

# Collecting Multiple Data Linkage Consents in a Mixed-mode Survey: Evidence from a large-scale longitudinal study in the UK

## Survey Methods: Insights from the Field

**Marie Thornby, formerly UCL Institute of Education, UK**

**Lisa Calderwood, UCL Institute of Education, UK**

**Mehul Kotecha, NatCen Social Research, UK**

**Kelsey Beninger, Kantar Public, formerly NatCen Social Research, UK**

**Alessandra Gaia, City, University of London, formerly UCL Institute of Education, UK**

**How to cite this article :** Thornby, M., Calderwood L., Kotecha M., Beninger K. & Gaia A. (2018). Collecting Multiple Data Linkage Consents in a Mixed-mode Survey: Evidence from a large-scale longitudinal study in the UK Survey Methods: Insights from the Field. Retrieved from <https://surveyinsights.org/?p=9734>

**DOI :** 10.13094/SMIF-2018-00004

**Copyright :** © the authors 2018. This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0)

**Abstract :** Linking survey responses with administrative data is a promising practice to increase the range of research questions to be explored, at a limited interview burden, both for respondents and interviewers. We describe the protocol for asking consent to data linkage on nine different sources in a large-scale nationally representative longitudinal survey of young adults in England: the Next Steps Age 25 Survey. We present empirical evidence on consent to data linkage from qualitative interviews, a pilot study, and the mainstage survey. To the best of our knowledge, this is the first study that discusses the practicalities of implementing a data linkage protocol asking consent both retrospectively and prospectively, on multiple domains, and in the context of a mixed-mode survey.

## Introduction

Data linkage is a promising practice. It allows researchers to enhance survey data with detailed information at a low survey cost and interviewer and respondent burden. In some contexts, data can be linked both retrospectively and prospectively – adding information also for cohort members that have not participated to previous survey waves, or that may attrite in the future.

Besides its potential benefits, data linkage presents methodological and practical challenges. In several countries, consent needs to be asked to respondents before linking their records to survey responses (Sakshaug *et al.* 2017). A substantial proportion of sample members may not consent to data linkage (Sakshaug and Kreuter, 2012) and consenters may differ from non-consenters on key characteristics, leading to consent bias (Al Baghal, Knies and Burton, 2014).

To tackle these challenges, the methodological literature has mainly focused: i. on the respondents' and

interviewers' characteristics associated with consent; ii. on how the interviewer behaviour, the interviewer-respondents' rapport, the interviewers' attitudes toward sharing personal information, influence the likelihood of obtaining consent; iii. on consent bias, and iv. on which wording and positioning of consent questions maximises consent rates. Recent reviews on these topics are presented elsewhere – e.g. Al Baghal and Burton (2016), Al Baghal, Knies and Burton (2014), Korbmacher and Schroeder (2013), Sakshaug and Kreuter (2012) Sala, Knies and Burton (2014).

Little empirical evidence is available on best practices to implement data linkage protocols and on designing data linkage accompanying materials. This lack of knowledge is particularly problematic, since new challenges are arising in these areas.

The increase in adoption of less expensive self-completion modes of data collection (e.g. web), either alone or in conjunction with other modes, urges survey methodologists to understand how to optimise the collection of data linkage consent in self-completion modes. This task presents the challenge of simulating the interviewer persuasion in a self-completion context; not surprisingly, recent experimental research found lower consent rates in self-administered modes (web and mail) compared to interviewer administered modes – face-to-face (Burton, 2016; Sakshaug et al. 2017).

Also, collecting data linkage in mixed mode contexts entails logistical issues, since collecting signed consent forms is not practical in web and telephone surveys. However, there is little empirical evidence on the design of consent protocols in mixed-mode contexts.

Moreover, while many surveys attempt linking data from multiple records, and for future records, consent research has mainly focused on single consent requests and on existing records.

This study addresses these research gaps. We report our experience of developing a procedure to collect data linkage consents on Next Steps: a large scale longitudinal study in England of people born in 1989-90. We use data from: qualitative interviews, the Next Steps pilot study, the mainstage study, and interviewer debriefings.

## **The Next Steps study**

Next Steps is a longitudinal study of people born in 1989-90. Cohort members were originally recruited from schools in England in 2004, and interviewed annually between 2004 and 2010. In 2015/2016 the Next Steps Age 25 Survey was implemented. It is a multi-purpose survey, collecting information on family life, economic circumstances, education, employment, etc.

Next Steps Age 25 survey adopted a sequential mixed-mode design. Eligible sample members were firstly invited to participate in the survey by web; non-respondents in the web phase (who participated in the previous survey wave) were followed-up by a telephone interview. After the telephone fieldwork, all eligible sample members that had not yet taken part were assigned to face-to-face interview.

## **The data linkage preparatory work: qualitative interviews and pilot study**

In order to evaluate the data linkage consent materials and protocols, a qualitative study and a pilot study were implemented. The pilot was considered as an appropriate research design for collecting

feedback on: fieldwork design and protocols, fieldwork materials, the ease of questionnaire administration, implementation of the data linkage protocol and consent rates, as well as the survey overall. However, it was considered that the pilot was not the best tool for in-depth exploration of specific issues relating to data linkage; respondents' fatigue after a long interview would not have allowed in-depth exploration of specific topics, and interrupting the natural flow of the interview to include probes was not considered advisable. Thus, the findings from the pilot were supplemented with in-depth qualitative interviews aimed at exploring the practical and ethical issues around data linkage and to gain more detailed feedback on the proposed protocol and materials.

The sample for the qualitative study was composed of twenty individuals, aged 23-27 and recruited from the general public with the aim of including a diverse group of respondents in terms of gender, educational level, and working status.

Data were collected through face-to-face in-depth and cognitive interviews, lasting up to one hour and fifteen minutes. Interviews took place in participants' homes over a two week period in September 2014. Participants received an incentive of £25 for their cooperation.

