

## Public Preferences regarding Data Linkage for Health Research: A Discrete Choice Experiment

Aitken, M<sup>1\*</sup>, McAteer, G<sup>2</sup>, Davidson, S<sup>2</sup>, Frostick, C<sup>2</sup>, and Cunningham-Burley, S<sup>1</sup>

### Submission History

Submitted:	27/10/2017
Accepted:	07/03/2018
Published:	26/06/2018

<sup>1</sup>University of Edinburgh<sup>2</sup>Ipsos MORI

### Abstract

The potential for data collected in the public and private sector to be linked and used in research has led to increasing interest in public acceptability of data sharing and data linkage. The literature has identified a range of factors that are important for shaping public responses and in particular has noted that public support for research conducted through data linkage or data sharing is contingent on a number of conditions being met. In order to examine the relative importance of these conditions a Discrete Choice Experiment (DCE) was conducted via an online questionnaire among members of Ipsos MORI's online panel in Scotland. The survey was completed by 1,004 respondents. Overall the two most influential factors shaping respondents' preferences are: the type of data being linked; and, how profits are managed and shared. The type of data being linked is roughly twice as important as who the researchers are. There were slight differences across age groups and between genders and slight differences when comparing respondents with and without long term health conditions. The most notable differences between respondents were found when comparing respondents according to employment and working sector. This study provides much needed evidence regarding the relative importance of various conditions which may be essential for securing and sustaining public support for data-linkage in health research. This may be useful for indicating which factors to focus on in future public engagement and has important implications for the design and delivery of research and public engagement activities. The continuously evolving nature of the field means it will be necessary to revisit the key conditions for public support on an ongoing basis and to examine the contexts and circumstances in which these might change.

## Introduction

Increasing amounts of health research are conducted through data-linkage, which is defined as: "the bringing together from two or more different sources, data that relate to the same individual, family, place or event" (1). This has enabled many important new insights, relating to, among other things, examination of relationships between social factors and health or access to health services. The potential for data collected in the public and private sector to be linked and used in research has led to increasing interest in public acceptability of data sharing and data linkage practices (e.g. 2,3), leading to broad recognition of the importance of public engagement both for understanding public preferences and concerns and for developing socially acceptable and ethically robust research and governance processes (2, 4). Recent highly publicised controversies over data use in research, such as with national data records systems in England (5) and Australia (6) have drawn attention to the importance of ensuring public support for the ways that data are used. Thus, there is increasing attention to public acceptability of secondary uses of data and to ensuring that these uses are understood and supported by the

wider public in order to develop and maintain a social license for data linkage in health research (6).

This has resulted in a growing body of evidence highlighting a range of factors shaping public views on data linkage for research purposes and public preferences for how this happens. Studies have explored public attitudes towards secondary uses of data, often focussing on issues relating to the anonymisation of data or (lack of) consent mechanisms (7 - 12). Additionally, the literature is increasingly emphasising the importance of trust in shaping public attitudes (3, 7, 13 - 15). A recent systematic review and thematic synthesis of qualitative studies (4) found that there is widespread support for data linkage and data sharing in health research but that this support is never unconditional and depends on a range of considerations. The review identified key factors influencing public attitudes including what data are used in research, who the researchers are, whether there is commercial/private sector involvement and to what extent or in what ways the research has public benefits. However, it noted that the extant literature does not "point to clear relationships or hierarchies between particular areas of concern or conditions for support and there is a lack of evidence relating to the ways in which trade-offs might be made

\*Corresponding Author:

Email Address: [maitken1@ed.ac.uk](mailto:maitken1@ed.ac.uk) (M Aitken)

or how preferences would be formed in reality" (4). While a range of factors are known to influence public responses to data-linkage, and a number of conditions underpin public support it is unclear which are most important or which ought to be prioritised in developing publically acceptable systems and processes for data-linkage research.

In order to examine the relative importance of factors influencing public preferences a Discrete Choice Experiment (DCE) was conducted. DCEs provide a means to understand choices people make between different fixed options – in this case options around the sharing and linkage of personal information for research purposes – and, more specifically, to establish the relative importance of different factors and considerations that constitute each option. The approach provides an alternative to simple ranking exercises in which people are asked to rate uni-dimensional options in order of preference. DCEs more accurately reflect 'real world' choices people face by inviting them to pick between multidimensional scenarios, some aspects of which they may favour and others of which they may not. DCE analysis has its origin in market research, where it was originally used to identify factors influencing demand for different products, however it has come to be used in a wide range of research areas including public health and health care (e.g. 16, 17).

In a DCE, scenarios are devised comprising a number of attributes and levels. Attributes are the main variables of the scenarios (e.g. the type of data being shared, the researchers doing the sharing etc.). Levels are the variants of each attribute (e.g. for the researchers doing the research, levels include 'only university researchers' and 'only university researchers or NHS staff'). Research participants are randomly assigned a scenario pack, comprising multiple pairs of scenarios and, from each pair, they are asked to pick the scenario they most prefer. Aggregate analysis of participants' preferences across the various pairs provides a measure of the relative importance of different attributes and levels in influencing people's preferences.

## Objectives

This study had the following objectives:

1. To understand the relative importance of previously identified key factors in shaping public preferences for the ways that data are linked and used in health research;
2. To identify the relative importance of key factors in shaping public preferences of different groups (i.e. according to socio-demographic variables).

## Methods

The DCE was conducted via an online questionnaire among members of Ipsos MORI's online panel in Scotland. This panel is made up of 238,577 members of the public across the UK including around 18,000 in Scotland from which the participants would be invited. The questionnaire was co-designed by the University of Edinburgh and Ipsos MORI, and was informed by previous qualitative research on public attitudes to data linkage.

