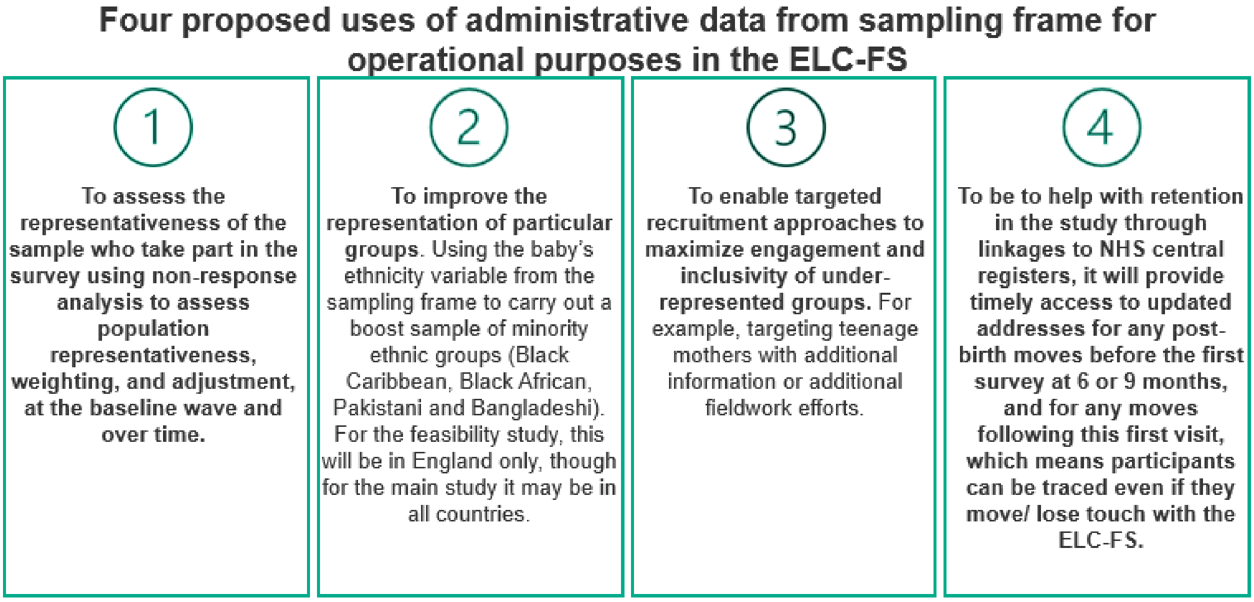
Two step approach: Overall, stakeholders widely had a preference for the two step approach. This was because this approach lessens the potential concern from the general public over their data being passed to a third party without their prior knowledge. This was despite stakeholders acknowledging at some point a third party would likely be involved **to send out the initial opt-out** before data could be moved from the data controller to the study team. Some stakeholders thought it would also be beneficial for the study team to have initial communications sent from NHS trusts the general public are familiar with, as parents would have just gone through these trusts' maternity services so would be more familiar and (potentially) trusting of them.

Conversely, a minority of stakeholders also had ethical concerns around the two step approach. These stakeholders queried whether because you have given the public the option to opt out, it does not necessarily mean they choose to opt in and felt an **opt in process at the first stage** was more **ethically robust**. Stakeholders also appreciated that, compared to the one step approach, the two step approach would narrow down the number of people the study team would receive from the data controller, thereby limiting the comprehensiveness and representativeness of the sample.

2.4 Views towards proposed uses of administrative data for operational purposes in the ELC-FS

Four uses of administrative data from sampling frame data sources to help with survey operational aspects of the ELC-FS were tested with stakeholders, these are shown in Figure.1. Stakeholders reported that these uses were standard in longitudinal research and in line with what they expected.

Figure 3 Proposed uses of administrative data from sampling frame for operational purposes.



Four proposed uses of administrative data from sampling frame for operational purposes in the ELC-FS

1. **To assess the representativeness of the sample who take part in the survey using non-response analysis to assess population representativeness, weighting, and adjustment, at the baseline wave and over time.**
2. **To improve the representation of particular groups.** Using the baby's ethnicity variable from the sampling frame to carry out a boost sample of minority ethnic groups (Black Caribbean, Black African, Pakistani and Bangladeshi). For the feasibility study, this will be in England only, though for the main study it may be in all countries.
3. **To enable targeted recruitment approaches to maximize engagement and inclusivity of under-represented groups.** For example, targeting teenage mothers with additional information or additional fieldwork efforts.
4. **To be to help with retention in the study through linkages to NHS central registers, it will provide timely access to updated addresses for any post-birth moves before the first survey at 6 or 9 months, and for any moves following this first visit, which means participants can be traced even if they move/ lose touch with the ELC-FS.**

Below stakeholders' responses to each proposed use of administrative data are outlined.

2.4.1 To assess the representativeness of the sample who take part in the survey using non-response analysis to assess population representativeness, weighting, and adjustment, at the baseline wave and over time.

Overall stakeholders thought this was an acceptable use of administrative data and given the importance of determining and attempting to **ensure the representativeness of the sample**, should be undertaken. Stakeholders noted that a key benefit of this approach is it can highlight marginalised groups that are less likely to take part in birth cohort studies. This was seen to be important so that the study team can **determine whether their findings are biased** and how **accurate** they are. Stakeholders also raised questions around how this approach would be administered and noted that data controllers would either stipulate or prefer for this use to take place in a **secure environment** to help protect the data.

Stakeholders did raise some concerns and limitations around the use of administrative data to assess the representativeness of the sample, although broadly these were not seen as too concerning. This included several stakeholders raising concerns around the use of administrative data of those who had **opted out of the study** and felt this was potentially **unethical.** However, one other stakeholder felt this was less of a concern as data controllers were unlikely to provide administrative data on an individual case basis, but rather as counts.

> "*This is all data that's collected and analysed by ONS and happens routinely, you're never looking at data on an individual / personal level. It's attribute data that you're analysing*." (England, Stakeholder)

Additionally, stakeholders raised the limits of this approach due to the a forementioned challenges of administrative data, that it is not complete and has some missing groups. These groups are likely to be those more marginalised in society. Therefore, the administrative data the study team are using to compare to their participant sample is itself not fully complete.

2.4.2 The use of administrative data to improve the representation of particular groups.

Stakeholders thought undertaking the use of administrative data to improve the representation of particular groups including those from minority ethnic groups (such as Black Caribbean, Black African, Pakistani and Bangladeshi) was important to help improve the inclusivity of the sample.

Particular benefits of the approach were noted that would enable the study team to **target groups who are less likely to take part** which is important to improve **the inclusivity of the study**. From other research studies, researchers are already aware that ethnicity, health and social disadvantage all intersect. Through ensuring these groups are included, and in large enough numbers, the study team will be able to carry out subgroup analysis and draw out meaningful findings on these groups.

However, stakeholders did voice concerns around the analytical and practical constraints of this approach. Notable practical challenges included how the study would attempt to include other groups who are likely to be low responders such as those who are **homeless** and **those living in insecure accommodation**, where this data is not in the sample frame and overall, how effective this approach is, is dependent on the accuracy of data in the records.