In terms of study design, the interview process was conducted using a topic guide, which replicated the interview stages with respect to the data linkage consent requests. Specifically, participants were asked to review the information leaflet, use flashcards presenting the different data linkage consent questions, discuss the information leaflet, page by page, to express views on whether content was clear/unclear, whether any content was missing or leading to misunderstanding. Interviewers then explored views on framing the introductory text to the survey.

Participants were split in two groups:

- Group 1, which were shown an overview of the questionnaire topics at the start of the interview.
- Group 2, which were given the questionnaire overview after the consent questions and information leaflet had been explored with them.

Table 1 summarises the interview process.

**Table 1: Summary of interview process**

Stage of Interview	Overview of section topics
1. Introducing NatCen, the study, Next Steps, and ease participants into study	<ul style="list-style-type: none"> <li>• Purpose of study, what participation involves, reassurances</li> <li>• Next Steps content, target population, frequency of survey, survey modes</li> <li>• Establish survey mode preference of study participants</li> <li>• Deliver the information leaflet</li> </ul>
2. Exploring views and comprehension of consent questions for nine sources of records, using flashcards	<ul style="list-style-type: none"> <li>• Health</li> <li>• Economic</li> <li>• Education</li> <li>• Police and criminal justice</li> </ul>
3. Exploring views and comprehension of the information leaflet	<ul style="list-style-type: none"> <li>• Explore comprehension, concerns, suggestions, and changes in views on consent for each data source</li> </ul>
4. Exploring views and comprehension on impact of consent wording, on benefits of data linkage, and on combining education questions, using different flashcards	<ul style="list-style-type: none"> <li>• Explore preferences(if any) between positively and negatively worded consent introductions</li> <li>• Explore preferences from a list of benefits of data linkage, including reasons for preference</li> <li>• Explore views on having just one consent request for education to capture the three sources, including preferences</li> </ul>
5. Exploring views on how consent should be recorded and confirmed, using a copy of the confirmation of consent letter	<ul style="list-style-type: none"> <li>• Explore preferences for confirming consent in either email or post and explore views on adequacy of content in proposed letter.</li> </ul>
6. Exploring impact of knowing questionnaire content on informed consent, using a questionnaire overview	<ul style="list-style-type: none"> <li>• <u>Group 1</u> (those that received the survey overview): explore whether this impacted on their decision to consent</li> <li>• <u>Group 2</u> (those that did NOT receive the survey overview): explore whether this impacted on their decision to consent</li> </ul>
7. Views on the overall components of consent	<ul style="list-style-type: none"> <li>• <b>Component 1:</b> retrospective and prospective nature of consent</li> <li>• <b>Component 2:</b> consent at a click of a button</li> <li>• <b>Component 3:</b> end date for consent not being offered</li> </ul>
8. Close the interview	<ul style="list-style-type: none"> <li>• additional comments, reassurances about confidentiality and provide incentive</li> </ul>

Qualitative interviewers used the following materials: a topic guide; an overview of the questionnaire content; a set of flashcards each presenting the consent questions; a list of benefits associated with data linkage; two versions of question wording; a flashcard to assess views on combining education consent questions.

In the mainstage study the wording of consent questions was adapted to the different modes of data collection (web, telephone, and face-to-face) taking into account that web respondents read the questions themselves while telephone and face-to-face respondents have the questions read out to them by an interviewer. In the qualitative interviews, the web versions of the consent questions were used throughout. The pilot study allowed collection of feedback from interviewers, from participants in a post-

interview questionnaire, as well as from a small number of participants who directly contacted the office.

The data linkage section of the pilot study was aimed at answering the following research questions: was it useful and appropriate to send a detailed leaflet about data linkage as part of the advance mailing? Could informed consent be effectively gained (in terms of consent levels and acceptability from respondents)? Was gaining consent without paper forms feasible and acceptable? And was it feasible and acceptable to send post-interview confirmation of consents by email or letter? Were there any specific challenges of implementing data linkage consents in different modes – web, telephone, and face-to-face?

The pilot study took place in October and November 2014; 120 participants aged 23-27 were recruited from the general public in three areas of England with a quota sampling approach taken in order to include a diverse group of respondents in terms of gender, presence of children, cohabitation and employment status (as well as ethnicity in London). The number of participants who completed the data linkage section was 89 (of the 96 fully productive interviews). Respondents were given a £20 incentive for participation. Participants were randomly allocated to complete the survey in different survey modes with 35 participants taking part on the web, 33 by telephone, and 28 face-to-face.

## **The protocol for asking consent to data linkage**

In the mainstage Next Steps survey, cohort members were asked for consent to link their survey data with nine separate administrative data records, covering multiple domains (i.e. education, economics, health, and criminal justice), and held by several government departments and non-governmental bodies (Table 2).

### **Table 2: Data holder institutions and administrative records**

<b>Data holder institution</b>	<b>Administrative data record</b>
Student Loans Company	Amount taken out in loans and institution attended
Department for Education	School participation, attainment and pupil characteristics
Department for Business Innovation and Skills (Now known as the 'Department for Business, Energy and Industrial Strategy)	Participation in further education and attainment
Higher Education Statistics Agency	University participation and attainment
Universities and College Admissions Service	Higher education applications and offers
Department for Work and Pensions	Benefit and employment programs
Her Majesty's customs and Revenue	Employment, earnings, tax credits, occupational pensions and National Insurance Contributions
National Health Service	Health records including Primary Care data (visits to family doctor and other health professionals) and Hospital Episode Statistics (admissions and hospital attendance)
Ministry of Justice	Arrests, cautions and sentences

### Consent at the "click of a button"

The protocol varied by mode of data collection. Web respondents recorded their consent at the "click of a button", on a page within the web questionnaire. Consent was provided verbally in the telephone and face-to-face interviews.

In all modes respondents were not required to provide signed consent, for three main reasons: i. a higher response burden (since respondents in the telephone and web fieldwork would need to send to the office signed consent forms), ii. a negative impact on consent rates (since some consenting respondents may fail to send back the signed consent forms), and iii. an increase in survey costs (associated with dispatching, chasing, receiving and processing paper forms).