The questionnaire comprised four main sections: 1) A page of narrative introducing the concepts of data sharing and linkage, including the reason it might be done and a statement that only anonymised personal data are used; 2) a set of introductory questions in which the different attributes (and associated levels) were introduced to participants one at a time as a 'warm up' to the DCE; 3) The DCE itself; and 4) a small number of questions to collect socio-demographic information on participants. In order to allow a focus on the conditions known to underpin public support for data-linkage in health research, in each question contained in section two any respondents who said they felt data linkage should not be permitted under any circumstances were routed out of the survey. Routing out respondents who stated they did not approve of data linkage in any circumstance was done to improve the quality of the data obtained through the discrete choice experiment: by ensuring we were only asking those who did not fundamentally object to data linkage, the different choices between different attributes and levels were more representative of what people would prioritise in that situation. If we included those who did not approve of it under any circumstances, the difference between the levels and attributes would matter less, as they were opposed to the whole idea of data linkage. More people would have chosen 'neither of these is acceptable to me' which would have reduced the overall significance of each attribute in the experiment. Moreover, previous qualitative and deliberative research has demonstrated that there is widespread conditional support for data linkage in research and that members of the public come to express quite nuanced views as they engage in discussion of the issues (2, 4, 13). We were therefore interested to explore further the nuances of public preferences and the relative importance of various conditions underpinning this conditional support, as this will be useful in developing good governance in the future. In short, by ensuring our population was comprised of people who did not state that they were not fundamentally opposed to data linkage at the outset, the final dataset was more likely to provide accurate information about the nuanced attitudes to data linkage and the conditions necessary for maintaining public support.

The discrete choice module of the questionnaire was developed from themes identified through previous qualitative research and public engagement work. This identified the relevant attributes to include in the DCE. Public engagement, including discussions with the Scottish public panel associated with the Farr Institute of Health Informatics Research was valuable for refining the questionnaire and developing the discrete choice module. The final discrete choice module comprised five attributes: who the researcher would be accessing linked data; the types of information being linked; the purpose of the research; options for profit making as a result of the research; and who would oversee the research. For each attribute there were four levels (See Table 1). A total of 240 different scenario pairs were developed from the attributes and levels, and 20 scenario packs were created to present to participants, each comprising 12 pairs of scenarios. Table 2 shows an example of two scenarios respondents were asked to choose between.

The questionnaire was piloted through cognitive testing among 20 randomly selected members of the public. A hall test approach was used whereby members of the public were

Table 1: DCE Matrix

Attribute	Levels			
	1	2	3	4
Researchers	Only university researchers.	Only university researchers or NHS staff.	Only university researchers, NHS staff or government researchers.	University researchers, NHS staff, government researchers and commercial researchers such as market research organisations or pharmaceutical companies.
Type of information	Information from your GP records being linked with information from your other NHS health records (e.g. hospital records).	Information from your NHS health records being linked with information from your social care or education records.	Information from your NHS health records being linked with information from your social care, education records or from your employment and benefits records.	Information from your NHS health records being linked with information from your social care, education, employment, and benefits records, as well as information collected about you in the private sector (e.g. through online shopping accounts).
Purpose	Research using linked information should only be conducted if it will have direct benefits for the people whose information is being used.	Research using linked information should only be conducted if it will have general public benefits.	Research using linked information should be allowed for any reason.	
Profit-Making	Nobody should be allowed to profit from research carried out using linked information.	Any profit made from research carried out using linked information should be shared with the public.	Any profit made from research carried out using linked information should be invested into public services.	Any profit made from research carried out using linked information should be kept by those carrying out the research.
Oversight	The process should be overseen by a non-governmental independent body.	The process should be overseen by the relevant public service(s); for example, research that uses information from people's health records should be overseen by the NHS.	The process should be overseen by the Scottish Government.	The process should be overseen by the organisations undertaking the research.

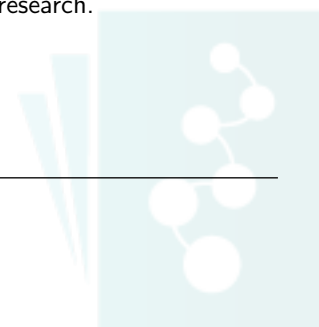


Table 2: Example of scenario choice

	Q) Which of these scenarios would you prefer?	
	1	2
The researchers are:	Only university researchers.	Only university researchers, NHS staff or government researchers.
The type of data being linked: N.B. All data would be anonymous	Information from your GP records being linked with information from your other NHS health records (e.g. hospital records).	Information from your NHS health records being linked with information from your social care, education records or from your employment and benefits records.
The purpose of the research:	Research using linked information should be allowed for any reason.	Research using linked information should only be conducted if it will have direct benefits for the people whose information is being used.
Profit-Making:	Any profit made from research carried out using linked information should be invested into public services.	Any profit made from research carried out using linked information should be shared with the public.
Oversight:	The process should be overseen by the Scottish Government.	The process should be overseen by the relevant public service(s); for example, research that uses information from people's health records should be overseen by the NHS.
	I prefer this scenario <input type="checkbox"/>	I prefer this scenario <input type="checkbox"/> Neither scenario is acceptable to me <input type="checkbox"/>



recruited on the street and brought into a building to one of two rooms which were set up with the questionnaire on a computer. Cognitive testing is a valuable method for ensuring that a questionnaire is accessible and easily understood by respondents and to understand how people from different perspectives and backgrounds interact with the questionnaire (18). In our pilot a member of the research team sat with respondents while they completed the questionnaire on a computer. After each section of the questionnaire respondents were asked for their reflections on it (e.g. was it clear what they were meant to do, were the questions clear, how easy was it to choose their answer). The aim of the pilot was to ensure respondents understood the instructions and the language used in the questionnaire and to test how they interacted with the survey and how long they took to complete it. In light of the pilot findings, minor changes were made to the ordering and wording of some questions for improved clarity. Additionally, the number of pairs of scenarios in each scenario pack was reduced from 12 to six to reduce the likelihood of respondent fatigue; a degree of which was evident during the cognitive testing.