> *Garbage in vs garbage out!"* (Wales, Stakeholder)

Additionally, stakeholders had practical concerns for the study on the quality of data on **ethnicity in Northern Ireland**, ethnicity data is not collected in GP data and is ‘patchy’ in The Northern Ireland Maternity (NIMSAT) system, so the study team will have to devise another way to collect this information such as ‘country of birth’ of the parent instead, although this would still miss people who are from an Ethnic Minority group and were born in Northern Ireland.

Analytically, stakeholders also raised that the study team may have issues **determining the exact weights needed to take into account the boost sample**. This is because the data on Ethnic Minorites by local area is estimated and as it is inexact it is **challenging to work out the exact weights** needed. One stakeholder questioned how the **sampling for the boost sample would be completed** and thought the common approach through **purposively sampling in clusters**, which is usually undertaken for efficiency and cost reasons, may not be helpful as when people cluster this tends to be due to likeness, which doesn't help to **ensure diversity within a subgroup**.

Additionally, the use of administrative data to improve the representation of particular groups may cause analytical challenges later on with the **more deviations from a simple random sample** the **greater the sampling errors** are in overall results.

2.4.3 The use of administrative data for targeted recruitment approaches to maximize engagement and inclusivity of under-represented groups.

Overall, stakeholders thought this was an acceptable use of administrative data as it offers a further way to **improve the inclusivity of the research**. However, stakeholders did raise concerns around the delivery of this approach and thought the study team should demonstrate how they intended to recruit under-represented groups in a sensitive manner.

A key benefit of this approach was that it can **increase the likelihood smaller groups are represented** within the sample, which will improve the representativeness of the study, which makes it more likely researchers can make **meaningful statements about them too**, which includes teenage mothers and those belonging to lower socioeconomic groups.

However, stakeholders did have concerns around how this approach would be **delivered to ensure sensitivity.** In particular, stakeholders raised attempting to **recruit teenage mothers is a sensitive topic,** especially depending on their age, and especially if they are under 16. For example, even the sending of recruitment letters to the home could have potential implications for the mother, such as causing discord to the household if the mother's parents are unhappy with her having a child under the age of consent. Mothers who were **under the age of consent at delivery** may also be dealing with legal implications too. One stakeholder queried how the study team plan to mitigate for this issue and had the study team considered **how their communications may impact on potential participants, and how to limit negative impacts**.

Also, if adopting this approach, stakeholders queried how much detail the study team intend to use in their initial communications, as if the information was too detailed it may alarm potential participants, who may be uncomfortable with the study team holding too much sensitive information on them or leave them feeling stigmatised.

> *"I think the principle is, at least until you have some sort of consent, to think about how much that initial mail out carries info about that person and you need to be very careful….We know this about you and we want you to take part because you have a pre-term baby'. I think that can start to become surprising for some people.*" (England, Stakeholder)
>
> "*You need to be careful in how communicate to 'why have I been selected?"* (England, Stakeholder)

Moreover, stakeholders raised targeting specific groups is acceptable but needs to be done in a highly sensitive way as some groups **are more sensitised and anxious to receiving communications from government institutions**, which the study team need to be conscious of when creating communications. For example, certain ethnic minority groups. Part of this includes how the study team communicate to these groups **why they in particular have been selected t**o avoid them feeling stigmatised or targeted.

A further practical concern was also raised by one stakeholder who noted with this approach there is the potential to **over complicate the sample structure by targeting too many diverse subgroups for inclusion.** This means the study team may lose the flexibility to turn the study towards investigating future research questions which are not currently envisaged. This may constrain **the flexibility of the sample because they have put in these more rigid overrepresented groups.**

2.4.4 The use of administrative data to help with retention in the study through linkages to NHS central registers, providing timely access to updated addresses for any post-birth moves before the first survey at 6 or 9 months, and for any moves following this first visit.

Overall, stakeholders felt the use of administrative data to help with the retention of the study through tracing participants was seen as an acceptable. Although after the first interviewer visit this should be undertaken after other means of contact e.g. using contact details collected have first been attempted Stakeholders widely recommended the study team collect participant details which enable them to **contact participants through multiple channels**, including a secondary contact, before resorting to using administrative data.

Stakeholders were familiar with this use of administrative data and noted this approach was used on other studies and therefore, there was a **precedent** to do so. This approach was understood to be efficient and maximised data quality and without it could lead to **greater participant attrition,** which could compromise the **quality of the data**. This approach was also seen to potentially benefit the participant too, as it reduces the burden on the participants to notify the study team if they move house. One stakeholder also thought this approach helped to maintain **participant's privacy** as through verifying where participants lived routinely this limits the potential for the **study team to send communications to the wrong address and sharing participants' personal information**.

However, similar to the other proposed uses of administrative data for operational purposes, stakeholders did raise concerns and limitations of this approach. This included how **effective this approach** is dependent on the **quality of the data used to make direct contact with families**. For example, some information held by data controllers is not fully up to date. Many of the ways to trace participants rely on the participant updating their GP records, but often there is a delay when participants update their address with the GP or registering with a new GP practice, this means the data the study team are accessing may not be wholly accurate and limited its usefulness.

Stakeholders felt how **this approach is delivered is critical.** For example, stakeholders thought this potential use needs to be explained from the **outset of recruitment** such as in the opt-out letter and participants need to **consent** in order for the study team to do this. Also, there are other contact details that can be collected after the first visit from the interviewer such as **phone, email, address, contact details of other family members** (if happy to share), that can also be used to help trace participants before identifiable administrative data needs to be used to locate participants.

A further practical consideration for the study team raised by stakeholders was they will need a **data sharing agreement with the data controller to allow them to access this information.** Stakeholders had concerns around **how frequently the study team would be recollecting this consent** to carry out this use. They felt this should be an **ongoing activity**, which reflects that **giving consent is not fixed** and participants may change their mind over the course of the study.

2.5 Views towards the proposed data linkages

A further use of de-identified administrative data which was explored with stakeholders was the use of **data linkage for substantive and analytical purposes**. Stakeholders were informed the ELC-FS would like to link to administrative data to enrich the survey data collected and to enable the study team to compare outcomes among children in the cohort to the wider population. Due to the varying availability of administrative data to link by nation and differing bodies in each nation holding this data, different linkages to be completed for each nation were proposed.[8] Therefore within the interviews stakeholders were shown different stimulus accordingly, these are shown in Appendix C. Linkages were proposed that included both sampling frame and other types of data.

Overall, stakeholders widely thought data linkage was an acceptable use of administrative data as it uses de-identified data and knew there was a precedent for doing so. For example, Department for Education data was used for Next Steps study and Department for Work and Pensions (DWP) data has been used for a range of other studies. Stakeholders had concerns around this use related to challenges collecting informed consent but felt these could be mitigated through the study team collecting the necessary consents from participants to do so in a thorough and transparent manner.

