NatCen
Social Research that works for society

Public Attitudes to Data Linkage

A report prepared for University College London by NatCen Social Research

March 2018

Table of Contents

Executive summary

This report describes the findings from deliberative discussion groups on public attitudes to data linkage commissioned by UCL (University College London). The overall objective was to explore people's understanding and perceptions of data linkage, particularly linkage between health examination survey data and administrative records. The findings will help inform approaches to linking data for research, and communication strategies with survey participants regarding data linkage.

Three deliberative discussion groups were carried out between November 2017 and January 2018. Respondents had all previously taken part in the 2016 Health Survey for England (HSE) and included a mixture of individuals who had and had not provided consent for data linkage at the time of their HSE interview.

Understanding data collection and linkage

Respondents were very aware that their data is constantly being collected and used for a variety of purposes (examples of data collectors included local councils, the passport office and travel card providers). In general, individuals could see the benefit of providing personal data *if there was either a personal or societal benefit to doing so*. Respondents felt that data collection was useful and important in some circumstances, for instance for government planning, but that it was misused at other times (e.g. personal details being passed on to third parties for commercial purposes).

Respondents' initial level of understanding of data linkage varied, and understanding changed substantially throughout the course of the discussion. There were some individuals who had a clear understanding of what data linkage entailed whereas others' understanding was poor. There were respondents who did not have a clear understanding of what data linkage really meant and others had an incorrect understanding. For many, understanding grew throughout the course of the discussion, but this was not universal. However there was a group of individuals who were not able to distinguish between data linkage and the sharing of personal contact details between organisations.

A number of concerns were raised regarding the data linkage process. There were mixed views on whether **consent** should be sought on an on-going basis for each use of the data. Respondents also questioned the **de-identification** process and whether the data could be completely anonymous. There was also concern around the **secure transfer** of data**.**

Several factors underpinned respondents' views and concerns regarding data linkage.

- **Trust and legitimacy of organisations**. Respondents reported that they were more likely to consent and be at ease with data collection and linkage if it was being undertaken for societal benefit.
- **Timeframe for consent to data linkage.** There was a sense across all three groups that consent to data linkage specifically had a 'shelf life' and that while individuals may have prospectively agreed to data linkage, this decision may change over time.
- **Transparency**. Findings suggest that individuals felt more at ease about providing their details for data collection and data linkage if they knew specifically what it was going to be used for.

Linking HSE data to administrative health records

Recollection of providing consent, or not, to data linkage during the 2016 HSE interview was generally poor. However, individuals who did not remember providing consent were not

concerned that they had done so. Similarly, recall was poor about *why* they had or hadn't consented to data linkage at the time. Respondents suggested reasons for agreeing to data linkage, which included the perceived **societal benefits** (particularly in the health sector), assurance that the **linked data would not identify them** and trust in NatCen as a **reputable organisation**. Reasons suggested for not agreeing to data linkage included the lack of detail about exactly **how the data would be used** and the **indefinite timeframe** for which it could be used.

Respondents had a mixture of views regarding the current difficulties researchers face when attempting to process and use linked data. Some felt reassured that the process was being reviewed and that appropriate procedures and security measures were being put in place. Others said that the restrictions were frustrating because no-one was able to benefit from the linked data. There were also individuals who explained that knowledge of the current restrictions added to their existing cynicism around appropriate data protection and suggested that the delays were probably due to a data breach.

Case study (HEAT): linking HSE data to other data sources

UCL's Energy Institute and Department for Epidemiology and Public Health have secured time-limited funding to carry out analysis of linked HSE data to examine the Health Impacts of Energy Efficiency (HEAT).

Following a short presentation about the project, respondents held generally positive views about the idea of using their HSE data, linked to other data sources, for the HEAT study. Individuals across the groups recognised the wider societal benefits of the study and some felt that they might benefit directly, through government policy decisions regarding energy and housing. Respondents also said that the explanation of the process of data linkage in this specific study reassured them that appropriate controls were in place to protect their personal data. A group of respondents did raise concerns about data linkage for the HEAT study. In particular, there were some questions about the potential future use of the linked data for other purposes and scepticism around who might be provided with access in the future.

The HEAT project proposes to link administrative datasets that participants did not provide specific consent for during the HSE interview, namely **Met Office data** about outdoor temperature (matched at a geographical level based on the date of HSE participation), and the **National Energy Efficiency Data-Framework** (NEED), which contains information about energy efficiency measures installed in homes (and matched at the address level). Respondents were asked to consider these two sources of data separately.

Respondents understood the rationale for linking both these sets of data onto their HSE survey answers. On the whole, respondents were more comfortable with the idea of the Met Office data being linked than NEED. Respondents raised concerns about who would have access to the information about their home's energy efficiency and about whether they would be identifiable from this data source.

Views regarding consent to match on these data sources varied. One view was that it was not necessary to ask HSE participants for consent to add these additional administrative datasets, but some of these individuals noted that this view was a result of having learned more about data linkage during the course of the deliberative discussion. Other respondents felt that other HSE participants should at least be informed about linkage of additional sources and offered an option to opt out. Finally there was a group who said that they would not want to be notified every time a researcher wants to link a new source of administrative data, for fear of being inundated with notifications and consent forms.