The target number of interviews for the main stage survey was 1,000. A total of 13,275 members of the Ipsos MORI panel in Scotland were invited (via an email invitation) to take part, with quotas set on age, gender and working status with the aim of ensuring a sample that was representative of the Scottish population. Survey invites were sent in batches until the desired quotas were met. This was done both to make efficient use of the (limited) panel and also to reduce the likelihood of panellists trying to enter a survey, only to find it has been closed and becoming frustrated as a result.

The survey was launched on August 12th 2016 and was live for a period of 14 days. In total, 1,004 respondents completed the full survey. An additional 461 began the survey but were routed out at the introductory questions because they stated that data linkage was unacceptable under any conditions.

The survey data was weighted by sex, age and working status, using 2011 Census data. The data was analysed using a logit modelling technique to identify the underlying utility for each level of each attribute. Values were selected using maximum likelihood estimation, and this was run separately for each subgroup. The relative importance of each attribute is determined by comparing the ranges of utility (maximum-minimum) for each attribute, and calculating a simple share of the total of these ranges. This can then be interpreted as an estimate of the relative share of variation in choices which is explained by each attribute. In addition, a Hierarchical Bayesian method was used (in addition to the aggregate model) to provide respondent-level utility estimates, thus providing a measure of the variation in estimates and the consequent reliability. The determination of the importance of attributes and the various levels that attributes can take, is made by estimating underlying utilities for each level of each attribute. The estimation process is Bayesian, and includes estimates at respondent level. We have quoted an average value across all respondents as the best estimate of each parameter. The reliability in that estimate can be measured by considering the variation in the variable based on the variation across the respondents. We estimated both mean and variation for each parameter, and these values can be (and have been) used to determine confidence intervals or significance testing. When

testing the difference between two levels of an attribute, we have considered the differences, as it would be wrong to treat the two levels as independent variables.

## Results

### Responses to initial questions

The first question asked respondents for their preferences regarding the purpose of data linkage. By far the most common response was: "Research using linked information should only be used if it will have general public benefits". This option was chosen by 57 per cent of respondents. The responses to this question resonate with previous studies which have highlighted the importance of public benefits for public acceptability of data-linkage research (4).

The second question asked which types of researchers respondents felt comfortable with having access to anonymised linked data. Here the most common answer was: "Only university researchers, NHS staff or government researchers" which was chosen by 31 per cent of respondents. The responses to this question largely reflect the findings of previous research which have shown greater support for public sector uses of data compared to private sector uses (4).

Question three asked which types of information respondents would be happy with being linked for research purposes. The responses to this question suggest that respondents may have been uneasy with cross-sectoral data-linkage and that there was a preference for research to be conducted using solely health data. Almost half (48 per cent) of respondents chose: "Information from your GP records being linked with information from your other NHS health records (e.g. hospital records)".

Question four asked for respondents' preferences for the ways in which potential profits arising from research might be managed. The majority of respondents (62 per cent) chose: "Any profit made from research carried out using linked information should be invested into public services". Only 8 per cent chose "Any profit made from research carried out using linked information should be kept by those carrying out the research".

Question five asked for respondents preferences on how data linkage processes are overseen or monitored. 35 per cent of respondents chose "The process should be overseen by another [non-governmental] independent body" and a similar number (32 per cent) chose "The process should be overseen by the relevant public service(s); for example, research that uses information from people's health records should be overseen by the NHS". Just 6 per cent chose "The process should be overseen by the organisations undertaking the research".

As noted above, each question included an option to answer "Research using linked information should not be allowed under any circumstances", respondents who selected this answer were not asked to undertake the DCE. As a result a total of 457 respondents were routed out of the DCE. A summary of these respondents including their age and gender and the questions at which they were routed out is given in table four. The two questions which had the most respondents selecting this response related to the type of information that is linked and how profits are managed. Roughly half of respondents routed

Table 3: Achieved Sample profile

	Achieved Sample (unweighted)	Achieved Sample (weighted)	Achieved Sample Unweighted (percentage)	Achieved Sample Weighted (percentage)
Total				
Male	421	481	41.9%	47.9%
Female	583	523	58.1%	52.1%
	1004	1004		
18 – 24	86	119	8.6%	11.9%
25 – 34	189	159	18.8%	15.8%
35 – 54	358	358	35.7%	35.7%
55+	371	368	37.0%	36.6%
	1004	1004		
Working full time	557	420	55.5%	41.8%
Not working full time	447	584	44.5%	58.2%
<b>Total:</b>	1004	1004		

out were male (46 per cent) and half female (54 per cent), respondents were more likely to choose “Research using linked information should not be allowed under any circumstances” in older age groups (42 per cent of respondents routed out were aged 55 and over, 34 per cent were aged between 35 and 54 and, 24 per cent were aged between 18 and 34).

## Responses to the DCE

While the initial questions provided insights into public preferences on each of the key considerations, the DCE enabled an exploration of the relative influence of each of the variables on overall preferences and acceptability of data linkage arrangements.

Unsurprisingly, overall the preferred scenario reflected each of the preferred options from the first round of questions (as outlined in Table 5). In this scenario research was conducted only by university researchers or NHS staff and only with health data. The research must lead to public benefits and any profits arising should be invested into public services. Finally, the process should be overseen by a non-governmental independent body.

The least preferred option (as outlined in Table 5), included the largest number of possible researchers as well as the widest range of types of data being linked. It also had no restrictions on the purpose of research and allowed for any profits arising to be kept by the researchers. In this scenario the processes were overseen by the organisations undertaking the research.

