Citation: Thornby, M., Calderwood, L., Kotecha, M., Beninger, K. & Gaia, A. (2017). Collecting multiple data linkage consents in a mixed mode survey: evidence and lessons learnt from next steps. London: Centre for Longitudinal Studies.

Institute of Education

UCL

Collecting Multiple Data Linkage Consents in a Mixed Mode Survey: Evidence and Lessons Learnt from Next Steps

By Marie Thornby, Lisa Calderwood, Mehul Kotecha, Kelsey Beninger and Alessandra Gaia

Centre for Longitudinal Studies
Working paper 2017/13

CENTRE FOR LONGITUDINAL STUDIES

E·S·R·C
ECONOMIC & SOCIAL RESEARCH COUNCIL

Contact the author
Lisa Calderwood
Principal Investigator
Centre for Longitudinal Studies
University College London
l.calderwood@ucl.ac.uk

This working paper was first published in November, 2017
by the Centre for Longitudinal Studies,
UCL Institute of Education
University College London
20 Bedford Way
London WC1H 0AL
www.cls.ioe.ac.uk

The Centre for Longitudinal Studies (CLS) is an Economic and Social Research Council (ESRC) Resource Centre based at the UCL Institution of Education (IOE). It manages four internationally-renowned cohort studies: the 1958 National Child Development Study, the 1970 British Cohort Study, Next Steps, and the Millennium Cohort Study. For more information, visit www.cls.ioe.ac.uk.

Contents

Acknowledgements

The Next Steps Age 25 Survey is funded by the Economic and Social Research Council (ESRC) and is run by the UCL Centre for Longitudinal Studies (CLS). It was previously funded and managed by the Department for Education and known as the Longitudinal Study of Young People in England (LSPYE).

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. This paper describes the protocol for asking consent to data linkage on nine different sources in a large-scale nationally representative survey of young adults in England: the Next Steps Age 25 Survey. Evidence from preparatory qualitative research, piloting and from the main stage of the study is presented. This research constitutes a novel contribution to the literature, discussing 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.

Keywords

Data linkage, longitudinal, Next Steps study, Mixed modes.

Non-technical summary

Data linkage is a technique to add other data to survey data which enhances its usefulness for research and policy. It consists of linking the survey responses with administrative data; these are, for example, the records held by government departments, such as the Department for Work and Pensions.

In order to implement data linkage, study members need to be informed about this procedure, and are free to give or refuse their consent for their administrative records to be linked to their survey data.

The procedure to request consent needs to be carefully planned and tested to ensure that respondents are fully informed about what is involved, and are able to take an informed decision about this.

This research describes the procedure that has been implemented to ask consent to data linkage in Next Steps, which is a study of people in England born in 1989-90, interviewed eight times so far at key ages during their lives.

In developing this procedure we had to consider several methodological challenges. Indeed, the study asked participants for consent to link their survey data to administrative records from many different sources, on several different domains (health, economic, education, and criminal justice).

Consent was requested both for the past as well as for the future, meaning that administrative records will be added to the respondents' survey data with no end date, unless the consent is withdrawn by the respondent. It was therefore important that study members understood the retrospective and prospective nature of their consent.

Also, some study members were interviewed face-to-face, others by telephone, and others completed the survey on the web; thus, it was crucial that all respondents could access relevant information on data linkage and could give/refuse consent, irrespectively of how they filled in the survey.

This paper investigates the challenges and the proposed solutions. This evidence may help survey researchers and practitioners in planning a survey which includes data linkage requests, especially in a mixed-mode context.

Introduction

Data linkage is a promising practice, as it allows researchers to enhance survey data at a reduced survey cost and low interviewer and respondent burden (Sakshaug and Kreuter, 2012; Sala, Burton and Knies 2012; Korbmacher and Schroeder, 2013; Sakshaug, Tutz and Kreuyter, 2013; Sala, Knies and Burton, 2015); also, linked data are a useful source to inform the methodological literature on survey error (Kreuter, Muller and Trappmann, 2010; Olson, 2006; Sakshaug, Couper, Ofstedal and Weir, 2012).

There are many benefits to data linkage. The level of detail that can be acquired from administrative records may be difficult to gather through survey questions. Additionally, the prospective nature of consent to data linkage allows adding data from future records: this information could otherwise be collected only with survey follow-ups. Also, while current and past information can be collected by including additional survey questions, in order to have similar levels of detail of that acquired with data linkage, the respondents' survey burden needs to be substantially increased, and the quality of data depends on the accuracy with which the respondent recalls the information.

However, besides its potential benefits, data linkage presents methodological concerns and practical challenges. The literature has shown that a substantial proportion of sample members may not agree to consent to data linkage, with the level of consent varying widely, from 19.0% as reported by McCarthy *et al.* (1999) to 96.5% as recorder by Rhoades and Fung (2004)[1]. Thus, it is particularly important to design consent requests and supporting materials in order to provide respondents with all the necessary tools to give (or refuse) informed consent.

The methodological literature on data linkage has focused on four main strands: firstly on the respondents', interviewers', and respondents' household characteristics associated with consent to data linkage; secondly, on how the interviewer behaviour, the interviewer-respondents' rapport, the interviewers' attitudes toward sharing personal information, influence the likelihood of obtaining data linkage consent. Thirdly, on the assessment of consent bias, and fourthly on which wording, framing, and

---

[1] For a review please see de Silva et al. 2012, Kho et al. 2009, and Sakshaug et al. 2012.

positioning of consent questions would maximise consent rates. Recent reviews on the state of the literature on each of these topics are presented elsewhere – please see Sala, Knies and Burton (2014), Al Baghal, Knies and Burton (2014), Korbmacher and Schroeder (2013), and Sakshaug, J. W., & Kreuter, F. (2012), Al Baghal and Burton (2016), Jenkins *et al.*, (2006), Korbmacher and Schroeder (2013).

To the best of our knowledge, there is little evidence on the practical implementation of asking consent to data linkage, particularly in a mixed mode context and in relation to asking consent to linkage to data from different domains. In this respect, we believe that this study would constitute a novel contribution to the literature.

This paper is aimed at survey practitioners willing to include data linkage consent in a survey; it reports 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.

More specifically the paper discusses best practice in asking consent to data linkage, focusing on three main aspects: asking consent to data linkage on multiple domains in one single survey; asking consent to data linkage in a mixed mode design, when web is one of the modes of data collection; and designing effective materials to promote consent.

In particular we will answer the following research questions: What are the challenges of asking consent to data linkage in a mixed modes context? Is it feasible to ask consent to multiple domains at the same time? How to best design materials to promote consent?

More specifically, based on the Next Steps experience, we will describe how the interviewer persuasion can be simulated in a mixed mode context, and how to manage the logistics of asking consent without the presence of an interviewer. Furthermore, we will describe the experience of collecting consent at the “click of a button”, without a signature. Also, we will describe which web specific features can be implemented to enhance the information provided to respondents (e.g. the use of hyperlinks).

Furthermore, in asking consent to data linkage in multiple domains, we will discuss participants’ evaluations of the topic sensitivity.

Finally, we will provide guidance on how to phrase and word the data linkage consent materials.

Next Steps: description of the study

Next Steps is a longitudinal study of people born in 1989-90. Cohort members were originally recruited from schools in England in 2004. They were interviewed annually between 2004 and 2010. There was then a gap in the study follow-up, and following a change in the management and funding of the study, in 2015/2016 the Next Steps Age 25 Survey was implemented. This is a follow-up study on the whole original sample of 16,000 pupils. It is a multi-purpose survey, collecting information on family life, economic circumstances, education, employment, health, and wellbeing.

Next Steps Age 25 survey adopts a sequential mixed mode design. Eligible sample members are firstly invited to participate to the survey by Web Self Interview (WSI);

followed by a telephone interview. After the telephone fieldwork period, all eligible sample members that did not yet take part in the survey are assigned to face-to-face interview.

The web option remains available throughout the whole survey fieldwork. During telephone and face-to-face phases, if participants are adamant in their preference for the web or in the case of an impending refusal they are invited by telephone and face-to-face interviewers to complete the survey by web. If this is the case, interviewers were instructed to follow-up the respondent in the same mode (telephone or face-to-face) when web completion is not observed within one week, with the aim of avoiding masked refusal through cohort members saying they will complete via web.

All sample members are entitled to an incentive for having completed the survey. This incentive was £20 for completion of the web survey within the first 3 weeks of fieldwork, and £10 incentive for completion of a fully productive interview at any other point in time and in any other mode.

Within the preparatory work undertaken for the Next Steps Age 25 survey, qualitative research on data linkage and a pilot study were undertaken. Both of these informed the data linkage protocol, and evidence from these two studies will be described in this paper along with the final consent rates achieved in the main stage.

The protocol for asking consent to data linkage in the Age 25 Next Steps survey

Cohort members are requested to give consent to link their survey data with nine separate administrative data records. These cover multiple domains, i.e. education, economics, health, and criminal justice. The records are held by several different government departments and non-governmental bodies. Table 1 below, shows the administrative records to be linked and the holder institutions.

Table 1: Data holder institutions and administrative records

Data holder institution	Administrative data record
Student Loans Company (SLC)	amount taken out in loans and institution attended
Department for Education – National Pupil Database	school participation and attainment and pupil characteristics
Department for Business Innovation and Skills[2] – Individualised Learner Records	information about participation in further education and attainment
Higher Education Statistics Agency (HESA)	university participation and attainment

---

[2] Now known as the ‘Department for Business, Energy and Industrial Strategy’

Universities and College Admissions Service (UCAS)	higher education applications and offers
Department for Work and Pensions (DWP)	benefit and employment programs
Her Majesty’s customs and Revenue (HMRC)	employment, earnings, tax credits, occupational pensions and National Insurance Contributions
National Health Service (NHS)	health records including Primary Care data covering visits to family doctor and other health professionals and Hospital Episode Statistics (HES) covering admissions and attendance at hospital
Police National Computer (PNC) held by the Ministry of Justice	records covering arrests, cautions and sentences

This was not the first occasion of data linkage in the life of the Next Steps cohort study. Indeed, survey data have been already linked to administrative records from the National Pupil Database (NPD).