Stakeholders thought the linkages provided through **HMRC data were particularly valuable** as it will enable the research team to look at **socio-economic impacts on health and potential life outcomes**, which stakeholders thought would be useful to explore in a birth cohort study.

In England, the number of data linkages proposed was seen as exciting by stakeholders as they would enable researchers to explore a **wider number of research questions**.

However, stakeholders also raised some **ethical** and **practical concerns** around the proposed data linkages. From an ethical perspective stakeholders anticipated HMRC and DWP data would likely be perceived as the most sensitive by the general public. Stakeholders queried how this data would be formatted and to what level of specificity. One suggestion to help mitigate concerns was for linkages to health data to be separated from any personal identifiers.

Stakeholders also questioned whether **it is possible to collect informed consent on this topic**. This is because of the detail and **technical aspect of the linkages** it is challenging to explain to the public, in a meaningful way, what the study is aiming to do with the data linkages to ensure comprehension.

Stakeholders further anticipated the general public may also have some concerns around some of the proposed linkages which **feel intrusive**, such as the data on self-harm. Stakeholders thought asking participants for their consent around these specific linkages will likely put off participants from agreeing to them, which means the study team will collect fewer consents.

Practical concerns raised by data controllers included that the number of linkages looked ambitious and **will take a long time for the study team to collect**. Also due to the sensitivity of some data sets; these already have specific arrangements with data controllers under which conditions they can be accessed, which could pose further practical challenges for the study team and extend the time it takes for them to gain access to key data sets. For example, DWP share data sets which contain detailed assessment on health and disability e.g., Disability Living Allowance (DLA), Personal Independence Payment (PIP), Employment and Support Allowance (ESA) with the ONS. Under the current agreement, the ONS is not permitted to share without further agreement from DWP. Therefore, to obtain these types of data sets, the study team will need

---

[8] Participants in the public dialogue were shown different stimulus to stakeholders for the proposed data linkages in the ELC-FS.

to provide these reassurances as well as collect the necessary agreements, which may make the process of collecting the data longer and more cumbersome.

In **Northern Ireland**, data linkages would require gaining permission from all five NI trusts for health data. This puts an additional burden on the study team who will need to come to arrangements with all 5 trusts. Guidance from the Northern Ireland Privacy Advisory Committee would also need to be obtained. Although, once these permissions have ben gained the data received will be near complete. Whilst in **Wales,** stakeholders raised that not all data sets will be available through SAIL. Additionally, there is not complete coverage for all the datasets suggested for Wales. Stakeholders also flagged the proposed linkages are missing primary care data **in Scotland** which helps provide rich data of service utilisation and contact with primary care to understand child development.

> *"I know we [at X] can sometimes get carried away with what we can link but there is a difference between what we can and what we should link that needs to be remembered as well."* (Northern Ireland, Stakeholder)

2.5.1 Views towards proposed data linkage to child, parents and other family members data

As part of the data linkage, the ELC-FS would like to link data for the child, parents and potentially other family members, including pre-birth records, to enrich the survey data collected and enable them to compare outcomes. Stakeholder views towards the proposed linking for each are outlined in the table below.

Data Linkage Type	Benefits of approach	Concerns/ considerations
Child	Stakeholders agreed linkages are necessary in a birth cohort study as you are following participants across the course of their lives.	Any data linkage comes with potential risks, including the risk of re-identification and potential data leaks if not undertaken in a secure research unit. Parents are likely to feel more sensitive and worried about linkages for their children.
Parent	Stakeholders were aware of the value of linking to data on the parents and that to understand the outcomes of the birth cohort, researchers often need to look at the generation before. For example, maternal health history can be used to understand child outcomes, and parental income is important for household income. This approach has been done on other studies so there is a precedent. For example, in the Millennium Cohort Study, consent was obtained from participants well in advance of data linking.	When seeking these consents this needs to be achieved in a meaningful way – which as it is a technical and complex topic can be challenging. Timing of consent is important, in past longitudinal studies often consent is asked for in different waves (often when it becomes useful to do the linking) rather than asked for all at one time. The study team need to outline when and how frequently they plan on collecting these consents.
	Provides analytical benefits and has been undertaken on other studies where sibling controls are widely used on audit trace and to adjust for unmeasured compounding in shared households. This approach can help provide insight into areas such as understanding the impact of disadvantaged health problems and involvement in crime for interventions in social care and early	As the study team are proposing long term linkage, they need to demonstrate how they are factoring in that family relations will be changing over time. For example, when the sibling is young it is acceptable for their parents to give permission on their behalf, but the study team will need to consider at what age they have the opportunity to agree to these linkages themselves.

Other family members	health care, which are architected in family not just the child. For linkages with other family members who are already adults this is easier as the study team can collect their consent directly.	It is easier to gain consent from full siblings, but it can be more challenging to link to half siblings or steps siblings, some of whom will live with the main study child and others who will not. Some of these concerns may be overcome through gaining consent from the parents from the outset of the study. Less data may be available the further back in time the study team go especially if trying to link two generations back or further, the study team will get diminishing returns.

> *"A sensitive and powerful dataset including issues on self harm."* (Northern Ireland, Stakeholder)
>
> *"If you really want to understand child health outcomes and health and wellbeing, you need to understand something about the generation before that."* (Scotland, Stakeholder)
>
> *"You should always have the participants involved in the co-creation and management and dissemination of your research - always! . . . You lose your social licence if you do things that people don't expect you to do.*" (Northern Ireland, Stakeholder)

2.5.2 Views towards implicit and explicit consent models for data linkage

What type of consent model the ELC-FS should use to carry out these data linkages to survey responses was also explored with stakeholders, which included both **implicit and explicit models**.[9] Definitions of implicit and explicit consent used in the interviews are shown in Figure 4.

Figure 4 Definitions of implicit and explicit consent

Implicit consent	Explicit consent
By implicit consent we mean that participants would agree to data linkage by signing up to the study.	Explicit consent would involve the use of consent questions in the survey and enabling participants to select which specific data linkages they do and do not consent to.

Stakeholders were familiar with both types of consent models and similar to other aspects of the ELC-FS, stakeholder views centred around the **ethical and practical considerations** of each approach.

Implicit consent

Stakeholders were aware that using implicit consent was possible in research, but this was seen as **unusual** and is only undertaken where there is a **justified reason** to do so. From a practical perspective, this approach **was seen as preferential for the study team** and for the research as it was likely to ensure linkages can be completed for more people and would be easier for researchers to manage. However, this approach was seen as **less ethical** and was unlikely to be granted by data controllers because of the potential **reputational risk** it poses.