Data linkage communication

Understanding of data linkage was generally very low, even among survey participants who have provided consent for their data to be linked. Any communication regarding the topic must explain the reasons for linking data and the process of doing it very clearly.

A brief exploration at the end of the deliberative discussion groups allowed respondents to provide spontaneous suggestions about what information should be included in a public notice (i.e. on the survey's website). Respondents suggested the following ideas:

- Details of individual research projects that will be using the linked data including how and why the data is being used.
- The names of organisations and individuals who have access to linked data.
- Specify the length of time data can be linked and how long the linked data is stored
- Provide reassurance that individuals will not be contacted by third party organisations

Respondents suggested that **real examples** of data linkage helped them understand why and how it was done. **Visual information**, such as the animation and diagrams used during the deliberative aspects of the discussion, were useful for understanding the linkage processes. They also suggested that they would like **written notifications**, by mail or email, to inform them what is being done with their data.

1 Introduction

Background

This report presents evidence from research exploring Health Survey for England (HSE) participants' attitudes to data linkage.

Data linkage is the process of joining together two or more streams of data that relate to the same person, household, area or event. Data linkage can refer to matching two or more administrative datasets together (for example, health records and benefit receipt records) or, more commonly in social research, the linkage of administrative records onto social survey data. Linked data allows researchers to make use of additional information that was not collected in the original surveys.

UCL's Energy Institute and Department for Epidemiology and Public Health have secured time-limited funding from the Department for Business, Energy, and Industrial Strategy to conduct a study called Health Impacts of Energy Efficiency (HEAT). The HEAT study involves secondary analysis of HSE data and, in particular, linking health data from HSE to other data, such as Met Office and home insulation data, to explore issues around the link between temperature and health.

The three specific aims of the HEAT study are to assess:

- Relationship between energy use and energy performance of housing on health outcomes, such as general health, cardiovascular and respiratory conditions;
- Impact of a range of energy efficiency retrofits and eligibility for fuel payments on health outcomes; and
- Associations between indoor temperature and health outcomes.

Before any data linkage can take place UCL need to obtain the permission of NHS Digital, the arms-length body of the Department of Health and the HSE data controller, to make use of HSE data. During initial discussions, NHS Digital's Confidentiality Advisory Group (CAG) has expressed concern over the public's and HSE respondents' understanding about data linkage and their views on its acceptability.

Previous qualitative research, conducted by Cameron et al (2014)[1], explored the public's understanding and views of administrative data and data linking and found participants in the study had a low level of awareness and understanding of social research and data linkage. Particular concerns about data linkage were driven by worries about data security and importance of de-identification, ensuring data is stored and transmitted securely, and that data linkage is being undertaken for socially beneficial purposes.

Similar findings were presented in a report published by the Scottish Government (2014)[2] which also explored the public's attitudes to data linkage using deliberative groups and found

[1] Cameron, D., Pope, S., Clemence, M. (2014) Exploring the public's views on using administrative data for research purposes. Ipsos Mori.

[2] The Scottish Government. Public Acceptability of Cross-Sectoral Data Linkage, 2012. Accessed on 14th February 2018. Available online: http://www.gov.scot/Publications/2012/08/9455

on the whole participants were positive about data linkage as long as it was done for social purposes such as for medical advancement or to improve services.

There has however been limited research focusing on public attitudes to linking health data specifically.

UCL commissioned NatCen Social Research to conduct qualitative research with HSE participants to explore their understanding of and attitudes towards linking health and lifestyle data collected during HSE with other data.

Objectives of the study

There were three specific research objectives of the research:

- To map perceptions and understanding of data linkage and, in particular, data linkage involving health records and survey data from HSE;
- To identify concerns and perceived benefits of linking data that is: specific to them as an individual (e.g. health record data) or to their household (e.g. the presence of loft insulation, Smart meter data); information that is not specific to them (e.g. external temperature); and information that can be used for studying national policies but is not directly related to health (e.g. which type of energy efficiency intervention affects indoor temperature best); and
- To generate suggestions for what information should be put on the HSE website and identify the factors that would help persuade people to agree to health (and other) data linkage.

Methodology

A qualitative approach using deliberative discussion groups was used. This method was chosen for two reasons:

- First, the use of discussion groups allowed for a greater range of views to be explored and allowed for individuals to be prompted by the views of others in the group. This is particularly useful when exploring potentially complex subjects such as data linkage.
- Second, the use of a deliberative approach, rather than a traditional focus group was chosen because of the complexity and abstract nature of data linkage. Deliberative events allow for the researcher guiding the discussion to present specific information about a topic area to inform individuals and facilitate a discussion around information presented.

Three deliberative discussion groups were conducted in two locations with HSE 2016 participants.

A summary of the sampling and recruitment of participants is detailed below. The topic guide used to guide the discussion can be found in Appendix A.