However, data linkage in the Next Steps Age 25 survey presents specific challenges; firstly, the survey has a sequential mixed mode design. Not only, the mixed-mode nature of the survey implies specific consent protocols for the different modes; but also, the different protocols may lead to different consent propensities.

Secondly, the survey requests consent to data linkage from several different data sources, and on a wide timespan, as consent is requested both retrospectively as well as for the future, with no end date.

As it will be discussed in more detail in the subsequent paragraphs, several mitigation strategies were implemented to overcome these challenges. The web instrument was developed to attempt to simulate the interviewer persuasion and thus reduce the mode differences in the consent request. Also, the wording of the consent questions was changed, in order to clarify the wide timespan of consent.

Furthermore, the consent requests on multiple domains were carefully worded and evidence from the qualitative study, the pilot study and the main study show that asking consent to link multiple records from multiple domains is generally considered acceptable by respondents.

Finally, even though asking multiple consents to data linkage simultaneously is challenging as it increases interviewers’ and respondents’ burden, the evidence from the qualitative study suggested an “efficiency effect” of asking consent on multiple domains: respondents capitalise from previous questions and the cognitive effort of understanding subsequent consent questions is lower at each subsequent request. More details of this phenomenon are discussed in the section which summarises the results on the evidence from the qualitative study.

The data linkage protocol: consent at the click of a button

The protocol for asking consent to data linkage varies by mode of data collection. In all modes respondents were not required to provide signed consent. Cohort members filling in the web survey recorded their consent at the "click of a button", on a page within the web questionnaire. In the Computer Assisted Telephone Interview (CATI), consent is provided verbally over the telephone, and interviewers record it electronically in the Computer Assisted Interview instrument. In the face-to-face Computer Assisted Personal Interview (CAPI), consent is provided verbally and recorded electronically in the Computer Assisted Interview instrument by the interviewer.

Evidence from the qualitative study and from the pilot study signalled that this approach was acceptable. In a mixed-mode context, signed consent was not advisable, for three main reasons: a higher response burden, a negative impact on consent rates due to not all respondents returning the consent forms, and an increase in survey costs.

Firstly, in order to collect signed consent in the telephone and web survey cohort members would have needed to send to the office a signed paper form. This step would increase the time and effort necessary to participate in the study, and respondents may view this additional step as unnecessary as consent has been already expressed during the interview. In face-to-face the administration of paper forms would have been less burdensome to respondents than in other modes, but still more burdensome than collecting the consent verbally; thus, also in this mode, the click of a button procedure would also minimise respondent burden.

Secondly, it was perceived that some web and telephone respondents may not return the signed consent form, even if an effective reminder strategy would have been put in place. Respondents for whom a consent form would have not been received in office would have been classified as not consenting, even if they would have been happy to consent.

Finally, dispatching, chasing, receiving and processing paper forms, and reconciling the information provided in the CAI instrument and on the paper forms, would have required additional logistical resourcing, increasing the survey costs as well as the burden of the survey on the environment.

The data linkage protocol in a mixed mode design

The adoption of different protocols by mode of data collection may influence consent rates; face-to-face is expected to be the mode of data collection that allows for higher consent rate, followed by telephone interview and subsequently by web interview. Indeed, consent rates may be higher in modes where an interviewer can attempt to persuade the respondent and the respondent have the chance to ask questions/clarifications, compared with modes where there is no interviewer-respondent interaction.

However, it should be noticed that the empirical evidence from the Next Steps Age 25 survey cannot prove whether this theoretical argument holds. In fact, as allocation into mode is not random, it is not possible to disentangle whether potential differences across modes are mode effects or due to self-section into mode. Conversely, experimental evidence is provided in the Next Steps pilot study, where participants were randomly allocated to fill in the survey in different modes. While the sample size of this small experiment does not allow us to determine whether there are mode effects in consent rates, the results seem consistent with the hypothesis of a higher consent rate in face-to-face, followed by telephone and finally by web. Details of consent rates in the pilot study are discussed later in this paper.

Several materials and information sources on data linkage were made available to Next Steps cohort members. These were aimed at: guiding cohort members in deciding whether they want to consent to data linkage, inform cohort members that consent is voluntary, and that they can revoke consent at any time.

Before the survey, respondents received advance materials providing information on data linkage. These are: the survey advance letter and the data linkage information leaflet. The advance letter mentioned the data linkage questions in the questionnaire and signposts to further information.

The data linkage leaflet provided full information on each of the data linkages being sought, the purpose of linkages, the linkage process, how linkage has been used on other studies, the voluntary nature of consent and ways to revoke consent, and refer to the website, email and telephone contact details for further information and/or queries. Some extracts of the data linkage information leaflet are provided in the appendix[3]. The content and wording of the data linkage leaflet was developed during the qualitative interviews and during the Next Steps pilot study.

The support provided during the interview is specific to the mode of data collection. In interviewer-administered modes (telephone and face-to-face), respondents can ask further questions and clarifications to the interviewer. Interviewers have had detailed briefings on data linkage; indeed, a section of the project briefings was focused on data linkage, which included a group exercise in which interviewers had the opportunity to simulate a scenario in which a respondent would not agree to linkage. Furthermore, the project instructions included a section on data linkage with information on different consent types and procedures. In addition to this, interviewers have been asked to familiarise themselves with the data linkage information leaflet. Moreover, interviewers could use the help screens embedded in the computer administered questionnaire to gather further reference information; also, they could refer to a laminated ‘Data linkage FAQs’ sheet.

In the web questionnaire, the interviewer persuasion is not possible as this is a self-completion mode. On one hand, several mitigation strategies were put in place to simulate the role of the interviewer in the web survey – for example, a video about data linkage addressed to participants was developed to try to simulate interviewers’ persuasion; on the other hand, the web instrument allowed the adoption of web specific features that could increase the respondents’ understanding of data linkage and that were inapplicable in other modes – for example, the request to data linkage was accompanied by hyperlinks to the data holders websites (e.g. NHS, *etc.*).

The web instrument provided an opportunity to enhance the information provided to respondents, through the use of hyperlinks; as a result, web respondents had access

---

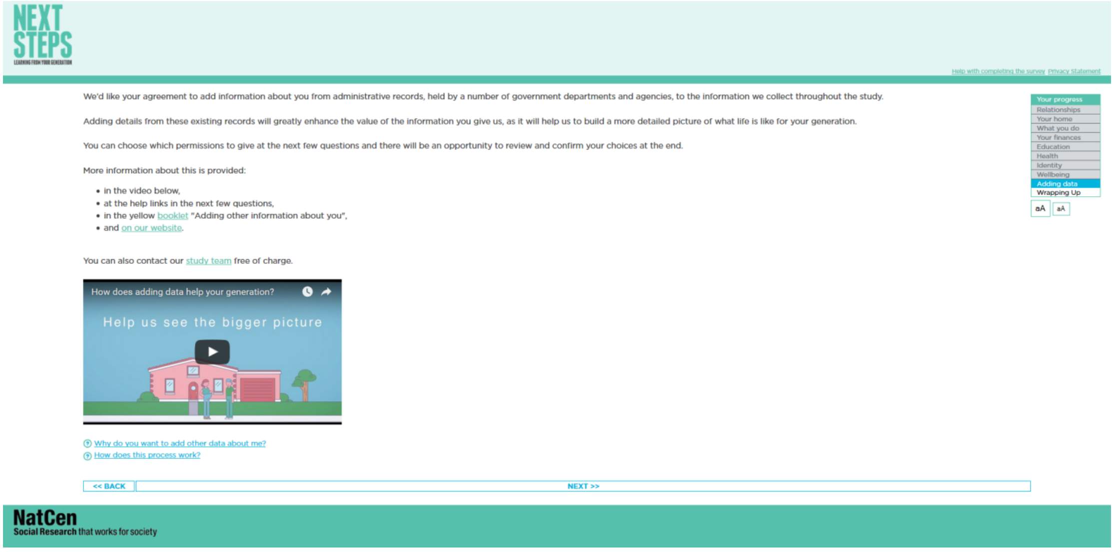
[3] The complete leaflet is available at http://www.cls.ioe.ac.uk/nextsteps/library-media%5Cdocuments%5CNS8_MAINSTAGE_DATA%20LINKAGE%20LEAFLET_WEB.PDF

to more information than face to face and telephone respondents, as, for example, the hyperlinks to data holders websites.

In the web instrument, respondents could also access the “Frequently Asked Questions” on data linkage by clicking on a hyperlink embedded in the web instrument: the hyperlink opens (as a new window) the “frequently asked questions” page on the study website. Also the respondents could access the Freephone telephone number for clarifications on data linkage.

The image below shows the first page in the data linkage section; besides the explanation of data linkage, the website included a video embedded in the page, which would provide an overview of the data linkage procedure.