Most participants to the qualitative work had no concerns about the absence of signed consent; only in rare circumstances respondents expressed concerns that could result in the decision to not provide consent, unless a written signature was collected.

### The data linkage leaflets

Before the survey, respondents received an advance letter - mentioning the data linkage questions and signposting to further information - and a data linkage information leaflet providing information on the linkages being sought, their purpose, the linkage process, how linkage has been used on other studies, the voluntary nature of consent, and ways to revoke consent (see Figure 1).

**Figure 1: extracts from the data linkage leaflet**

1.

## ADDING OTHER INFORMATION

Government departments and agencies routinely collect administrative information about all of us to help plan and provide the services we need.

We'd like to add some of this information about you held by government departments and agencies to the information we collect about you during the study on an ongoing basis.

The information we would like to add is kept in your health, education, work and benefits records, as well as any police and criminal justice records you may have.



Watch our video to find out more:

[www.nextstepsstudy.org.uk](http://www.nextstepsstudy.org.uk)

## WHY WE WANT TO DO THIS

Adding extra details from administrative data opens up new possibilities for researchers from universities, charities and within government who all use the Next Steps data to understand the experiences of your generation and to make the services you use and the places you live better.

We learn a lot about your lives from the questions we ask in the surveys, but adding extra information from administrative records helps us to build a more complete picture of what life is really like for you.

It also means we can make the Next Steps data as valuable and accurate as possible, as it allows us to fill in the blanks for any details you may not know or remember (such as the date of a hospital visit) and to avoid asking you for other details during the survey.



## WE NEED YOUR PERMISSION TO DO THIS

You can agree for us to add information from all of the records we ask about, from just some of the records or to add nothing at all – **it's your choice.**

Whatever decision you make, we would like you to take part in the survey.

7.

## INFORMATION WE WOULD LIKE TO ADD

### HEALTH

#### WHICH RECORDS?

The National Health Service (NHS) routinely updates medical and health-related records for all patients accessing NHS health services throughout their lives.

This information is centrally managed by the Health and Social Care Information Centre. Cosmetic and other procedures provided through private health care are not included in these records.

#### WHAT DO THEY INCLUDE?

National Health Service records include:

- use of NHS health services; such as visits to the doctor, nurse or midwife, hospital attendance or admission and the dates and waiting times of these visits
- health diagnoses or conditions
- medicines, surgical procedures or other treatments received
- NHS number

#### WHY ADD THIS INFORMATION?

This study has huge potential to advance our understanding of health and illness and to change the way that conditions are screened, managed and treated.

This information can help researchers understand:

- what factors prevent or contribute to poor health
- how health conditions can be treated or managed effectively
- how early life experiences affect health later in life

13.

## WHEN DOES YOUR PERMISSION EXPIRE?

We'd like to add information relating to your past, present and future circumstances to the information that we collect as part of the study.

We have not put an end date on the permissions that you give as we do not know exactly when we will receive or add the information. So any permission you give will remain valid and we will collect these records on an ongoing basis unless you contact us to withdraw your permission.

You can change your mind about adding information from these records or withdraw any of your permissions to add information from these records at any time, without giving us any reason.



6.

## SOME ASSURANCES

We are not asking to access information that relates to your parents, partners, siblings or children.



Government departments and agencies will only receive the details they need to establish an accurate match to your records, nothing more.



Your decision about whether or not to agree to add data from your records will not affect your benefits, tax position or employment, your health treatment or any health insurance.



Your details can not be used to look at whether you are paying the right amount of tax or loan repayments, claiming the right type of benefits or by the police to find things out about you that they do not know.



## WATCH OUR VIDEO TO FIND OUT MORE



[www.nextstepsstudy.org.uk](http://www.nextstepsstudy.org.uk)



FIND OUT MORE AT...  
[nextstepsstudy.org.uk](http://nextstepsstudy.org.uk)

Based on evidence from the pilot and the qualitative study, we advise survey practitioners to highlight the



voluntary nature of linkage, include reassurances on data security, stress that non-consenters can still participate in the survey, and highlight the prospective nature of the linkages.

Consistently with the literature, we suggest to keep the leaflet short and concise. Our research found that some participants only “skim read” and then ask the interviewer general questions about the procedure.

Also, we advise to avoid wording that may be unclear or ambiguous, to provide definitions for unfamiliar expressions, to include examples, wording the leaflet as participant centred, and visualising the process using graphics and diagrams. Some respondents interpreted the term “withdrawal” as withdrawal from the whole survey (instead of withdrawal consent). Participants found confusing the use of “administrative records”, “administrative data”, and “records” as synonymous. Also, it was suggested to include the full department names instead of their acronyms.

Given that the advance mailings may not arrive to all participants, as, for example, some may have moved, we advise equipping face-to-face interviewers with spare leaflets, and instruct telephone interviewers to direct participants to leaflets on the survey website.

### *The data linkage protocol in a mixed-mode design*

The adoption of different protocols by mode of data collection influences consent rates; consistently with experimental evidence (Burton, 2016; Sakshaug *et al.*, 2017) we expect self-completion modes (web) to lead to lower consent rates than interviewer assisted modes (face-to-face and telephone), where an interviewer can attempt to persuade the respondent and the respondent has the chance to ask questions/clarifications.

Telephone and face-to-face interviewers received extensive training on data linkage (e.g. thorough simulation exercises and detailed project instructions). Additionally, interviewers were asked to familiarise themselves with the data linkage leaflet. Moreover, interviewers could use the help screens embedded in CASI to gather further reference information; also, they could refer to a laminated ‘Data linkage FAQs’ sheet.

In the web questionnaire, several mitigation strategies were put in place to simulate the role of the interviewer – e.g., a video about data linkage addressed to participants.

The web instrument allowed the adoption of web-specific features that could increase respondents’ understanding and that were inapplicable in other modes – e.g. hyperlinks to the data holders’ websites.

