

Online deliberative engagement

A pilot study

A report for
The Wellcome Trust



University of Surrey

Dr Julie Barnett

Professor Chris Fife-Schaw

Professor Richard Shepherd

Dr Lada Timotijevic

Brook Lyndhurst

Jon Fletcher

White October

David Fletcher

Contents

1	Executive summary	2
2	Background & methodology	8
3	The development of the website	15
4	Results.....	23
5	Discussion and Implications	46
	Appendix A - The text	51
	Appendix B – The questionnaire	54
	Appendix C - Example replies to participants	68

© University of Surrey / Brook Lyndhurst / White October 2008

This report has been produced by the University of Surrey, Brook Lyndhurst and White October under/as part of a contract placed by the Wellcome Trust. Any views expressed in it are not necessarily those of the Wellcome Trust. The University of Surrey, Brook Lyndhurst and White October warrant that all reasonable skill and care has been used in preparing this report. Notwithstanding this warranty, the University of Surrey, Brook Lyndhurst and White October, singly or collectively, shall not be under any liability for loss of profit, business, revenues or any special indirect or consequential damage of any nature whatsoever or loss of anticipated saving or for any increased costs sustained by the client or his or her servants or agents arising in any way whether directly or indirectly as a result of reliance on this report or of any error or defect in this report.

1 Executive summary

The Wellcome Trust invited a team from the University of Surrey to explore the feasibility of developing an online tool that could be used to facilitate people's interest, deliberation and awareness around biomedical science. The key principles that we sought to embody in the tool were to enable people to engage on their terms, to address the questions they are asking and yet to embody a genuinely two way process. In a move away from 'event-based' deliberation this tool rather sought to engage people in a way that approximates the more everyday processes of information seeking and sense making. In theory such a tool could be highly resource effective and could be used with large numbers of people.

Along with partners from Brook Lyndhurst and White October, and in line with the Wellcome Trust objectives, an online tool was thus developed that was intended to have the capability to

- a. provide an environment where participants felt able to consider new information and to feel able to provide both enquiry and comment about this;
- b. provide feedback that was relevant and responsive to the particular comments and questions that people had made
- c. build in methods of capturing the ways in which people had engaged with the tool.

1.1 The structure of the study

An online engagement tool was developed by the project team tool in relation to the topic of personal data for use in biomedical research. It was extensively piloted before 85 incentivised participants took part in a study to test the feasibility of using the tool to encourage participant engagement with the topic.

There were four waves to the study which were conducted sequentially. Waves 1 and 2 received a flat rate incentive of £25. The Waves 3 and 4 incentive was paid in relation to the time spent on the task and, on average, was £7.

Participants registered on the site and completed a short series of pre-text questions. They then read 9 short chunks of text. Glossary terms could be highlighted by the participant if they wished to gain further information. Alongside this text participants had the opportunity to ask questions or to comment on the text. The first stage of the study finished with participants completing a series of post-text questions. The project team emailed each respondent with confirmation of the questions and comments they had asked for responses to. The team conducted the necessary research to provide feedback on all questions participants had submitted and this was emailed to participants. They were asked to read this before logging back onto the project website. Upon logging in they completed a final set of post-task questions.

The software we developed collected and recorded all survey responses, details of all questions and comments that had been produced and information about how long each participant spent in each of the reading, glossary and response windows. These data could be downloaded in a Microsoft Excel-compatible CSV file and were transferred to SPSS for statistical analyses.

1.2 Results

Who took part in the study?

The response rate in relation to the standard invitation of the recruitment company was 6.6%. Of the 85 participants that took part in the first part of the study (completing the pre- and the post-text questions), 74 (87%) of them returned to complete the final set of questions. Demographic measures showed a tendency for older participants to choose to complete the study and SES measures unsurprisingly showed the lowest participation among Grades A and E. The gender profile of participants was evenly split.

Did participants ask questions and make comments?

One of the key questions of interest in this study was whether we could design a tool that would elicit questions or comments from the participants. 57 participants (67%) produced at least one question or comment; 28 (33%) did not ask a question or make a comment. The number of questions or comments that were made ranged from none to twenty and in total 348 questions and comments were submitted.

How much time did participants spend on the exercise?