Figure 1: 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:

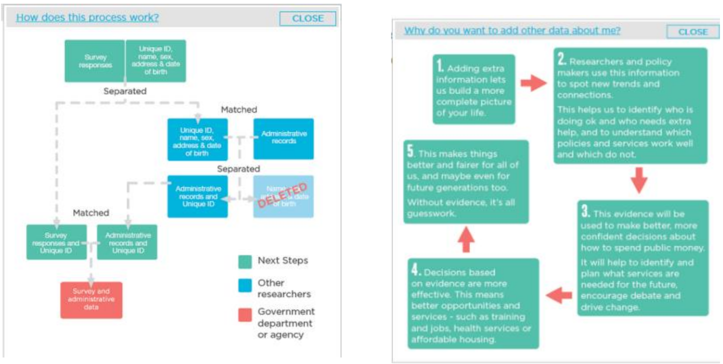
- in the video below,
- at the help links in the next few questions,
- in the yellow booklet "Adding other information about you",
- and on our website.

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?

Below the embedded data linkage video, two hyperlinks were provided: “Why do you want to add other data about me?” and “how does this process work?” The two hyperlinks opened pop-up windows, showed in the images below.

Figure 2: Pop-up windows embedded in the web questionnaire, with information on data linkage



1. Adding extra information lets us build a more complete picture of your life.

2. Researchers and policy makers use this information to spot new trends and connections.
This helps us to identify who is doing ok and who needs extra help, and to understand which policies and services work well and which do not.

3. This evidence will be used to make better, more confident decisions about how to spend public money.
It will help to identify and plan what services are needed for the future, encourage debate and drive change.

4. Decisions based on evidence are more effective. This means better opportunities and services - such as training and jobs, health services or affordable housing.

5. This makes things better and fairer for all of us, and maybe even for future generations too.
Without evidence, it's all guesswork.

Care was taken so that all respondents had a chance to consult the data linkage information leaflet. In the face-to-face mode, a leaflet was offered to respondents if they haven't had the chance to consult one yet. In the telephone mode, the interviewer read the salient content of the leaflet over the phone or directed the respondent to the study website to read the leaflet, in PDF format. In the web mode, a link was provided with the information in the leaflet on the study website or to view the leaflet itself in PDF format.

In order to maximise the level of information that the cohort member had at their disposal when expressing consent to data linkage, the request was asked at the end of the questionnaire, when respondents have a more complete picture of all the other survey information that are collected on him/her.

Furthermore, asking consent at the end of the questionnaire would not disrupt the flow of the questionnaire with potential administration of the leaflet and looking up of further information. Presenting the questions earlier in the questionnaire could also risk putting some respondents off completing the rest of the questionnaire, if they were to have a negative reaction to being asked consent to link to administrative data.

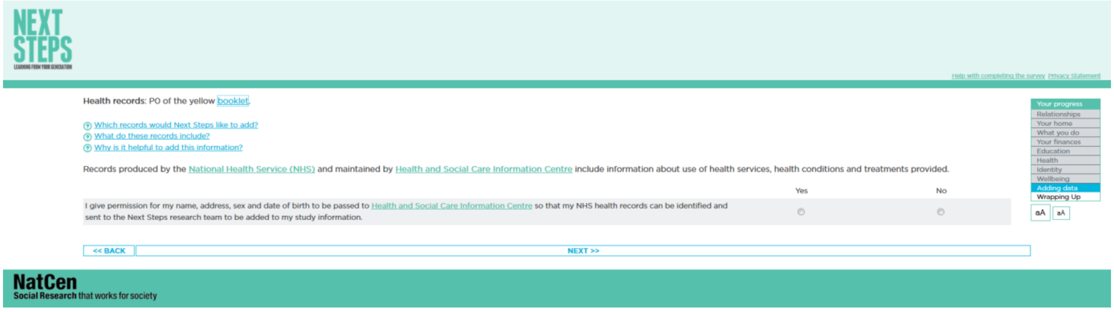
As already mentioned, consent to data linkage is requested for nine different records, plus a tenth question on National Insurance Number for those who gave consent to DWP/HMRC data linkage. These are grouped in four consent areas, which are:

- Health records (NHS consent)
- Economic records (DWP and HMRC consents – with follow-up National Insurance Number question)
- Education records (DfE, BIS, HESA, UCAS, SLC consents)
- Police records (PNC consent)

The questions included the following content: a title which give a short introduction to the topic area of consent (e.g. "Economic records"), a consent to link question (e.g. "Records kept by Her Majesty's Revenue and Customs include information about employment, earnings, tax credits, occupational pensions and National Insurance contributions. I give permission for my name, address, sex and date of birth to be passed to Her Majesty's Revenue and Customs so that my records can be identified and sent to the Next Steps research team to be added to my study information."); and two answer options ("yes", "no").

The figure below shows an example of a data linkage request page, specifically the one requesting consent to link data with health records.

Figure 3: Data linkage request page for health records



Health records: P0 of the yellow booklet.

Which records would Next Steps like to add?
What do these records include?
Why is it helpful to add this information?

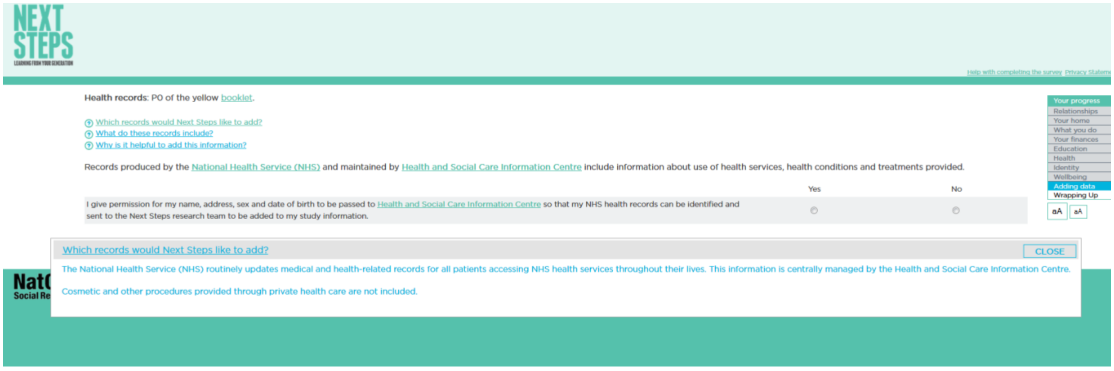
Records produced by the National Health Service (NHS) and maintained by Health and Social Care Information Centre include information about use of health services, health conditions and treatments provided.

	Yes	No
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.	○	○

The web implementation of the data linkage section allowed for the inclusion of several hyperlinks. Specifically, the “booklet” hyperlink opens the “resource” section of the website. The “National Health Service (NHS)” hyperlink opens the website to the National Health Service, and the “Health and Social Care Information Centre” hyperlink opens the Health and Social Care Information Centre website[4].

The hyperlinks: “Which records would Next Steps like to add?” “What do these records include?” and “Why is it helpful to add this information?” open a pop up window each with additional information – for an example, please see the figure below.

Figure 4: A pop up window with information on: “Which records would Next Steps like to add?”



Health records: P0 of the yellow booklet.

- Which records would Next Steps like to add?
- What do these records include?
- 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.

	Yes	No
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.	○	○

Which records would Next Steps like to add?

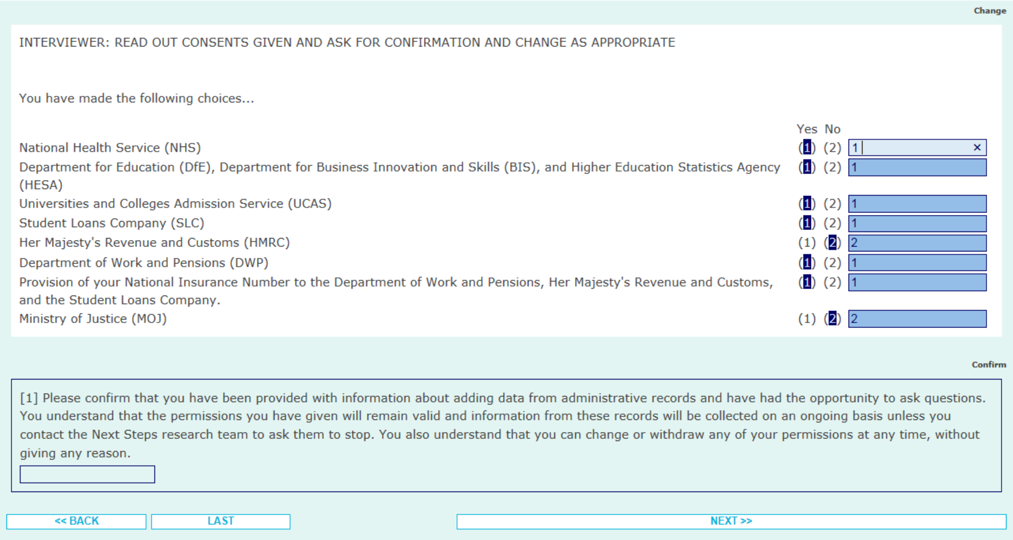
The National Health Service (NHS) routinely updates medical and health-related records for all patients accessing NHS health services throughout their lives. This information is centrally managed by the Health and Social Care Information Centre.

Cosmetic and other procedures provided through private health care are not included.

In all modes respondents were asked to confirm that they have read the data linkage information leaflet, have been given an opportunity to ask any question they had and understand that data will be treated in accordance with the Data Protection Act and used for research purposes only.

At the end of the data linkage section respondents (in web) and interviewers (in telephone and face-to-face) were presented with a screen summarising the permissions given during the interview (see figure 5 below).

Figure 5: Data linkage request confirmation page in CATI and CAPI



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

[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.

