Centre for Longitudinal Studies: Early Life Cohort Feasibility Study - Public Dialogue Report

Stakeholder activation workshop topline findings

November 2022

Bethany Dokal, Lucy Williams, Alice Coulter and Danny Price

40322420

Contents

1. Background and Method

1.1 Background

Funded by the Economic and Social Research Council (ESRC) and led by the Centre for Longitudinal Studies (CLS), the Early Life Cohort Feasibility Study (ELC-FS) will test the feasibility of a new UK-wide birth cohort study. Administrative data is intended to be used in a number of ways to help improve the representativeness and inclusivity in the ELC-FS – for sampling, recruitment and retention of participants and to enhance the study with data linkages to administrative records, including embedding the ELC-FS in administrative data to a greater degree than previous birth cohort studies. The study team has undertaken an extensive consultation and development phase to inform the design of the ELC-FS and as part of this Kantar Public was commissioned in 2021 to undertake stakeholder interviews and a public dialogue, followed by a stakeholder activation workshop.

This topline report summarises findings from the stakeholder activation workshop. The purpose of the workshop was to discuss the findings from the stakeholder interviews and public dialogue research conducted in 2021 exploring attitudes towards CLS's proposed uses of administrative data in the ELC-FS, to reflect on what the findings imply for key design decisions for the study and how best to balance the findings with wider scientific and operational considerations.

Three topics were discussed with stakeholders:

- Recruitment models for the study
- Use of admin data for non-response analysis and adjustment
- Administrative data linkages and how participant consent should be sought

1.2 Methodology

In total, 20 stakeholders attended the workshop. The workshop consisted of presentations and discussions held in four breakout groups. For each topic, Kantar Public presented the findings from the research and CLS presented wider scientific and operational considerations. Each breakout group was led by a Kantar Public moderator and included at least one member of the ELC-FS study team.

Stakeholder type	Number
Academics who make up the ELC-FS study team	8
Academics not part of the ELC-FS study team	4
Fieldwork delivery partners	2
Research council and agencies	3
Government and companies relating to government	3
Total	20

The workshop took place on Wednesday 5 October 2022 via Zoom.

2. Findings

2.1 Recruitment models

Within the activation workshop, stakeholders were presented with findings from the public dialogue and stakeholder research, which included responses to the proposed one step and two-step recruitment models. Findings included that most parents did not support the one step approach as they perceived it as untransparent and offered less participant control. Additionally, stakeholders had a preference for the two-step approach and thought it was unlikely that data controllers would agree to the one step approach. This was followed by an overview of the scientific needs and operational considerations around recruitment, which included the one step approach would be possible in England, Wales and Scotland but the two-step approach would be required in Northern Ireland. The materials used in the stakeholder activation workshop can be found in the appendix.

The presentation of findings was followed up with discussions in breakout groups. Key discussion points included the benefits and risks of using either the one step or two-step recruitment model and how the chosen model could be aligned with public needs and preferences.

In the breakout group discussions, stakeholders acknowledged that there was not a significant difference between the two recruitment models (one step / two-step) in terms of how they would work in England, Wales and Scotland in practice. A key characteristic of the initial two step approach (as presented in the public dialogue) was data controllers conducting an opt out exercise with potential participants prior to sharing their details with the fieldwork agency or study team. This has now been deemed unfeasible in England, Wales and Scotland as the data controllers are not able to do this, so the opt out exercise would need to be carried out by the fieldwork agency after the data had already been shared. Therefore, in the stakeholder activation workshop, , discussions around the two models focused on whether potential participants would prefer/ respond differently to receiving one or two letters ahead of their first visit from the interviewer, and the tone, branding and content of those letters.

Discussions acknowledged parents' concerns about their data being shared with a third party without their permission and that the prevalence of concerns about data privacy has increased over time. Now that data will be passed from the data controller to the fieldwork agency without first consulting potential participants in the proposed two-step approach, , stakeholders flagged the importance to addressing parents' concerns by being transparent about the legal basis of data sharing and the scrutiny applied.

Below we outline stakeholders' views about each recruitment model.

Views about the one step approach

- If potential participants only receive one letter ahead of their interviewer visit, stakeholders thought it was imperative to ensure they have other opportunities to opt out of the study and this is communicated with them.
- Stakeholders also warned that if participants are unhappy with their data being passed to a third party without their permission this has potential ramifications for the study. The Scottish Government's Health and Wellbeing Census was cited as an example where public concerns were amplified in the national press and ultimately damaged the study. Avoiding a similar scenario prompted some stakeholders to prioritise a risk averse two-step approach which would likely lead to fewer complaints.

- A further limitation of the one step approach highlighted by stakeholders was the risk of overwhelming parents if all information about the study is supplied in a single letter. Although one suggested mitigation for this could be through providing more detailed information for those who want it via FAQs or QR codes and keeping the recruitment letter to essential information only.
- Additionally, if participants only receive one letter it increases the chance, they may miss initial communications.
- However, one break out group thought that because of the national opt-out in England and Wales in these countries it is effectively a two-step approach, even if participants only receive one recruitment letter from the study team.

Views about the two-step approach

- Stakeholders acknowledged that the two-step approach was the preferred recruitment model during the public dialogue and is therefore likely to be more acceptable to the wider public than the one step model.
- As noted above, stakeholders commented that sending multiple letters provides an opportunity to provide greater research context whilst not 'overloading' potential participants in the first letter. And also, that potential participants would be less likely to miss two contact letters. Stakeholders from the fieldwork agency thought this may make it easier for the interviewers as those who wish to opt out would be more likely to have already communicated this.
- There was discussion about the format of the letters and in particular which headings and logos would be used in the initial opt out letter. Stakeholders commented on the importance of the first letter, and several felt it should be branded as and/or signed by the data controller. Understanding Society was cited as adopting a similar approach where they are looking to add the data controller's logo to see if this increases response rates. However, there were mixed views about this approach, with several stakeholders stressing the need for caution. They warned against creating the impression the letter was sent by the data controller, which may obscure that the data has already moved to the fieldwork agency. In some discussions the ELC-FS study team confirmed the data holders' signature cannot be used on the letters as that may seem the letter has come from the data controller rather than they may have simply approved the research.
- A further suggestion was to design the first letter, so it does not look like it comes from the fieldwork agency. However, given other concerns raised by stakeholders around misleading participants around who communications are from, this may not be seen as an ethical approach.
- Stakeholders also suggested using individual sample frame fields to maximum efforts on those who would otherwise be less well represented.

Whether different recruitment approaches should be used for different nations

Overall, stakeholders did not think the study needed to follow one recruitment approach across nations. Stakeholders did not think inconsistency in approach across nations would have an impact on the study from a scientific perspective. The Growing Up in Scotland (GUS) study was cited as an example of using different recruitment approaches across waves without affecting participation rates. Initial communications were sent from DWP in the first wave and the study team in the second wave, with participation rates being similar in both waves. However, they did not look at the composition of sample population between the two waves which may be a risk.

Other stakeholders queried the relevance of the GUS survey to the ELC-FS. One stakeholder warned that the wider context for carrying out research with the general public has changed, which may mean operational findings from the GUS survey are less applicable and due to digital developments where there is now a need to offer mechanisms for easier opts outs, which may mean the study team have to facilitate for a higher number of refusals.

Which recruitment model should be used for the ELC-FS

As mentioned above, the limitations of the two-step model in England, Wales and Scotland prompted stakeholders to focus on the practical differences between the two models, specifically about sending one or two initial contact letters. This resulted in mixed views about which model to recommend for the ELC-FS study.

- Only one stakeholder recommended the one step approach. They thought as it was possible the study team may as well use this approach.
- Those who preferred the two-step approach thought the one step model may cause unnecessary complaints from the public. Additionally, they thought the two-step model could reduce the possibility of an interviewer appearing at the door of a participant unexpectedly as they were less likely to miss one of the recruitment letters.
- One breakout group thought the study team should use the one step approach for England, Wales and Scotland and the two-step approach for Northern Ireland in the ELC-FS as a natural experiment and use findings from how this affect participation to inform the approach used in the mainstage.
- A further break out group did not reach a conclusion on which recruitment model they thought the study team should use.

Regardless of which approach is used stakeholders felt that maintaining a positive public perception of the study was important. Stakeholders suggested the following to achieve this:

- Adopting the most cautious approach (without compromising the study) and communicating this to participants. This could include details about who gave permissions for the study team to hold the data and the scrutiny applied to the research. It is important the research team strike a balance between normalising and explaining the research.
- Offering a compelling narrative in the participant communications including how and why the study team has people’s data.
- Providing assurances that any data sharing was minimal and only essential information was shared to the study team.
- Ensuring key ethical considerations are met, such as information being provided in a clear and transparent way, allowing for longer windows for participants to opt out of the study, and a single data holder handling participant information.
- Providing information in a layered format so as not to overwhelm participants; for example, starting with the overview of the study before going into specific information in later communications.

Other considerations for recruitment approach

Stakeholders also had other observations when considering recruitment approach for the ELC-FS, these included:

- Stakeholders raised concerns around the role of the interviewer, which reflected concerns raised during the public dialogue in which parents did not want to receive an unexpected visit. Stakeholders thought it was important to explain to participants what the interviewer visit would involve and to ensure sufficient time for people to opt out before that visit. One stakeholder also suggested that interviewers pre-send a postcard informing participants they will be arriving tomorrow as a polite courtesy.
- Stakeholders also emphasised the imperative to remove from the sample any babies who have died – although it was noted often some cases will likely be missed as seen in letters sent out as part of large scale cancer audits.

2.2 Non- response analysis

Within the activation workshop, stakeholders were also presented with findings from the public dialogue and stakeholder research on non-response analysis. Initially parents in the dialogue found the use of administrative data for non-response analysis as widely acceptable. However, as discussions developed, parents had mixed views about the acceptability of using data of those who do not respond to the study team’s invite letter or later recruitment attempts. In comparison, stakeholders found the use of de-identified data for non-response analysis and weighting acceptable as they help to ensure a representative sample. This was followed by an overview of the scientific needs and operational considerations such as the data flows included in ethics and data controllers' agreements and the overview of the proposed approach for the ELC-FS. The materials used in the stakeholder activation workshop can be found in appendix A.

The presentation of findings was followed up with discussions in breakout groups. Key discussion points included views on using de-identified data for non- response analysis and adjustment, the proposed uses of sampling frame data for those who opt-out or otherwise do not participate in the study, and what can be done to align approaches to what the public feel is acceptable.

Overall, stakeholders recognised the necessity of analysing opt-out data for non-response bias, which will contribute towards ensuring the quality of the data overall. **They accepted the use of administrative data for this purpose** and recognised that there was no equivalent alternative for achieving this. Moreover, due to the low risk of potential harm to the individual associated with using deidentified data they felt it was a suitable use and acceptable trade off.

Stakeholders agreed using the **privacy notice to highlight the use of deidentified admin data for non-response analysis was appropriate.** However, stakeholders raised concerns about the potential to 'bury' anything in the privacy notice and felt this would undermine the ethical basis for including it. They discussed concerns about reputational damage if the study was perceived to be attempting to deceive participants. One breakout group mentioned that the National Pupil Database (NPD) team had received criticism about an insufficient privacy notice from the Information Commissioner's Office (ICO). Stakeholders emphasised the need to prioritise ethical principles regardless of legality and one stakeholder stressed the need to avoid seeing the privacy notice as a 'tick box' exercise and carefully consider how the issue is communicated to participants.

In order to **mitigate public concerns around the use of deidentified data for non-response analysis**, stakeholders suggested the following:

- Explain why the use of deidentified administrative data is necessary and the implications of this use for both study participants and the ELC-FS. Stakeholders acknowledged this was a complex issue and a balance needs to be reached between transparency and prompting a disproportionate response to a proposed use which is low risk and standard in research.
- Trial and test different ways of communicating the issue so that it supports participant comprehension whilst avoiding drawing undue attention and alarm. This could involve visual examples, focusing on outcomes (what will happen to your data) rather than process. Also, one stakeholder recommended avoiding the word 'analysis', which they thought may be mis-interpreted as being used for substantive research.
- Signpost participants to further information (e.g. website) to read about the proposed use in more detail rather than overload the privacy notice.