Sampling and recruitment

The sample frame used to recruit from included adult respondents to HSE 2016. The HSE is a series of annual surveys and is conducted with a random sample of 10,000 individuals (8,000 adults, 2,000 children) from the general population of England each year. In order to capture a range and diversity of views across two different locations, a sample was selected

on the basis of geography and whether participants had consented to having their health data linked during their participation in HSE 2016. The following sample frames and participant characteristics were drawn on for recruitment:

- HSE 2016 participants in Birmingham who consented to data linkage
- HSE 2016 participants in London who did not consent to data linkage
- HSE 2016 participants in London who consented to data linkage.

Selected participants were sent an advance letter providing them with the opportunity to opt-out of being contacted about the research. A maximum of one adult per household was selected from the HSE participants (whereas multiple household members may have participated in the HSE). A one week opt-out period was given, after which calls were made by the research team to invite participants to take part in a discussion group.

Initially, priority calls were made to respondents who had received a visit from the nurse during HSE. As the HEAT study is linking data collected via a nurse it was felt appropriate to focus recruitment efforts on those who provided this data. Once this sample frame was exhausted the research team broadened recruitment calls to those who did not receive a nurse's visit.

The sample of respondents included participants from different genders and a range of ages. More details of the final sample can be found in Appendix B.

Data collection

Deliberative discussion groups were designed using a step by step approach to take participants on a journey which aimed to increase their knowledge and understanding of the issue, and also to gather views of participants as they became more informed about data linkage. Each discussion group followed the same structure:

- **Warm up**: background and introductions, views on data collection and why data is collected
- **Initial understanding of data linkage**: this involved exploring participants' pre-existing understanding of data linkage and the principles that underpin it
- **Explanation of data linkage**: this involved a NatCen video which explains the basic principles of data linkage as carried out at NatCen and government bodies
- **Informed discussion on data linkage**: this involved a participant discussion in light of data linkage video
- **Explanation of the HEAT study:** this involved a presentation from UCL which demonstrated which health data it proposes to link with other data and how UCL proposed to undertake data linkage
- **Informed discussion on HEAT**: this involved exploring participants' views on the HEAT study
- **HEAT Scenarios**: participants were given scenarios that drew on the specific ways the HEAT study will link data to explore specific views and attitudes toward the work being conducted.
- **Recommendations:** this involved exploring ideas and recommendations on what information about data linkage should be presented on the HSE website

Discussion groups took place between November 2017 and January 2018 and lasted up to two hours. With permission, discussion groups were audio recorded and transcribed

verbatim. Each participant received £30 cash as a thank you to them for their time and contribution.

Data management and analysis

NatCen uses the Framework approach to manage qualitative data. Framework uses a 'matrix' approach that organises data according to both cases (e.g. deliberative discussion groups) and themes (e.g. views on linking outdoor heat temperature to HSE survey responses). This results in a series of thematic matrices, each representing one key theme. The column headings on each matrix relate to key sub-topics and the rows to each deliberative discussion group. Data from each case is then summarised ('charting') in the relevant cell, while retaining participants' language.

Analysis within the Framework approach involved interpreting the managed data and providing explanations. The process of interpretation had three steps.

- **Detection:** This involves familiarisation with the managed data and extracting the full diversity of answers to the analytical question.
- **Categorisation**: Meaningful conceptual boxes are then created and all the data detected in the previous step is assigned a category.
- **Classification**: Similar to categorisation, this creates higher order categories and assesses the relationship between categories.

Interpreting the findings

Verbatim quotations are used to illuminate findings. They are labelled to indicate which group participants attended. Further information is not given in order to protect the anonymity of research participants. Quotes are drawn from across the discussion groups.

The report avoids giving numerical findings, since qualitative research cannot support numerical analysis. This is because purposive sampling seeks to achieve range and diversity among sample members rather than to build a statistically representative sample, and because the questioning methods used are designed to explore issues in depth within individual contexts rather than to generate data that can be analysed numerically. What qualitative research does do is to provide in-depth insight into the range of experiences, views and recommendations. Wider inference can be drawn on these bases rather than on the basis of prevalence.

2 Understanding data collection and linkage

This section first outlines individuals' understanding and views of the types of personal data collected, organisations that collect it, and the perceived reasons why personal data is collected and used. It then discusses individuals' initial comprehension, understanding and views on data linkage, including before and after stimulus was used to explain the purpose and method of data linkage.

Initial comprehension and understanding of data collection and use

Exploring individuals' understanding of what personal data is collected, by who and why was important as it illuminated individuals' general attitudes toward the collection of personal data and tended to influence individuals' understanding and views of data linkage.

Individuals were aware that personal data was collected from a multitude of organisations, and that the collection and use of data was often outside of their control. For example, individuals discussed that in order to create social media profiles or apply for insurance, they were required to give organisations their personal information. They also felt they had limited control on how this data was used.

Views of data collection and use

Respondents were able to provide a long list of organisations that collect personal data and set out the various ways in which data was collected and potentially used. Some examples included:

- Travel data collected through Oyster cards and potentially held by Transport for London;
- Household composition and other data collected through council tax information and other sources, held by local councils;
- Passport and other information about individuals and their families, including travel and visa information, collected by Government departments; and
- Health information collected by charities such as Cancer Research through questionnaires with individuals who have a family history of cancer, to track health behaviours and outcomes over time.