---

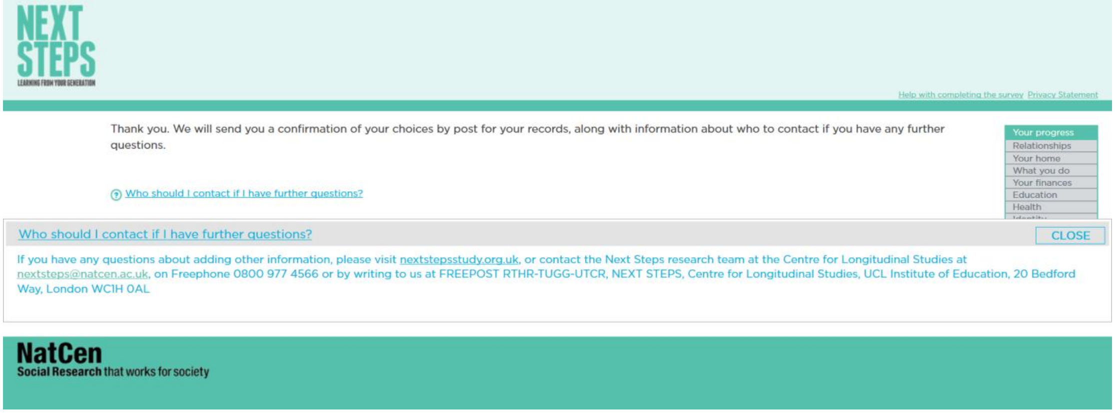
[4] Now known as ‘NHS Digital’

The respondent has an opportunity to confirm the consent provided, and to change any consent given. In the CAPI and CATI interview, the interviewer read out each listed record type and the response given by the cohort member for them to confirm, and if needed, s/he changes the responses provided by the cohort member 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 their consent choices, the respondent is asked to give confirmation, ticking a confirmation box in the web survey or accepting the following sentence in the CAPI and CATI interview: “Please confirm that you have 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 changed or withdraw any of your permissions at any future time, without giving any reason”.

In the web survey, after this confirmation page, an additional page was displayed to the respondent stating that written confirmation would be sent by post, and with an additional hyperlink which gave the respondents some contact details for further information.

Figure 6: Thank you page



Thank you. We will send you a confirmation of your choices by post for your records, along with information about who to contact if you have any further questions.

Who should I contact if I have further questions?

If you have any questions about adding other information, please visit nextstepsstudy.org.uk, or contact the Next Steps research team at the Centre for Longitudinal Studies at nextsteps@natcen.ac.uk, on Freephone 0800 977 4566 or by writing to us at FREEPOST RTHR-TUGG-UTCR, NEXT STEPS, Centre for Longitudinal Studies, UCL Institute of Education, 20 Bedford Way, London WC1H 0AL

Finally, in all modes, written confirmation of data linkage consent choices was sent to respondents after the survey. The confirmation was embedded in a “Thank you” mailing, which also included the incentive and a change of details card for future survey waves. The letter was sent within two weeks of the respondents’ full completion of the survey, or, in case of partial completion with the cohort member completing as far as the data linkage section, the letter was sent at the end of face-to-face fieldwork. This confirmation was sent to all participants, regardless of the mode of data collection.

In this mailing the respondent is provided with information on how to change or withdraw their consent(s) if they wish. Also, study contact details are supplied so that participants can get in touch with further questions or concerns.

Evidence from the qualitative study suggested that a post-survey confirmation of consent in hard copy was perceived by respondents as helpful, in part in order to keep for future reference, and also to give sample members another chance to check that their preferred consents have been recorded accurately.

Research ethics and data linkage

The data linkage materials and consent protocol was reviewed by the Multicentre Research Ethics Committee as a guarantee that it was conducted in accordance with the requirements and best practice in research ethics.

The committee serve to safeguard the rights, safety, dignity and well-being of research participants, and review applications for research in order to give an opinion about the proposed participant involvement and whether the research is ethical. The committee was entirely independent of research sponsors, funders and investigators.

Exploring consent to data linkage on Next Steps: evidence from a qualitative study and a pilot study

The adopted data linkage protocol was a result of extensive development work which included a qualitative research and a pilot study. In this section we review the evidence from both of these preparatory phases. More specifically, we describe the design and the aims of the qualitative and pilot study; we outline the results, focusing on participants' general reaction and understanding of the data linkage consent protocol and participants' evaluation of the quality of the information provided on data linkage.

In particular, we focus on the lifespan of consent, the procedure for providing consent and for consent confirmation, the acceptability of the consent process overall, the comparison of different ways of framing the consent questions and the clarity of the information leaflet.

Finally we show the consent rates obtained in the pilot study.

The aim and design of the qualitative study

The qualitative study was aimed at evaluating the data linkage consent materials and protocol. More specifically, this work attempted to answer the following research questions:

1. Do the data linkage consent materials promote understanding of data linkage?
2. Is the proposed protocol acceptable to participants (i.e. consent "at the click of a button", consent to a wide range of linkages, and email versus written confirmation of consent letter)? Is the protocol feasible: i.e. do the participants understand what is expected of them and can they complete the tasks?
3. Do the current materials and protocol elicit informed consent – especially for web mode where there is no interviewer support?
4. How can the participant (and interviewer) experience of the data linkage protocol be best supported?
5. Do factors (such as question placement and framing) impact on consent propensity amongst this group, and how can consent rates be maximised and response bias be minimised?

The sample for the qualitative work was composed of twenty individuals, aged 23-27. These young adults were not cohort members; instead the group was recruited from the general public by a recruitment agency with the aim of broadly representing the study population.

More specifically, selection criteria were set as to include 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. Interviews lasted 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 as a thank you for their time and cooperation.

The aim and design of the pilot study

The pilot study took place in October/November 2014 with 120 participants in aged 23-27 in three areas of England. These young adults were not cohort members; instead the group was recruited from the general public by NatCen interviewers with the aim of broadly representing the study population. More specifically, selection criteria were set as to include a diverse group of respondents in terms of gender, presence of children, cohabitation status and employment status (as well as ethnicity in London). Respondents were given a £20 incentive for participation.

The pilot trialled the data linkage protocol and provided an opportunity to obtain feedback on the data linkage protocol from telephone and face-to-face interviewers, from participants in a post-interview feedback questions, as well as from a small number of participants who directly contacted the office. Also, data from the pilot are analysed to explore consent to data linkage.

In the pilot the nine consent requests were grouped into four types:

- Health records (NHS consent)
- Economic records (DWP and HMRC consents – with follow-up National Insurance Number question)
- Education records (DfE, BIS, HESA, UCAS, SLC consents)
- Police records (PNC consent)

The protocol for consent to data linkage was different from the protocol adopted in the mainstage: in the pilot, respondents were asked a "pre-consent confirmation question", where they could confirm they have read the leaflet, have had the opportunity to discuss any question they may have had, and understood that the information was treated confidentially, and in accordance with the Data Protection Act; if no confirmation was provided participants were not asked the consent question. Conversely, in the mainstage, respondents were asked confirmation after answering the consent items, as evidence from the pilot showed that asking this up-front was off-putting and confusing.

Acceptability and feasibility of consent process

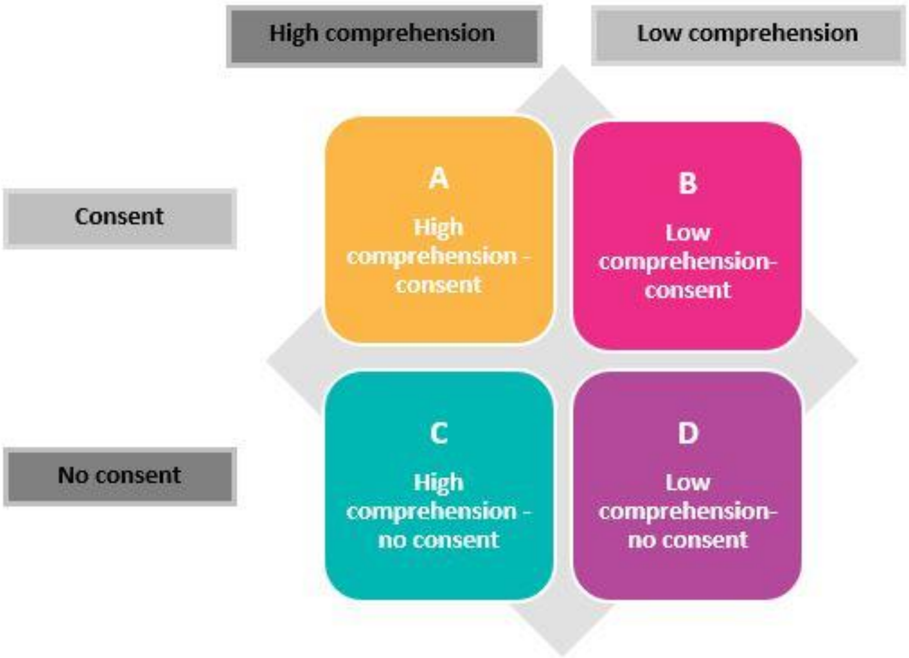
In general, evidence from the qualitative work showed that the protocol was considered acceptable, and the transparency of asking each of the nine different consent questions separately instead of one single "catch-all" item was valued.

However, the participant’s 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. Concern was expressed 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 stated: “[I] don't mind doing study but not prepared to link data as that's scary” (Quotation from telephone participant, Pilot telephone interviewer feedback form).

In some circumstances, the level of trust was not sufficient to guarantee that the respondents consent to data linkage. As one telephone participant 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 from telephone participants, Pilot telephone interviewer feedback form)”

While the consent procedure was considered easy, the comprehension of what was being asked was limited. The qualitative research showed that participants can be clustered in four groups according to their comprehension and willingness to provide consent to data linkage, as shown in Figure 7.

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



Membership to these groups was not static as participants could belong to different groups across different consent questions; also, 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 reasons behind data linkage and of the benefits both in terms of policies and of participants' survey experience.