Stakeholders were asked to consider **other approaches to understanding non-response and bias**. Although most of these were ultimately disregarded, suggestions included:

- Comparing the sampling frame data against the whole population, from which the study team creates the weights to correct for bias. In this scenario the study team does not take forward the data of those who have opted out. This approach was seen as beneficial as although the data of those who have opted out are still in the whole population data proportionately, they only make up a small number – which stakeholders thought may be preferable for this group.
- One group suggested to keep participant data within NHS Digital and use the NHS Trusted Research Environment (TRE) to analysis participant data. Stakeholders felt the benefit of this approach was that it would offer a more complete data set, allow comparison against the whole population and avoid needing to share data with a third party.
    - For this to be undertaken at the individual level using whole population data in this setting, a 'flag' of the invitation to be part of the sample and participation status would need to be transferred into the TRE (including the opting out non-responders). This would need to be done as an amendment to the NHS-D application, for example at the NHS-D application stage for the main study. Depending on who exactly is picking the sample (i.e. NHS -D or fieldwork agency) they would need to create this as an extra field, which could

  then be imported into the whole population data. This is an additional complexity but could be of enduring value.
  - A possible alternative could be high fidelity synthetic data/ directly replicate the selection criteria in the whole population data.
  - Whether or not the flags could be sent for the opting-out non-responders is a main topic of conversation and depends on their opportunity to consent/ opt-out of use of their data for non-response analysis.
  - Ultimately, this suggestion was deemed unfeasible as it is now too late/ complicated to do this now for the Feasibility Study. Additionally, the individual level data will also not have people who have opted out from the National Opt out, which will be c. 8% and so, a proper non-response analysis would still need to fall back on aggregate statistics on the composition of the whole population, that NHS-D would need to be requested to provide.
- Non response and sample analysis from Digi-Trials.
- Look at web forums and social media to establish whether people are referencing the ELC-FS. Although this was understood to provide insight on non-response motivations and not non-response bias.
- Asking NHS Digital for the aggregate statistics on the sample, calculated prior to the national-opt out, for non-response analysis

Stakeholders discussed **other mitigations which would allow wider or longer use of sampling frame data.** Stakeholders appreciated that study team may wish to hold on to raw data in case there is an issue with weighting. However, they acknowledged that GDPR requirements meant this needed a reasonably constrained time limit just sufficient to allow for the discovery of any issues around weighting. Stakeholders felt that deleting data as early as possible tended to reassure the public around the proposed use of data. Whilst keeping hold of data for a specific use was generally seen as acceptable, holding on to data ‘just in case’ was seen as less acceptable.

Further comments for non-response analysis:

Stakeholders were engaged by this discussion on non-response analysis and had other questions in relation to the design of the ELC-FS, these included:

- Whether further research will be undertaken to explore why participants did not wish to participate in the ELC-FS such as qualitative in-depth interviewing.
- Whether interviewer observations of participant homes will be included in the non-response data – which some stakeholders thought was potentially unethical.
- One stakeholder also queried that he was fully sure how participants can opt out from non-response analysis as they would still be included in the counts of the wider population.

2.3 Consent models

The final set of findings stakeholders were presented with in the activation workshop were on the topic of consent models for administrative data linkage. Findings included that parents were supportive of data linkage to survey response but had greater reservations around more sensitive types of data. Initially, parents preferred Model 4b (opt-in consent for each linkage) as it offered most transparency and choice to study participants. However, as the models were explored in greater depth, some parents shifted their position and thought Model 1 (not informed) was acceptable. Additionally, another small group of parents emerged who thought Model 2 (informed part of study) or Model 3 (informed, need to proactively opt-out) was acceptable. Whilst stakeholders thought explicit consent from participants was needed because it is more ethical, and therefore more appealing to data controllers. This was followed by an overview of the scientific needs and operational considerations including the proposed approach for ELC-FS, including which linkage consents will be sought. The materials used in the stakeholder activation workshop can be found in the appendix A.

The presentation of findings was followed up with discussions in breakout groups. Key discussion points included the benefits of using different consent models and how these should inform the approach of the ELC-FS, as well as the limitations and risks of the different models and what can be done to mitigate them and make it more acceptable to the public.

Across discussions around consent models, stakeholders appreciated the need to balance the scientific needs of the ELC-FS against ethical considerations and participants' potential concerns. Ultimately, most stakeholders thought an explicit consent model approach was preferable to collect consent for administrative data linkages initially in the study and potentially shifting to implicit consent for subsequent contact.

Views about explicit consent models

Explicit consent was generally viewed as a more ethical option as it offers participants choice and transparency. However, stakeholders were unsure what the impact of collecting explicit consent would have on the number of consents collected, particularly for social care and economic records data where there is less precedent. The Next Steps study was highlighted as an example where approximately 70% participants agreed to linkages, and this was 10-20% lower for economic linkages. Although one stakeholder warned that findings from Next Steps may be less applicable to the ELC-FS, as in society there is now a greater wariness about data linkage, becoming more salient post Covid-19. They warned collecting fewer consents would have a scientific bearing on the ELC-FS. For example, a potential negative impact could be collecting fewer consents from those who are in social care who will likely be a small group in the sample anyway. Therefore, they highlighted the need to balance providing participants with information and giving them the opportunity to agree to data linkage versus collecting data which is less common/ readily available and from groups less used to sharing. Emphasis should also be placed on building trust with participants, so they feel comfortable providing consent. In addition, explicit consent was noted to place greater burden on the participant, and they may feel pressurised to make a decision in the moment.

Stakeholders also warned that an explicit consent approach may be essential for data controllers to agree to their records being linked, and that the wording for each consent will likely need to be agreed with the data controller beforehand. Also, Randomised Controlled Trials (RCT) guidance requires explicit opt-in consent so this will be most familiar to health-related review panels. Stakeholders warned it is a risk that the Independent Group Advising on the Release of Data (IGARD) of NHDS Digital would not honour an opt-out consent.

Views about implicit consent models

Stakeholders were familiar with the implicit consent approach. Model 3 is currently being used in the GUS study with 17-year-olds, which checks awareness of data linkage but does not collect consent to do so. However, these participants have been part of the study for a long time and the study team have built trust with the participants. Stakeholders emphasised the importance of building trust by explaining why data is valuable to participants and the study team being respectful when participants to not want to share information/ participate.

Stakeholders noted that implicit consent could be less burdensome for participants. Administrative data could be used in place of survey questions, removing the need to explain potentially traumatic events from their past or recall mundane information (e.g. *how many times they have visited the GP in the last 12 months).* However, stakeholders also noted that participants may prefer disclosing this information themselves in their own words especially if they feel misrepresented by the data or have been treated unjustly by the service holding their data.

Further perceived limitations of an implicit consent model included potential reputational damage to the ELC-FS if participants are not certain about what they have agreed to and become aware later. This type of reputational damage could lead to participants opting out. This could also be caused be having descriptions of the data types which are too generic. Although one stakeholder flagged that this risk was also possible for explicit consent if participants forget they have previously consented.

Whether different consent models should be used for different linkages

Stakeholders generally thought it was acceptable to have different consent models for different linkages and appreciated that for more sensitive types of data participants would prefer a choice. Where stakeholders disagreed, this was due to concern that different consent models implied certain data was more or less special. One stakeholder highlighted that perceived sensitivity was a personal issue, reflecting participants’ circumstances and/or experiences with particular organisations. Another stakeholder acknowledged that different consent models also risk complicating the research process.

Preferred model for the proposed linkages

Stakeholders had mixed views on which consent model should be used for the data linkages – reflecting key concerns about ethical considerations and participant burden.

- One view that emerged among stakeholders was that **explicit consent was the most ethical approach** and potentially failing to collect this type of consent could lead to negative publicity for the ELC-FS. Stakeholders who shared this view recommended model 4b. They emphasised that the use of participant data should follow the principles of ‘no surprises, reasonable expectations and [constitute a] fair use of data.’ To minimise participant burden of model 4b, one stakeholder referenced a study which allowed participants to either agree to all data linkages or respond separately for each different data type. In this study, 70% of participants agreed to sharing all data and 30% selected by data type.
- A second view was that the use of **implicit or explicit consent should be informed by the proposed data linkage**. Stakeholders who shared this view thought that grouping less sensitive types of data together would be sufficient and would place less burden on the participant. However, they still recommended asking for individual consent to link to more sensitive types of data together such as mental health, social services and criminal record data.
- A third view was **uncertainty around the impact of asking about data linkage for more sensitive types of data**. They thought the feasibility study presented a good opportunity to determine the impact of consent model approaches and the impact on opt in/outs to study and in gaining permission to link to survey responses.

Despite these mixed views, on balance, three of the breakout groups ultimately **recommended using an explicit consent model approach in the initial recruitment such as model 4b**, which could then be amended to an implicit consent model such as model 3 over time. The fourth break out group did not reach a conclusion on which consent model should be used for the data linkages.

Collecting ongoing consents over time for study participants who continue to take part and those who do not

Stakeholders appreciated the challenge of achieving on going consent over time and, regardless of which model is used, emphasised the importance of building trust with participants to maintain linkages over time.

For those who continue to take part: Stakeholders warned that if the study team regularly requested explicit consent for linkages this would be too burdensome for participants. Instead, they recommended collecting initial explicit consent (model 4b) at the outset then implicit consent (model 3) throughout the study with participant actively having to opt out. This was seen as an approach that showed ‘proportionate caution’. Stakeholders recommended recollecting consent when the child reaches the ‘threshold of adulthood’.

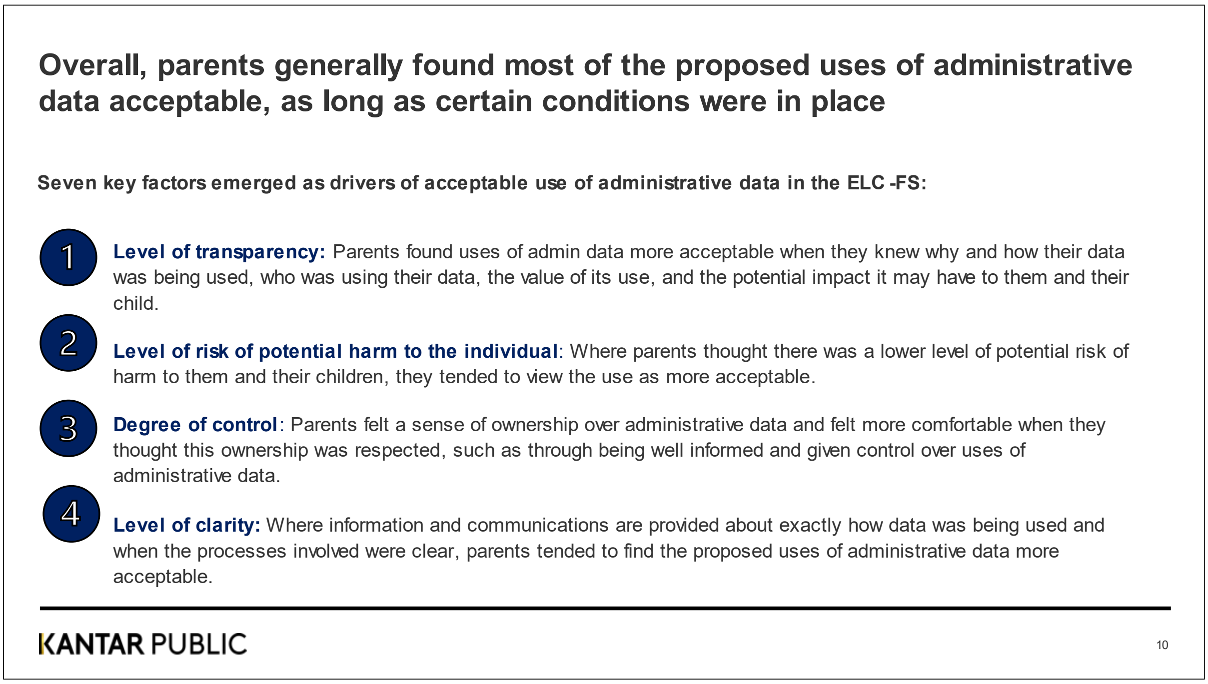
However, this was not a universal view and some stakeholders cautioned that greatest participant attrition happens between first and second contact. Therefore, providing too much information around data linkages in initial communications may be off-putting and lead to fewer consents collected and higher numbers of opt outs. One stakeholder also cautioned that reminding participants about linkages can provoke a negative response from those who have forgotten what they agreed to, resulting in considerable effort for the study team to address their concerns.

For those who do not continue to take part: Stakeholders recommended a layered approach to withdrawing from the study which would involve firstly checking if participants wish to no longer fill in the survey, followed up with a question whether participants no longer wish for the study team to use their administrative data.

3. Appendix A: Materials used in Workshop

Powerpoint slides were used to throughout the workshop to explain the findings from the Public Dialogue and Stakeholder findings research and CLS' wider scientific and operational considerations.

3.1 General dialogue findings materials



Overall, parents generally found most of the proposed uses of administrative data acceptable, as long as certain conditions were in place

Seven key factors emerged as drivers of acceptable use of administrative data in the ELC -FS:

1. **Level of transparency:** Parents found uses of admin data more acceptable when they knew why and how their data was being used, who was using their data, the value of its use, and the potential impact it may have to them and their child.
2. **Level of risk of potential harm to the individual**: Where parents thought there was a lower level of potential risk of harm to them and their children, they tended to view the use as more acceptable.
3. **Degree of control**: Parents felt a sense of ownership over administrative data and felt more comfortable when they thought this ownership was respected, such as through being well informed and given control over uses of administrative data.
4. **Level of clarity:** Where information and communications are provided about exactly how data was being used and when the processes involved were clear, parents tended to find the proposed uses of administrative data more acceptable.

KANTAR PUBLIC

Overall, parents generally found most of the proposed uses of administrative data acceptable, as long as certain conditions were in place

Seven key factors emerged as drivers of acceptable use of administrative data in the ELC -FS:

5 **Level of data security in place:** Reassurances around data security increased the acceptability of the proposed uses of administrative data.