> "*It can be necessary but needs to be properly justified*." (Wales, Stakeholder)

Explicit consent

Comparatively, explicit consent was seen as less practical as you have to gain the individual consent of participants and apply accordingly, which puts a **greater burden on the study team**. The process of

---

[9] Implicit and explicit consent models were explored in the following way to stakeholders. "By implicit consent we mean that participants would agree to data linkage by signing up to the study. Explicit consent would involve the use of consent questions in the survey and enabling participants to select which specific data linkages they do and do not consent to."

introducing participants to the potential linkages was also likely to **deter participants from agreeing to the linkages** depending on how it is framed. In particular, listing all proposed linkages can seem overwhelming to participants and that they are sharing a lot of personal data. Despite only partial consents being collected, stakeholders tended to think **explicit consent needs to be gained to ensure transparency** with participants, this was seen as the more ethical approach, and therefore, **more appealing to data controllers too as it poses less reputational risk.**

Mitigating for concerns

Regardless of at which point consent was collected to complete the data linkages, stakeholders emphasised the importance of using **plain English** to participants. This would help make the communications more **accessible** to participants, so they **understand** what the study team intends to do with the data. Although, some stakeholders felt that because of the complexity of the linkages involved it was difficult to ever collect fully informed consent.

Given the challenges of creating accessible communications on a technical process, stakeholders recommended **running a feasibility test** of how to collect consents to balance the **need for comprehension with the need to collect a sufficient number of consents** for the data linkage to take place robustly.

2.6 Views towards using administrative data to understand non-response bias

The final use of administrative data the study team wants to use which was explored with the stakeholders was using the data to understand non-response bias. This study team wishes to do this in two ways: using identifiable administrative data to find those no longer actively participating in the study and through analysing the de-identified administrative data for those who do not take part.

Using identifiable administrative data to find those no longer actively participating in the study

Stakeholders noted this was a good approach to **combat survey attrition** and is a relatively **standard approach** in longitudinal research. A benefit of this approach is the study team can **gain consent** from participants to do so from the outset of the study.

However, stakeholders did raise consideration around how this consent should be treated. For example, they thought the study team should stipulate for what **time period** it will be valid for and whether this will be communicated to the participant. Additionally, stakeholders felt **how ethical** and **legitimate** this approach is, is contingent on the **consent wording used** and needs to be transparent. This consent should also be recollected across sweeps.

> *“[Using administrative data to find participants no longer actively participating in the study] would work if the consent is there and explicit enough to say 'if you take part we continue to follow you and your child over time unless you tell us not to” (England,* Stakeholder*).*
>
> "*What is critical here [when using administrative data to find those no longer actively participating in the study] is what you told people at the outset about how their data would be used*." (Wales, Stakeholder)

Analysing de-identified administrative data for those who do not take part

Stakeholders appreciated why the study team would want to undertake this approach and had fewer concerns as it involves the use of **de-identified data**. A further benefit of this use is the study team will not necessarily need to collect consent from participants to administer it.

However, other stakeholders raised concerns around the **proposed blending of different sets of data** as linkage between multiple sets could make **participants more identifiable**. This could be mitigated through functional separation in the systems used or through using an independent researcher to carry out the analysis who is not part of the study team. For some data controllers, they would only allow researchers to access this information through their **secure research service** to access anonymised record level data to help keep participants unidentifiable.

3. Conclusion

This chapter sets out the conclusions and some recommendations on the use of administrative data in the ELC-FS.

This section starts with key overarching themes that arose across the stakeholder interviews and suggested topics to cover in the public dialogue. It concludes with recommended principles to help support the acceptability of any future proposed use of administrative data in the ELC-FS.

Key findings:

- *Topics stakeholders wanted to see covered in the public dialogue: The sampling approach in particular the construction of sampling frame, recruitment method, data linkage, views on using administrative data to understand non-response bias, participants general reservations around sharing their data and awareness of GDPR options.*
- *Demonstrating the following principles could help make the use of administrative data in the ELC-FS more acceptable to both the general public and to stakeholders.*
    - *Communicating the value of the research*
    - *Collecting fully informed consent*
    - *Re-establishing consent at regular intervals*
    - *Ensuring the necessary agreements and legal frameworks are in place with data controllers*

3.1 Key overarching themes

Across the stakeholder interviews several key overarching themes emerged from the proposed uses. In relation to the design of key components ELC-FS, stakeholders appreciated **the practical and scientific benefits from a research perspective** of using a one step approach in recruitment and using an implicit consent approach to linkages. However, in most cases, **the benefits gained for the research team and the scientific benefits through improved representativeness and inclusivity were felt not be sufficient to justify these approaches.**

Instead, stakeholders generally had a preference for the two step recruitment approach and the explicit consent approach to linkages as they felt these gave **more information to the public to be informed** and **presented more opportunities to make choices** about how their data is used. This was especially apparent for **data controllers, who were highly conscious of the reputational risks** associated with sharing identifiable administrative data for sampling purposes without an initial opt-out with researchers.

3.2 Topics to cover within the public dialogue

Across the interviews, stakeholders highlighted key proposed uses of administrative data in the ELC-FS which they thought were important to cover in the public dialogue and to get parents views on. These included public views on:

- **The sampling approach:** in particular their views around consent and at what point in the identifiable administrative data is transferred to the study team and whether an opt out is needed first. One stakeholder also felt how this was framed was particularly important and should not be the study team attempting to persuade people when testing different sampling options. Other stakeholders noted this information needs to be contextualised to ensure participants understand not just the intended use but also what has already taken place in research and the safeguards in place to protect their data.
- **Recruitment method:** whether the study team use a one or two step recruitment approach and in particular, if the one step approach is seen as sufficiently ethical by the public/parents.

- **Data linkage:** Where consent is collected to carry out this use, understanding from parents how long is this consent valid for; e.g. how many sweeps. Also, when would parents like to be asked for their consent to data linkage at the beginning of the study or later on and whether the study team need to collect consent for linkage to every type of data set.
- Parents views on using administrative data to **understand non-response bias**.
- **Using administrative data to find those no longer actively participating in the study using administrative data:** parents views to this use and in particular in incidences where the study team cannot get in touch whether it is acceptable for the study team to continue to use their data.
- Whether participants have reservations around sharing their data and whether they are aware of GDPR options to opt out of their data being shared/ used and how visible they feel this option is to them.

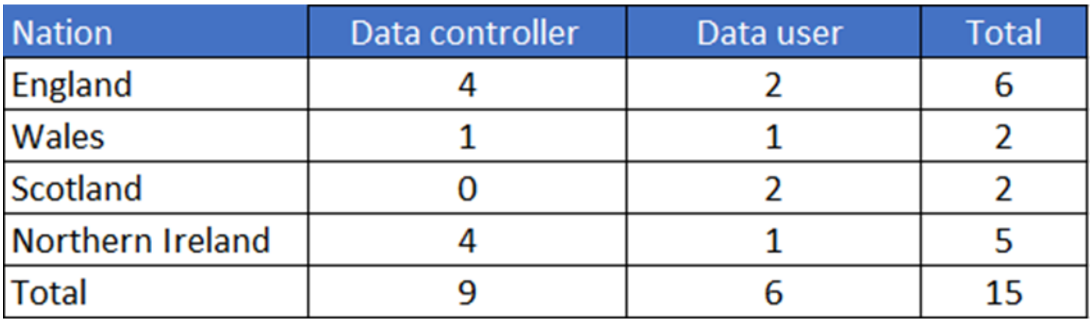
These findings helped to inform the materials that were created for the public dialogue which was conducted in autumn 2021.