Given the responses to the initial questions, the most and least preferred scenarios are not surprising. What is more interesting to consider is the range of different possibilities and combinations of variables in between. The range of choices made through the DCE enables an examination of the relative importance of each of the variables in shaping public preferences. The relative importance of one attribute against another is derived from the difference between the utilities for the “best” and “worst” level within each attribute – i.e. if there

is a large difference between the most and least desirable level within an attribute, it implies that that attribute is a driving factor in people’s scenario decisions. Conversely, if there is a low difference between the two, it implies the attribute is less of an important factor in overall scenario decision making. Figure 1 illustrates the relative importance of each of the factors. From this analysis it is possible to rank the importance of each of the factors in terms of their influence on public preferences, with number 1 being the most influential:

1. The type of information being linked (30%);
2. How profits are managed/shared (24%);
3. The purpose of research (18%);
4. Who the researchers are (16%);
5. How the processes are overseen (12%).

## Differences according to key demographic variables

As has been noted previously (4), it is important to recognise that within the public there will be a range of views and perspectives and that demographic as well as social and experiential factors are likely to be important for shaping public preferences. For this reason our analysis sought to examine the relative importance of the various attributes for different groups within our sample. In doing so we focussed on four key variables: Age; Gender; Working Status and; Long-term health conditions.

### Age

Previous research on attitudes or preferences relating to data linkage or data sharing has presented varying and, at times, contradictory findings regarding differences across age groups (4). This has pointed to the need for greater research to explore the variations in perceptions and opinions across age groups.

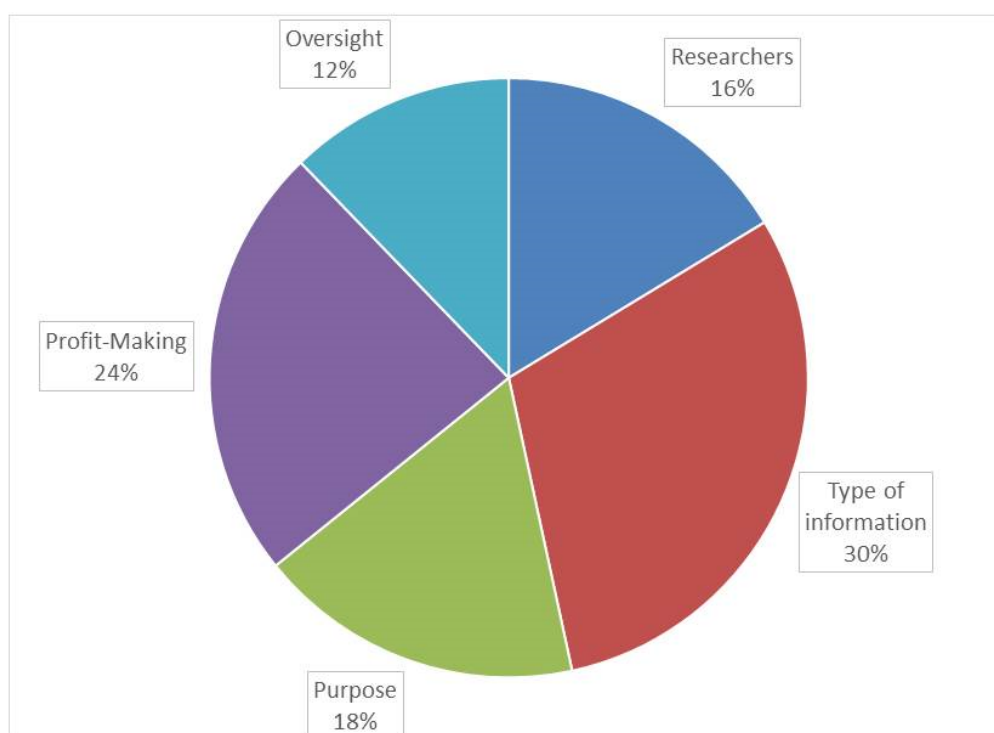
Table 4: Respondents routed out of questionnaire

Question	Respondents Routed Out (No.)	Male	Female	Age Distribution	
Q1: The purpose of the research	66	50%	50%	18 – 34:	32%
				35 – 54:	30%
				55 +:	38%
Q2: Who the researchers are	103	45%	55%	18 – 34:	27%
				35 – 54:	35%
				55 +:	38%
Q3: What types of information are linked	133	71%	29%	18 – 34:	20%
				35 – 54:	38%
				55 +:	42%
Q4: What happens with profits	145	56%	44%	18 – 34:	19%
				35 – 54:	33%
				55 +:	48%
Q5: Who oversees the process	8	63%	37%	18 – 34:	50%
				35 – 54:	38%
				55 +:	12%
Q6: What role should the public play	2	50%	50%	18 – 34:	100%
				35 – 54:	0%
				55 +:	0%
<b>Overall</b>	<b>457</b>	<b>46%</b>	<b>54%</b>	18 – 34:	24%
				35 – 54:	34%
				55 +:	42%

Table 5: Overall Most and Least Preferred Scenarios

	Overall Preferred Scenario	Overall Least Preferred Scenario
The researchers are:	Only university researchers or NHS staff.	University researchers, NHS staff, government researchers and commercial researchers such as market research organisations or pharmaceutical companies.
The type of data being linked:	Information from your GP records being linked with information from your other NHS health records (e.g. hospital records).	Information from your NHS health records being linked with information from your social care, education, employment, and benefits records, as well as information collected about you in the private sector (e.g. through online shopping accounts).
The purpose of the research:	Research using linked information should only be conducted if it will have general public benefits.	Research using linked information should be allowed for any reason.
Profit-Making:	Any profit made from research carried out using linked information should be invested into public services.	Any profit made from research carried out using linked information should be kept by those carrying out the research.
Oversight:	The process should be overseen by a non-governmental independent body.	The process should be overseen by the organisations undertaking the research.

Figure 1: Relative Importance of Attributes



As shown in figure 2, in this DCE there was very little difference in the relative importance of attributes across the three age groups examined. However, the one area where a difference was identified relates to who the researchers are. Concern with this factor lessens as the age range increases (19.2 per cent for age 18-34, 16.2 per cent for age 35-54 and 15.5 per cent for age 55 plus).

### Gender

As shown in figure 3, the findings of the DCE are largely consistent between male and female respondents. However, male respondents were slightly more concerned with oversight arrangements while female respondents were slightly more concerned with the type of information being linked.