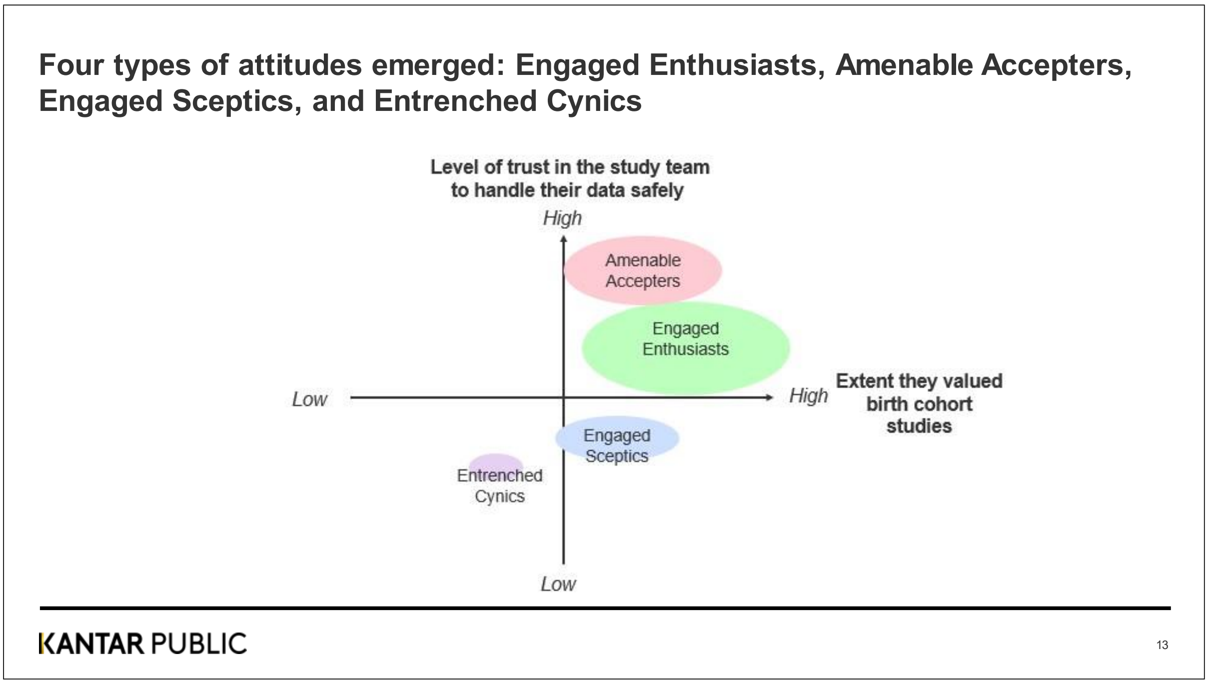
6 **Level of necessity:** Where the proposed use of administrative data was understood to be essential to the implementation of the study or a particular aim it was more acceptable to parents.

7 **Degree of normalisation:** Uses of administrative data were seen as more acceptable when parents knew data was used routinely in this way.

Despite general acceptability, there was variation in views across the sample and some parents retained higher levels of concern than others throughout the dialogue

Across the sample, how acceptable parents found the proposed uses of administrative data was driven by two key factors:

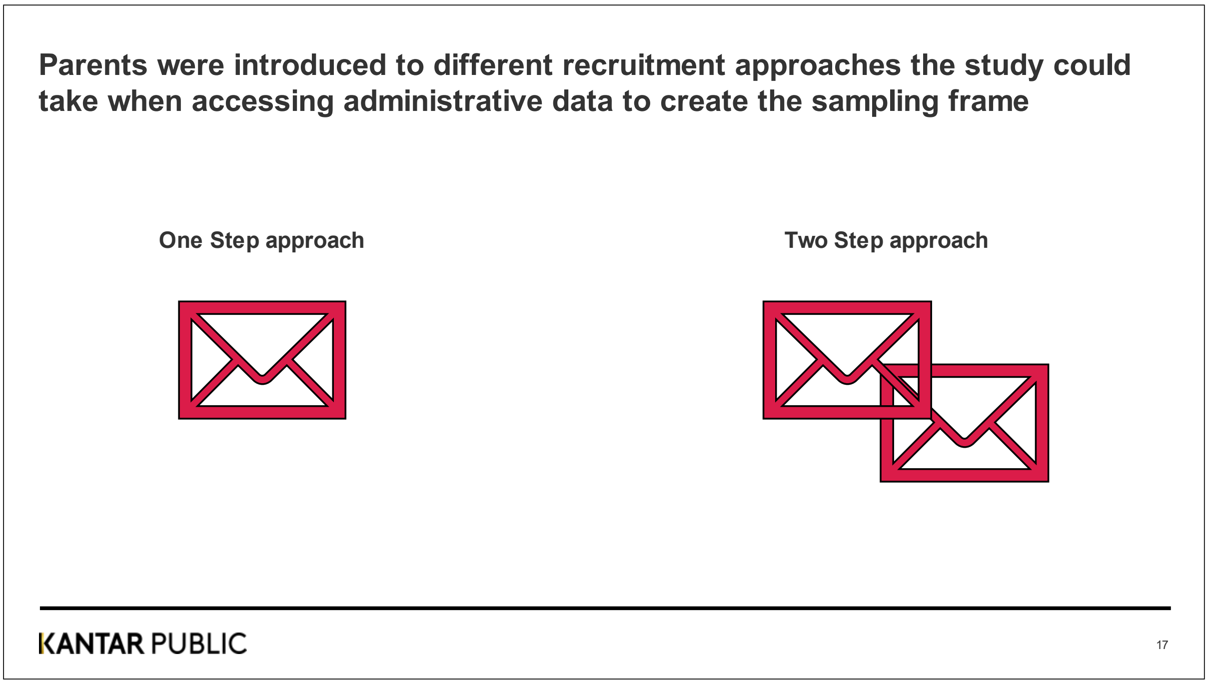
1. **The extent to which they valued birth cohort studies**
2. **the level of trust they had in the study team to handle their data safely**



Four types of attitudes emerged: Engaged Enthusiasts, Amenable Accepters, Engaged Sceptics, and Entrenched Cynics

KANTAR PUBLIC 13

3.2 Public Dialogue and stakeholder findings materials for recruitment models

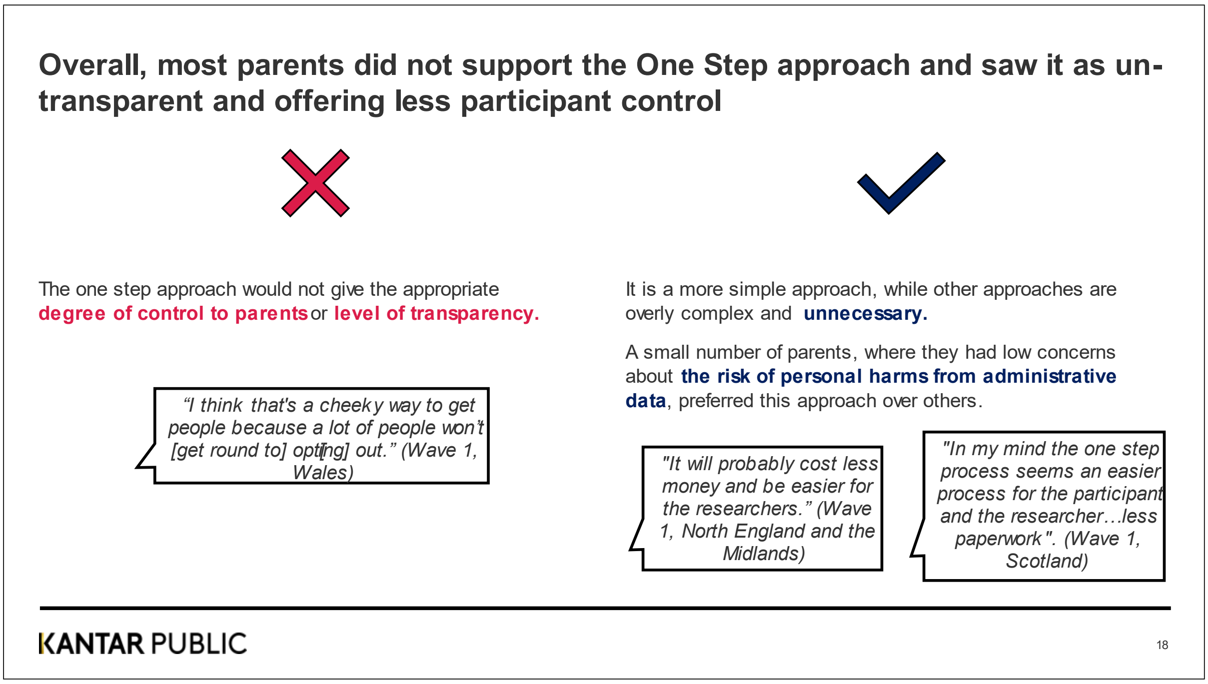


Parents were introduced to different recruitment approaches the study could take when accessing administrative data to create the sampling frame

One Step approach

Two Step approach

KANTAR PUBLIC 17



Overall, most parents did not support the One Step approach and saw it as un-transparent and offering less participant control

✗

The one step approach would not give the appropriate **degree of control to parents** or **level of transparency.**

> *"I think that's a cheeky way to get people because a lot of people won't [get round to] opt[ing] out." (Wave 1, Wales)*

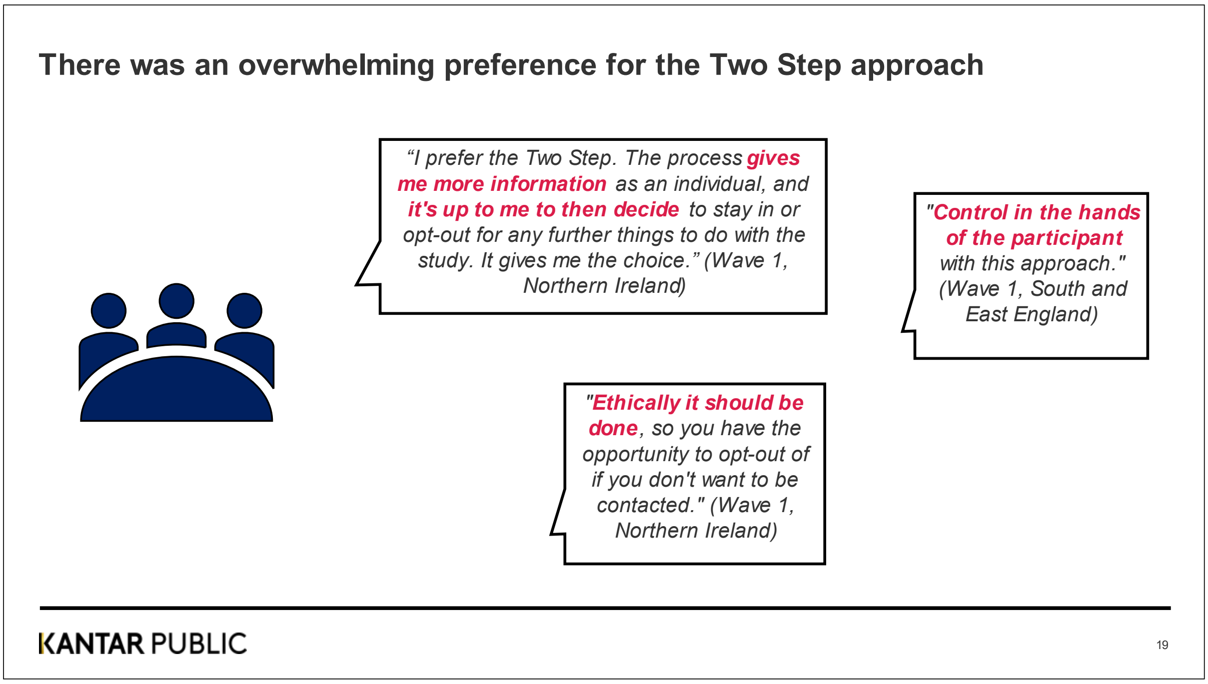
✓

It is a more simple approach, while other approaches are overly complex and **unnecessary.**

A small number of parents, where they had low concerns about **the risk of personal harms from administrative data**, preferred this approach over others.

> *"It will probably cost less money and be easier for the researchers." (Wave 1, North England and the Midlands)*

> *"In my mind the one step process seems an easier process for the participant and the researcher…less paperwork". (Wave 1, Scotland)*



There was an overwhelming preference for the Two Step approach

> *"I prefer the Two Step. The process **gives me more information** as an individual, and **it's up to me to then decide** to stay in or opt-out for any further things to do with the study. It gives me the choice." (Wave 1, Northern Ireland)*

> *"**Control in the hands of the participant** with this approach." (Wave 1, South and East England)*

> *"**Ethically it should be done**, so you have the opportunity to opt-out of if you don't want to be contacted." (Wave 1, Northern Ireland)*

Stakeholders widely thought it was unlikely data controllers would agree to the one-step approach, despite thinking it was generally acceptable

✗

The one- step approach carries **greater reputational risk for data controllers** and has the potential to alarm the general public.