Generally, individuals could see the benefit of data collection and were willing to share personal data if they could see the personal or societal benefit of it. Individuals were more willing and trusted some organisations to use their personal data appropriately (e.g. storing it securely and not sharing it with other organisations) than others. For example, the government collecting data via the Census was viewed as important as it facilitates government planning for the future. Also data collected by the NHS was considered important for similar reasons:

> *"I think they sometimes provide a better service, if they have a better understanding of who the patients are."*
>
> (Discussion group two)

On the other hand, individuals felt the use commercial organisations, such as supermarkets and marketing companies, made of personal data was less trustworthy, and they only wanted individuals' personal data to use for commercial gain. There was also a perception among individuals across the discussion groups that commercial organisations were not to be trusted with personal data and that they would share data with other organisations. This is highlighted by a respondent who explained that providing personal information to an organisation led to multiple calls from other organisations:

> *"[If] I fill out one survey and then my phone just never stops or you click yes for helping out and providers are sharing it and you're going to get all the other calls coming through."*
>
> (Discussion group one)

Comprehension and understanding of data linkage

As explained above, discussion groups were structured to take individuals on a journey aiming to increase their understanding and awareness of data linkage. During each group, individuals were first asked what they understood data linkage to mean and what the purpose of it is.

Individuals can be organised according to their initial and ongoing comprehension and understanding of data linkage. Comprehension represents both their initial understanding of the purposes and different stages involved in data linkage and their understanding once stimulus materials explaining the purpose and process had been provided.

Findings indicate that that across and within discussion groups, individuals fell into one of two groups in relation to their comprehension and understanding of data linkage – those with a good understanding and those with poor understanding.

Individuals with good understanding of data linkage

Prior to any stimulus or information about data linkage, there were individuals in this group who had a good pre-existing understanding of data linkage and were able to explain what they perceived the main purposes of data linkage were. In more limited circumstances, this group could also identify and explain the different stages involved in the data linkage process.

> *"Individual level, the data from one source can be matched to your data from another source and they can match it on an individual level."*
>
> (Discussion group three)

There were also individuals in this group who began the discussion group poorly informed about the purposes and stages involved in data linkage. This group's understanding and comprehension of data linkage and stages involved improved once they were provided with information from the facilitator. At the end of the discussion group they had a good understanding of the purpose and stages undertaken to link data.

Individuals in this group varied in age. Those with a good pre-existing understanding of data linkage had previously worked or were working in a research-related field. For example, one individual in this group was a retired academic.

Individuals with poor understanding of data linkage

Individuals with a poor understanding were not able to provide any insight on the purpose of data linkage and the stages involved prior to receiving any stimulus. In some circumstances, individuals in this group made incorrect assumptions about what data linkage meant based on the terminology. For example, one respondent thought data linkage meant using previous research to answer new research questions to save money and resources repeating a study:

> *"So if you're [going to carry] out this research and there's another group that are similar to yours and want to carry out similar research, they might be better off linking to your data as opposed to do it again themselves and spend the money and time and resources and getting the people and finding the people, [be]cause you've already done that."*
>
> (Discussion group two)

This group also confused data linkage with the sharing of personal contact details from one organisation to another. Respondents discussed concerns about this form of information sharing and highlighted the risks that data could be hacked if not shared securely. This concern related to the transfer stage of the data linkage process. Individuals who could not move beyond their initial misinterpretation were unable to distinguish between data linkage and the sharing of personal details between organisations. Researchers found no terminology that made a clear distinction between data linkage and sharing.

There were no discerning characteristics of the individuals in this group.

Impact of stimulus on comprehension and understanding

Researchers showed a video demonstrating the process of linking survey and administrative data to each discussion group, to provide respondents with an overview of the purpose and stages involved in data linkage.

In some circumstances, the video did not improve the comprehension of the stages involved in data linkage or its purposes for individuals with a poor understanding. This group continued to conflate the sharing of personal details and data linkage during discussions.

Respondents who had poor pre-existing knowledge, but eventually had a good understanding either explicitly expressed that they felt more informed or demonstrated a better comprehension through their involvement in further discussions about the stages involved in data linkage.

Overall, individuals from both groups could see the benefit of linking data within the context of it being done to benefit society (as shown in the video):

> *"If it was going to help other people, but not, you know if you were looking at the way people live and the way people … and also to give people a better life but not necessarily just shopping and things like that if you know what I mean."*
>
> (Discussion group one)

Views and concerns about the data linkage process

While respondents could see the benefit, people in both groups (with a good and poor understanding of data linkage) raised concerns about certain aspects of the stages of data linkage.

Consent

There were individuals who felt that in principle consent was essential before any data linkage could happen. The key reason respondents wanted to provide consent to having their data linked was so they could find out what the data linkage was being used for:

"I think they should state what they're going to use it for, because if you've not got any knowledge of what it's going to be used for, then you know, you, you may not agree […]. You may worry it's not what they could use it for. So if you know that they were going to use it for a specific thing and perhaps if they changed their mind later down the line and wanted to use it for something else they could just contact you and just say, 'Is that okay still?"