3.3 Key principles to follow from the stakeholder interviews

Ultimately, most stakeholders thought the proposed uses of administrative data in the ELC-FS were acceptable, and generally thought they would be acceptable to the public too, as long as the study team could ensure and demonstrate the following principles.

- **Communicate the value of the research:** where participants know what is being done and why they tend to be more accepting.
- **Collect informed consent:** to collect this the study team need to explain to participants in plain English what they intend to do with the data.
- **Re-establish consent at regular intervals:** regardless of whether an implicit or explicit consent model is used for record linkage, this consent needs to be collected every few years to reflect that participants change their minds and in birth cohort studies children also come of age to make their own decisions about how their data is used.
- **Ensure the necessary agreements and legal frameworks are in place with data controllers:** this was seen as essential and would help to ensure that the study team are handling and storing administrative data securely.

4. Appendix A: Achieved sample



Nation	Data controller	Data user	Total
England	4	2	6
Wales	1	1	2
Scotland	0	2	2
Northern Ireland	4	1	5
Total	9	6	15

5. Appendix B: Stakeholder Topic guide

The interviewers followed a topic guide for each stakeholder interview. The topic guide ensured topics and activities were consistently covered across interviews. Topic guides are designed for interviewers and not participants. They are not a script and topics do not necessarily have to be covered in the order outlined.

40322420 Centre for Longitudinal Studies Early Life Cohort - Feasibility Study Dialogue Research

Stakeholder Interviews - Topic Guide

Background

Funded by the ESRC and being led by the Centre for Longitudinal Studies, 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 will recruit several thousand new babies in their first year of life, collecting information about their families and their development through in-home interviews, observations, assessments, and collection of biological samples. The feasibility study will be evaluated in early 2023, and if judged to be successful, commissioning of a new, larger main study is anticipated in 2023.

The Feasibility Study has multiple strands and includes a public dialogue. This will explore the views of parents of young children and administrative data controllers about some of the planned uses of administrative data in the Feasibility Study, how acceptable they find these and what, if anything, can be done to mitigate any concerns they have. The dialogue will discuss plans to use administrative data for sampling, recruitment, and retention as well as planned data linkages and consent models for this.

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

- Firstly, it will be used to sample mothers and fathers, including OHFs. Ideally this will involve linking birth registrations and NHS maternity records to sample participants and build a more comprehensive sample frame. Securing access to a high-quality sample-frame will be crucial to the feasibility of the study.

- 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 nonresponse analysis and adjustment (e.g. weighting).

- 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. A range of consents will be required, and maximising consent rates will be key to the success of the study.

The Centre for Longitudinal Studies (CLS) need input from **data controllers and data users and parents of young children** into the proposed uses of admin data in the ELC-FS specifically, the sampling, operational and substantive uses. These interviews will explore the views of data controllers and data users about the plans.

Research Aims

The objectives of these stakeholder interviews are to explore the attitudes of **data controllers and users** to the following proposed uses of admin 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 admin 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, parents and other family members and potentially pre-birth records (including potentially health, education, HMRC, DWP, LA and social services data)
    b. Proposed consent models (implicit/ explicit, opt in/ opt out models) – and proposed consent wording
    c. Substantive use of sampling data and linked data to counter non response bias, including:
        i. Using data to ‘follow’ study participants via admin linkages even if ELC-FS lose track of them.
        ii. For those that are invited but choose not to take part in the study, the ELC - FS would like to carry out research using their anonymised admin data.

The findings from this research will inform the design of the ELC-FS, information provided to participants, consent models, and linkages sought as well as CLS’ negotiations with data holders.

Stimulus list

Guide, stimulus and recorder

Key contacts

Bethany Dokal (Bethany.dokal@kantar.com)
Sam Bond (Samantha.bond@kantar.com)

Amy Busby (Amy.busby@kantar.com)

Please note, this guide is not a script and is intended to be used flexibly, with the flow of conversation likely to be shaped by the participant's role (e.g. data controller or data user or both), their experience and the nation they work in.

1. Introduction (5 minutes)

Introduce research, reassure about confidentiality, and set tone of discussion

- **Warm up and introduction**
  - Introduce moderator and Kantar Public
    - RESEARCH ASSOCIATES– introduce yourselves as 'working **on behalf of** Kantar Public' (not 'from Kantar')
  - Funded by the ESRC, and being led by the Centre for Longitudinal Studies, 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 will recruit several thousand new babies, collecting information about their families and their development through in-home interviews, observations, assessments, and collection of biological samples. The feasibility study will be evaluated in early 2023, and if judged to be successful, commissioning of a new, larger main study for the UK is anticipated in 2023.
  - The study will be conducted through face-to-face in-home interviews with both parents, this will include a 70 minute interview with the primary informant, usually the mother, and an additional informant, who is usually the father.
  - Kantar Public is carrying out public dialogue research on behalf of the CLS, to explore the views of parents with young children and data controllers and data users, about the planned uses of administrative data in the ELC-FS. This interview is one of 15 stakeholder interviews which are part of the dialogue research.
  - Aim of the discussion is to explore their views about the proposed uses of administrative data and data linkages for sampling, operational purposes and substantive purposes in the ELC-FS. In particular, we want to understand what concerns they have and what challenges there might be for the study, as well as how they would like to see any concerns mitigated if possible.
  - The insights produced will be used to inform the design of the feasibility study.
  - Interview length – 60 minutes
  - Research is voluntary – can stop participating any time
  - Research is confidential– we will not use your name or your organisation's name in the reporting.
  - Thank you leaflet – check if received beforehand (in appointment confirmation email), else email after interview
    - Kantar's privacy policy can be accessed on our website: https://www.kantar.com/uki/surveys
  - Any questions?

- **Recording**

Ask participant for permission to record, then start recording and confirm consent

- **Participant introduction**

  - Their role and day-to-day responsibilities
  - Career history

- **Data handling**
  - Types of administrative data they work with
  - Experience of sharing/ using admin data in research

2. Views on using admin data for social research (5 minutes)

This section establishes participant's view on using admin data for social research purposes.

- **Explore participant thoughts on using admin data <u>itself</u> for social research**
  - Familiarity with using admin data for social research
  - How important / valuable using admin data is for social research
  - Benefits of using admin data
    - Spontaneous
    - Probe for quality, coverage, time, cost, inclusivity
  - Concerns/ risks of using admin data particularly linked to survey data and for longitudinal research
    - Spontaneous
    - Probe for any legal and ethical concerns
    - Probe for any concerns around consent rates/consent process and linkage rates

- **Explore participant thoughts on using admin data <u>to link to survey data</u> for social research**
  - Familiarity with using admin data linked to survey data for social research
    - Probe for familiarity for use in longitudinal research
  - How important / valuable is using admin data linked to survey data for social research
    - Probe for importance/ value in longitudinal research
  - Benefits of using admin data linked to survey data
    - Spontaneous
    - Probe for quality, coverage, time, cost, inclusivity
  - Concerns/ risks of using admin data linked to survey data
    - Spontaneous
    - Probe for any legal and ethical concerns
    - Probe for any concerns around consent rates/consent process and linkage rates
    - Whether any concerns are greater/ different for longitudinal research

3. Perceptions of the ELC-FS (5 minutes)

To establish participant's awareness, understanding and perceptions of the ELC-FS

Moderator to first check participant's awareness of and relationship with the ELC-FS.