- However, this was less of a concern for some stakeholders who noted ONS, and NHS Digital, would likely use third party organisations to send initial participant communications.
- Some felt they would be reassured about potential risks if they knew the study team would keep household attribution information separately from identifiers.

✓

Preferable for the study team as it enables access to the most complete sample.

Meets legal requirements.

Overall, stakeholders widely had a preference for the two step approach

✓

Would lessen the potential concern from the general public over their data being passed to a third party without their prior knowledge.

✗

Stakeholders appreciated that the two step approach **would limit the comprehensiveness and representativeness of the sample** in comparison to the one step approach.

A minority of stakeholders felt an opt in process at the first stage was **more ethically robust**.

Some stakeholders thought it would be beneficial for **initial communications to be sent from NHS trusts** as parents would have recently gone through their maternity services so they would be more familiar and (potentially) trusting of them in comparison to other data holders.

3.3 Scientific needs and Operational considerations materials for recruitment models

Recruitment models – ethics and data controllers

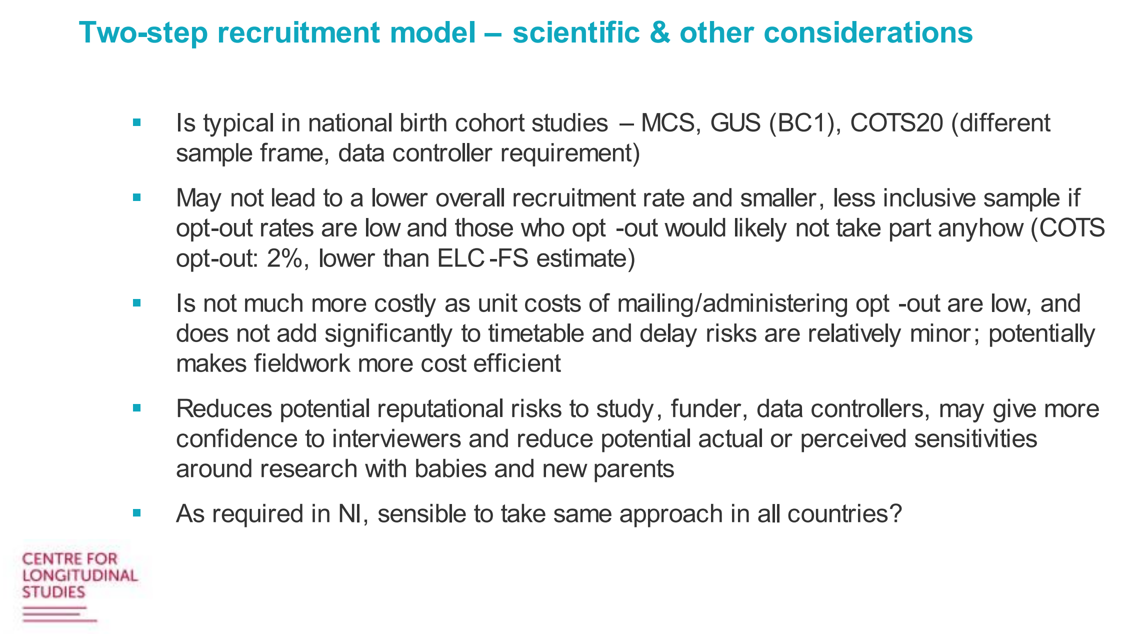
- UK wide ethical approval for one or two step approach
- England, Wales and Scotland
  - One step approach approved by CAG & Scotland PBPP (Two step approach would likely also be approved, likely requires amendment)
  - Data controllers not resourced to send opt -out letters (or to commission a third party)
  - Fieldwork agency could still send a prior opt out letter (or a third party)
  - Parents (and stakeholder) preference for two step approach was mainly driven by desire to have opportunity for opt out before the data controller shares their personal details with the study/another organisation. But in practice, as opt -out wouldn't be done by data controller, this would have to happen anyway, even for two step approach.
- Northern Ireland
  - Two step is expected, with data controller sending prior opt out letter

CENTRE FOR
LONGITUDINAL
STUDIES

One step recruitment model – scientific & other considerations

- May lead to higher overall recruitment rate and larger, more inclusive recruited sample (opt-out rate estimated: 8%)
- Is typical for most social surveys, and many health surveys and clinical trials
- Is quicker, fewer costs & resources, may minimise risks of fieldwork delay
- Would still allow refusals prior to interviewer visit, and would still give transparency and assurance to participants about data controller role
- Would mean less complexity around individual vs family opt -outs, as sample frame has two named parents (including own -household fathers)

CENTRE FOR
LONGITUDINAL
STUDIES

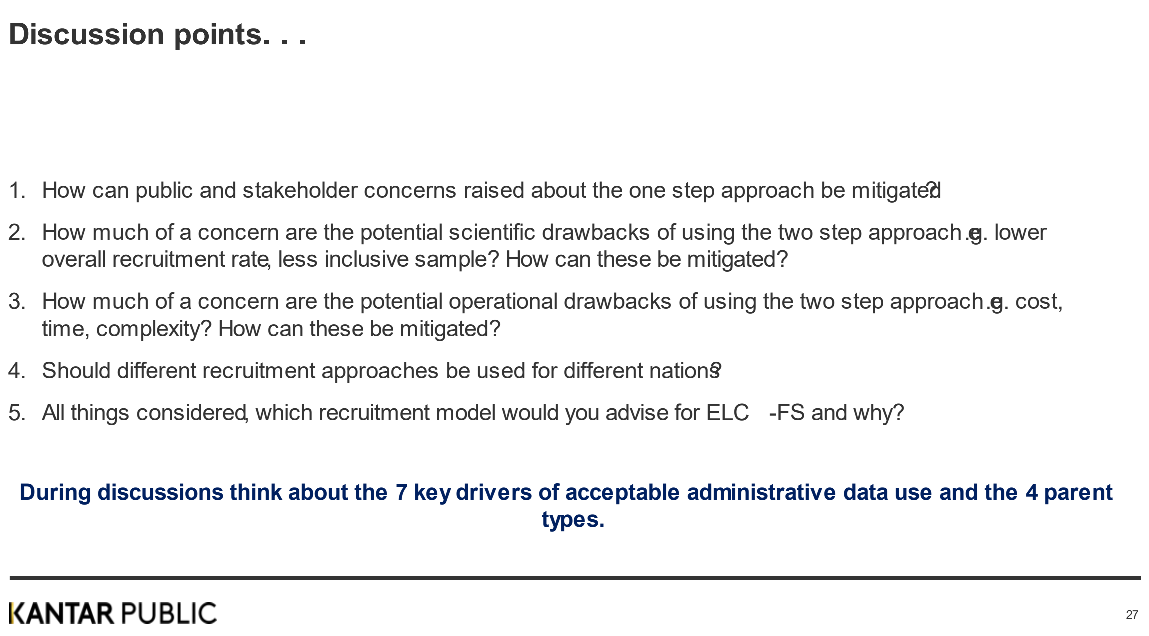


Two-step recruitment model – scientific & other considerations

- Is typical in national birth cohort studies – MCS, GUS (BC1), COTS20 (different sample frame, data controller requirement)
- May not lead to a lower overall recruitment rate and smaller, less inclusive sample if opt-out rates are low and those who opt-out would likely not take part anyhow (COTS opt-out: 2%, lower than ELC-FS estimate)
- Is not much more costly as unit costs of mailing/administering opt-out are low, and does not add significantly to timetable and delay risks are relatively minor; potentially makes fieldwork more cost efficient
- Reduces potential reputational risks to study, funder, data controllers, may give more confidence to interviewers and reduce potential actual or perceived sensitivities around research with babies and new parents
- As required in NI, sensible to take same approach in all countries?

CENTRE FOR LONGITUDINAL STUDIES

3.4 Recruitment models discussion points



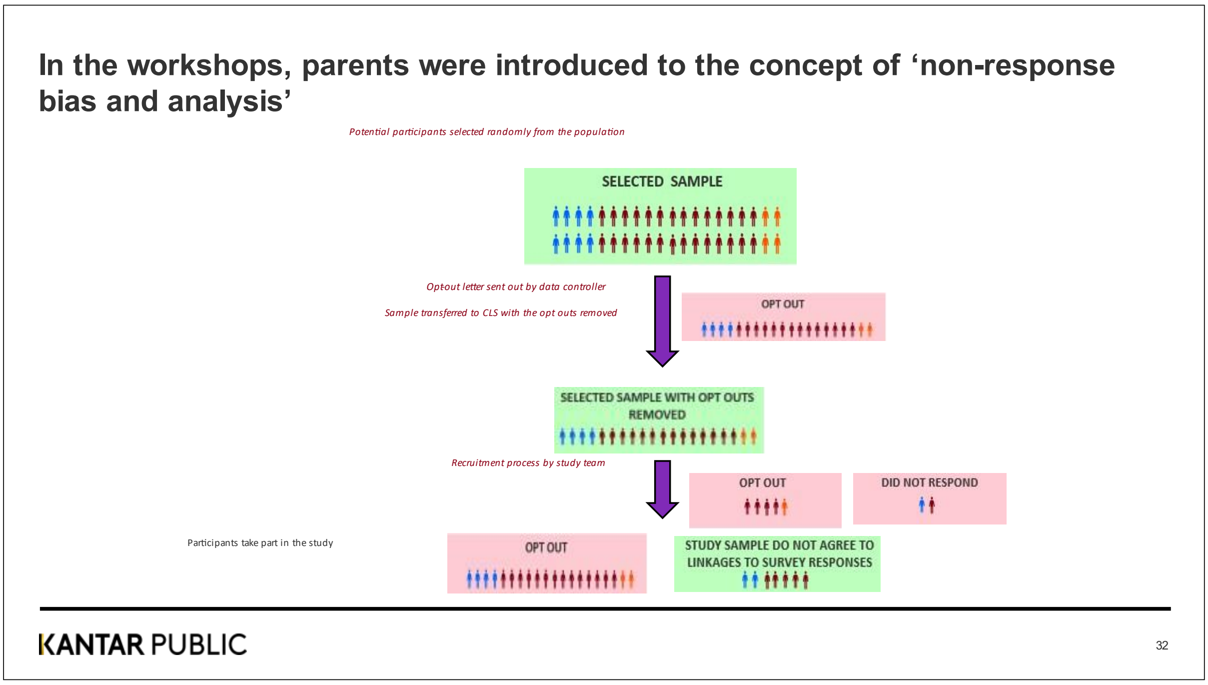
Discussion points. . .

1. How can public and stakeholder concerns raised about the one step approach be mitigated?
2. How much of a concern are the potential scientific drawbacks of using the two step approach e.g. lower overall recruitment rate, less inclusive sample? How can these be mitigated?
3. How much of a concern are the potential operational drawbacks of using the two step approach e.g. cost, time, complexity? How can these be mitigated?
4. Should different recruitment approaches be used for different nations?
5. All things considered, which recruitment model would you advise for ELC-FS and why?

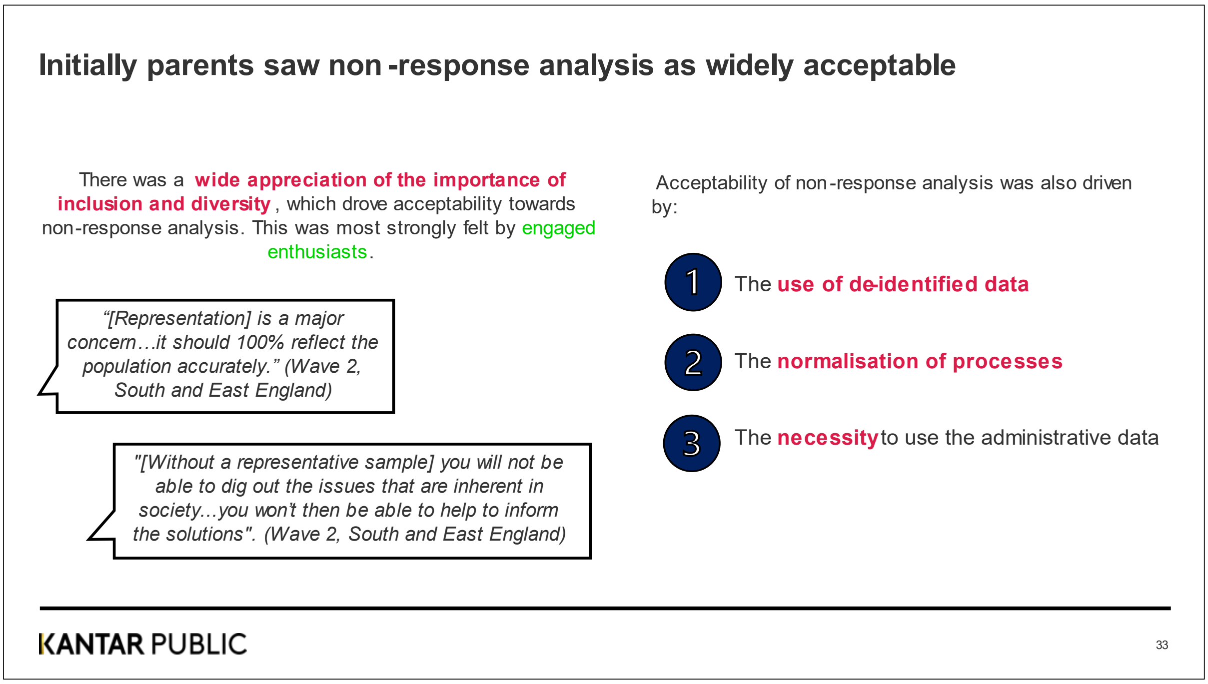
During discussions think about the 7 key drivers of acceptable administrative data use and the 4 parent types.