(Discussion group one)

There were also individuals that thought it important that there must be a process for which consent was always sought each time personal data was going to be used for data linkage, as it gave people the opportunity to reflect each time about whether they want to opt-out of specific data being linked. Respondents reported that their decision to having data linked might change over time:

"Because what you decide today, might not be what you want in a year's time, two years' time, or even tomorrow. You might change your mind and knowing that you can actually, right, I signed a consent form today, but actually I've changed my mind a week of, it's something that, that would put me at ease in to do that linking thing. Otherwise, if it was me, I would say, 'No way', but knowing that you can come back and say, 'Actually I've changed my mind, I want all my data to be scrapped', or whatever, I think that's a very good idea".

(Discussion group two)

De-identification

The discussion groups talked about anonymisation, rather than pseudonymisation, since the term is more commonly used. However technically linked survey data are pseudonymised. Definitions, according to the Information Commissioner's Office are as follows:

Anonymisation is "the process of turning data into a form which does not identify individuals and where identification is not likely to take place".

Pseudonymisation is "the processing of personal data in such a a way that the data can no longer be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution to an identified or identifiable person".

Respondents generally felt that "anonymisation" was a key part of the data linkage process and felt it was of great importance that this part happened at the earliest point possible. Individuals however found it difficult to articulate explicitly why this was the case.

A group of respondents with a good understanding of data linkage were (correctly) not convinced that individual records would not be fully anonymous, as they discussed that there might always be a dataset with their individual name on it:

"They exist as three different…separate bits of information so your name's always [going to] be associated with that data…it's just no one's actually held the three things together, isn't it? So [you] might see it as it's A, B and C, you know when you put A, B and C together, but you've still got A has still got your name on, B's got your name on, C's got your name on. So you're never fully anonymised."

(Discussion group three)

This demonstrates the need to use appropriate language when explaining the data linkage process to research participants. For example, rather than using the word 'anonymous', explaining that the *linked data* will not contain any information that can identify them and specifying the point at which personal details are removed from individual data sets.

Data transfer

A key concern across all discussion groups centred on the security of data while it was being transferred, the concern being that data may become vulnerable to hacking or get lost in transit. The views on this of individuals with a poor understanding of data linkage appeared to be shaped by news stories they had seen about hacking incidents that made data vulnerable to interception.

Data quality

Individuals with a good understanding of data linkage also raised concerns about the accuracy of the administrative data (for example, NHS hospital records will not have information about private health care users) and they questioned the representativeness of the linked data, suggesting that people who do consent may be fundamentally different to those who do not consent.

Factors underpinning views about data collection and linkage

There were three factors that underpinned individuals' views and concerns about whether they would make their personal details available for data collection and/or consent to data linkage:

- **Trust and legitimacy of organisations**. Respondents reported that they were more likely to consent and be at ease with data collection and linkage if it was being undertaken for societal benefit by a body such as the NHS. It was felt that if data was being collected to benefit society, those holding the personal data were more likely to take care of individuals' personal data and were less likely to share it with others.
- **Timeframe for consent to data linkage.** There was a sense across all three groups that consent to data linkage specifically had a 'shelf life' and that while individuals may have prospectively agreed to data linkage, this decision may change over time. Changes to individuals personal circumstances such as their health, may impact on whether or not they want their data linked.
- **Transparency**. Findings suggest that individuals felt more at ease about providing their details for data collection and data linkage if they knew specifically what it was going to be used for.

3 Linking HSE data to administrative health records

This section discusses individuals' recall and reasons for their decision to consent, or not, to data linkage during their participation in HSE 2016. It then goes on to set out individuals' views on the delays in linking HSE data with other data.

Recall on consenting to data linkage

To stimulate discussion around consenting to data linkage, respondents were given copies of the data linkage consent form usually given to respondents towards the end of the HSE interview. Across the three discussion groups recall around completing this consent form was poor for both individuals who had consented and had not consented to data linkage.

On the whole, individuals had just assumed they had consented to having their data linked, this included people who had not consented during participation in HSE 2016. Individuals who were more likely to recall the consent form were those with a good understanding of data linkage.

Consenting to data linkage during HSE 2016

Consenting to data linkage

While recall for consent was poor, those who could remember consenting gave the following reasons as to why they agreed:

- A perception that their personal data would be used for a wider purpose to benefit society, for example future planning within the health sector;
- Reassurance that their data would be linked anonymously; and
- That they trusted the interviewer conducting the survey and/or felt NatCen was a reputable organisation that they trusted.

Not consenting to data linkage

As explained above, recall amongst those who had not consented to data linkage was poor and within this group there were individuals who thought they had agreed to it when they hadn't. For those who could recall disagreeing to having their data linked, the main reason for this was because the list of ways data might be linked was not comprehensive enough:

> *"It gave examples of how, of the type of data that might be linked, or the purposes that it might be linked for, but examples can never be exhausted, exhaustive list. So it said, you know, it might be used for this the purpose, or such stuff for data, including, but not limited to, type of language, and I thought, okay, that's not really the whole list and I'm not sure where it might end up."*
>
> (Discussion group two)

Reviewing the consent form again stimulated discussion among individuals who felt that there should not be an unlimited time period for storing personal data for the purposes of data linkage. As explained above, individuals felt that a defined time period and process would reassure them in giving consent. Respondents were favourable towards the concept of having flexibility to choose which datasets they did or didn't want to provide consent for (as indicated in the data linkage video where health, education and benefit records were separate consents).