Figure 2 shows the first page in the CAWI data linkage section; it includes the explanation of data linkage, an embedded video, which overviews the procedure, and two hyperlinks, which opened pop-up windows (Figure 3).

### **Figure 2: The introduction to the data linkage page**

We'd like your agreement to add information about you from administrative records, held by a number of government departments and agencies, to the information we collect throughout the study.

Adding details from these existing records will greatly enhance the value of the information you give us, as it will help us to build a more detailed picture of what life is like for your generation.

You can choose which permissions to give at the next few questions and there will be an opportunity to review and confirm your choices at the end.

More information about this is provided:

- in the video below,
- at the help links in the next few questions,
- in the yellow booklet "Adding other information about you",
- and [on our website](#).

You can also contact our [study team](#) free of charge.



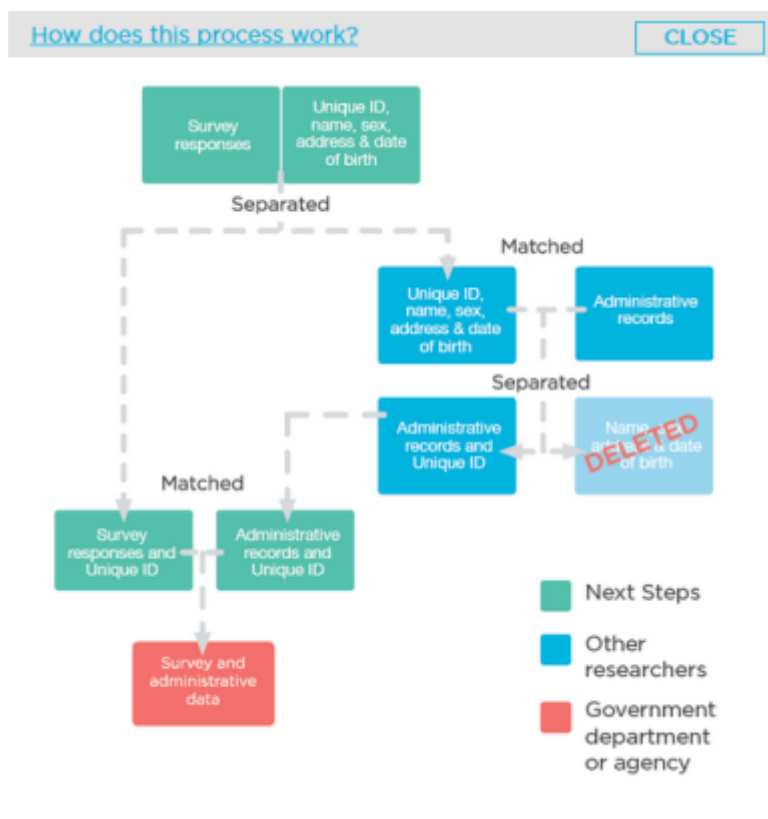
- [Why do you want to add other data about me?](#)
- [How does this process work?](#)

<< BACK NEXT >>

Your progress
Relationships
Your home
What you do
Your finances
Education
Health
Identity
Wellbeing
<b>Adding data</b>
Wrapping Up

aA

**Figure 3: Pop-up windows embedded in the web questionnaire**





### Positive and negative framing

Two different wordings were tested on the introduction to the data linkage questions. One wording was framed positively (i.e. “The information you have already given us will be more useful if information about you can be added from these other records”) and one negatively (“The information you have already given us will be less useful if information about you cannot be added from these other records”).

Participants to the qualitative study were asked to elicit which of the two versions they favoured. These wordings were not further tested in the pilot study.

The overwhelming majority of participants to the qualitative study preferred the positively worded version; it was perceived that this acknowledged better participants’ contribution, it avoided a sense of moral obligation that participants may feel in the negatively worded version, and it was overall felt as more welcoming and inviting.

### The data linkage questions

The questions included the following content: a title, a consent question, and two answer options (Figure 4).

**Figure 4: Data linkage request page for health records**

Health records: PO of the yellow [book list](#)

- ④ [Which records would Next Steps like to add?](#)
- ④ [What do these records include?](#)
- ④ [Why is it helpful to add this information?](#)

Records produced by the [National Health Service \(NHS\)](#) and maintained by [Health and Social Care Information Centre](#) include information about use of health services, health conditions and treatments provided.

I give permission for my name, address, sex and date of birth to be passed to [Health and Social Care Information Centre](#) so that my NHS health records can be identified and sent to the Next Steps research team to be added to my study information.

Yes	No
<input type="radio"/>	<input type="radio"/>

**Your progress**

- Relationships
- Your home
- What you do
- Your finances
- Education
- Health
- Identity
- Wellbeing
- Adding data
- Wrapping Up

aA aA

<< BACK

NEXT >>

The web implementation of the data linkage section allowed for the inclusion of several hyperlinks with additional information. For example, in the consent question displayed in Figure 3, the “National Health Service (NHS)” hyperlink opens the website to the National Health Service and the hyperlink: “Which records would Next Steps like to add?” “open a pop-up window with additional information.

This step was not implemented in the mainstage. At the end of the section respondents (in web) and interviewers (in telephone and face-to-face) were presented with a screen summarising the permissions given (see Figure 5).

**Figure 5: Confirmation page in CATI and CAPI**

Change

INTERVIEWER: READ OUT CONSENTS GIVEN AND ASK FOR CONFIRMATION AND CHANGE AS APPROPRIATE

You have made the following choices...

	Yes	No	
National Health Service (NHS)	(1)	(2)	1
Department for Education (DfE), Department for Business Innovation and Skills (BIS), and Higher Education Statistics Agency (HESA)	(1)	(2)	1
Universities and Colleges Admission Service (UCAS)	(1)	(2)	1
Student Loans Company (SLC)	(1)	(2)	1
Her Majesty's Revenue and Customs (HMRC)	(1)	(2)	2
Department of Work and Pensions (DWP)	(1)	(2)	1
Provision of your National Insurance Number to the Department of Work and Pensions, Her Majesty's Revenue and Customs, and the Student Loans Company.	(1)	(2)	1
Ministry of Justice (MOJ)	(1)	(2)	2

Confirm

[1] Please confirm that you have been provided with information about adding data from administrative records and have had the opportunity to ask questions. You understand that the permissions you have given will remain valid and information from these records will be collected on an ongoing basis unless you contact the Next Steps research team to ask them to stop. You also understand that you can change or withdraw any of your permissions at any time, without giving any reason.