### Working status

Examining the responses of respondents who are in full time work compared to those who are not revealed interesting differences. As shown in figure 4, respondents who were not working full time were more concerned with oversight arrangements and the type of information being linked, compared to those who were working full-time. Those working full-time were less concerned with oversight arrangements and more concerned with the purpose of data-linkage, who the researchers were and how profits were managed/shared.

In order to understand to what extent professional experience relevant to the subject of the DCE impacted on preferences, we identified a number of key working sectors. These were:

- Human health/social work activities;

- Government/public administration/social security;
- Financial/insurance activities;
- Information technology;
- Research

Additionally, all respondents who had stated that they were in employment (full or part-time) were asked: "In your line of work, are you involved in handling or managing data, or in data security?" All respondents who worked in one of the key sectors or who answered yes to this screening question were included in a sub-group identified as "Working in Key Sector". As illustrated in figure 5, there were notable differences in the relative importance of attributes when comparing individuals "Working in key sectors" with those who were working in non-key sector areas. Those working in key sectors were more concerned with oversight arrangements (16 per cent, compared with 7.7 per cent for non-key sector respondents) and the purpose of data linkage (21.7 per cent, compared to 16.9 per cent for non-key sector respondents). Those not working in key sectors were more concerned with who the researchers were (18.9 per cent, compared to 12.1 per cent for key sector workers) and what information was being linked (29.7 per cent, compared to 25.1 per cent for key sector workers). There was little difference between the two groups in relation to management/sharing of profits.

### Long term health conditions

Previous research has suggested that individuals with long term health conditions may be more supportive of health related research (e.g. 12). Therefore, a question was included to



Figure 2: Relative Importance of Attributes for Respondents across Different Age Groups

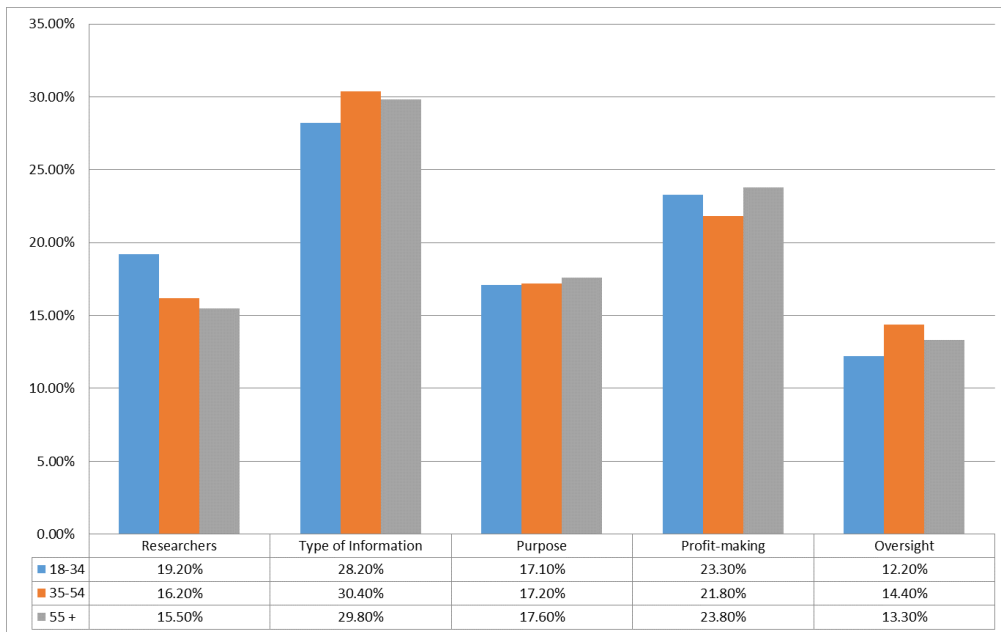


Figure 3: Relative Importance of Attributes for Male and Female Respondents

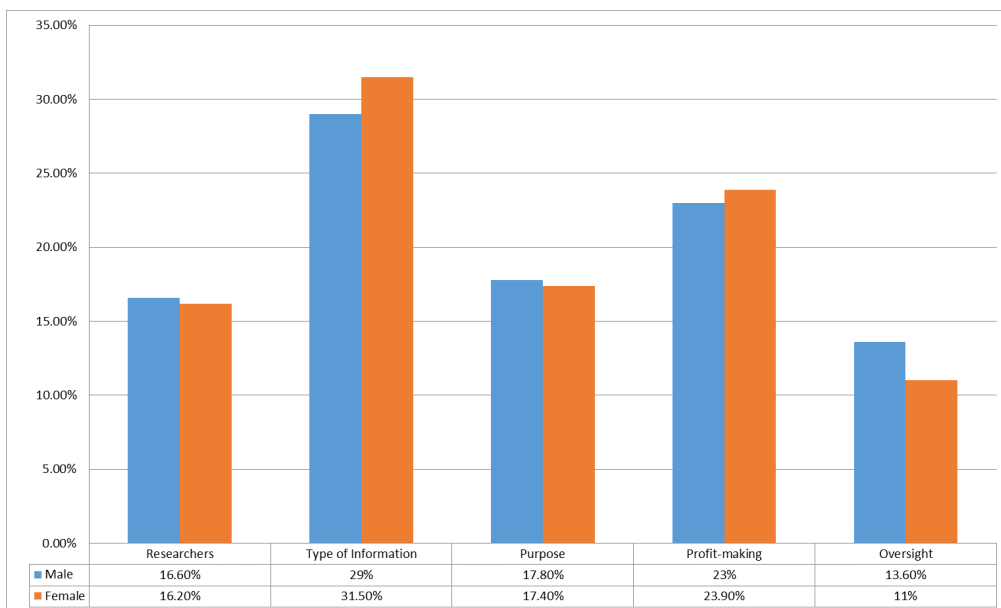


Figure 4: Relative Importance of Attributes for Participants according to Working Status