The two groups of concern were the groups showing a low comprehension, regardless of whether the consent level was high or low. The group showing high level of comprehension and low consent is particularly problematic as no additional information is likely to influence their consent to data linkage.

The qualitative evidence identified six factors underpinning the typology of comprehension and consent, highlighted in Figure 8.

Figure 8: Six factors underpinning the typology of comprehension and consent

1

Perceived benefits

- Willingness to provide consent improved as participants' awareness of why the information was needed, the uses it would be put to and the perceived benefits it would have on society.

2

Sensitivity of question

- Perceived personal risks. E.g. misuse of data, reputational risks (e.g. morally judged), whether the data also affected others (e.g. parents).
- Relevance to participants.
- Routine sharing of information E.g. ducation information in CVs.

3

Trust

- Llikelihood of consent increased, regardless of comprehension, if participants felt they could trust the survey that was requesting the information and/or the organisation from which data was requested E.g. the NHS was seen to be trustworthy.

4

Assurances

- In lieu of trust (and sometimes to cement it), participants sought assurances that their information will be kept confidential. An understanding of trust was particularly the case for questions perceived to be sensitive.

5

Feelings about personal disclosure

- "Nothing to hide". Participants open about their data .
- 'Not being judged for past actions'. This included people who did not want to share past information if it reflected unfavourably on them (e.g. debt or checkered university record).

6

Timeframe

- Participants' comprehension tended to focus on providing consent for data to be linked retrospectively. Those that did acknowledge the prospective element of data linkage tended to be more guarded in giving consent.

Furthermore, as mentioned earlier, the qualitative research provided evidence of an “incremental effect” of each data linkage question within a particular subject area. Indeed, participants were prompted to reflect on the current question in light of the previous ones.

For example, in the qualitative study the question to link data from HMRC followed the question on linkage from the DWP; respondents had a greater understanding at the second question compared to the first; also, they were prompted from the second question to read back the first question, and compare the text to it, in an attempt to determine in which respects these two differ.

The "incremental effect" of asking consent on multiple domains affected the respondents' experience of data linkage in different ways: it had a positive impact on consent, a positive impact on comprehension, and raised awareness on both the volume of information that participants were invited to share, and that various organizations held on them.

Firstly, asking consent on multiple domains had a positive impact on consent; this is because 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 the subsequent questions.

Secondly, asking consent on multiple domains impacted positively on comprehension, as the details provided at any additional question helped clarifying previous consent requests.

Thirdly, reading multiple consent requests made participants gradually aware of the volume of information that they were asked to share and of the volume of information that are held on them by various organization. It should be noticed, however, that the awareness on the quantity of information held on study participants did not necessarily impact negatively on consent.

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

Analysing the participants' understanding of the various benefits of data linkage

Participants to the qualitative development work were presented with eight different benefits to data linkage (Table 2). Understanding which of these benefits are the most salient for the target population is particularly important; in fact, the most salient benefits may be used as leverages to increase the consent to data linkage.

Among the eight proposed benefits to data linkage the one that was more often appreciated by participant was: data linkage "builds a more complete picture".

Other benefits – such as "improves the accuracy and the value", "enhances the information by adding more depth", "saves time", "easily completes information we already have" – were all liked by participants, and no reason for disliking them was identified.

Conversely, other benefits such as "makes better use of existing information", "adds value by helping policy makers plan and improve services", and "enhances the core information already given to us" were not always understood and considered relevant by participants. Table 2 below provide a summary of the eight different benefits proposed and participants' reactions.

Table 2: Summary of views on benefits of data linkage

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 the point	
Builds a more complete picture	Simple way of explaining reasons for data linkage Already comfortable with giving information because of participation in survey	
Enhances the information by adding more depth	Improves the information captured in the survey; ‘adds value’	
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)	Highlights the benefit for the individual participant-saves their time in future surveys; saves time in looking up information	
Makes better use of existing information (identifying population trends, understanding specific needs of this generation)	Provides new perspective of people in their mid-twenties Demonstrates usefulness of maximising existing data	Poor understanding of term ‘population trends’
Adds value by helping policy makers plan and improve services	Use of the word ‘value’ is persuasive and indicates the participant is helpful	Not relevant to individual participants Poor understanding of term ‘policy maker’. This phrase described by one participant as “Is it like improving insurance policies?” (Female, low education, in work, Group 2).
Enhances the core information already given to us (opening up research opportunities)	Demonstrates usefulness of maximising existing data	Not relevant to individual participants

Consent at the click of a button

As discussed, consent to data linkage was asked differently in different modes. The qualitative work reviewed whether consent "at the click of a button" was considered acceptable; the majority of respondents expressed no concerns, and only in rare circumstances it was expressed a view of concern that could result in the decision to not provide consent, unless a written signature was collected.

As a result of this evidence, some mitigating factors were implemented in the web questionnaire in order to minimize the concerns of participants to give consent "at the click of a button". The web instrument provided three ways for respondents to seek further clarification: the "Frequently Asked Questions" on the help screen via hyperlinks displayed on each page; the project Freephone telephone number displayed for telephone clarification; and the link to a dedicated section on the study website with further information and with a web contact form. In addition, it was decided that a record of consents which had been given would be displayed at the end of the data linkage section for respondents' to review, and that a record of which consents had been given would be sent by post to all respondents.

The lifespan of consent

Regarding the lifespan of consent, the evidence from the qualitative study suggested that linking survey data with past individual record was understood and considered acceptable. Conversely, concern was raised about linkage to future records; participants initially understood that survey responses would be added to past and present administrative records, but didn't consider the possibility of their survey answers to be linked to future records. The information leaflet used at that stage did not clarify this aspect.

Not only in the qualitative research, but also in the pilot study, participants asked information on the lifespan of consent, as well as querying the motivations to ask consent.

Once it was clarified that records would be added to survey data, participants asked whether they could chose to limit their consent in the future. Participants suggested adding expressions as: "you can opt out at any time", "we would like to add records from your past and present", or "we would also like to add any future records to your survey responses to provide a more complete picture".

Also, some respondents claimed that an annual reminder to sample members about their on-going consent would be beneficial; they considered this practice particularly useful if there are gaps in running the survey.

These issues were addressed by amending the questionnaire wording; in the confirmation of consent an explicit reference to consent been valid in the future was added; this was worded as following: *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.*

Moreover, the data linkage leaflet was amended to make the prospective nature of consent more prominent, and the data linkage FAQ leaflet amended to explicitly include a question and answer paragraph on the lifespan of consent; this was worded as following: *How long will the permission last? All permission you give will remain valid and administrative data will be collected on an ongoing basis unless you withdraw your permission. We have not put an end date on the permission that you give as we do not know exactly when we will receive or add the information.”*

Comparing positive and negative framing of data linkage questions

Questions on data linkage can be framed positively (i.e. emphasising the advantages to data quality obtained through data linkage) or negatively (i.e. emphasising the loss for the survey from not obtaining the link to the data).

Two different wordings, one positive and one negative, were tested on the introduction to the data linkage question in the qualitative study. The positive wording included the sentence: *The information you have already given us will be more useful if information about you can be added from these other records*; while the negative wording included the sentence: *The information you have already given us will be less useful if information about you cannot be added from these other records*.

There was an overwhelming preference for 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 sensitivity of the consent to data linkage requests

One factor that may influence consent to data linkage is to which extent the consent request is considered sensitive. As in survey questions in general, whether a consent request is considered sensitive or not depends, among other things, on whether the sample member engages in any socially undesirable behaviour or has a socially undesirable characteristic associated with the request.

For example, regarding health records, 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, sexual health problems, or cosmetic surgeries).

A similar reasoning applies to economic records. One participant stated that he doesn’t have any objection and don’t perceive any harm to share his benefits records, because he does not receive benefits and he is not part of a benefit programme; thus, implying that other respondents with these characteristics may be less comfortable than him in giving consent to data linkage on these records.

These participants’ considerations are consistent with the theory on sensitive questions in surveys, as the level of sensitivity is expected to vary depending on whether the respondent engaged in the socially undesirable behaviour or have the socially undesirable characteristic or not.

Moreover, participants in the qualitative study did not consider all consent requests as being sensitive to the same degree; conversely, some consent requests were considered more sensitive than others, depending on the topic area.

As stated above, Next Steps requested data linkage on four different dimensions: health, economics, education, and criminal justice.

Asking data linkage on health questions was considered sensitive; as already mentioned, participants expressed concerns for sample members that may have had health conditions that are socially considered as embarrassing, or sample members that had health treatments that are not perceived as socially desirable.

Other records for which it is considered sensitive to ask consent to data linkage are employment records; as one participant stated referring to the request to link data to HMRC records: “You just don’t know what someone could do with that information… I don’t think many people would want that. It’s very private” (female, low education, in work). However, the opinions on the sensitivity of the HMRC question varies widely; in fact, another respondent argues: “A lot of that is probably on record somewhere anyway. Your earnings, employment, national insurance… so it is not really anything new you are giving away that you haven’t probably given somewhere before” (male, medium education, in work).

Within the educational area, the only question that raised concerns on sensitivity 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, this is considered as a more sensitive area; as one participant states: “Like the benefit thing, this is more of a picky subject” (female, low education, in work).

Information on student loans are collected only for sample members that attend university, and, thus, participants felt that this data consent request should not be considered sensitive for the whole sample, but only for those for which it is applicable. As with other consent requests, some participants were not concerned about sharing their records, and some reiterated the concept that financial information are routinely collected and shared by organisations.

However, some participants were not willing to give consent to data linkage; reasons quoted for non-consent were: previous negative experiences with the student loan company, the fear that the information held on their parents’ financial situation may not be treated confidentially, and the consideration that having a loan to repay in their records was social undesirable. On this latter point, one participant stated: “I could be really a good person but off with paying loans and that could put me off” (Female, low education, not in work).