One of the capabilities of the tool we developed is to record behavioural measures that can serve as proxies for the extent to which people have engaged with it. One such measure is time. The software allowed us to measure the times that participants spent within different components of the tool. These measures are of interest in themselves but more importantly we are interested in the way in which they relate to:

- a. Our other main behavioural proxy of engagement: the numbers of questions asked and comments made
- b. Attitudinal measures
- c. Self reported knowledge
- d. Changes in attitudes and knowledge over the course of the online deliberation

The time taken by participants at time 1 ranged between 6 and 55 minutes with a mean time of 24 minutes. 3 scores were classed as outliers and these participants were excluded from this and all statistical analyses involving measures of time.

We considered 3 components of our Total Time measure in more detail:

- The time taken to produce questions or comments: this varied from 31 seconds for Section 3 (on Anonymity) to 1min: 23secs for Section 7 on Problems with Personal Data.
- The mean time taken to read each Section varied from 1min:12secs for section 1 to 22 seconds for sections 2 and 4. We acknowledge the way in which reading time is in part

a product of the length of the section and we use time scores that are weighted for this in any analyses involving reading time.

- Forty-five participants did not click into any of the highlighted glossary terms and 40 did. For those that did click into at least one glossary term the time they spent on the 10 terms in total ranged between two seconds and 1 min:42 secs with a median time of 22 seconds.

Exploring the relationships between the time based measures of engagement revealed a coherent set of relationships – people asking more questions and making more comments spent longer reading the text and investigating the text via the click-on glossary terms.

How did participants' attitudes relate to the nature and extent of their engagement?

Another dimension of considering the way in which the tool worked is to examine the relationship between people's views about personal data and the nature and extent of their engagement with the study: What difference does engaging with the text and with the research team make to people's views?

- Those who were less confident that they understood the phrase 'personal data' were more likely to ask questions and make comments than those that were more confident. Those that were more reticent about what they knew in this area seemed more willing to actively engage with the tool by questioning and commenting.
- The people that had engaged by asking questions showed a different pattern of changes in their self assessed knowledge than those who did not: at the post-task wave they had increased in their self assessed knowledge and those who had not engaged had not.
- There was a very similar pattern of results in relation to another measure of engagement: reading time. Those who took longest over reading the text became more positive about their self assessed knowledge than those who took the shortest time.
- There was a very different pattern of results where we looked at how comfortable people were about their personal data being used for research: engaging with the tool by producing comments and questions or by taking longer over reading the text did not make any difference to how comfortable those participants felt.
- We asked how important it is in medical research that personal data remain anonymous. Those that had produced questions or comments said that it was less important that data should remain anonymous than those that did not.

Did the extent to which people were incentivised relate to the extent and nature of their engagement?

There were no significant differences between the high incentive and the low incentive group in terms of reading time, time producing questions or comments or in time looking at glossary terms. However those promised the higher incentive produced significantly more comments/questions.

What did people think of the exercise?

The task was assessed very positively by all participants. Understanding of the instructions and the purpose of the research was high and interacting with the text adjudged as easy. Those who thought it more difficult to ask questions actually produced fewer questions. Ratings of the way in which the tool held participant interest and was enjoyable were very high. People were more positive about the exercise when they had engaged with it to a greater extent by asking questions. Those who had received feedback were very positive about how useful it was and how satisfied they were with it.

Sixty four percent of participants requested a report of the study¹ and 87% said that they would be happy to be contacted in the future to take part in a further study in this area.

What were people's comments and questions?

The comments and questions that were contributed by participants are indicative that at least three of the core criteria of deliberation are facilitated in the design of the tool

- a. There was evidence that the tool and the information that was presented stimulated a search for new information.
- b. There was evidence that the tool allowed the expression of argumentation: the questions and comments suggested that people were able to use the tool to interrogate the text for assumptions and implications and to probe the value of the concepts that were being explored in relation to the wider social and political context.
- c. There was evidence that the tool allowed for reflection: participants expressed preferences and used detail (or lack of it) from the text to justify or to provide caveats to the positions they took.
- d. The deliberative context of this tool is stripped down in terms of the social interactional context it provides. Even in this thin social context however it is noteworthy that participants' responses were often written in a conversational way and were predicated on the assumption that there was a dialogue being conducted with the research team.