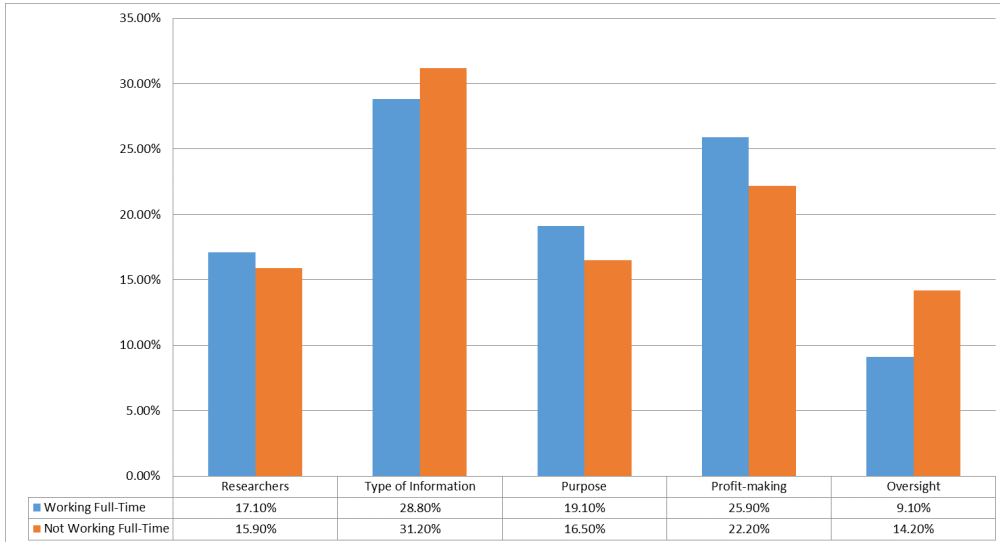
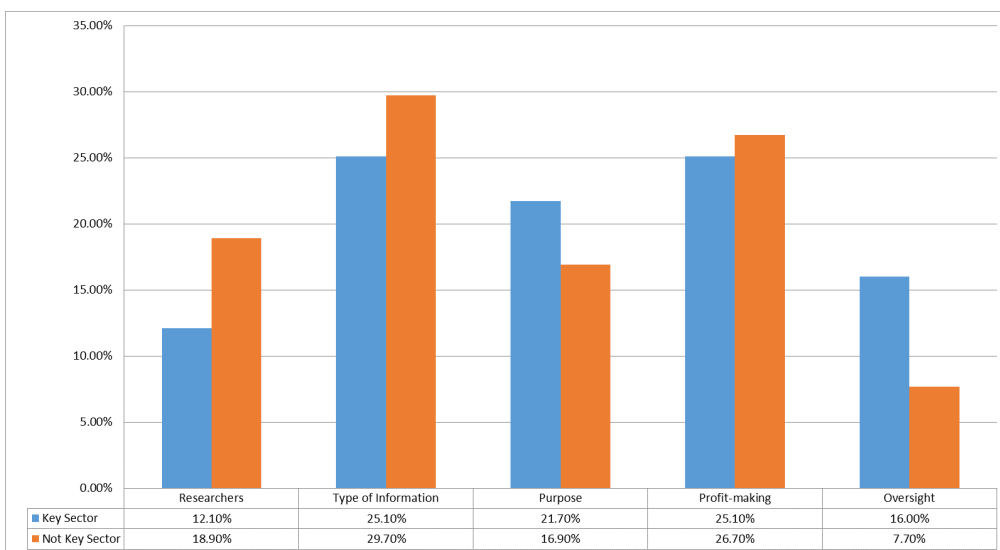


Figure 5: Relative Importance of Attributes for those Working in Key Sectors and Non-Key Sectors



identify individuals with long-term health conditions. Respondents were asked: “Do you have any physical or mental health condition or illness lasting or expected to last 12 months or more?” Respondents who answered yes to this question were included in the sub-group “with long-term health condition”. As illustrated in figure 6, respondents with long-term health conditions were slightly less concerned with arrangements for managing/sharing profits (20.3 per cent, compared with 24.9 per cent for respondents with no long-term health condition) and slightly more concerned with oversight mechanisms (14.8 per cent, compared with 11.4 per cent for those with no long-term health condition). Respondents with long-term health conditions were also slightly less concerned with the purpose of research (16.9 per cent compared with 18.5 per cent for those with no long term health conditions).

## Discussion

Discrete Choice Experiments remain a novel method in the context of data linkage preferences. In the pilot, we found that the unfamiliarity of this method led to both positive and negative responses. Some respondents clearly enjoyed the exercise and approached it with curiosity and a sense of fun, while others approached it with hesitancy and were unsure of its purpose. Nevertheless, response rates were good and the target sample was achieved without difficulty, indicating that while respondents may have been unfamiliar with the method it did not put them off completing the questionnaire. Indeed, in our pilot respondents tended to become more confident as they progressed through the scenario choices.

Nevertheless, as noted above, 461 respondents were routed out of the survey before the DCE element due to choosing an answer in Section 3 of the questionnaire which stated that data linkage for research was unacceptable under any circumstances. Respondents did not know that selecting such an option would have ended the questionnaire and so there is no evidence that such options were chosen as a way of avoiding completing the questionnaire, but instead indicate a substantial proportion of respondents who seemed to have significant concerns about data linkage for research. It is therefore important to consider which questions appeared to have raised most concerns.

The numbers of people choosing “Research using linked information should not be allowed under any circumstances” at each of the questions in Section 3 are listed below.