Views about delays in data linkage

During discussion groups researchers explained that there had been delays in the ethical approval to allow HSE data to be linked to other forms of health data.

Views about the delays to linking data were divided among respondents. One the one hand there were individuals who considered this delay a **positive step** in ensuring that the process is ethical and appropriate, especially as the original dataset holds personal and confidential information. Individuals reflected that it gave them "*peace of mind*" that their personal data was going to be used appropriately.

> *"It's not a bad thing though is it really? I don't think so. [An] ethics committee is a safeguard that things are being done as they should be done."*
>
> (Discussion group one)

Others found the restriction **frustrating** and reported that data collection could potentially be considered "*a waste of all that time and resources"* as no-one is benefitting from the linked data.

> *"Absolutely, [be]cause it's like saying, yeah, you have all my information, collect it all, but just keep it sitting there for however long it's [going to] be. Three years, five years, ten years. So another way of putting it, for me is, our opinions have been taken because you need, you need the information, but that information is just [going to] be s[a]t there, stagnant, with nobody benefiting from that information and by the time they do benefit, it could be old information […] so it could be useless."*
>
> (Discussion group two)

Finally there was a group who felt that the delays added to a sense of **cynicism** that their data was potentially **not being protected** or used appropriately. Even after reassurances by researchers that the delays were not due to any breaches in data security; this did not fully allay respondents concerns.

Information about the delays to linking data stimulated discussions about the reason why there had been delays in the permissions to link the data. The following reasons were speculated:

- Potential policy changes may have caused delays;
- There had been changes to ethical approval process; and
- That there may had been a previous mistake during data linkage or the application process which triggered the delay.

4 A case study (HEAT): linking HSE data to other data sources

This section discusses individuals' views of the HEAT study and initial reactions to how individuals' health data will be used within the study.

The final deliberative element of the discussion group involved a presentation of the proposed HEAT project. The purpose of this was to provide context and meaning to the more conceptual discussions individuals had been having about data linkage at the beginning of the discussion groups.

Views on the HEAT study

Positive views of the study

Overall both those who had good and poor understanding of data linkage were positive about linking their health data collected via HSE to inform the HEAT study objectives. There appeared to be two key reasons for the positive viewpoint. First, individuals recognised the wider societal benefits of the study:

> *"I think in terms of the energy efficiency and the cost-effectiveness, I think it is, it needs to happen. Purely because if the deaths are not going down as fast as they could, or possibly could, well we need to find ways of preventing them and if this sort of information is what does that then it's, it's a good thing, because you, as you said, you'll pass it on to Age UK, local governments, then maybe yes, right, we need to step up our game, I find, it difficult to argue against."*
>
> (Discussion group two)

Personal benefits of the study were also realised. For example, one respondent identified a personal experience relating to heat and health and felt their family members would be likely to be those who directly benefit from government and local authorities making use of the research findings.

Second, further insight into the actual data linkage process, specifically that one individual is responsible for each step of the process allayed concerns individuals had regarding data linkage (see section 2 for more details):

> "*I actually kind of like that it's only one person dealing with the separation, because then it, this, there's less likely for human error*."
>
> (Discussion group two)

The main concern here was that having a bigger group of people linking the data might leave more room for error and lead to data breaches.

Concerns about the HEAT study

There were no concerns about the principles, and objectives and indeed the purpose of this example of data linkage. Concerns centred on how the research might be used in future and there were also some concerns raised about the methodology used (for example, the coverage of the administrative data sets).

There were a group of individuals who were sceptical about whether any government or policy changes would be made based on the findings. This included people from both groups who had a good and poor understanding of data linkage. Those who had a poor understanding raised another concern that their personal health data might be shared with other charities or government departments as a result of data linkage. This type of concern highlights the continued misinterpretation by some individuals of what data linkage meant.

Views on HEAT scenarios

During the discussion groups individuals were given two different scenarios. The scenarios set out how the HEAT study will link HSE health data with external temperatures and information specific to individuals' homes, as well as the reasons for these types of data linkage (for more detail on scenarios see Appendix A). The purpose of the scenarios was to provide specific examples of how the HEAT study will work in practice, to draw out any further views and attitudes on data linkage to geographical and household level data.

Scenario one – linking home temperatures with external temperatures

Overall respondents could see the benefits of linking the temperature within their home with external temperatures. Views about both the purpose and the process of data linkage differed between individuals who had a good and poor understanding of data linkage.