Besides this item, other requests within the educational area were not labelled as sensitive; also, participants felt that these information are routinely shared in the everyday life (e.g. in generic conversations, social media, CVs).

Some participants, however, mentioned the fear of social judgment on the data linkage consent question on educational records. Specifically one respondent stated “This one is a bit risky… pupil characteristics.. that can affect people later in life. When you’re younger everybody does stupid stuff so it shouldn’t follow you through life” (male, low education, in work). Conversely, other participants seem particularly enthusiastic about the opportunity to link their records as they were proud of their school marks, and they somehow considered this as an opportunity to share their achievements; as one participant stated: “(…) I have really good educational records so I’d like to show it off” (Male, Medium education, in work).

Finally, there were participants in the qualitative study who perceived giving consent to data linkage to educational records as an opportunity to get access to their own records, to recollect their own achievements.

In assessing the feasibility of asking data linkage consent to criminal justice records, all participants were asked to express their views the sensitivity level of the political and criminal justice question. In terms of sensitivity of the question, it was considered feasible to include it.

Participants stated that the consent request could concern sample members having had an arrest, caution, or conviction, as this would have a general negative impact on someone’s reputation, and concrete consequences (e.g. in the job market) if data were misused.

Some participants stressed the sensitive nature of this topic noticing that these information are generally considered private, as one participant stated: “If you are not going to talk about it with your friends then it is not the sort of thing you’d just let anyone know”.

The information leaflet

The feedback on the data linkage leaflet provided in the pilot study was overall positive. Most interviewers felt that participants who have read the leaflet understood the data linkage process.

The thoroughness of the materials was generally appreciated but some participants suggested that the leaflet could have been more concise. Indeed, while the thoroughness of the materials were appreciated by some participants, as it contributed to the understanding of the process, not every participant read it in detail. For example, one telephone interviewer noted that some participants had only “skim read” the leaflet, probably because of the length of the booklet. As a result of this behaviour, some have asked general questions about the data linkage procedure putting the onus the interviewer.

Survey practitioners wishing to design data linkage leaflet may bear in mind the suggestions from the pilot study. First, we advise to highlight the voluntary nature of data linkage, and stress that the respondent not consenting to data linkage can still participate in the survey – this aspect was stressed by interviewers.

Second, consistently with the general advice from the literature on how to design survey materials, we suggest to keep the leaflet short and concise; also, we advise to highlight that the data is kept confidential, as suggested by participants to the pilot study.

Third, we suggest to use graphics and diagrams to visualise the data linkage process; the data linkage diagram in the leaflet was valued by interviewers as informative on how the process works and how the data is kept anonymous.

Fourth, we advise to stress the timeline of consent to data linkage; in particular, both interviewers and respondents highlighted the need for information on the prospective nature of the linkages.

Finally, a criticism outlined in the pilot study that survey practitioners should bear in mind is that the advance mailings with the data linkage leaflet may not arrive to all participants, as, for example, some may have moved. Appropriate mitigating strategies were implemented in Next Steps, and we advise to follow these procedures in other surveys: face to face interviewers were equipped with spare data linkage leaflet for participants who have not received/read them, and telephone interviewers were instructed to direct participants to the survey website where a link to the leaflet was provided.

Survey practitioners designing data linkage leaflets should bear in mind that, as probably with other survey materials, not all participants read them in detail. Evidence from the qualitative study suggests that participants may be divided in two groups according to their reading behaviour of the data linkage leaflet: the skimmers and the diggers. While skimmers focused on headers and images spending between 1 to 3 minutes on the leaflet, diggers read the leaflet in depth spending approximately 5 minutes on the leaflet.

A similar behaviour was observed in the pilot study. Indeed, while the thoroughness of the materials were appreciated by some participants as it contributed to the understanding of the process, the length of the leaflet may have prevented other participants from reading it carefully. Indeed, some participants only “skim read” and have then asked the interviewer general questions about the data linkage procedure. While reducing the length of the booklet may decrease the number of “skim readers”, interviewers felt that all information provided where necessary. Survey practitioners designing these materials should bear in mind the trade-off between the level of details provided and having the participants reading and fully understanding them. This should be borne in mind when drafting the leaflet e.g. ensuring that different sections are self-contained and stand-alone, using clear section heading and sign-posting so respondents can easily find the information they are looking for.

Regarding the content we advise survey practitioners to include examples, to word the leaflet as participant centred, and to include reassurances on data security. These were the aspects which participants in the qualitative research valued the most. All participants found that the most effective and reassuring sections were: “Why is it helpful to add this information”, “Some assurances”, and “Did you know?”

Also, we advise survey practitioners to avoid wording that may result unclear, vague or inconsistent. For example, the expression: “…destroy the file containing your personal details” may be interpreted by participants as the intention to destroy personal records, rather than identifiable details (namely: name, sex, address and date of birth); also the term “withdrawal” was understood ambiguously, as some respondents interpreted it as withdrawal from the whole survey[5]; participants found confusing the use "administrative records", "administrative data", "records" and "information" as synonymous, and the use of the term "information" inconsistently, meaning both "survey responses" and "records"; other terms considered as unclear were: "any sensitive info", "survey questions", and "anonymous identifier".

Given the misunderstanding of some terms presented in the information leaflet, we suggest survey practitioners provide definition for expressions that participants may not be familiar with and avoid the use of multiple terms for the same concept. Also, it was suggested to include the spelling of the full department names instead of their acronyms (e.g. DWP and HMRC).

Overall, the evidence for reactions to the information leaflet suggests that the general understanding of the process of linkage and the anonymity varied across participants.

Consent confirmation

Evidence from the qualitative work shows that the post survey confirmation of consent received by respondents by post is perceived as helpful as it gives a chance to double-check that the right consents have been recorded and serves as a record of consents given.

The qualitative work also explored preferences for the mode of confirmation of consent. Some participants would have preferred to receive a confirmation letter, as paper records are considered easier to keep record of and participants valued that this medium would convey higher formality; 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 in the confirmation email. Finally, a group of participants valued receiving both an email and a letter of confirmation of consent. In the Next Steps age 25 survey it was decided to send a paper confirmation letter to all respondents.

In addition to these communications, participants expressed the desire to receive an additional leaflet which would describe how linked data contributed to research: for example, an intra-wave mailing or a "findings hand-out". Although it was not possible to implement this procedure in the current survey, this is an aspect that may be considered in other studies and by Next Steps in the future.

Consent rates from the pilot study

In the pilot study, the number of participants who completed the data linkage section was 89 (of the 96 fully productive interviews). Depending on the mode of data collection, and on the consent type, consent rates range from 47% to 89%.

In analysing consents rates to data linkage by mode of data collection, it should be noted that, in the pilot study, participants were randomly allocated to different survey modes – thus, selection into mode does not undermine the comparison across different

---

[5] Clarification on this was provided in the leaflet but respondents felt they needed it earlier than page 9, where it was positioned in the data linkage leaflet.

modes; however, 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; indeed, the experience in collecting consent differ for face to face compared to telephone interviewers: in the telephone mode, some participants were hesitant and reported that this was an excessive request and it was too intrusive; despite the reassurances of data security and the voluntary nature of consent, the request put some participants off taking part altogether. Face to face interviewers reported more positive feedback than telephone interviewers; they 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.

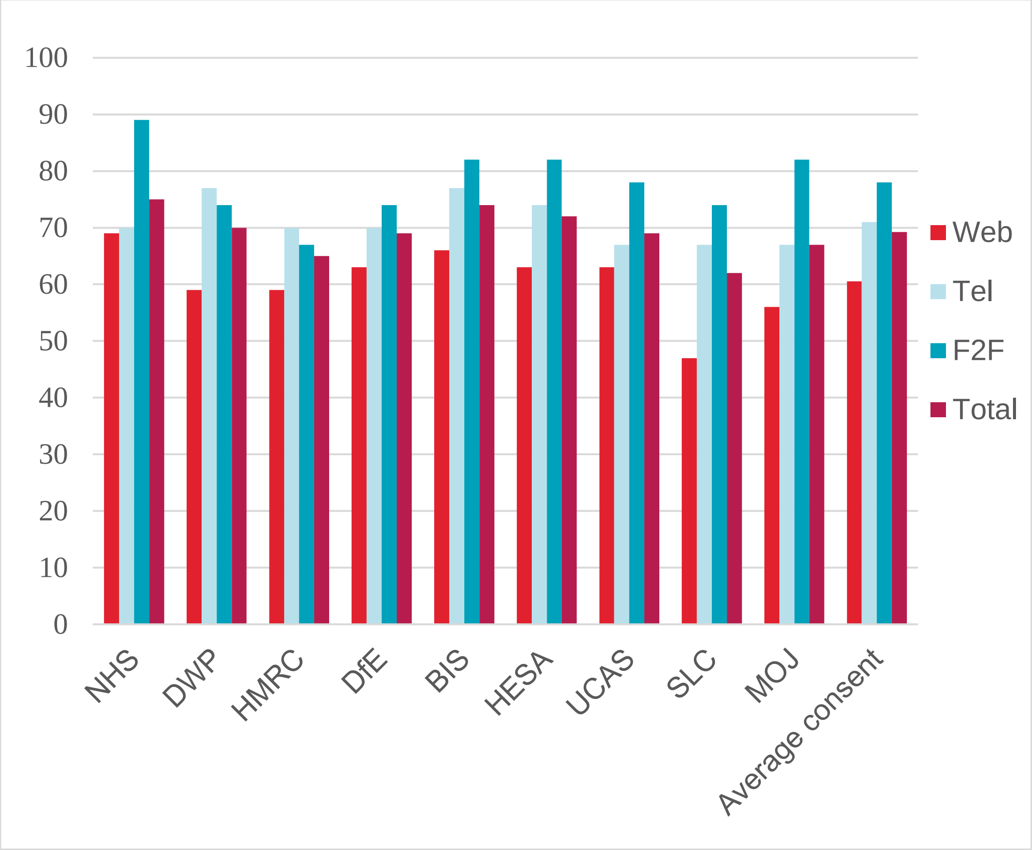
Looking at the overall response rate per consent type, the lowest consent rates where those related to loans (Student Loan Company) and the economic area (especially HMRC).