KANTAR PUBLIC 27

3.5 Public Dialogue and Stakeholder findings materials for non-response analysis and adjustment



In the workshops, parents were introduced to the concept of 'non-response bias and analysis'



Initially parents saw non-response analysis as widely acceptable

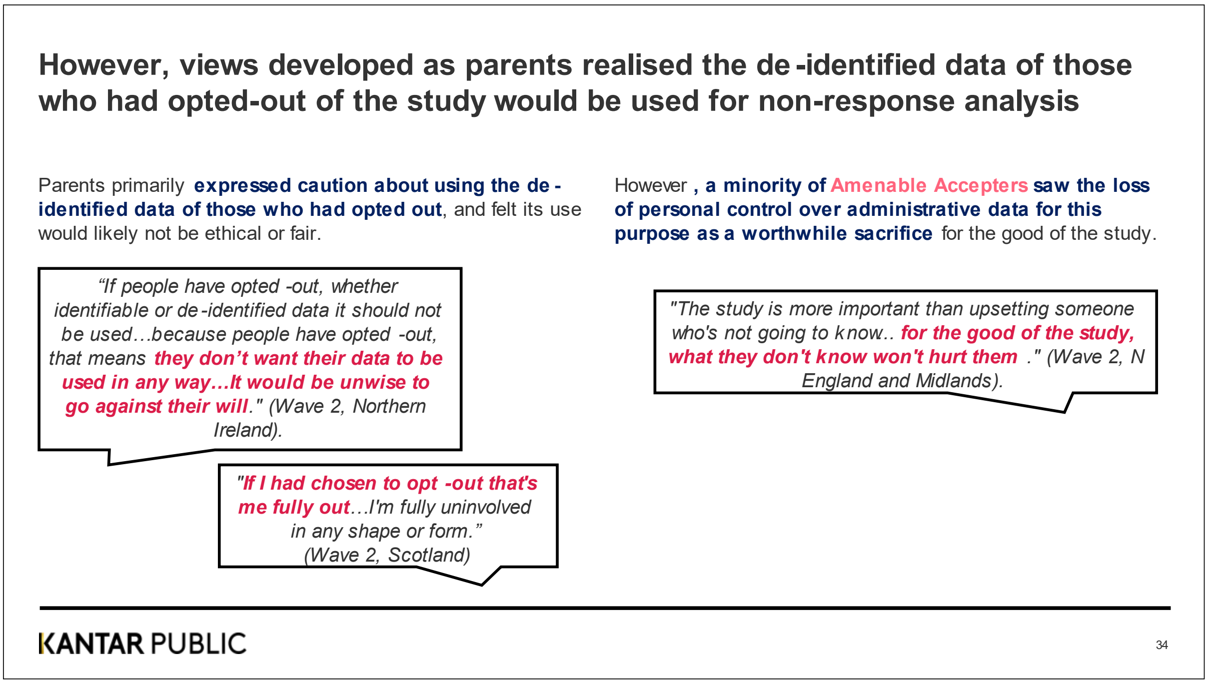
There was a **wide appreciation of the importance of inclusion and diversity**, which drove acceptability towards non-response analysis. This was most strongly felt by engaged enthusiasts.

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

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

Acceptability of non-response analysis was also driven by:

1. The **use of de-identified data**
2. The **normalisation of processes**
3. The **necessity** to use the administrative data



However, views developed as parents realised the de-identified data of those who had opted-out of the study would be used for non-response analysis

Parents primarily **expressed caution about using the de-identified data of those who had opted out**, and felt its use would likely not be ethical or fair.

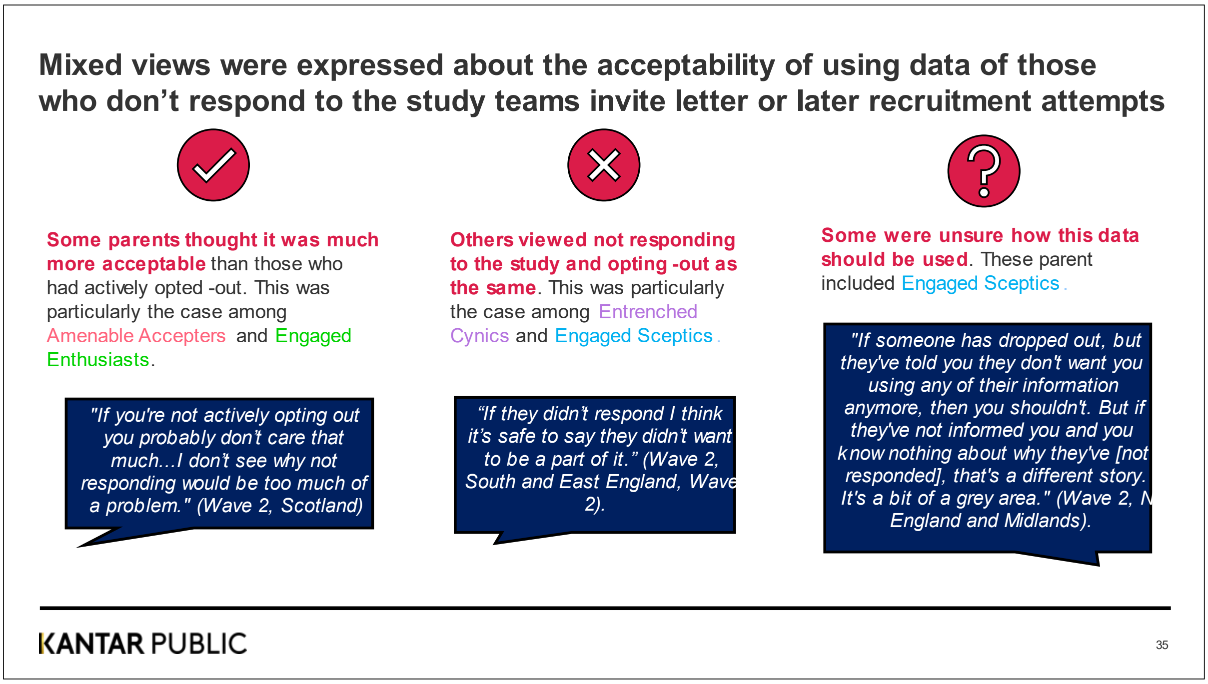
"If people have opted-out, whether identifiable or de-identified data it should not be used…because people have opted -out, that means ***they don't want their data to be used in any way…It would be unwise to go against their will****." (Wave 2, Northern Ireland).*

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

However**, a minority of** Amenable Accepters **saw the loss of personal control over administrative data for this purpose as a worthwhile sacrifice** for the good of the study.

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

KANTAR PUBLIC 34



Mixed views were expressed about the acceptability of using data of those who don't respond to the study teams invite letter or later recruitment attempts

✓

Some parents thought it was much more acceptable than those who had actively opted-out. This was particularly the case among Amenable Accepters and Engaged Enthusiasts.

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

✗

Others viewed not responding to the study and opting-out as the same. This was particularly the case among Entrenched Cynics and Engaged Sceptics.

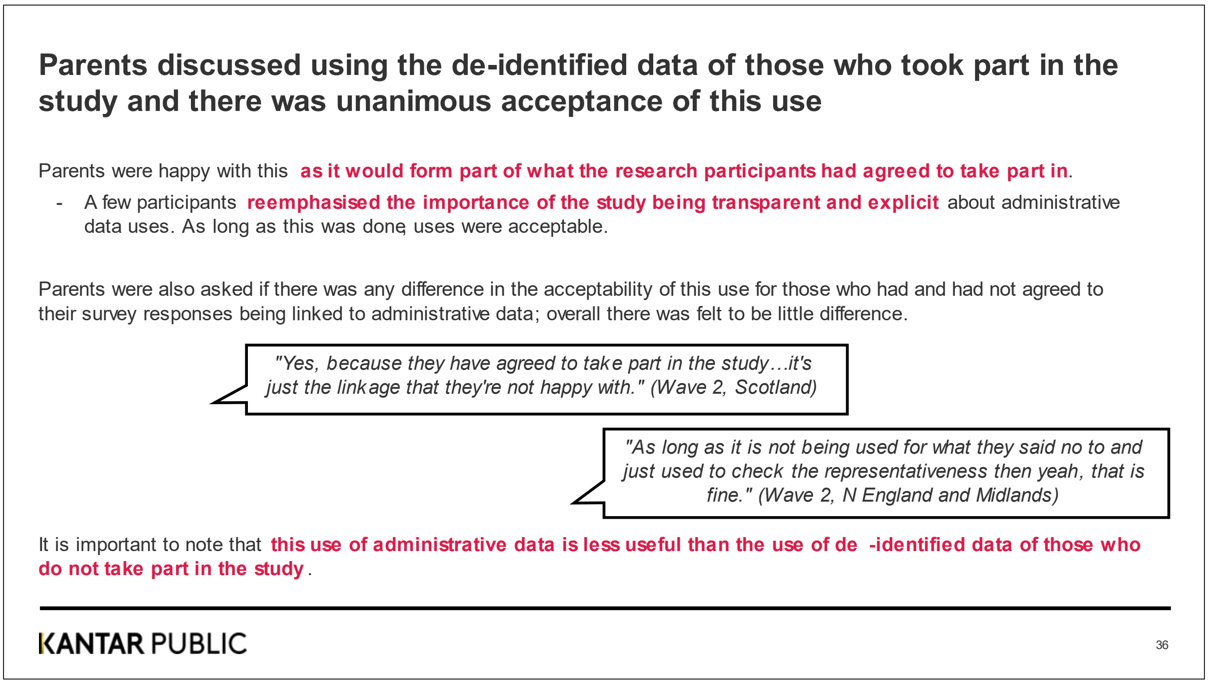
"If they didn't respond I think it's safe to say they didn't want to be a part of it." (Wave 2, South and East England, Wave 2).

?

Some were unsure how this data should be used. These parent included Engaged Sceptics.

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

KANTAR PUBLIC 35



Parents discussed using the de-identified data of those who took part in the study and there was unanimous acceptance of this use

Parents were happy with this **as it would form part of what the research participants had agreed to take part in**.

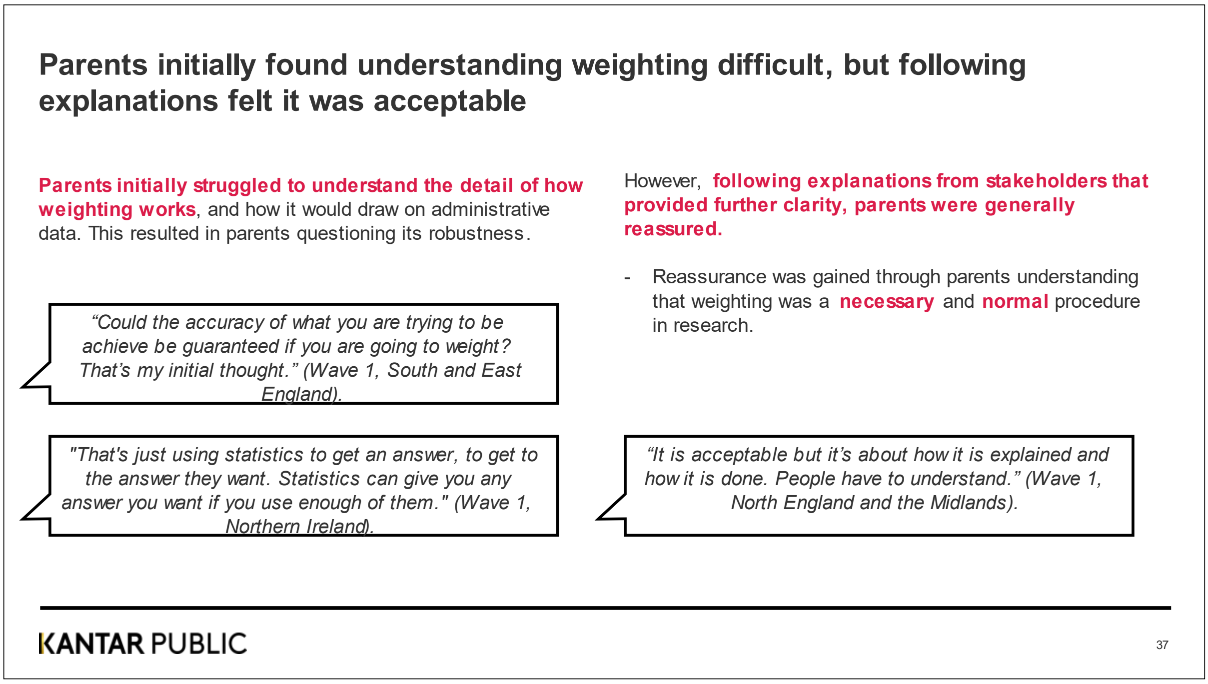
- A few participants **reemphasised the importance of the study being transparent and explicit** about administrative data uses. As long as this was done, uses were acceptable.