<< BACK      LAST      NEXT >>

S8MAINV013 Web/CAPI 2057 3:30

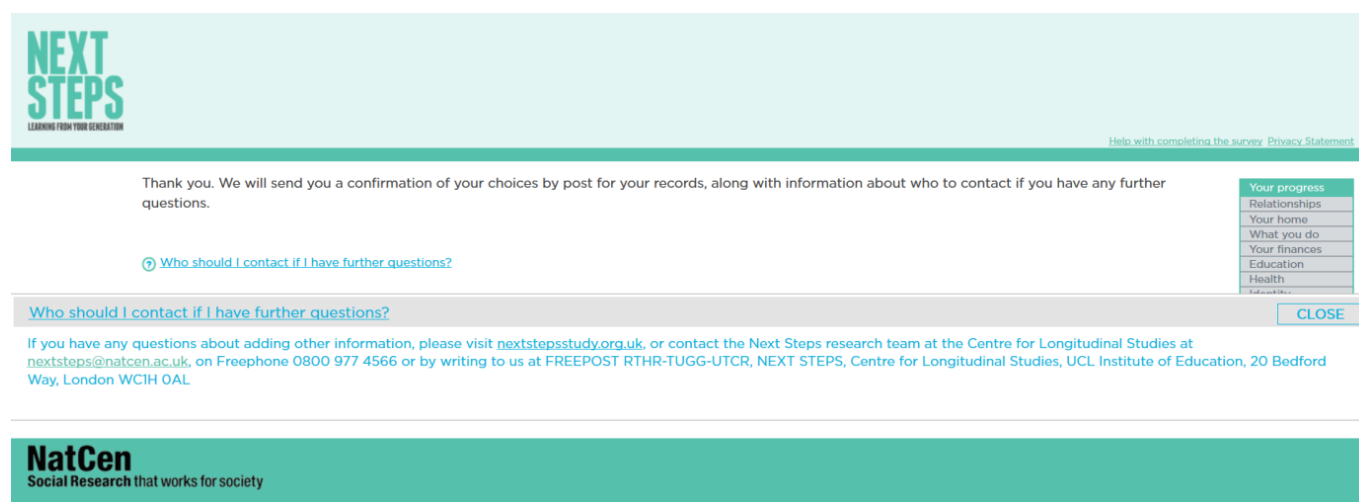
aA aA

The respondent has an opportunity to confirm the consent provided, and to change any consent given. In the face-to-face and telephone interview, the interviewer read out each listed record type and the response; if needed, the interviewer changes the responses provided in this same screen, without going back to the original question. Similarly, in the web interview respondents were asked to review and confirm the consent provided.

After reviewing all consent choices, the respondent is asked to give confirmation, ticking a confirmation box in the web survey or accepting a confirmation statement in the face-to-face and telephone interview.

In the web survey, an additional page was displayed to the respondent stating that written confirmation would be sent by post, and with an additional hyperlink with contact details for further information (see Figure 6).

**Figure 6: Thank you page**



### *Hard copy consent confirmation and intra-wave mailing*

Written confirmation of the consent choices was sent to respondents in a “Thank you” mailing, which also included the incentive and a change of details card for future survey waves. Respondents were provided with information on how to withdraw their consent(s), and study contact details were supplied so that participants could get in touch with further questions/concerns.

A post-survey confirmation of consent in hard copy worked well at the pilot, and the research team felt it was important from an ethical perspective to give respondents another chance to check that their consents have been recorded accurately and to keep for future reference.

Some participants preferred a paper record (easier to keep and more formal); others preferred an email confirmation, on the grounds of environmental concerns and on a perceived easiness to withdraw consent, if an unsubscribe hyperlink is included.

Participants expressed the desire to receive an intra-wave mailing or a “findings hand-out” describing how linked data contributed to research.

## **The acceptability of the consent process**

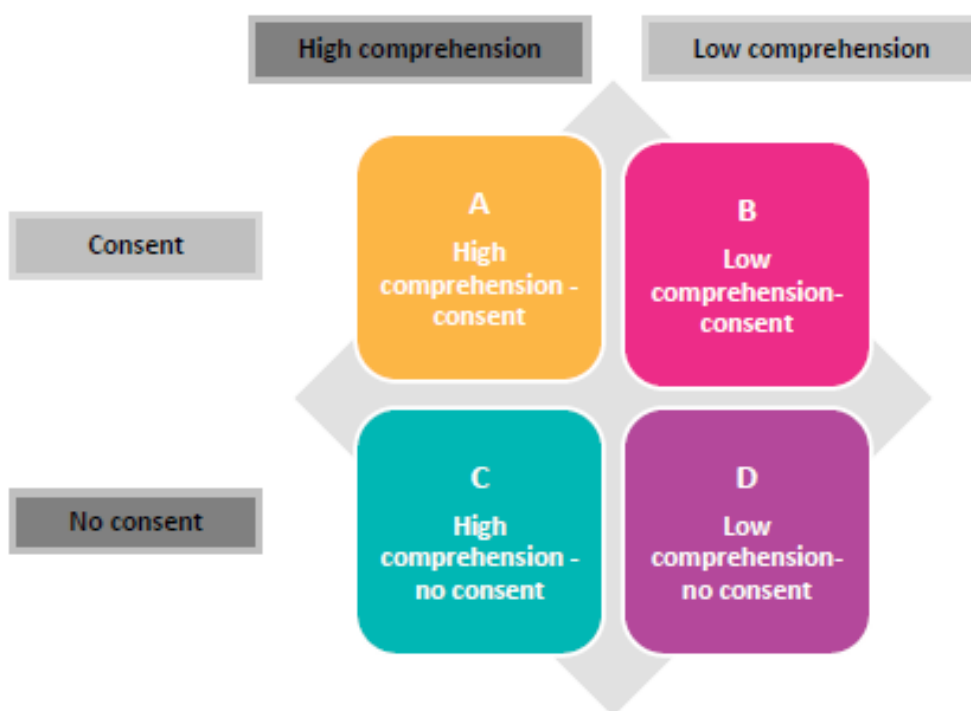
Evidence from the qualitative interviews showed that the protocol was considered acceptable. Specifically, participants considered that the protocol appropriate to the complexity and sensitivity of the data linkage request, and not excessively burdensome. Furthermore, participants understood the necessity of asking nine different consent questions.