1.3 Discussion and implications

A small pilot study was conducted to develop a web based tool suitable for encouraging citizen deliberation around science and technology. The way in which people engaged with it was monitored. Uniquely, behavioural and attitudinal measures were captured alongside extensive qualitative material. Analyses have been conducted that consider some of the relationships between these data. This chapter contains some final reflections on the results of the pilot study and some suggestions as to possible ways of taking this work forward.

The objective behavioural measures of engagement that were captured related to each other in coherent and consistent ways. Unlike most methods of public engagement, the development of this tool has focused upon eliciting the *questions* that people have about a particular topic. The

¹ We are preparing a suitable results summary for this purpose.

results suggest that this focus is likely to be a productive one insofar as people who were less confident that they knew about personal data generated significantly more questions and comments. It may be that the privacy afforded by the online environment enables people to seek out information more readily. Second, greater engagement with the tool led to a significantly greater increase in self-assessed knowledge. This is important as the change in this measure seems to represent these peoples' genuine sense that they have learned something.

The lack of any relationship between engaging with the material and assessments of how comfortable people were with the idea of their personal data being used for medical research serves to highlight the importance of not considering increasing knowledge instrumentally as a simple route to acceptance.

Examination of the nature of the questions and comments produced by participants suggested that they embodied some of the key indicators of deliberation. The information that was provided about personal data stimulated a search for new information; provided a context for the deployment of argumentation and allowed people to reflect on, and state, their preferences.

Although the focus of this study was on the feasibility of the tool as a way to engage people, on other occasions there is likely to be a clear focus upon the insights it facilitates about the sense that people make about the topic itself. The output of the tool provided an array of both qualitative and quantitative data about participant understandings of the use of personal data in medical research.

Although this was a small study we designed it in order identify differences in the behavioural indicators of engagement between participants receiving the high and low incentives. Although there were some important differences, for example those on the lower rate produced less questions and comments than those on the higher rate, the implications of these differences for future use of the tool are unclear and are likely to vary in relation to the topic area, the exact nature of the task and how participants are recruited.

There are a range of possibilities for further developing the tool's potential. For example this could involve participants interacting with images, video or audio in addition to text. In addition to structuring a conversation with participants over time the tool could be used to provide a snapshot of the way in which people's questions about a topic do or don't change over time, to see what questions media coverage of a particular topic stimulates or for assessing the effects of subtle changes in the language used in the text.

It would also be possible to return to participants in the current study and see whether the different profiles of self assessed knowledge of those who engaged more or less actively are maintained over a longer period of time.

Using the tool in its current form makes considerable demands upon researcher time in terms of providing information in response to the questions participant questions ask.

In conclusion it would seem that this pilot study has developed and tested a tool that has considerable potential for future use. It can be used as a tool for online deliberation and learning and simultaneously allow us to address research questions about the relationships between behavioural measures of engagement, self reported knowledge and attitudes, and the questions that people ask.

2 Background & methodology

This project originated with a bid to the Wellcome Trust in March 2007 to carry out a much larger piece of work: Attitudes to Biomedical Science: Social Trust, Affect and Deliberation. While the Trust was supportive of the overall aims of the project, they raised some quite reasonable questions about the proposed method, which was central to the proposal and relied upon a previously untested system of online engagement. As a result, the University of Surrey, in partnership with Brook Lyndhurst and web developers White October, were commissioned by the Trust to do a pilot study to explore the feasibility of this approach. This report represents the key findings from that pilot study.

2.1 The aims of the project

The aim of the project was to explore the feasibility of developing an online tool that could be used to facilitate people's interest, deliberation and awareness around biomedical science. The key principles that we sought to embody in the tool were to enable people to engage on their terms, to address the questions they are asking and yet to embody a genuinely two way process. In a move away from 'event-based' deliberation this tool rather sought to engage people in a way that approximates the more everyday processes of information seeking and sense making. In theory such a tool could be highly resource effective and used with large numbers of people.

The online deliberation tool was developed in order to embody the following characteristics:

- a. To provide an environment where participants felt able to consider new information and to feel able to provide both enquiry and comment about this;
- b. To provide feedback that was relevant and responsive to the particular comments and questions that people had made
- c. To build in methods of capturing the ways in which people had engaged with the tool.