Parents were also asked if there was any difference in the acceptability of this use for those who had and had not agreed to their survey responses being linked to administrative data; overall there was felt to be little difference.

> *"Yes, because they have agreed to take part in the study…it's just the linkage that they're not happy with." (Wave 2, Scotland)*

> *"As long as it is not being used for what they said no to and just used to check the representativeness then yeah, that is fine." (Wave 2, N England and Midlands)*

It is important to note that **this use of administrative data is less useful than the use of de-identified data of those who do not take part in the study**.



Parents initially found understanding weighting difficult, but following explanations felt it was acceptable

Parents initially struggled to understand the detail of how weighting works, and how it would draw on administrative data. This resulted in parents questioning its robustness.

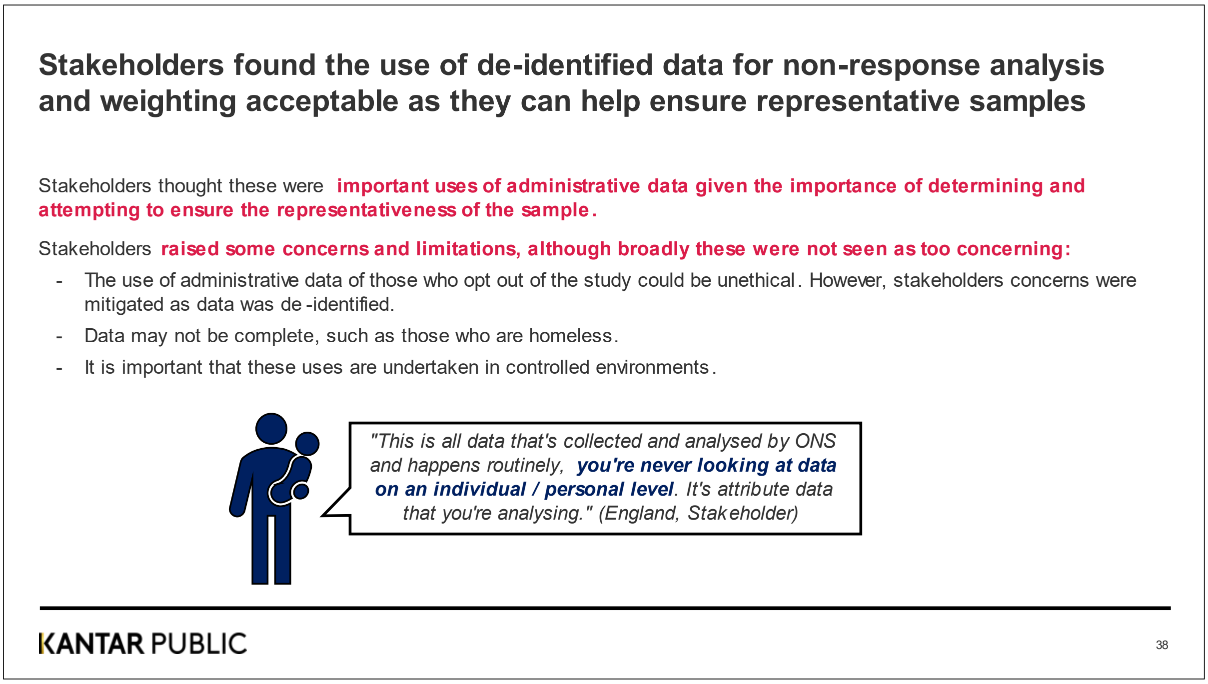
> *"Could the accuracy of what you are trying to be achieve be guaranteed if you are going to weight? That's my initial thought." (Wave 1, South and East England).*

> *"That's just using statistics to get an answer, to get to the answer they want. Statistics can give you any answer you want if you use enough of them." (Wave 1, Northern Ireland).*

However, **following explanations from stakeholders that provided further clarity, parents were generally reassured.**

- Reassurance was gained through parents understanding that weighting was a **necessary** and **normal** procedure in research.

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



Stakeholders found the use of de-identified data for non-response analysis and weighting acceptable as they can help ensure representative samples

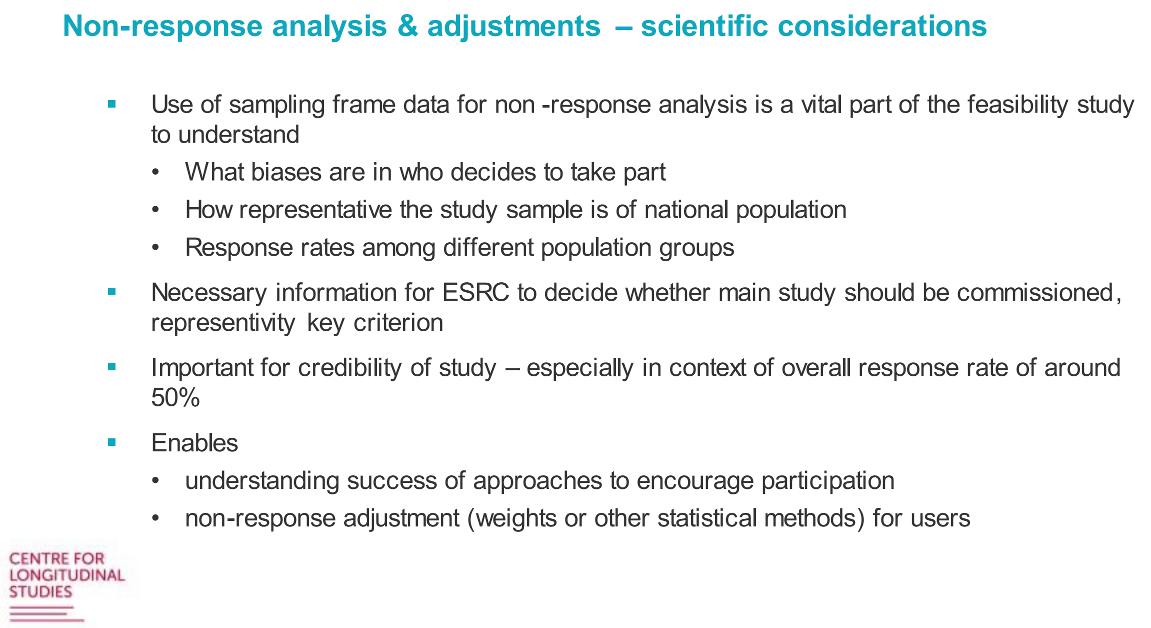
Stakeholders thought these were **important uses of administrative data given the importance of determining and attempting to ensure the representativeness of the sample.**

Stakeholders **raised some concerns and limitations, although broadly these were not seen as too concerning:**

- The use of administrative data of those who opt out of the study could be unethical. However, stakeholders concerns were mitigated as data was de-identified.
- Data may not be complete, such as those who are homeless.
- It is important that these uses are undertaken in controlled environments.

"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)*

3.6 Scientific need and Operational considerations materials for non-response analysis and adjustment



Non-response analysis & adjustments – scientific considerations

- Use of sampling frame data for non-response analysis is a vital part of the feasibility study to understand
  - What biases are in who decides to take part
  - How representative the study sample is of national population
  - Response rates among different population groups
- Necessary information for ESRC to decide whether main study should be commissioned, representivity key criterion
- Important for credibility of study – especially in context of overall response rate of around 50%
- Enables
  - understanding success of approaches to encourage participation
  - non-response adjustment (weights or other statistical methods) for users

CENTRE FOR LONGITUDINAL STUDIES

Non-response analysis & adjustments – proposed mitigations

- Public concerns around study receiving de -identified data from sampling frame about non - participants. Stakeholders mostly thought acceptable, with mitigations
- Similar concerns raised by Scotland PBPP and by CAG, have approved with mitigations
- Study aims to mitigate concerns –
  - link to sampling frame privacy notice in opt-out/advance letters - informs all selected sample members of data processing for non-response analysis and allows them to object to this(GDPR requirement)
  - use of sampling frame data of non-participants would be
    - used only to evaluate the representivity of study respondents (and to produce non-response weights/missing data guidance for users)
    - shared only within narrow project team
    - destroyed when analysis has taken place
  - sampling fields to go direct to fieldwork agency who would send the study team
    - de-identified sampling fields for the full sample, flagged according to participation
    - for non-participants, effectively anonymised
- There are precedents from other surveys sampled via administrative records (e.g. COSMO sampled via NPD) for this use of sampling frame data

CENTRE FOR LONGITUDINAL STUDIES

Non-response analysis & adjustments – inherent tension between public concerns and scientific needs?

- Use of sampling frame data for all non -participants is necessary to carry out non-response analysis.
- Proposed mitigations of inclusion in privacy notice gives transparency and opportunity to object to this processing
- However, in practice unlikely many people will object to this processing as this option would not be given strong prominence in letters, and would need to proactively get in touch to do this

CENTRE FOR LONGITUDINAL STUDIES

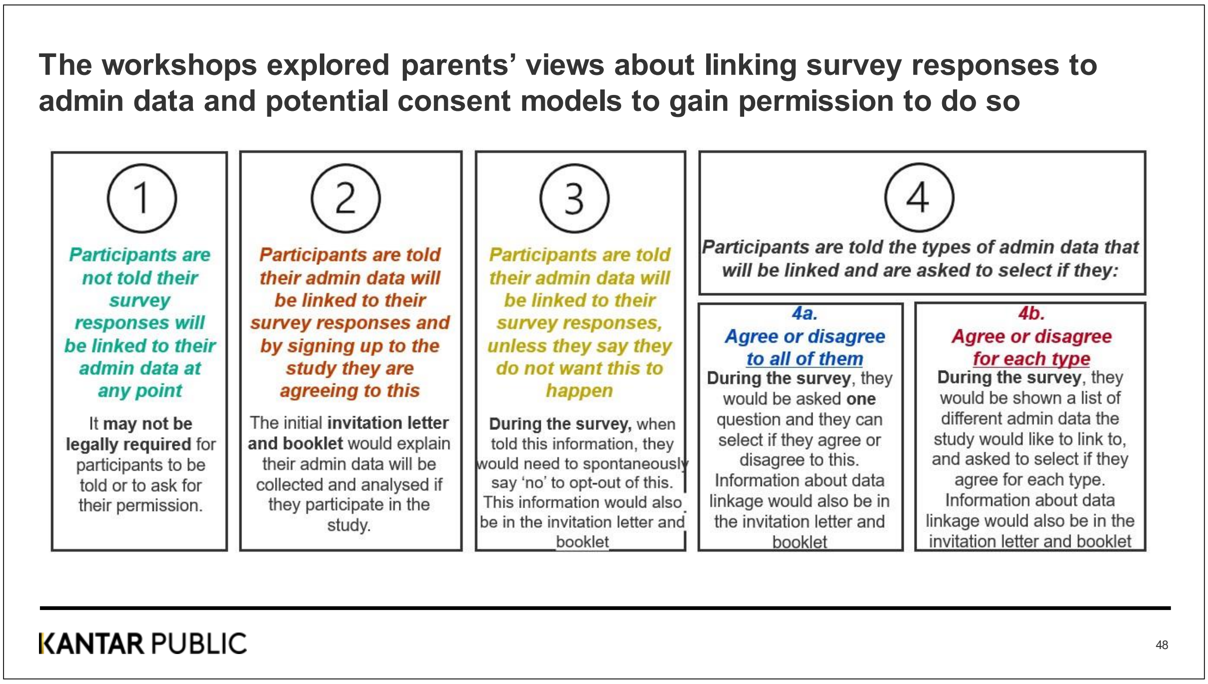
3.7 Non-response analysis and adjustment analysis

Discussion points. . .

1. How can public concerns raised about the use of de-identified data for non-response analysis be mitigated? Are these mitigations satisfactory?
2. Are there other approaches to understanding non-response and bias using administrative data?
3. Are there other mitigations which would allow wider or longer use of sampling frame data?
4. All things considered, would you say that the proposed approach is acceptable, bearing in mind public concerns? Why?