However, the participants' reaction to the consent request varied. In the pilot study, while some respondents did not have major (if any) concerns, others expressed strong negative reactions about the level of information collected, with a "big brother-ish" fear of being controlled, especially by the police and government bodies collecting taxes, and supplying pensions and benefits. As one participant to the telephone pilot study stated: "[I d]on't mind doing study but not prepared to link data as that's scary" (Quotation reported in the interviewer feedback form).

In some circumstances, respondents did not have sufficient trust to consent. As one telephone participant (in the pilot study) stated: "I don't know if I can trust who you are. Really I only have your say so, too many things happen these days." (Quotations reported in the interviewer feedback form).

While the consent procedure was considered easy, the comprehension of what was being asked was limited. Participants can be clustered in four groups according to their comprehension and willingness to provide consent (Figure 7).

**Figure 7: Typology of participants based on their comprehension and willingness to give consent**



Evidence from the qualitative interviews showed that participants could belong to different groups across different consent questions; the level of comprehension often changed during the qualitative interview, with participants moving from a lower to a higher comprehension group.

An improvement in comprehension was often associated with a higher likelihood to provide consent, driven by an increased understanding of the benefits of data linkage for society and for the participants' survey experience.

We identified six factors underpinning comprehension and consent (Figure 8).

**Figure 8: Factors underpinning comprehension and consent**



Overall, asking consent to data linkage on multiple domains leads to an efficiency gain, as participants capitalise from each question and the comprehension of the request requires less effort for each additional question.

Participants were more likely to give consent if they have already given consent to a request in the same domain, in order to be consistent with their previous choice or because they (mis)believed that consent to a current question presupposed consent to subsequent questions.

While participants became gradually aware of the volume of information that they were asked to share and that are held on them by various organisations, this awareness did not necessarily impact negatively on consent.

#### Participants' understanding of the data linkage benefits

Participants in the qualitative interviews were presented with eight different benefits to data linkage. Understanding which of these benefits are the most salient is important: these may be used as leverages to increase consent. Table 3 presents a summary of the proposed benefits and the participants' reactions.

**Table 3: Benefits of data linkage and participants' reactions**

Benefit description	Reasons liked	Reasons disliked
Improves the accuracy and the value (adding details you might not know or be able to remember)	includes two benefits within one sentence	
builds a more complete picture	simple reminds respondents that they already shared information	
enhances the information by adding more depth	stresses the 'added value' of linkage	
saves time (in future interviews, may be able to ask fewer questions)	highlights the benefit for the individual participant	
easily completes information already collected (not having to ask you to consult documents)		
makes better use of existing information (identifying population trends, understanding specific needs of this generation)	clarifies the meaning of "maximising existing data"	poor understanding of term "population trends"
adds value by helping policy makers plan and improve services	the word "value" is persuasive and indicates the participant is helpful	not relevant to individual participants poor understanding of term "policy maker".
enhances the core information already given to us (opening up research opportunities)	demonstrates usefulness	not relevant to individual participants

*The lifespan of consent*

The qualitative and pilot study showed that linking survey data with past individual record was understood and considered acceptable. Conversely, participants didn't initially consider the possibility of their survey answers to be linked to future records. For example, one participant stated: "It wouldn't change my opinion on that, I would still say yes, but I was just thinking up to the present" (Male, medium education, in work).



They expressed a preference to limit their consent in the future and claimed that an annual reminder about their on-going consent would be beneficial, especially if there are gaps in running the survey.

### *The sensitivity of the data linkage requests*

Data linkage may be influenced by the sensitivity of the consent request. As in survey questions in general, whether a consent request is considered sensitive or not depends on whether the sample member engages in any socially undesirable behaviour or has a socially undesirable characteristic associated with the request.

Participants anticipated that study members may have concerns about sharing their records if they have had a health condition or treatment that they are not willing to share with others (e.g. mental health problems).

Participants in the qualitative study did not consider all consent requests as being sensitive to the same degree. For example, within the educational area, the only question that raised concerns was the consent to link data from the Student Loan Company; since this institution does not only deal with schooling but also with financial information.