Poor understanding of data linkage

Individuals in this group were happy with this scenario in principle and did not think explicit consent was necessary to link home and external temperatures. However, respondents explained that they held this view because they had engaged in the discussion group and felt more informed about the purpose of the study and data linkage. Respondents did feel that other HSE survey respondents should be informed of this additional data linkage, so that they have an awareness of how their data is being used beyond the original consent form and have an option to opt out:

> *"Not in this case because we just had it explained to us, but other people that haven't been involved in this discussion group they might want to know. They might just want you to send them, I don't know just an email or a leaflet just to say, 'This is what we want to do and would you be happy with it?"*
>
> (Discussion group one)

Good understanding of data linkage

Individuals with a good understanding of data linkage were more wary of linking HSE data with external temperatures, but on the whole were supportive of the idea. Amongst this group views were divided on whether it was necessary to seek explicit consent. Those who felt individuals should be informed were influenced by the earlier discussions that data linkage processes had been delayed being approved by ethics. Although it was explained this was not down to any wrong doing by the researchers, this was still raised as an issue. Others felt there was no need to ask for explicit consent because of the nature of the administrative data the researchers will be linking.

A sub-set of this group included individuals with a stronger understanding and comprehension of data linkage who were not concerned about consent and reported they did not want to be overloaded with data linkage requests. They also raised two methodological concerns. First, that seeking further requests may reduce the sample size of

HSE respondents who agreed to data linkage. Second that it could lead to a biased sample, because not all respondents will be contactable, due to the likelihood of some moving address or dying. Respondents felt both these issues would be detrimental to the study.

Scenario two – linking HSE data to specific information about individuals homes

There was less of a consensus across and within discussion groups that individuals would support this linkage and also whether they felt explicit consent was necessary. It was also not possible to divide attitudes by individuals' level of comprehension and understanding of data linkage.

Across all individuals, there were two concerns about the scenario that centred on the stages of data linkage that would be involved:

- First, individuals reported that they would be concerned about who had access to the personal information about their home's energy efficiencies. Individuals felt that they were more likely to agree now that they have been provided with further information and can identify the individuals involved but without this information they would not agree to it taking place.
- Second, there was concern about how personal the individual home data would be. Individuals felt assured that as long as the de-identification process happened they would be comfortable consenting to this. Particular wording was also recommended, such as your 'specific address' rather than 'home' to prevent other survey respondents from worrying about their personal information.

Individuals also had two further concerns that related to the methodology proposed in the scenario:

- First, individuals suggested that other HSE respondents could have made changes to their home without using a government scheme.
- Second, individuals were concerned about the accuracy of the data and the sample size in terms of the specific research project. When discussing consent, this group were satisfied knowing that the additional data was coming from publically accessible datasets rather than direct communication with their energy providers.

5 Data linkage communication

The final section summarises individuals' views on information needed for the HSE website to help inform current and future survey participants about data linkage.

It was clear that participants' general understanding of data linkage and the processes involved was generally minimal at the outset of the deliberative discussions. This is despite the fact that they had all taken part in HSE during which they were asked to provide their permission to link their survey data to health records.

Respondents were asked to suggest what information they believed should be included on the HSE website about data linkage, as well as the format in which it should be included to help people learn and become more informed about data linkage:

Views on what should be included:

- **Explicit details on how, why and which individual research projects will be using the linked data.** Individuals felt that the more transparency about data linkage the better as it would provide individuals with reassurances that ethical and data security principles were being adhered to.
- **Provide details of the organisations and individual researchers**. Individuals reported that they wanted to know explicitly who would be handling the data and conducting the research, as well as having an explanation of the organisations involved. Again it was felt this added to the reassurance that the process would be done ethically and robustly.
- **Identify the length of time that the data is stored for and linked**. Individuals were concerned about the indefinite amount of time that the data is kept, as people could forget that this data exists about them and that it has been linked to other records about them. This links back to the need for organisations to be transparent about how personal data is being used. It was felt that if timeframes were clear, individuals might be more likely to agree to data linkage
- **Provide comprehensive information on the ways data might be linked with other data.** A barrier to consenting to data linkage was that information was limited about the different ways HSE data might be linked with external information. Having a more comprehensive overview of the ways data linkage will be linked might encourage more people to consent during participation in HSE.
- **Provide reassurance that individuals will not be contacted by other organisations.** Respondents feared that once they provided information or consented to an additional request, they would be overloaded with further requests in the future, which was considered a nuisance. If there were details on the HSE website which explicitly stated this would not happen, it may encourage consent.

Suggested formats for information sharing:

Below are some of the ways respondents would like to have information about data linkage communicated to them:

- Explicit examples or scenarios (like those used during the discussion groups) of what data is going to be linked and how. These provide real-life examples for people to relate to.

- Visual information such as the presentation slides. Individuals liked the visual demonstration used during the presentation of the HEAT study and thought these could be included on the HSE website.
- Contact individuals directly with emails or letters to inform them of what is happening with their data.