The successful development of such a tool would be in line with, and thus potentially valuable to assist with, the objectives of the Wellcome Trust for public engagement²:

- a. To promote interest, learning and excitement about biomedical science and its past, present and future impacts on society;
- b. To stimulate an informed debate to raise awareness and understanding of biomedical science, its achievements, applications and implications; and
- c. To provide data to inform the Trust's own – and the wider national – debates, research plans and policies in relation to public interests and concerns in order to balance the needs of the research endeavour with those of society.

² Wellcome Trust Strategic Plan 2005-2010: Making a Difference.

2.2 The structure of the study

In a nutshell, this project involved the recruitment of around 80 incentivised participants to test an online engagement tool developed by the project team. The participants were invited to take part through an online survey panel run by ICM and then redirected to the project website, held at <http://wbe.whiteoctober.co.uk>. We chose to develop the tool in relation to the topic of personal data for use in biomedical research. The University of Surrey had previously conducted a study for the Wellcome Trust in this area and it was also considered ideal insofar as it was likely to involve participants in assimilating new knowledge in an area that also had considerable personal relevance.

Conducting the research in waves

In order to allow us to review the effectiveness of the site and to make changes in light of emerging problems, participants were taken part in the exercise in four separate waves. Invitations for each wave contained a different password that participants used to log on to the website, allowing us to track which participants belonged to which wave. We applied a cap on the site to restrict the number of participants that could complete the exercise to the required number per wave.

Varying the incentives

In our original proposal we had envisaged some participants not being offered incentives at all in order to mimic a situation in which a larger study might rely upon people's natural interest in the subject matter to trigger the engagement. In the event, this turned out to be impossible – recruiting participants through conventional fieldwork agents was prohibitively expensive and the terms of the ICM panel required a minimum incentive to be paid to participants. This rate – £1 for every five minutes spent on a survey – replaced the 'zero incentive' option offered to waves three and four: the average payment made under these incentive conditions was £7. Participants in waves one and two received flat rate payments of £25 for completion of the exercise, irrespective of how long it took. We will refer to the former group as the Low Incentive group and to the latter as the High Incentive group.

Breaking up the participants into waves also gave us the opportunity to test, in a limited way, the impact of varying the incentives offered to participants: would participants who were paid according to the time they spent on the exercise record higher or lower levels of engagement?

Participants' route through the exercise

Participants' interaction with the site took place over a number of different stages and it is worth outlining what these were – and how we refer to them – in order to add clarity to the later stages of this report.

Part 1 – registration and pre-text questions

Participants arrived at the site, having been directed there from an invitation issued by ICM. After an introductory page explaining the nature and purpose of the exercise, participants were asked to enter a password, provided in their invitation email, which would identify which of the four waves they belonged to. On the same page, they were asked to provide some basic details – name, email address, age and gender.

Once registered, participants answered six pre-text baseline questions that were geared to assess their current attitudes towards, and understanding of, the use of personal data in medical research. These were:

- How confident are you that you understand the phrase ‘personal data’?
(Extremely confident; very confident; quite confident; not very confident; I don’t understand the phrase ‘personal data’ at all)
- On a scale of one to ten, how much would you say you know about who currently holds data about you?
(where one equals nothing and 10 equals everything you feel you need to know)
- How comfortable are you with the idea of some organisations holding your personal data?
(Extremely comfortable; very comfortable; quite comfortable; not very comfortable; not at all comfortable)
- Sometimes, personal data can help scientists in carrying out medical research. On a scale of one to 10, how much would you say you know about the ways in which personal data can be used in medical research?
(where one equals nothing and 10 equals everything you feel you need to know)
- How comfortable are you with the idea of your personal data being used for medical research?
(Extremely comfortable; very comfortable; quite comfortable; not very comfortable; not at all comfortable)
- On a scale of one to 10, how much would you say you know about the current limits on use of your personal data for medical research?
(where one equals nothing and 10 equals everything you feel you need to know)

Part 2 – Reading the text

Participants were then provided with the following instructions about how to read and respond to the prompt material:

“For the next few minutes we would like you to give your attention to some information about personal data and the way in which this may be used in relation to your health care and for medical research.