Table 3: Consent rate for individual consent types by mode in the Next Steps pilot

	Mode							
	Web		Tel		F2F		Total	
	N	%	*N*	%	*N*	%	*N*	%
NHS	22	69	21	70	24	89	67	75
DWP	19	59	23	77	20	74	62	70
HMRC	19	59	21	70	18	67	58	65
DfE	20	63	21	70	20	74	61	69
BIS	21	66	23	77	22	82	66	74
HESA	20	63	22	74	22	82	64	72
UCAS	20	63	20	67	21	78	61	69
SLC	15	47	20	67	20	74	55	62
MOJ	18	56	20	67	22	82	60	67
Average consent	*19*	*61*	*21*	*71*	*21*	*78*	*62*	*69*
Bases	*32*		*30*		*27*		*89*	

Note: the base is composed by all who answered consent questions. Source: Next Steps pilot

Graph 1: Consent rates by mode and consent request



Summary and discussion

In this paper we discuss three main aspects of asking consent to data linkage. First, we investigate what are the challenges of asking consent to data linkage in a mixed-mode context. Second, we analyse whether it is feasible to ask consent to data linkage to multiple domains simultaneously. Thirdly, we discuss the best practices in designing materials to promote consent.

The increase in adoption of web as a mode of data collection, either alone or in conjunction with other modes, urged survey methodologists to understand how to collect data linkage consent in web surveys or in mixed modes surveys including web. This task presents the challenge of simulating the interviewer persuasion in a self-completion context, and entails logistical issues, since collecting signed consent forms is not practical in web (and telephone) surveys.

As discussed in the paper, the Next Steps study adopted several mitigating strategies to promote consent in self-administered modes; the survey embedded "Frequently Asked Questions" on data linkage, a video describing the data linkage procedure, and included hyperlinks to the data holder institutions.

Overall, the evidence from the qualitative and pilot study showed that respondents considered acceptable to give consent at the "click of a button" in the web survey, and to express verbally consent in the telephone and face-to-face interview. As opposed to signed consent, this protocol minimises respondent burden and survey cost.

Since written confirmation of consent was valued by participants to pilot study and qualitative interviews, the Next Steps survey implemented a feature that would allow respondents to review their consent and receive confirmation of their data linkage choices by post, having an opportunity to check that the right records has been recorded and to serve as a record of consents given.

Further research is needed to investigate whether there are mode effects in the propensity to provide consent to data linkage. Experimental evidence from the Next Steps pilot study, where participants were randomly allocated to fill in the survey in different modes, 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 in the main stage to simulate the role of the interviewer in the web survey were insufficient to compensate fully for the lack of an interviewer. 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, it should be noted that as participants self-selected into mode, the descriptive analysis does not enable robust conclusions to be drawn about mode effects on data linkage consents.

Besides the challenge of asking consent in a mixed mode setting, the Next Steps Age 25 survey asked consent on multiple domains, and on a wide timespan – both retrospectively and for the future, with no end date – enhancing survey data with precious information, also for cohort members that have not participated to previous survey waves, or that may attrite in the future.

We find evidence 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.

We advise survey practitioners designing consent requests on a wide timespan to carefully word prospective consent requests, as cohort members may find it complicated to understand and welcome data linkage with future records.

In terms of question wording, we find evidence that emphasising the effects of data linkage on data quality (positive wording) is preferable to emphasizing the potential loss in data quality from not obtaining data linkage consent (negative wording).

Comparing the different domains, the questions in the area of health were considered sensitive, as participants were concerned that these requests may be asked to sample members that are affected by health condition which are socially considered as embarrassing. Conversely, in the area of education, the only question generally considered as sensitive is the request to link data from the Student Loan Company, as it refers to the respondents' financial situation. Consistently, questions in the domain of employment are generally considered sensitive. Finally, participants to the qualitative interviews considered sensitive, but overall feasible, to ask data linkage to criminal justice records.

In the pilot study and the main study 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, by using data from the Next Steps mainstage survey, to further inform survey practitioners of the factors affecting consent in different consent domains for this age group.

The third aspect investigated in this research is the design of effective data linkage materials for this age group. We advise survey practitioners to keep the leaflet short and concise, as in general for other survey materials; indeed, evidence from the preparatory work suggests that some survey participants only “skim read” the data linkage leaflet; thus, it is crucial to bear in mind the trade-off between the levels of details provided and the participants reading and fully understanding the content; we encourage survey practitioners designing data linkage materials to ensure that the different sections are self-contained and stand-alone, the use of clear section headings, and sign-posting, to allow respondents to easily find the information they are looking for. We also produced a separate ‘Frequently asked questions’ document with more detail regarding data linkage which interviewers had available and was also available on the study website.

We advise survey practitioners to stress the voluntary nature of data linkage consent, and to highlight the possibility of participating in the study even if the cohort member is not willing to provide data linkage consent.

Moreover, the use of graphics and flowcharts to visualise the data linkage process seems promising; these graphical tools may be useful also for clarifying the prospective nature of consent, which seemed one of the most challenging aspect to convey. More broadly, if the data linkage leaflets and materials are included in the advance mailings and there is a possibility that the cohort member doesn’t receive this communication (for example as a consequence of having an incorrect address) it is advisable, in face to face interviewing, to equip interviewers with spare leaflets, and, in CATI, to train telephone interviewers to direct cohort members to the data linkage section of the study websites.

In terms of wording, we advise the use of clear, precise, and consistent wording, and, when needed, the adoption of definitions and examples.

Further research may compare in an experimental setting different wording and formats of data linkage materials, as well as different protocols for informing respondents of data linkage.

Appendix

Figure A1: 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

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.

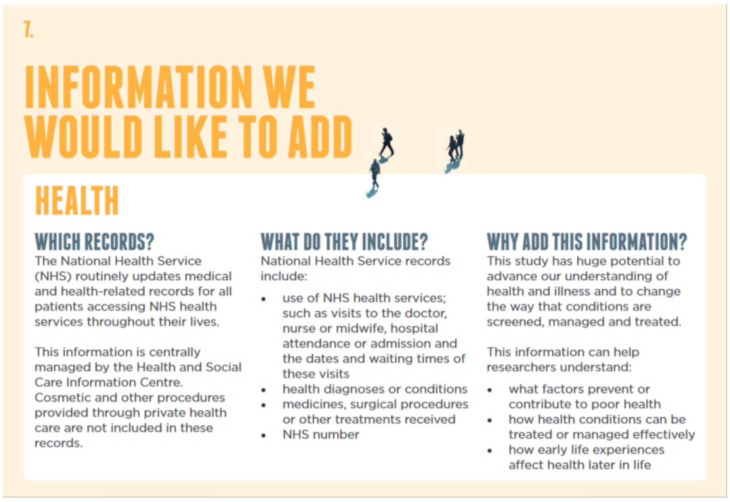
We learn a lot about your lives from the questions we ask in the surveys, but adding extra information from administrative records helps us to build a more complete picture of what life is really like for you.

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:

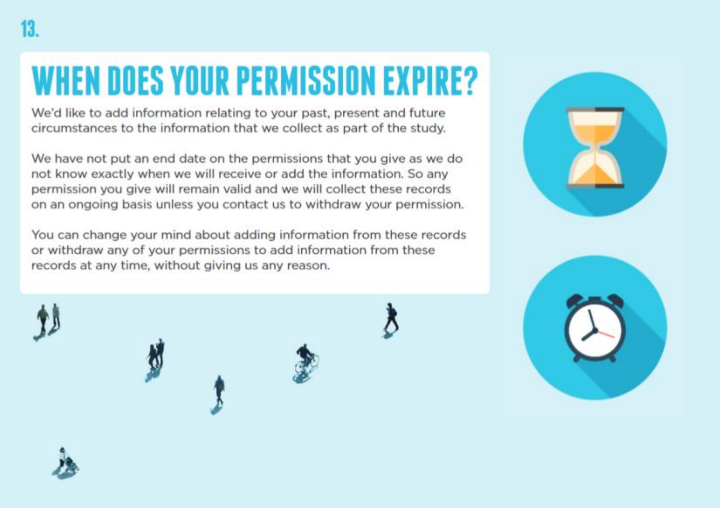
- use of NHS health services; such as visits to the doctor, nurse or midwife, hospital attendance or admission and the dates and waiting times of these visits
- health diagnoses or conditions
- medicines, surgical procedures or other treatments received
- NHS number

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.