Appendix A. Research materials – discussion group topic guide

Introduction to data collection

- Warm up exercise – split the group into pairs, two minute discussion to meet the other person and then introduce them to the group (name, where they live, one interesting fact about them)
- Explore types of organisations that might collect data
  - Government
  - Research agencies
  - Charities
  - Private businesses
  - Other
- Views on why different types of organisations collect data
  - Inform policy
  - Decision making
  - Advertising / marketing
  - Other
- Views on the different ways organisations collect data
  - Social media
  - Research e.g. surveys / polls
  - Application / monitoring forms
  - Other

Understanding and views on data linking

- Initial understanding of data linkage
  - Purpose
  - Benefits
  - Concerns
  - Perceived risks
- SHOW ANIMATION

https://www.youtube.com/watch?v=JRRRPUXQWbc

Interviewer note: explain that the Health Survey only asked to link to health records, not education or benefit records

Interviewer note: share example consent form with participants as stimulus for next discussion.

- Explore reasons why **did** give permission to have data linked after completing HSE survey
  - Confidence in future use
  - Benefit to wider society
  - Benefit future research
  - Reasons for feeling reassured, if applicable
  - Any potential concerns, explore each in turn
  - Other
- Explore participants views on current restrictions or delays for researchers wanting to use linked data
- Explore reasons why **did not** give permission to have data linked after completing HSE survey
  - Data is personal
  - Concern about data security
  - Concern data will be lost, sold or shared
  - Other
- Views on each stage of data linkage:
  - Consent,
  - De-identification,
  - Transferring data,
  - Future use of data
- For each explore stage:
  - Issues or concerns
  - Specific barriers / enablers to consenting
  - Elements of process less concerned about

Understanding data linkage using HEAT

- Present UCL slides
- Explore views on linking health data
  - Understanding of the purpose
  - Perceived benefits
  - Any initial concerns
  - Any perceived risks

In order to do this, NatCen would link publically available data from the Met Office (about the weather in your local area) to participants' HSE data.

To be able to do this:

- NatCen would use a unique ID plus your postcode and the date 'you' took part in the survey to find out the outdoor temperature at the time of day 'you' were seen by the nurse
- We would remove your address details and interview date
- We would then link the external temperature data to your survey answers

Researchers in the HEAT team at UCL would analyse the anonymised data and would not be able to identify who any of the participants were.

- Explore initial views on this type of data linkage
  - Whether something they would consent to
  - If not, why
- Views on each stage of data linkage:
  - Consent
  - De-identification
  - Transferring data
  - Future use of data
- For each stage explore:
  - Issues or concerns
  - Specific barriers / enablers to consenting
  - Whether lack of explicit consent is problematic
  - Elements of process less concerned about
- Explore information needed on HSE website to help inform participants of data linkage option
  - Format of information
  - Level of detail
  - Most important elements to include
  - Anything would not include
- Whether in principle if information discussed is shared they would consent to this type of data linkage
  - Enablers
  - Main concerns/ barriers
- The process for linking this data would be similar to the information about outdoor temperature. In this case, we would need the HSE participant's exact address but only an approximate date of the nurse visit not the exact date).
- Explore initial views on this type of data linkage
  - Whether something they would consent to
  - If not, why
- Views on each stage of data linkage:

  - Consent
  - De-identification
  - Transferring data
  - Future use of data
- For each stage explore:
  - Issues or concerns
    - Data loss
    - Identification
    - Security
  - Specific barriers / enablers to consenting
  - Whether lack of explicit consent is problematic
  - Elements of process less concerned about

Information requirements on HSE website

Interviewer note: For each element of information needed as participants to call it out and write it down on flipchart paper.

- Explore information needed on HSE website to help inform participants of data linkage option
  - Format of information
  - Level of detail
  - Most important elements to include
  - Anything would not include

- Whether in principle if information discussed is shared they would consent to this type of data linkage
  - Enablers
  - Main concerns/ barriers

Conclusion

- Explore overall views on data linkage
  - Whether they see it as important and why
  - Main concerns
  - Areas of less concern
  - Key reassurances needed from researchers, *probe on which is most important*
  - Key elements of information needed to help make decisions about consenting to data linkage or not, *probe on which is most important*

  - Thoughts about linking to information without specific consent
  - Whether information given had changed mind about group
- Anything else
- Questions

Scenario one – UCL want to link HSE survey responses to data about your local area, for example, the external temperature outside of your home when you took part in the survey. This will be relevant for HSE participants who received a visit from a nurse. The nurse measured the room temperature in the home before taking the participant's blood pressure. The purpose of measuring external temperature is to see what the difference is between the temperature outdoors and the temperature inside participants home. This information will help to know how warm or cool the home is compared to the outdoor temperature.

Scenario two - UCL could link HSE survey responses about your health with information specific to your home**,** for example home energy efficiency (e.g. loft insulation, double glazing, new boiler) and the amount of energy used for heating. The purpose of this is to see what effects these different ways of improving the energy efficiency of the home has on indoor temperatures, and what effects the efficiency improvements and heating energy use has on the long-term health of people living in those homes. UCL would want to know whether your home had any of those energy efficiency measures, and if so, whether it was installed in the years before or after the HSE nurse visit.

Appendix B. Achieved sample by characteristics

Overall 20 people participated in one of three focus groups. The table below provides a breakdown of the total number of participants by gender and age.

Gender	Male	8
	Female	12
Age range	16-24	0
	25-34	2
	35-44	3
	35-44	2
	45-54	3
	55-64	6
	65-74	1
	75+	1
	Unknown	2