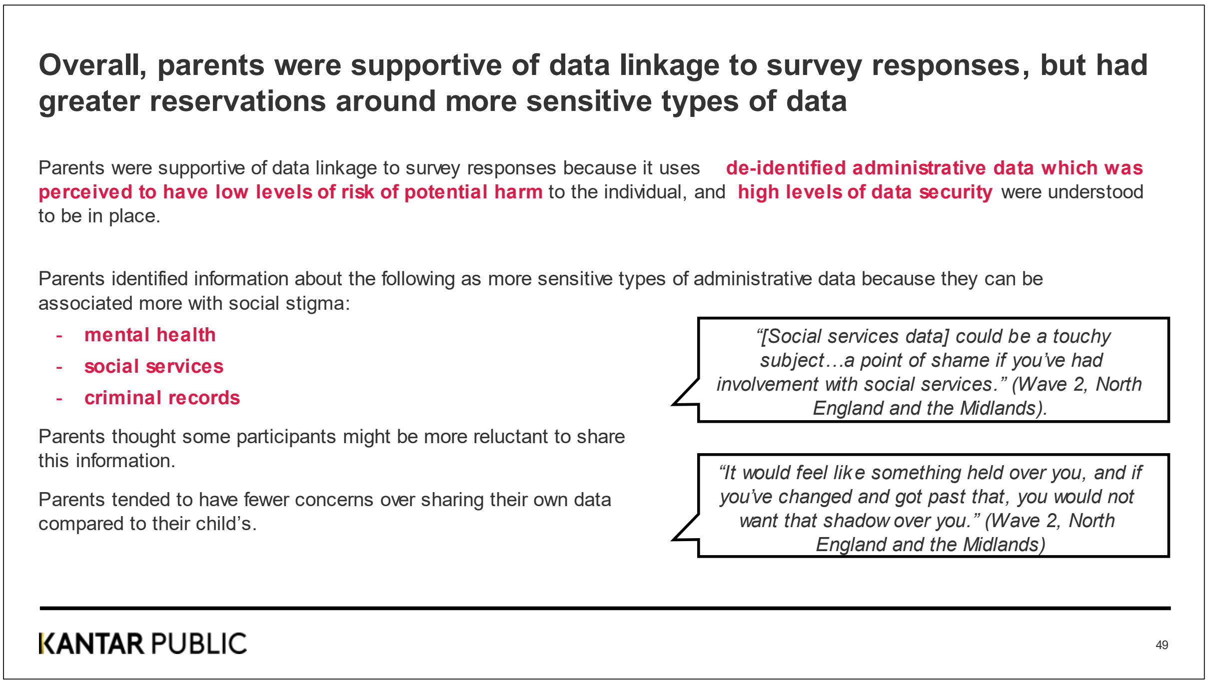
During discussions think about the 7 key drivers of acceptable administrative data use and the 4 parent types.

3.8 Public Dialogue and Stakeholder findings materials for consent models for administrative data linkage



The workshops explored parents' views about linking survey responses to admin data and potential consent models to gain permission to do so

1	2	3	4
Participants are not told their survey responses will be linked to their admin data at any point It **may not be legally required** for participants to be told or to ask for their permission.	***Participants are told their admin data will be linked to their survey responses and by signing up to the study they are agreeing to this*** The initial **invitation letter and booklet** would explain their admin data will be collected and analysed if they participate in the study.	***Participants are told their admin data will be linked to their survey responses, unless they say they do not want this to happen*** **During the survey,** when told this information, they would need to spontaneously say 'no' to opt-out of this. This information would also be in the invitation letter and booklet	***Participants are told the types of admin data that will be linked and are asked to select if they:*** ***4a. Agree or disagree to all of them*** **During the survey**, they would be asked **one** question and they can select if they agree or disagree to this. Information about data linkage would also be in the invitation letter and booklet ***4b. Agree or disagree for each type*** **During the survey**, they would be shown a list of different admin data the study would like to link to, and asked to select if they agree for each type. Information about data linkage would also be in the invitation letter and booklet



Overall, parents were supportive of data linkage to survey responses, but had greater reservations around more sensitive types of data

Parents were supportive of data linkage to survey responses because it uses **de-identified administrative data which was perceived to have low levels of risk of potential harm** to the individual, and **high levels of data security** were understood to be in place.

Parents identified information about the following as more sensitive types of administrative data because they can be associated more with social stigma:

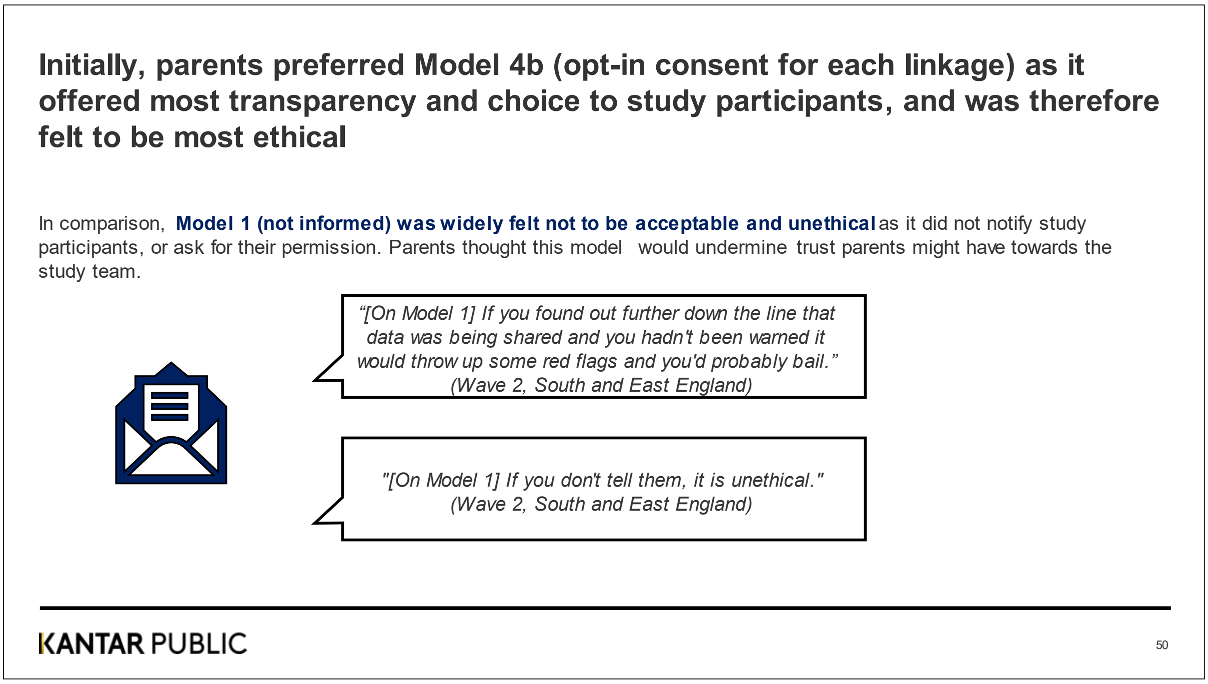
- **mental health**
- **social services**
- **criminal records**

Parents thought some participants might be more reluctant to share this information.

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

> *"[Social services data] could be a touchy subject…a point of shame if you've had involvement with social services." (Wave 2, North England and the Midlands).*

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



Initially, parents preferred Model 4b (opt-in consent for each linkage) as it offered most transparency and choice to study participants, and was therefore felt to be most ethical

In comparison, **Model 1 (not informed) was widely felt not to be acceptable and unethical** as it did not notify study participants, or ask for their permission. Parents thought this model would undermine trust parents might have towards the study team.

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

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

However, as the models were explored in greater depth, some parents shifted their position and thought Model 1 (not informed) was acceptable

Some Engaged Enthusiasts and Amenable Accepters thought Model 1 was acceptable to use to access important sensitive data in certain instances where participants might be less likely to consent.

- They felt these options could provide sufficient transparency and choice for the study participants.

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

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

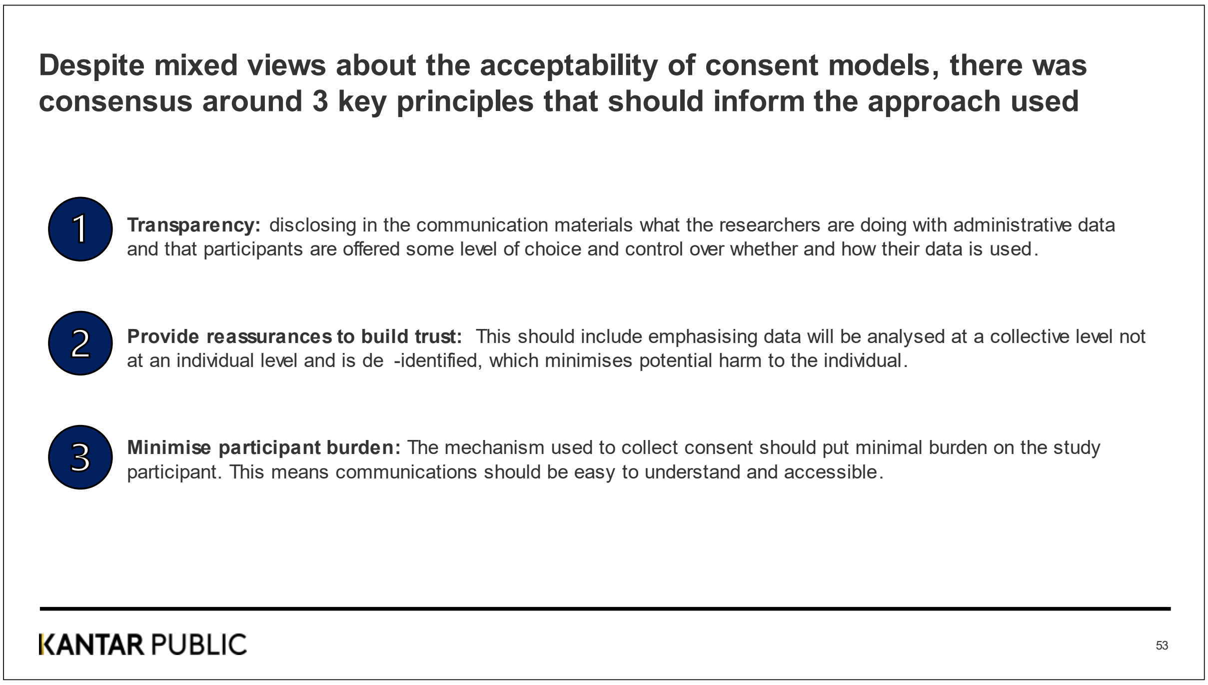
Additionally, another small group of parents emerged who thought Model 2 (informed part of study) or Model 3 (informed, need to proactively opt-out) was acceptable

During discussions, these parents went onto think that **Models 4a (opt-in consent for all together) and 4b (opt-in consent for each linkage) put an unnecessary burden on study participants**. It would also reduce consent to linkages. As a result, these parents went onto think that model 2 or 3 would be more suitable, and provide sufficient transparency and choice, while not burdening participants.

These parents were primarily a mix of Engaged Enthusiasts, Amenable Accepters and Engaged Sceptics who thought the study had high value.

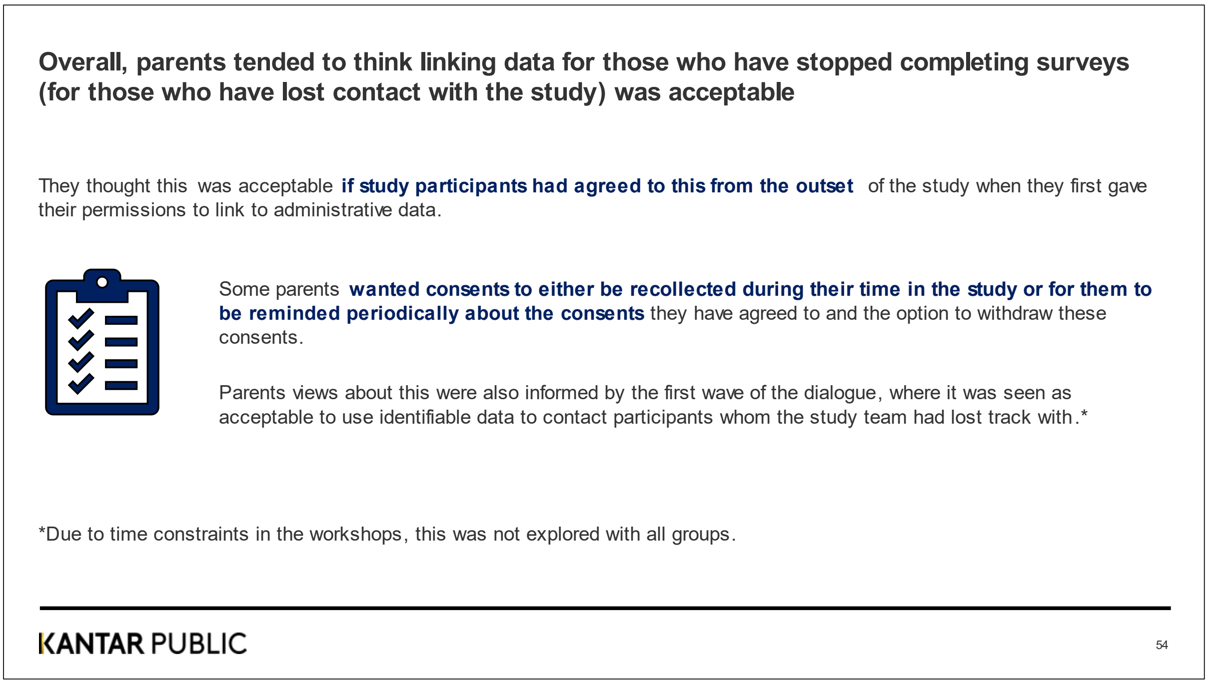
[On Model 2, there is an] element of transparency…would hopefully instil confidence that your data wouldn't be shared without your consent." (Wave 2, South and East England)

[On Model 2] It's better to say what you're doing, as people will likely find out at some stage anyway. It's the right thing to do and help to build trust with the study participants" (Wave 2, N England and Midlands).