- **Explore participant awareness and understanding of the ELC-FS**
  - Sources of awareness
  - Relationship with the Feasibility Study
  - How important they think it is

If needed, explain:

"Funded by the ESRC and being led by the 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 will recruit several thousand new babies, collecting information about their families and their development. The feasibility study will be evaluated in early 2023, and if judged to be successful, commissioning of a new, larger main study is anticipated in 2023. The Early Life Cohort Feasibility Study will collect rich data on a new generation of babies born across the UK between September – November 2021, capturing information about their economic and social environments, and their health, wellbeing and development during their first year of life. Recruitment will take place either at 6 months or at 9 months.

The study aims to paint a nationally representative picture of the circumstances and lives of a new cohort of babies born at a critical time in the UK's history.

The primary scientific aim of the study 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."

- **Explore how important they think a new birth cohort study for the UK is**
    - What value it will provide
    - Any challenges it might face

4. Views about the ELC-FS proposed sampling approach (5/10 minutes)

Test ELC-FS proposed approach to creating sample frame in their nation

Moderator to introduce sampling approach for specific nation. Moderator to explore this section in more depth with sampling frame data controllers (10 minutes) and less depth with other participants (5 minutes).

Moderator to explain:

"[FOR SAMPLING FRAME DATA CONTROLLERS]: although you may have already spoken to CLS about their proposed sampling approach, today we would like to discuss it in a bit more detail to understand any concerns you have in more depth and what information you would want to know from the public dialogue.*"*

" In order for the ELC-FS to build a representative and inclusive sample frame, the study team want to link birth and maternity records. Due to differing approaches in each nation, a different approach is needed to access this information in each nation."

Moderator to read out sampling approach for specific nation.

- *England:* "In England, the ELC-FS 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. The ELCL-FS– FS will also apply to NHS Digital for up-to-date addresses for England.”

- *Wales: “*In Wales the ELC-FS 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. They will also apply to Digital Health and Care Wales (DHCW) for up-to-date addresses*.*

- *Northern Ireland: “*In NI, ELC-FS will apply to the NI Business Services Organisation (BSO) to link National Health Application and Infrastructure Services (NHAIS) data for the feasibility study. and hope this will pave the way to use birth registrations held by Northern Ireland Statistics and Research Agency (NISRA) for the main study .*”*

- *Scotland:* “In Scotland, ELC-FS are intending that Public Health Scotland (PHS) send their maternity records data to National Records Scotland (NRS) to link with the birth registration data held by NRS. The NRS would prepare the sample for ELC-FS and check for infant and maternal deaths. The NHS Central Register (NHSCR) team would flag the sample on their system and would also conduct further checks for deaths before the sample goes into field. Up to date addresses would be collected from the Community Health Index”.

- **Explore participant’s views about the proposed sampling approach for their nation**
  - Spontaneous response
  - [For sampling frame data controllers] Views from their discussions with project team
  - Benefits of this approach

- **Explore potential concerns/ risks of approach**
  - Their concerns/ risks of this approach
  - What they think parents’ concerns might be
  - [For sampling frame data controllers] Probe for any issues relating to the age at recruitment e.g. how long does it take after birth for babies to appear in the administrative data records
  - [For sampling frame data controllers] Probe for any issues around accuracy and completeness of information held on sampling frame, and its suitability for this proposed use
  - Ways they suggest to mitigate risks/ concerns, if possible

- **Explore content for public dialogue workshops**
  - What they feel should be covered in the public dialogue workshops
  - What they would like to know from the public about this proposed approach

Moderator to explain possible one stage or two-stage recruitment process to participant. Moderator to read out:

“If the sample frame is secured, then either a one or two step recruitment process could be used. In either approach, it would be made clear to participants that taking part is voluntary.

- In a one step process, participants details would be passed from original data controller directly to ELC-FS team and then participants would be contacted directly by the data collection agency doing the survey.

- In a two step process, a postal opt out would first be conducted by the data controller before the details were shared with the ELC-FS and data collection agency.
- The information that would be shared would be participants' details and those of their baby (name, sex, date of birth, address etc), as well as other information on the sampling frame about the participants and baby i.e. parent's country of birth, gestational age of baby at birth, ethnic group of baby and/or parents, age of parents, occupation of parents."

- **Explore views of one or two step recruitment process**
  - Spontaneous response
  - Benefits of one-step approach
    - probe for timing, cost, inclusivity
  - Concerns/ potential risk for one-step approach
    - Legal and ethical concerns
  - Benefits of two-step approach
    - probe for timing, cost, inclusivity
  - Concerns/ potential risk of two-step approach
    - Legal and ethical concerns
  - Preference for approach and why
  - Precedents that they are aware of for different approaches

5. Views about proposed uses of admin data for operational purposes (10 minutes)

Exploring participant views about proposed approaches for using admin data for operational purposes including to improve the representativeness of the sample, to enable targeted recruitment, and to help with participant retention.

Moderator to read out:
"In the next section we want to explore your views about some proposed uses of admin data from the sampling frame data sources to help with survey operational aspects of the ELC-FS."

a) **"The first use of admin data would be to assess the representativeness of the sample who take part in the survey using non-response analysis to assess population representativeness, weighting, and adjustment, at the baseline wave and over time");**
b) **"The second use of admin data would be to improve the representation of particular groups. For the feasibility study, using the baby's ethnicity variable from the sampling frame to carry out a boost sample of minority ethnic groups (Black Caribbean, Black African, Pakistani and Bangladeshi). For the feasibility study, this will be in England only, though for the main study it may be in all countries. For the main study, boost samples may be carried out with other groups e.g. teenage parents, twins, parents with pre-term baby".**
c) **"The third use of admin data would be to enable targeted recruitment approaches to maximize engagement and inclusivity of under-represented groups. For example, targeting teenage mothers with additional information or additional fieldwork efforts."**

[if necessary/time: This could also include higher incentives given to ethnic minority families (based on admin data variable) and low-income families (based on local area).

d) ***"*****The fourth way admin data would be used would be to help with retention in the study through linkages to NHS central registers, it will provide timely access to updated addresses for any post-birth moves before the first survey at 6 or 9 months, and for any moves following this first visit, which means participants can be traced even if they move/ lose touch with the ELC-FS.*****"***

- **For each proposed use of admin data, explore:**
  - Spontaneous response
  - Perceived benefits of this use
  - Their concerns/ risks of this use
  - What they think parents' concerns might be
  - Ways to mitigate risks/ concerns, if possible
  - Whether they are other approaches researchers can/ should use to improve inclusivity instead or as well as these

6. Views about proposed data linkages and potential consent models (15 minutes)

Exploring participant views on proposed approaches for using admin data for substantive and analytical purposes in the ELC-FS.