We want to understand your views about this. We are particularly interested in what questions and comments come to your mind as you read and think about the information. As you read this through, you will see that some words are highlighted. If you want some further information about the highlighted words you can click on them.

As you read each short chunk of information, please try to be conscious of any questions that you may think of, or whether you find something particularly interesting or contentious. If you do, click on the 'ask a question' button. You will also be prompted to ask questions as you complete each section. If you get to the end of the exercise and find that some of your questions or comments have been dealt with, you can choose not to submit them to us.

At the end we have some final questions to ask about what you think of the use of personal data in medical research and also about how the design of *this* study can be improved. We will also tell you about what will happen in the rest of the study."

When they had read this and clicked 'continue' they moved to the first section of text. The full text, broken down by section, is provided in Appendix A. At the end of the text participants were asked whether they had any additional questions or comments and then asked to select which questions they wanted responses to.

Part 3 – Post-text questions

Participants were then taken through a series of post-text questions. Four of the questions asked at the pre-text phase were repeated here. This section also included some questions about their experience of completing the task. The full list of questions asked in this section was as follows (questions coloured blue were later repeated in Part 5 – see below):

- On a scale of one to 10, how much would you say you know about the ways in which personal data can be used in medical research?
(where one equals nothing and 10 equals a lot)
- How comfortable are you with the idea of your personal data being used for medical research?
(Extremely comfortable; very comfortable; quite comfortable; not very comfortable; not at all comfortable)
- On a scale of one to 10, how much would you say you know about the current limits on use of your personal data for medical research?
(where one equals nothing and 10 equals everything you feel you need to know)
- More specifically, how comfortable are you with the idea of the following types of personal data being used in medical research?
(Extremely comfortable; very comfortable; quite comfortable; not very comfortable; not at all comfortable)
 - Your postcode
 - How often you visit your GP
 - Your cholesterol levels
 - Your occupation
 - Your previous medical history
- [conditional] In the previous question, you said that you would not be comfortable with your XXXXXX being used in medical research. Please say why.
- Which of the following kinds of organisation, if any, regulate the use of personal data for medical research?
(General Medical Council; Universities UK; Research Ethics Committees; House of

Commons Committee on Standards and Privileges; county councils; The Academy of Medical Science)

- How important is it that personal data used in medical research remain 'anonymous'? (Extremely important; very important; quite important; not very important; not at all important; don't know)
- Should researchers ever be able to use medical data that identify the people they relate to without obtaining consent? (please explain your answer)
(Yes, always; yes, under certain circumstances; no, never; don't know)
- To what extent would you trust the following people or organisations to store your personal data?
(Completely; a great deal; quite a lot; not a lot; not at all)
 - Your local hospital
 - Your local GP (doctor) surgery
 - The Department of Health
 - Any other government department
 - A university
 - A pharmaceutical company
 - Another private company
 - The National Health Service (NHS)
- To what extent would you trust the following people or organisations to use your personal data responsibly for medical research?
(Completely; a great deal; quite a lot; not a lot; not at all)
 - Your local hospital
 - Your local GP (doctor) surgery
 - The Department of Health
 - Any other government department
 - A university
 - A pharmaceutical company
 - Another private company
 - The National Health Service (NHS)
- Do you think a system – such as the one currently being developed by the National Health Service, which allows medical staff to access your medical records from anywhere in the country – is a good idea? (please explain your answer)
(Yes, it's essential; I think the advantages outweigh the disadvantages; I have no feelings either way; I think the disadvantages outweigh the advantages; I think it's a terrible idea; it depends; don't know)
- How comfortable are you with the idea of personal data in the NHS system being used for medical research? (please explain your answer)
(Extremely comfortable; very comfortable; quite comfortable; not very comfortable; not at all comfortable; it depends)
- Thinking about the exercise you have just completed, how easy was it to...
(Extremely easy; very easy; quite easy; quite difficult; very difficult; extremely difficult)

- Understand the initial instructions
 - Understand what the research is for
 - Understand what you were expected to do at each point
 - Read the text
 - Understand the text
 - Make comments and ask questions
- [conditional] In the previous question, you said that it was difficult to XXXXXXXX. Please tell us a bit more about why this was the case.
 - Do you think the text was...
(Too long; too short; just the right length)
 - Do you think the text was...
(Too detailed; not detailed enough; just right)
 - On a scale of one to 10, overall how easy did it feel to think of questions about the personal data information?
(where one means that questions came easily to mind and where ten means that it was very difficult to think of anything to ask)
 - [As a way of informing the general public about a complicated issue and finding out about their questions and concerns, are there any ways you can think of in which this exercise could be improved?](#)
 - It may be that answering our questions has prompted some further questions or comments you have about personal data. If so, please note these in the space below.