Despite mixed views about the acceptability of consent models, there was consensus around 3 key principles that should inform the approach used

1. **Transparency:** disclosing in the communication materials what the researchers are doing with administrative data and that participants are offered some level of choice and control over whether and how their data is used.
2. **Provide reassurances to build trust:** This should include emphasising data will be analysed at a collective level not at an individual level and is de-identified, which minimises potential harm to the individual.
3. **Minimise participant burden:** The mechanism used to collect consent should put minimal burden on the study participant. This means communications should be easy to understand and accessible.



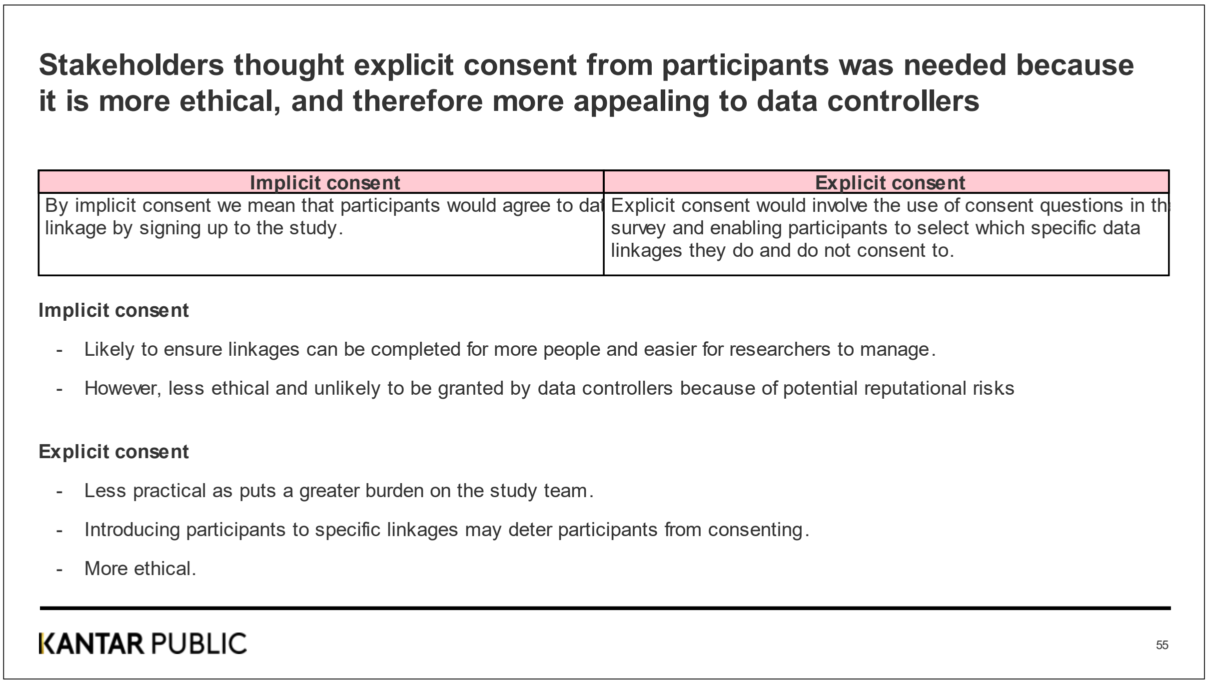
Overall, parents tended to think linking data for those who have stopped completing surveys (for those who have lost contact with the study) was acceptable

They thought this was acceptable **if study participants had agreed to this from the outset** of the study when they first gave their permissions to link to administrative data.

Some parents **wanted consents to either be recollected during their time in the study or for them to be reminded periodically about the consents** they have agreed to and the option to withdraw these consents.

Parents views about this were also informed by the first wave of the dialogue, where it was seen as acceptable to use identifiable data to contact participants whom the study team had lost track with.*

*Due to time constraints in the workshops, this was not explored with all groups.



Stakeholders thought explicit consent from participants was needed because it is more ethical, and therefore more appealing to data controllers

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

Implicit consent

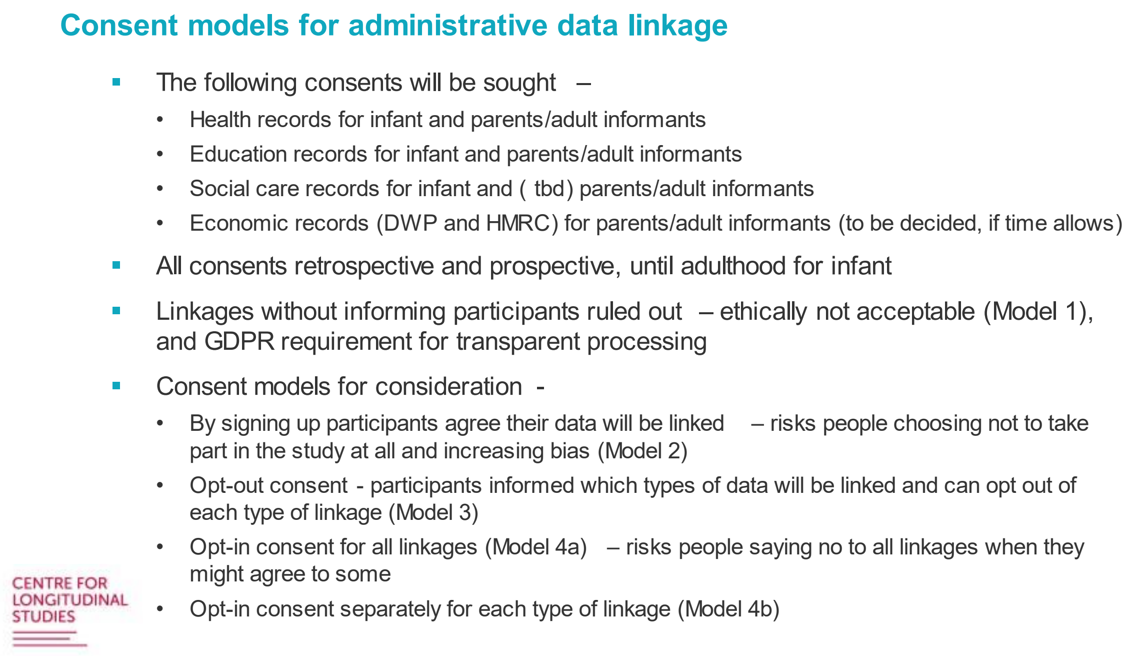
- Likely to ensure linkages can be completed for more people and easier for researchers to manage.
- However, less ethical and unlikely to be granted by data controllers because of potential reputational risks

Explicit consent

- Less practical as puts a greater burden on the study team.
- Introducing participants to specific linkages may deter participants from consenting.
- More ethical.

KANTAR PUBLIC 55

3.9 Operational consideration materials for Consent models for administrative data linkage



Consent models for administrative data linkage

- The following consents will be sought –
  - Health records for infant and parents/adult informants
  - Education records for infant and parents/adult informants
  - Social care records for infant and (tbd) parents/adult informants
  - Economic records (DWP and HMRC) for parents/adult informants (to be decided, if time allows)
- All consents retrospective and prospective, until adulthood for infant
- Linkages without informing participants ruled out – ethically not acceptable (Model 1), and GDPR requirement for transparent processing
- Consent models for consideration -
  - By signing up participants agree their data will be linked – risks people choosing not to take part in the study at all and increasing bias (Model 2)
  - Opt-out consent - participants informed which types of data will be linked and can opt out of each type of linkage (Model 3)
  - Opt-in consent for all linkages (Model 4a) – risks people saying no to all linkages when they might agree to some
  - Opt-in consent separately for each type of linkage (Model 4b)

CENTRE FOR LONGITUDINAL STUDIES

Consent models for administrative data linkage – background info

- Legal basis for study is public task, hence consents do not need to meet GDPR requirements
- Obtaining informed consent for linkages is primarily an ethical consideration (and may be a requirement – no ethics approvals as yet for this part of study)
- Informed consent is also likely to be required by data holders to agree to linkages
- Informed consent can be opt-out (implicit) or opt-in (explicit)
- Implicit consent likely to result in higher number of linkages and less bias, takes less time operationally
- Explicit consent is generally the model used for admin data linkages in social and health surveys and what ethics committees and data controllers usually expect
- However, there are some precedents from other studies for approaches that don't involve opt-in consent.

Consent models for administrative data linkage – COTS2020 example

- **Opt-in** consent for health linkages for infant and parents/adult informants – consent rates are high for primary caregiver and infant, collected during face-to-face interviewer visit – (lower for non-primary caregiver, via online survey)

  – opt-in approach taken as likely to be required by data controllers
- **Opt-out** consent for education linkages for infant and parents/adult informants
  - Information about linkage provided as part of study leaflet (and on website) and provided during the interview visit
  - Only one opt-out received so far – after 6797 interviews
  - Change from pilot where opt-out was presented more explicitly as part of interview – similar consent rates to an opt-in approach
  - Study funder (DfE) is data controller for education records

Consent models for administrative data linkage – are data holder requirements likely to constrain decision?

- Health records – likely to require opt-in consent?
- Education records – likely to require opt-in consent (as DfE not funder for ELC-FS)?
- Social care records – likely to require opt-in consent?
- Economic records – likely to require opt-in consent?

- As sampling frame is health records, opt-out consent may be more acceptable for this linkage?

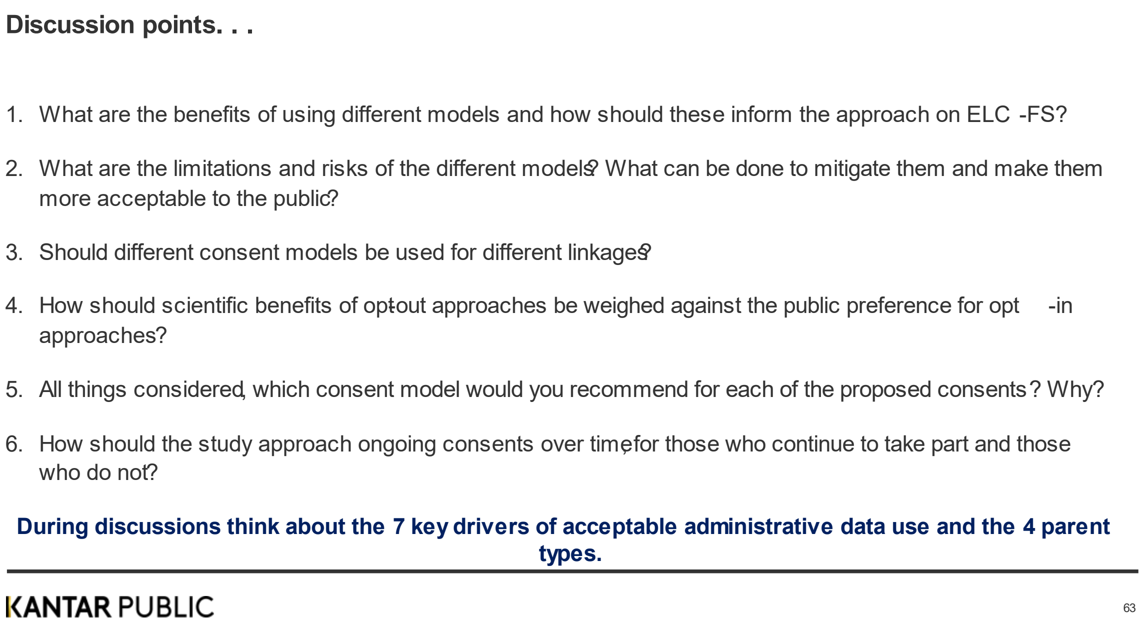
CENTRE FOR
LONGITUDINAL
STUDIES

Ongoing consent over time

- Strong scientific value of data linkage for those who drop-out of the study after consent is given
- Propose to continue to link unless consent explicitly withdrawn or request to stop processing personal details.
- Should we actively inform (e.g. via letter/email) those who withdraw from study that this will continue? What about those lost to follow-up?
- Should we inform study members when consents are collected that linkage will continue unless they withdraw consents explicitly, including if they withdraw from study/lose touch with study?
- How often (if at all) should we remind or re-collect consents from those who stay in study?

CENTRE FOR
LONGITUDINAL
STUDIES

3.10 Consent models for administrative data linkage discussion points



Discussion points. . .

1. What are the benefits of using different models and how should these inform the approach on ELC-FS?
2. What are the limitations and risks of the different models? What can be done to mitigate them and make them more acceptable to the public?
3. Should different consent models be used for different linkages?
4. How should scientific benefits of opt-out approaches be weighed against the public preference for opt-in approaches?
5. All things considered, which consent model would you recommend for each of the proposed consents? Why?
6. How should the study approach ongoing consents over time, for those who continue to take part and those who do not?

During discussions think about the 7 key drivers of acceptable administrative data use and the 4 parent types.