Moderator to explain:

*"*We now want to explore your views on ways the ELC-FS would like to use data linkage for substantive and analytical purposes. The ELC-FS would like to link data for the child, parents and potentially other family members, including pre-birth records, to enrich the survey data collected and enable them to compare outcomes (e.g. health and education) among children in the cohort to the wider population. Due to differing bodies in each nation holding this data a different approach is needed to access this information in each nation.*]*

Please note these are proposed linkages for the first survey and they are still under development so may change"

Moderator to introduce STIM 1 on proposed data linkages for each nation. Allow participant to read for a minute.

- **Explore views on proposed data linkages**
  - Spontaneous views
  - Whether any are surprising - and why
  - Which types of data linkages are most / least valuable to the study and why

- **Explore views around proposed data linkages to the child**
  - Whether they have concerns about the proposed linkages for the child and why,
  - Which linkage would be most beneficial
  - Which linkage would be the least beneficial
  - Any concerns around linking to pre-birth records,

  - Benefits of linking to pre-birth records
  - Whether they think the linkages will be concerning for parents and why
  - Whether they think there are any challenges ELC-FS will encounter accessing these data linkages
  - Ways to mitigate these challenges, if possible

- **Explore views around carrying out data linkages for parents**
  - Whether they have concerns about the proposed linkages for parents and why
  - Which linkage would be the most beneficial
  - Which linkage would be the least beneficial
  - Any concerns around linking to pre-birth records for parents
  - Benefits of linking to pre-birth records
  - Whether they think the linkages will be concerning for parents and why
  - Whether they think there are any challenges ELC-FS will encounter accessing these data linkages
  - Ways to mitigate these challenges, if possible

- **Explore any concerns around linking data for other family members (e.g. siblings)**
  - Whether they have concerns about the proposed linkages for other family members and why
  - Any concerns around linking to pre-birth records
  - Whether they think the linkages will be concerning for parents and why

Moderator to explain:

"ECL-FS are also exploring implicit versus explicit consent models for the proposed data linkages.,

- By implicit consent we mean that participants would agree to data linkage by signing up to the study.
- Explicit consent would involve the use of consent questions in the survey and enabling participants to select which specific data linkages they do and do not consent to*".*

- **Explore views on implicit and explicit consent models**
  - Spontaneous views
  - Benefits of each approach
  - Concerns/ potential risks of each approach
  - Ways to mitigate risks/ concerns, if possible
  - Establish their preference and why
  - Experience/knowledge of other studies which have used different consent models
    - What can be learned from these
    - Probe specifically for use of implicit consent.

7. Using admin data to understand non-response bias (5 minutes)

To explore stakeholder views on using sampling frame data for those not participating in the study to learn more about these people / groups in order to understand non response bias.

Moderator to read out:

"The ELC-FS would like to use sampling frame data to understand nonresponse bias in two ways*."*

Moderator to read out:

“The first way is for those who have participated and given their consent, which continues indefinitely until they actively withdraw their consent. This would enable ELC-FS to ‘follow’ them or their child via admin linkages even if ELC-FS lose track of them.”

  - Spontaneous views
  - Benefits of each approach
  - Concerns/ potential risks of each approach
  - Ways to mitigate risks/ concerns, if possible
  - Establish their preference and why

Moderator to read out:

*“*The second way is that for those who don’t take part, and/or for the whole cohort selected, the ELC - FS are considering the possibility of carrying out research using their anonymised admin data. *We note that plans in this area are developing and are not yet finalised and CLS would particularly welcome stakeholder views and suggestions on this as well as other suggestions regarding how administrative data could be used alongside the study”*

- **Explore views on use of sampling data for those not participating in the study**
  - Spontaneous views
  - Benefits of each approach
  - Concerns/ potential risks of each approach
  - Ways to mitigate risks/ concerns, if possible
  - Establish their preference and why

8. Reflections and Close (5 minutes)

Collect final thoughts on interview and bring interview to a close

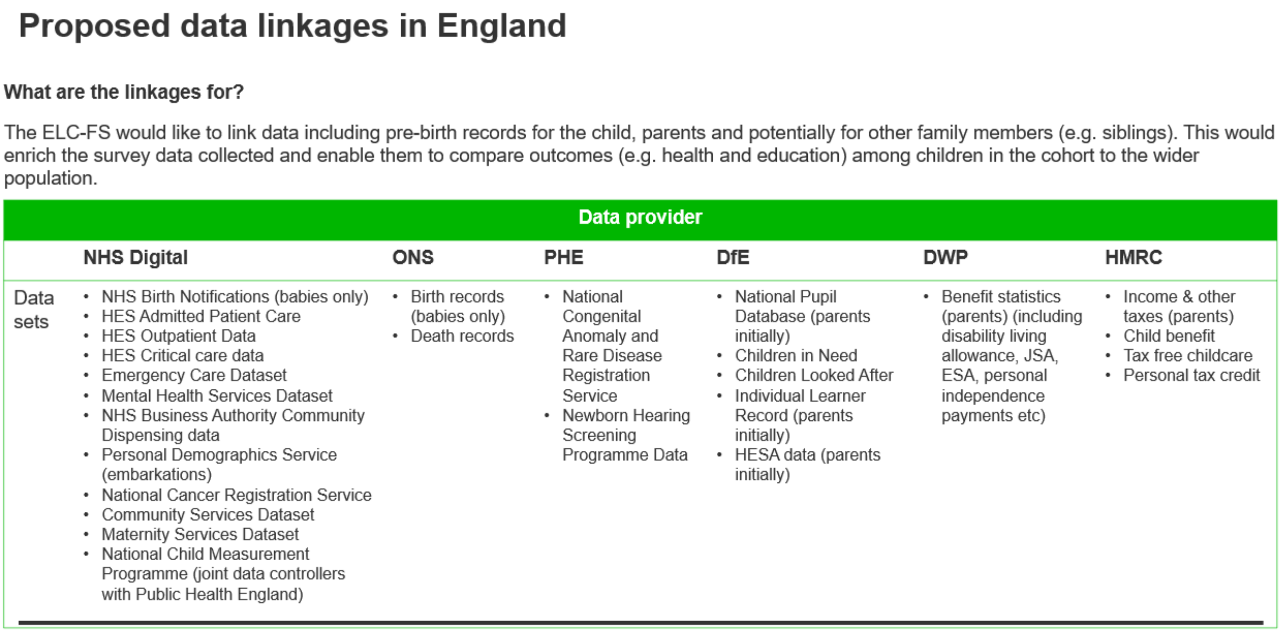
- **Collect final thoughts on interview**
  - After hearing about some of the proposed uses of admin data for the ELC-FS -what are their overall thoughts?
    - How important / valuable are these proposed uses
    - What are their key overall concerns
    - Any key mitigations they want to see in place overall
- **Thank and close**

Post-field work admin:

- Upload audio recording to secure project folder, labelled as follows: *40322252_CLS ELC-FS # [respondent number] AND TYPE RESEARCHER INITIALS.*
- Complete interview summary
- Complete analysis chart by 20/09/2021 and save in secure project folder

6. Appendix C: Stimulus on data linkages by nation

Stimulus materials were used in the interviews to present information on topics discussed. Stakeholders were shown a different slide depending on which nation they were recruited from.