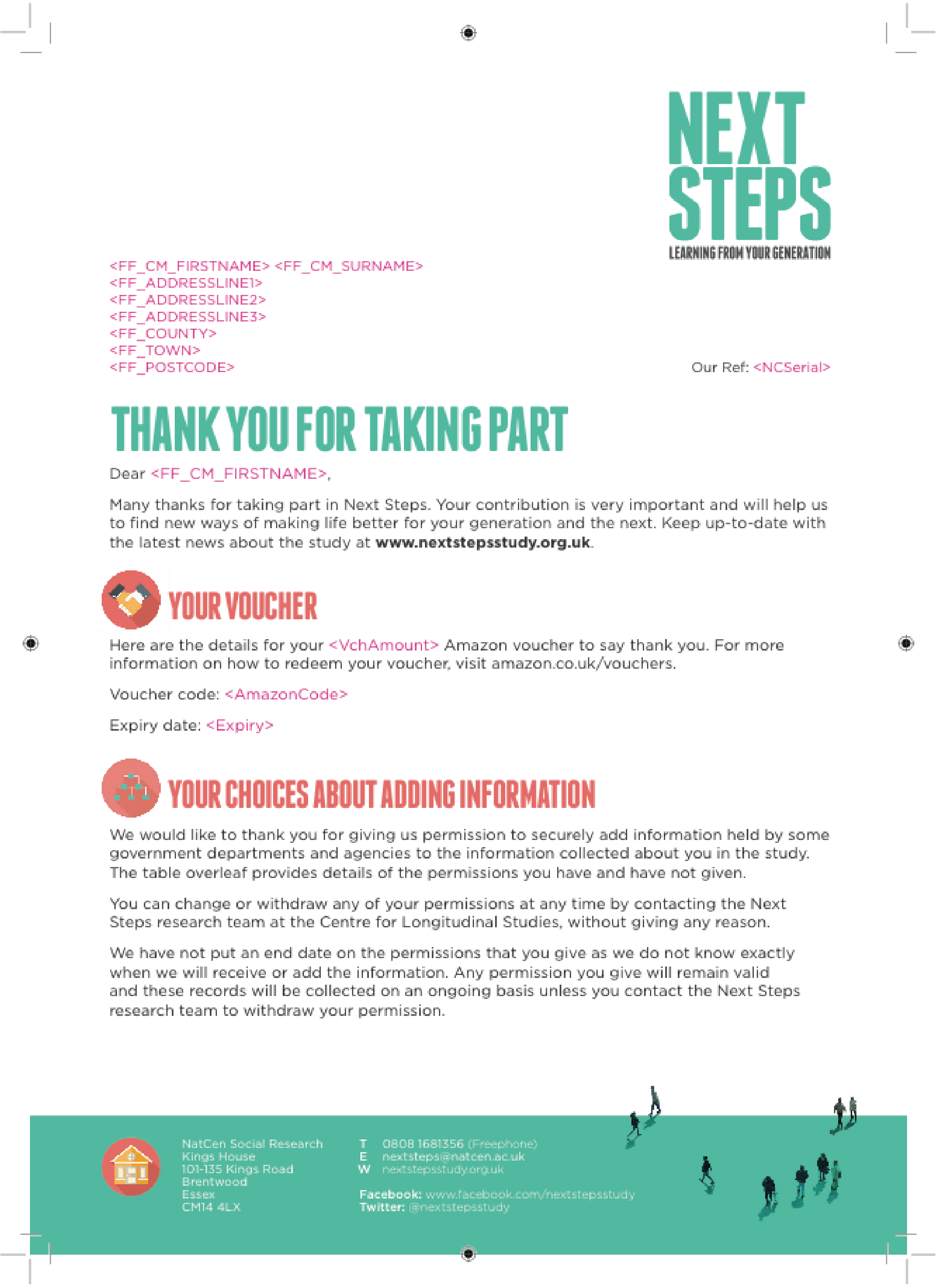
Your details can not be used to look at whether you are paying the right amount of tax or loan repayments, claiming the right type of benefits or by the police to find things out about you that they do not know.

WATCH OUR VIDEO TO FIND OUT MORE

www.nextstepsstudy.org.uk

FIND OUT MORE AT...
nextstepsstudy.org.uk

Figure A2: thank-you letter including confirmation of data linkage consents



NEXT STEPS
LEARNING FROM YOUR GENERATION

<FF_CM_FIRSTNAME> <FF_CM_SURNAME>
<FF_ADDRESSLINE1>
<FF_ADDRESSLINE2>
<FF_ADDRESSLINE3>
<FF_COUNTY>
<FF_TOWN>
<FF_POSTCODE>

Our Ref: <NCSerial>

THANK YOU FOR TAKING PART

Dear <FF_CM_FIRSTNAME>,

Many thanks for taking part in Next Steps. Your contribution is very important and will help us to find new ways of making life better for your generation and the next. Keep up-to-date with the latest news about the study at **www.nextstepsstudy.org.uk**.

YOUR VOUCHER

Here are the details for your <VchAmount> Amazon voucher to say thank you. For more information on how to redeem your voucher, visit amazon.co.uk/vouchers.

Voucher code: <AmazonCode>

Expiry date: <Expiry>

YOUR CHOICES ABOUT ADDING INFORMATION

We would like to thank you for giving us permission to securely add information held by some government departments and agencies to the information collected about you in the study. The table overleaf provides details of the permissions you have and have not given.

You can change or withdraw any of your permissions at any time by contacting the Next Steps research team at the Centre for Longitudinal Studies, without giving any reason.

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. Any permission you give will remain valid and these records will be collected on an ongoing basis unless you contact the Next Steps research team to withdraw your permission.

NatCen Social Research
Kings House
101-135 Kings Road
Brentwood
Essex
CM14 4LX

T 0808 1681356 (Freephone)
E nextsteps@natcen.ac.uk
W nextstepsstudy.org.uk

Facebook: www.facebook.com/nextstepsstudy
Twitter: @nextstepsstudy

WHERE CAN I FIND OUT MORE?

For any **queries about your voucher** please contact the NatCen research team at nextsteps@natcen.ac.uk, or on Freephone 0808 1681356.

For any **questions about adding other information or any other aspects of Next Steps**, please visit nextstepsstudy.org.uk, or contact the Next Steps research team at the Centre for Longitudinal Studies at nextsteps@ioe.ac.uk, on Freephone 0800 977 4566 or by writing to us at FREEPOST RTHR-TUGG-UTCR, NEXT STEPS, Centre for Longitudinal Studies, UCL Institute of Education, 20 Bedford Way, London WC1H 0AL.

CHANGE OF CONTACT DETAILS

Please also find enclosed a change of details card. If you move in the future, or if any of your other contact details change, please fill this in and return it to us. You can also update your details on the study website www.nextstepsstudy.org.uk .

L Calderwood

Lisa Calderwood
Study Director, Next Steps

PERMISSION TO ADD INFORMATION HELD BY GOVERNMENT DEPARTMENTS AND AGENCIES

National Health Service (NHS)	<NHS>
Department for Education (DfE)	<DfE>
Department for Business Innovation and Skills (BIS)	<BIS>
Higher Education Statistics Agency (HESA)	<HESA>
Universities and Colleges Admission Service (UCAS)	<UCAS>
Student Loans Company (SLC)	<SLC>
Her Majesty's Revenue and Customs (HMRC)	<HMRC>
Department of Work and Pensions (DWP)	<DWP>
Including provision of your National Insurance Number to DWP/HMRC/SLC	<NINOFlag>
Ministry of Justice (MoJ)	<MoJ>

Scan the code below with your smartphone to visit our website.

NatCen Social Research
Kings House
101-135 Kings Road
Brentwood
Essex
CM14 4LX

T 0808 1681356 (Freephone)
E nextsteps@natcen.ac.uk
W nextstepsstudy.org.uk

Facebook: www.facebook.com/nextstepsstudy
Twitter: @nextstepsstudy

NatCen
Social Research

P10591/TYLetter

References

Al Baghal, T. and J. Burton (2016). Does interviewers’ attitudes towards sharing personal information affect the consent rate they achieve? Background paper for the 5th Panel Survey Methods Workshop 2016

Fulton, J. A. (2012) Respondent consent to use administrative data (Unpublished PhD thesis). University of Maryland. http://hdl.handle.net/1903/13601

Jenkins, S. P., Cappellari,L., Lynn, P., Jaeckle, A ., & Sala, E. (2006). Patterns of Consent: Evidence from a General Household Survey Journal of the Royal Statistical Society (Series A) 169, 701-22.

Jenkins, S. P., Lynn, P., Jäckle, A., & Sala, E. (2008). The feasibility of linking household survey and administrative record data: New evidence from Britain. *International Journal of Social Research Methodology*, 11 ,29 –43.

Knies, G., Burton, J., & Sala, E. (2012) Consenting to health record linkage: Evidence from a multi-purpose longitudinal survey of a general population. *BMC Health Services Research*, 12, 52.

Korbmacher, J.M., and Schroeder, M. (2013) Consent when linking survey data with administrative records: The role of the interviewer. *Survey Research Methods* ,7, 115-131.

McKay, S. (2012) Evaluating approaches to family resources survey data linking. DWP working paper, no 110. London: Department for Work and Pensions.

Pascale, J. (2011). Requesting consent to link survey data to administrative records: Results from a split-sample ballot experiment in the survey of health insurance and program participation (SHIPP). U.S. Census Bureau Survey Methodology Study Series; 2011-03.

Sakshaug, J. W., & Kreuter, F. (2012) Assessing the magnitude of non-consent biases in linked survey and administrative data. *Survey Research Methods*, 6 ,113 – 122.

Sakshaug, J. W., Couper, M. P., Ofstedal, M. B., & Weir, D. (2012) Linking survey and administrative records: Mechanisms of consent. *Sociological Methods and Research*, 41, 535 – 569.

Sakshaug, J. W., Tutz, V., & Kreuter, F. (2013) Placement, wording and interviewers: Identifying correlates of consent to link survey and administrative data. *Survey Research Methods*, 7,33 – 144.

Sala, E., Burton, J., & Knies, G. (2012) Correlates of obtaining informed consent to data linkage. Respondent, interview, and interviewer characteristics. *Sociological Methods Research*, 41, 414 – 439.

Sala, E., Uhrig, N., & Lynn, P. (2011) It is time computers do clever things! The impact of dependent interviewing on interviewer burden. *Field Methods*, 23,3 – 23.