- Question 1 (Purposes of research): 67
- Question 2 (Who are the researchers): 104
- Question 3 (What types of information may be linked): 135
- Question 4 (Management of potential profits): 145
- Question 5 (Arrangements for oversight/monitoring): 8
- Question 6 (Public involvement): 2

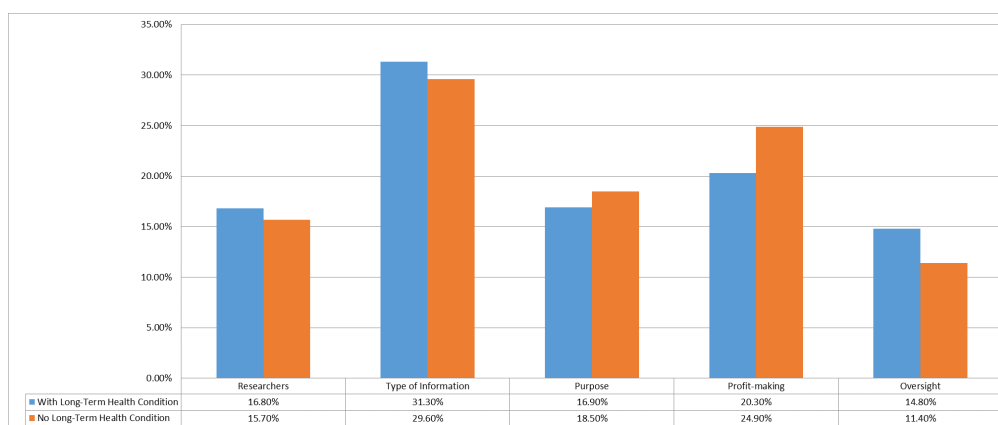
The most contentious questions related to the types of information being linked and the potential for profits to be created through research, and (to a slightly lesser extent) to who the researchers would be. It is noteworthy that a relatively small number of respondents (67) stated that “Research using linked information should not be allowed under any circumstances” at question one. This suggests that there was not

outright opposition to data linkage and that the purpose of research may be important for public acceptability (or avoiding public opposition), something reinforced by our qualitative work. In particular the inclusion of options that stressed research would only be done if it would have benefits for individuals or the wider public may have been important here. Conversely, for a significant number of individuals the options available in relation to limits on the types of information that could be linked, or how potential profits would be managed were unsatisfactory to address their concerns. The type of information appears to have been most concerning for male respondents (71 per cent of respondents routed out at this question were male). The management of profits appears to have been most concerning for older participants (48 per cent of respondents routed out at this question were aged 55 or over, compared to 19 per cent aged 18 – 34). In some cases it may have been the acknowledgement that profits could be made from this research which raised concerns, as previous research has shown (2, 3). However, our data does not enable clear identification of respondents’ reasoning. This would be a worthwhile area to explore in future research.

Combining the DCE with more traditional questionnaire questions enabled firstly insights into public preferences for each of the variables and then secondly an in-depth examination of how these variables relate to one another and their relative importance in shaping overall preferences and public acceptability. This represents the principal contribution of the study since while there is a growing body of literature evidencing a range of factors that influence public responses to data linkage these have largely not been able to indicate the relative significance of each of these factors or how they would be combined and/or traded off in practice. By examining the relative importance of each of the variables considered in our DCE we have been able to rank them in accordance with the extent to which they influence respondents’ preferences for data linkage scenarios. This has highlighted that overall the two most influential factors shaping respondents’ preferences are: the type of data being linked; and, how profits are managed and shared. The type of data being linked is roughly twice as important as who the researchers are.

It is important to note differences within our sample. There were slight differences across age groups (concern with who the researchers are lessened as the age range increased) and between genders (male respondents were slightly more concerned with oversight arrangements while female respondents were slightly more concerned with the type of information being linked). We also found slight differences when comparing respondents with and without long term health conditions. However, the most notable differences were observed in looking at differences in employment and working sector. Respondents who were not working full time were more concerned with oversight arrangements and the type of information being linked, compared to those who were working full-time. Those working full-time were less concerned with oversight arrangements and more concerned with the purpose of data-linkage, who the researchers were and how profits were managed/shared. Moreover, the biggest variations were found when looking at respondents working in identified key sectors. These respondents were more concerned with oversight arrangements and the purpose of data linkage compared to other respondents. This is likely to be a result of these respon-

Figure 6: Relative Importance of Attributes for Respondents with and without Long-Term Health Conditions



dents' experience or knowledge of data management/security or research practices. These differences suggest that individuals' experiences and expertise may be important for shaping preferences.

Throughout the DCE it was stressed that data linked and used for research would always be anonymous and that individuals would not be identified. In the pilot phase respondents commented that this had been an important consideration and in a number of instances it was stated that this assurance was the reason they did not select "Research using linked information should not be allowed under any circumstances" in the initial round of questions. Previous research has noted the importance of anonymization for public acceptability of secondary uses of data. However, anonymization is not straightforward and guarantees of absolute anonymity are often not possible. Therefore, consideration is paid to the ways in which individuals' confidentiality might be protected without perfect anonymization of data (4). Including a range of options relating to degrees of anonymization/pseudo-anonymisation or other mechanisms for protecting confidentiality may well have had a significant impact on responses to the questionnaire and DCE and future research might helpfully do this.

## Study Limitations

Given that the study was conducted with an online panel of members of the public who have signed up to take part in research questionnaires some limitations should be noted. In particular, respondents will all have had a reasonable level of confidence in using IT and the internet. It may be that individuals who are less confident using such technology would express different preferences. Similarly, since the respondents are all members of the panel and have expressed a willingness to take part in research surveys they may be more likely to be supportive of research than the wider general public. The DCE method could not accommodate those who expressed opposition to data linkage without compromising data quality. Further research would be required to explore those views and any scope for identifying mechanisms for acceptability.