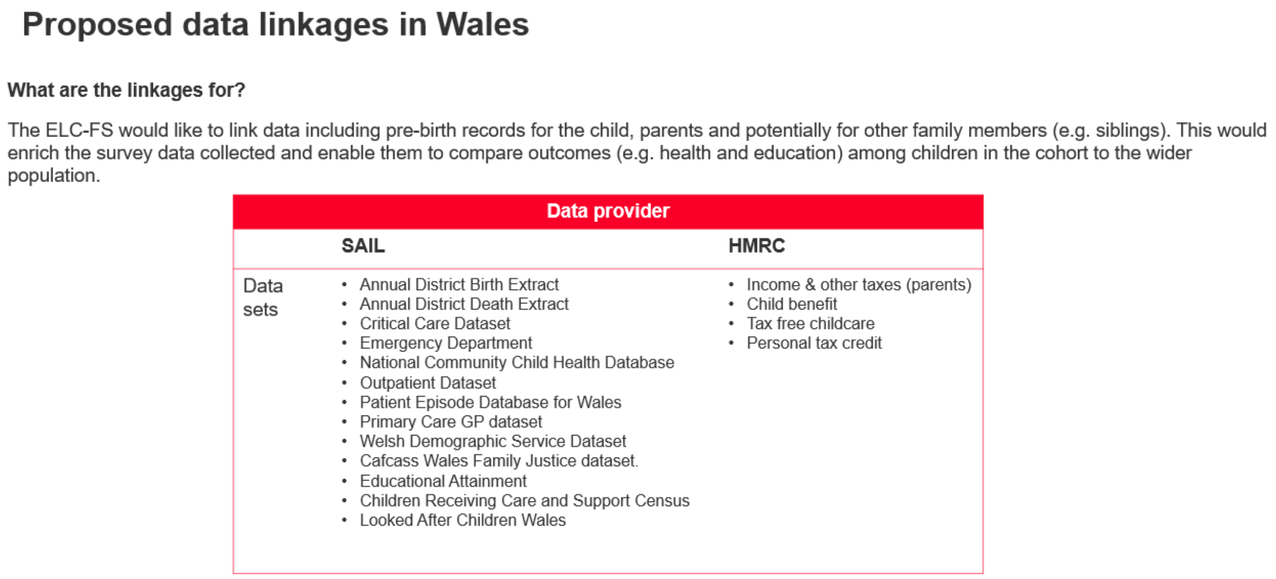


Proposed data linkages in England

What are the linkages for?

The ELC-FS would like to link data including pre-birth records for the child, parents and potentially for other family members (e.g. siblings). This would enrich the survey data collected and enable them to compare outcomes (e.g. health and education) among children in the cohort to the wider population.

	Data provider					
	NHS Digital	**ONS**	**PHE**	**DfE**	**DWP**	**HMRC**
Data sets	• NHS Birth Notifications (babies only) • HES Admitted Patient Care • HES Outpatient Data • HES Critical care data • Emergency Care Dataset • Mental Health Services Dataset • NHS Business Authority Community Dispensing data • Personal Demographics Service (embarkations) • National Cancer Registration Service • Community Services Dataset • Maternity Services Dataset • National Child Measurement Programme (joint data controllers with Public Health England)	• Birth records (babies only) • Death records	• National Congenital Anomaly and Rare Disease Registration Service • Newborn Hearing Screening Programme Data	• National Pupil Database (parents initially) • Children in Need • Children Looked After • Individual Learner Record (parents initially) • HESA data (parents initially)	• Benefit statistics (parents) (including disability living allowance, JSA, ESA, personal independence payments etc)	• Income & other taxes (parents) • Child benefit • Tax free childcare • Personal tax credit



Proposed data linkages in Wales

What are the linkages for?

The ELC-FS would like to link data including pre-birth records for the child, parents and potentially for other family members (e.g. siblings). This would enrich the survey data collected and enable them to compare outcomes (e.g. health and education) among children in the cohort to the wider population.

	Data provider	
	SAIL	**HMRC**
Data sets	• Annual District Birth Extract • Annual District Death Extract • Critical Care Dataset • Emergency Department • National Community Child Health Database • Outpatient Dataset • Patient Episode Database for Wales • Primary Care GP dataset • Welsh Demographic Service Dataset • Cafcass Wales Family Justice dataset. • Educational Attainment • Children Receiving Care and Support Census • Looked After Children Wales	• Income & other taxes (parents) • Child benefit • Tax free childcare • Personal tax credit

Proposed data linkages in Scotland

What are the linkages for?

The ELC-FS would like to link data including pre-birth records for the child, parents and potentially for other family members (e.g. siblings). This would enrich the survey data collected and enable them to compare outcomes (e.g. health and education) among children in the cohort to the wider population.

	Data provider		
	Public Health Scotland	**National Records for Scotland/ Scottish government**	**HMRC**
Data sets	• Maternity Records (SMR-02) • Birth registrations • Death registrations • Scottish birth records (neonatal discharge records) • Hospital admissions (SMR-01) • Mental Health data (SMR-04) • Cancer Registrations (SMR-06) • A&E attendances • Outpatient bookings (SMR-00) • Immunisation (SIRS) • Child Health Surveillance data Pre-School (CHS-PS) • Child Health Surveillance data – School (CHS-S) • Prescribing Information Systems	• Pupil Census and education datasets	• Income & other taxes (parents) • Child benefit • Personal tax credit • Tax free childcare

Proposed data linkages in Northern Ireland

What are the linkages for?

The ELC-FS would like to link data including pre-birth records for the child, parents and potentially for other family members (e.g. siblings). This would enrich the survey data collected and enable them to compare outcomes (e.g. health and education) among children in the cohort to the wider population.

	Data provider	
	Honest Broker Service Business Services Organisation (BSO)	**HMRC**
Data sets	• National Health Application and Information Structure (population spine) • NIMATS maternity records • Child Health System (newborn hearing screening, vaccinations and immunisation records, preschool developmental surveillance, school health, and special needs) • the Patient Administration System (PAS), which covers hospital admission and discharge data, outpatient treatment, etc. • SOSCARE/PARIS: centralised social service data • the Family Practitioner Services (FPS) Payment System, which contains information about payments for pharmaceutical services, dental services, ophthalmic services, and general medical services; • All dispensed medications (Enhanced Prescribing Database - EPD) • Emergency Department Systems, which covers treatment in EDs across NI • Linkages to other datasets (including cerebral palsy register; Self-Harm register	• Income & other taxes (parents) • Child benefit • Tax free childcare • Personal tax credit