## **Consent rates from the pilot study**

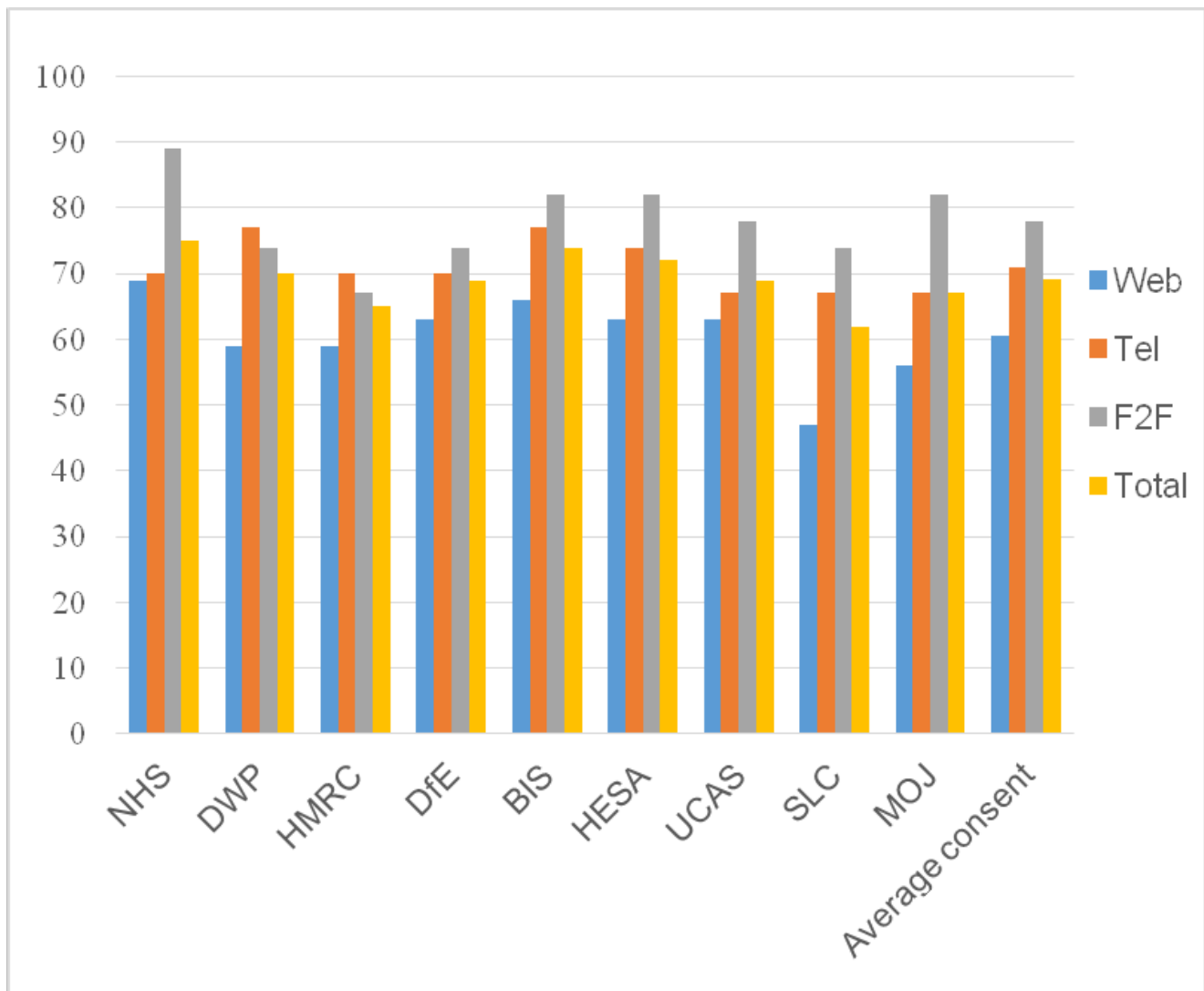
In the pilot study, depending on the mode of data collection, and on the consent type, consent rates range from 47% to 89% (graph 1).

Even though participants were randomly allocated to different survey modes, so that selection into mode does not undermine the mode comparison, given the small sample size, it is not possible to derive definite findings on mode effects.

Nevertheless, the evidence of a higher consent rates in face-to-face (78%), followed by telephone (71%) and finally by web (61%) is consistent with the hypothesis of higher consent rates in modes that allow for an interviewer persuasion, suggesting that with a larger sample size we might have been able to conclude that consent varies by mode of data collection.

Mode differences emerged in the feedback from interviewers in the pilot study. Face-to-face interviewers reported more positive feedback than telephone interviewers. In the telephone mode, some participants were hesitant and reported that this was an excessive and too intrusive request; despite the reassurances of data security and the voluntary nature of consent, the request put some participants off taking part altogether. Conversely, face-to-face interviewers stated that respondents had read the leaflet, and had no concerns in answering the question, even if some did not give consent to all the consent requests.

### **Graph 1: Consent rates by mode and consent reques**



## Consent rates from the mainstage of the study

In the mainstage of the study, the number of participants who completed the data linkage section was 7,502 (of the 7,707 productive interviews). Depending on the mode of data collection, and on the consent type, consent rates range from 44% to 90% (graph 2).

In the mainstage study, participants were not randomly allocated to different survey modes - thus, selection into mode means that differences in consent rates by mode may be driven by the characteristics of those who chose to participate in that mode.