## Conclusions

As ever-increasing amounts of data are collected and become available, and as there is continuous innovation in the ways that these data can be used, the importance of developing and maintaining a social license for health research conducted through data linkage becomes ever more prescient. Recent years have witnessed increasing interest in public acceptability of secondary uses of data, particularly for research, and a growing emphasis on public engagement as an important means through which to understand public preferences for the ways in which data are used in research. While previous research has indicated that there is widespread conditional support for data linkage in research and identified a range of conditions underpinning this support, the DCE has enabled insights into the relevant importance of each of these. This may be useful for indicating which factors to focus on in future public engagement and has important implications for the design and delivery of research and public engagement activities. This study has indicated that there is public support for the linking of health data and use by university and health service researchers without private sector involvement and with independent oversight. The policy challenge going forward is how to work with publics to extend that mandate given the increasing interest in and drive towards linking other kinds of data (from both the public and private sector) and the variety of ways that private sector actors are involved in research (e.g. as funders, collaborators or research partners). The continuously evolving nature of the field means it will be necessary to revisit the key conditions for public support on an ongoing basis and to examine the contexts and circumstances in which these might change in order to maintain a social license for current and future research practices. In particular, considering the terms under which private sector involvement might be acceptable will be an area of particular interest.

## Declarations

### Funding

This work was funded by the Wellcome Trust through the Scottish Health Informatics Programme Grant (Ref WT086113)

and further supported by The Farr Institute of Health Informatics Research.

The UK HIRN is supported by a 10-funder consortium: Arthritis Research UK, the British Heart Foundation, Cancer Research UK, the Economic and Social Research Council, the Engineering and Physical Sciences Research Council, the Medical Research Council, the National Institute of Health Research, the National Institute for Social Care and Health Research (Welsh Assembly Government), the Chief Scientist Office (Scottish Government Health Directorates), the Wellcome Trust, (MRC Grant No: MR/M501633/2).

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### References

- Holman CD, Bass JA, Rosman DL, Smith MB, Semmens JB, Glasson EJ, Brook EL, Trutwein B, Rouse IL, Watson CR, de Klerk NH. A decade of data linkage in Western Australia: strategic design, applications and benefits of the WA data linkage system. *Australian Health Review*. 2008;32(4):766-77. <https://doi.org/10.1071/ah080766>
- Wellcome Trust. *The one-way mirror: Public attitudes to commercial access to health data*. London: Wellcome Trust. 2016.
- Davidson S, McLean C, Treanor S, Aitken M, Cunningham-Burley S, Laurie G, Sethi N, Pagliari C. Public acceptability of data sharing between the public, private and third sectors for research purposes. *Scottish Government Social Research*. 2013
- Aitken M, Jorre JD, Pagliari C, Jepson R, Cunningham-Burley S. Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. *BMC medical ethics*. 2016 Dec;17(1):73.
- Carter P, Laurie GT, Dixon-Woods M. The social licence for research: why care. *data ran into trouble*. *Journal of medical ethics*. 2015 <https://doi.org/10.1136/medethics-2014-102374>
- Garrety K, McLoughlin I, Wilson R, Zelle G, Martin M. National electronic health records and the digital disruption of moral orders. *Social Science & Medicine*. 2014 Jan 1;101:70-7. <https://doi.org/10.1016/j.socscimed.2013.11.029>
- Damschroder LJ, Pritts JL, Neblo MA, Kalarickal RJ, Creswell JW, Hayward RA. Patients, privacy and trust: Patients' willingness to allow researchers to access their medical records. *Social science & medicine*. 2007 Jan 1;64(1):223-35. <https://doi.org/10.1016/j.socscimed.2006.08.045>
- Saxena N, Canadian Policy Research Networks. *Public Involvement Network. Understanding Canadians' attitudes and expectations: citizens' dialogue on privacy and the use of personal information for health research in Canada*. CPRN= RCRPP; 2006.
- Trinidad SB, Fullerton SM, Bares JM, Jarvik GP, Larson EB, Burke W. Informed consent in genome-scale research: what do prospective participants think?. *AJOB primary research*. 2012 Jul 1;3(3):3-11. <https://doi.org/10.1080/21507716.2012.662575>
- McGuire AL, Hamilton JA, Lunstroth R, McCullough LB, Goldman A. DNA data sharing: research participants' perspectives. *Genetics in Medicine*. 2008 Jan;10(1):46. <https://doi.org/10.1097/gim.0b013e31815f1e00>
- Willison DJ, Keshavjee K, Nair K, Goldsmith C, Holbrook AM. Patients' consent preferences for research uses of information in electronic medical records: interview and survey data. *Bmj*. 2003 Feb 15;326(7385):373. <https://doi.org/10.1136/bmj.326.7385.373>
- MRC & Ipsos Mori. *The use of personal health information in medical research: general public consultation*. Final report. London: Medical Research Council. 2007.
- Aitken M, Cunningham-Burley S, Pagliari C. Moving from trust to trustworthiness: Experiences of public engagement in the Scottish Health Informatics Programme. *Science and Public Policy*. 2016 Oct 1;43(5):713-23. <https://doi.org/10.1093/scipol/scv075>
- Ipsos MORI. *Dialogue on Data: Exploring the public's views on using linked administrative data for research purposes*. Ipsos Mori. [www.ipsos-mori.com](http://www.ipsos-mori.com). 2014.
- Trinidad SB, Fullerton SM, Bares JM, Jarvik GP, Larson EB, Burke W. Genomic research and wide data sharing: views of prospective participants. *Genetics in Medicine*. 2010 Aug;12(8):486. <https://doi.org/10.1097/gim.0b013e3181e38f9e>
- Ryan M, Bate A, Eastmond CJ, Ludbrook A. Use of discrete choice experiments to elicit preferences. *BMJ Quality & Safety*. 2001 Sep 1;10(suppl 1):i55-60.
- Haddow G, Cunningham-Burley S, Murray L. Can the governance of a population genetic data bank effect recruitment? Evidence from the public consultation of Generation Scotland. *Public Understanding of Science*. 2011 Jan;20(1):117-29. <https://doi.org/10.1177/0963662510361655>
- Collins D. Pretesting survey instruments: an overview of cognitive methods. *Quality of life research*. 2003 May 1;12(3):229-38. <https://doi.org/10.1023/a:1023254226592>