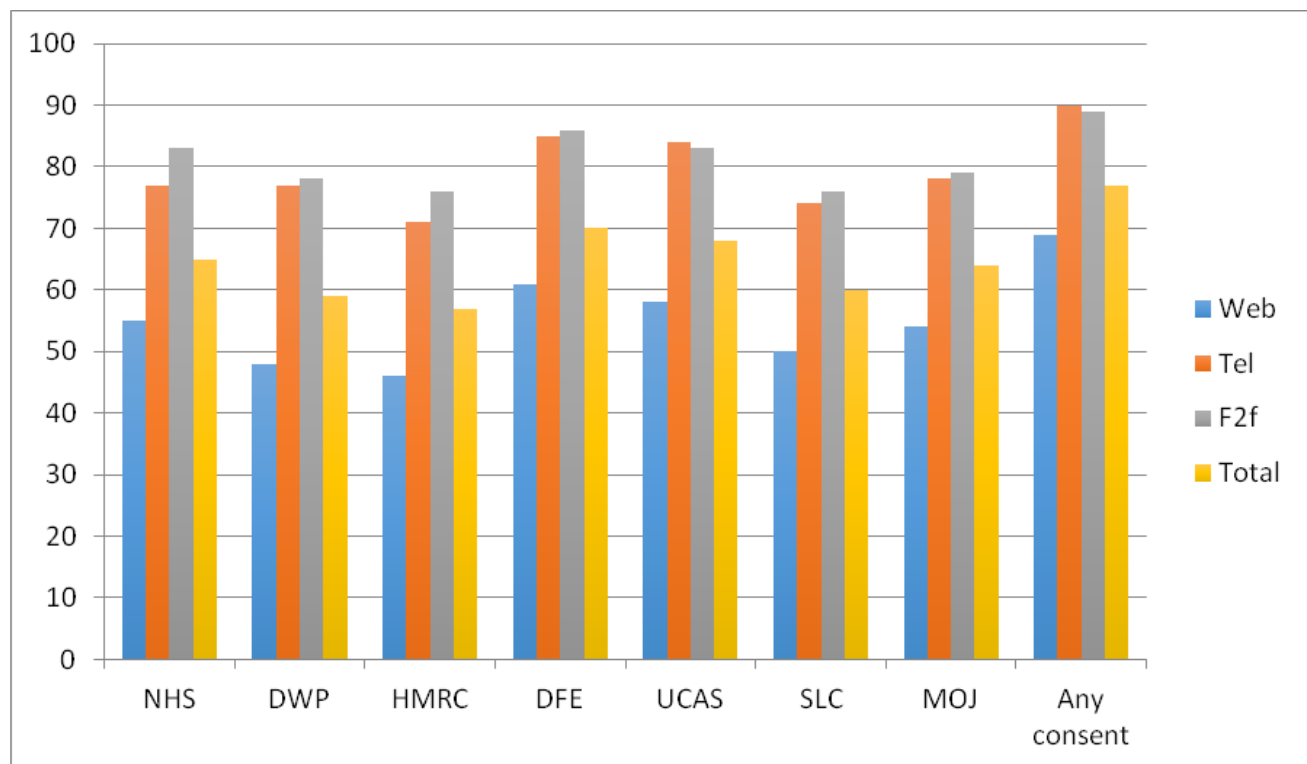
Nevertheless, the evidence of much higher consent rates in face-to-face (89%) and telephone (90%) than by web (69%) is consistent with the hypothesis of higher consent rates in modes that allow for an interviewer persuasion, and with findings from the pilot. For all modes, consent rates are higher for the mainstage of the study than in the pilot study.

Overall, despite the extensive efforts to incorporate features designed to maximise consent in the web mode, the consent rates for those completing the questionnaire on the web remained much lower than in face-to-face and telephone.

Looking at the overall response rate per consent type, the lowest consent rates were those related to

economic records (DWP, HMRC) and the student loans company (SLC).

**Graph 2: Consent rates by mode and consent request**



## Discussion

In this paper we investigate the challenges of asking consent to data linkage in a mixed-mode context; we analyse whether it is feasible to ask consent to multiple domains simultaneously and on future records; and we discuss the best practices in designing materials to promote consent.

Overall, respondents considered it acceptable to give consent without signing forms. As opposed to

signed consent, this protocol minimises respondent burden and survey cost.

Experimental evidence from the pilot study seems to suggest higher consent rates in face-to-face interview, followed by telephone and finally by web; although the small sample size of the experiment doesn't allow to derive conclusive evidence.

The descriptive analysis of the consent rates in the mainstage Next Steps Age 25 survey shows that consent rates were much lower in web than in telephone and face-to-face. This provides indicative evidence that the mitigating steps we implemented to simulate interviewer role in the web survey (e.g. a video describing the procedure, and hyperlinks to the data holder institutions) were insufficient. We could recommend that other studies implementing data linkage consents in a web survey consider further steps such as telephone call back for non-consenters. Having said that, as participants self-selected into mode, the descriptive analysis does not enable robust conclusions about mode effects on data linkage consents.

Qualitative interviews showed that, overall, asking consent to link records from multiple domains is considered acceptable, and separate questions are preferred to a unique "catch all" item; we also find evidence of an "incremental effect", with respondents capitalising from previous questions, leading to a lower cognitive effort, at each subsequent request.

Consent rates varied by domain. Data linkage in the domain of economic records and records held by the Student Loan Company obtained the lowest levels of consent. Further research may compare the response propensities on different domains by socio-demographic group.

Regarding the timespan of consent, we advise survey practitioners to carefully word prospective consent requests, as cohort members may find it complicated to understand and welcome linkage with future records.

One limitation of this study is that this evidence is limited to a specific age cohort; further research may replicate these findings on different age groups and/or in different countries.

## References

1. Al Baghal, T., Knies, G., & Burton, J. (2014). Linking administrative records to surveys: Differences in the correlates to consent decisions. *Institute for Social and Economic Research University of Essex. Understanding Society Working Paper Series*, (2014-09). Colchester: University of Essex.
2. Al Baghal, T. & Burton, J. (2016). "Does interviewers' attitudes towards sharing personal information affect the consent rate they achieve?" Background paper for the 5th Panel Survey Methods Workshop 2016, Berlin.
3. Burton, J. (2016). "Results for Web/Face-to-Face Linkage Consent Questions in the Innovation Panel." Presented at the Mixing Modes and Measurement Methods in Longitudinal Studies Workshop. London: CLOSER.
4. Jenkins, S. P., Cappellari, L., Lynn, P., Jäckle, A., & Sala, E. (2006). Patterns of Consent: Evidence from a General Household Survey. *Journal of the Royal Statistical Society (Series A)* 169 (4), 701-722.
5. Korbmacher, J. M., & Schroeder, M. (2013). Consent when linking survey data with administrative records: the role of the interviewer. *Survey Research Methods* 7 (2), 115-131.
6. Sakshaug, J. W., Hülle, S., Schmucker, A., & Liebig, S. (2017). Exploring the Effects of Interviewer-and Self-Administered Survey Modes on Record Linkage Consent Rates and Bias. In *Survey Research Methods* 11 (2), 171-188.

7. Sakshaug, J. W., & Kreuter, F. (2012) Assessing the magnitude of non-consent biases in linked survey and administrative data. *Survey Research Methods*, 6 (2), 113-122.
8. Sala, E., Knies, G., & Burton, J. (2014). Propensity to consent to data linkage: experimental evidence on the role of three survey design features in a UK longitudinal panel. *International Journal of Social Research Methodology*, 17 (5), 455-473.