Once participants had finished these questions, they were presented with a thank you page and informed that the project team would contact them shortly with further details.

Part 4 – Responding to questions and comments

After roughly a week, the project team emailed each respondent with confirmation of the questions and comments they had asked for responses to. Full responses followed 1-2 weeks later, which participants were asked to read through before logging back onto the project website.

Part 5 – Post-task questions

When they logged in for the second time, participants were taken through a final series of questions. As with the post-text questions, these covered both attitudes towards, and knowledge of the use of personal data in medical research and questions about the usefulness and navigability of the tool itself. In addition to those questions coloured blue Part 3 above, we asked the following questions:

- This study has provided you with the opportunity to ask questions that we have then tried to answer. How useful was the feedback you received?
(Extremely useful; very useful; quite useful; not very useful; not at all useful)

- To what extent were you satisfied with the feedback we provided?
(Extremely satisfied; very satisfied; quite satisfied; not very satisfied; not at all satisfied)
- [conditional] Why were you unsatisfied with the feedback we provided? Is there any way we could improve it?
- Overall, as a way of engaging people in a short discussion about the use of personal data in medical research, do you think this exercise has been...
(Extremely effective; very effective; quite effective; not very effective; not at all effective)
- Please rate the study between one and 10 according to the following criteria, where one is low/poor and 10 is high/good.
 - How enjoyable the exercise was
 - How well the exercise kept your interest
 - The objectivity of the information provided
 - How useful the feedback was
- Since you completed the first part of this exercise, you may have discussed the project with friends or family, or looked for additional information on personal data or the project itself. We're interested to know how often this has happened, so if you did follow up the exercise in any way, please give details.
- What, if anything, do you think is the most important thing you have learned through taking part in this exercise?

This completed the exercise.

3 The development of the website

The tool had to have the following features:

- Survey functionality to allow us to collect pre-text, post-text and post-task data from participants;
- The ability to display text to participants a section at a time and to allow users to respond to that text with questions and comments;
- The capacity to record data on participants' use of the site and to report that data.

The following sections cover our approach to each of these areas in more detail.

3.1 Survey functionality

While it had originally been our intention to use standard online survey software such as SurveyMonkey (www.surveymonkey.com) to collect baseline and follow-up data, after some consideration it was felt that directing participants to a different site would be disruptive and potentially lead to a lower conversion rate. White October therefore developed some basic questionnaire functionality (see Figure 1) that was added to the site to enable us to run the survey and text on the same platform. The full questionnaire is included in Appendix B.

The screenshot displays a web interface titled "Introductory Questions". It contains several questions with corresponding dropdown menus for user input. The first question is "How confident are you that you understand the phrase 'personal data'?" with a dropdown menu currently showing "Quite confident". The second question is "On a scale of one to ten, how much would you say you know about who currently holds data about you? (where one equals nothing and ten equals everything you feel you need to know)". Below this question is a dropdown menu with a list of options: "-- Please select --", "1 - nothing", "2", "3", "4", "5", "6", "7", "8", "9", and "10 - everything I need to know". The option "4" is currently selected. The third question is "How comfortable are you with the idea of your personal data being used for medical research?" with a dropdown menu currently showing "-- Please select --". The fourth question is "On a scale of one to ten, how much would you say you know about the current limits on use of your personal data for medical research? (where one equals nothing and ten equals everything you feel you need to know)". Below this question is a dropdown menu currently showing "-- Please select --".

Figure 1 The front end of the tool's survey functionality

3.2 Displaying text; inviting questions

The tool developed for this project was designed to present information to participants in bite-sized chunks and inviting participants to ask questions or make comments whenever a thought occurred to them. The aim was to collect these enquiries without constraining them with questions of our own, though of course the nature of the information does provide a frame for their responses. The tool also needed to collect questions as they occurred to participants. The hope was that this would give us an insight into the ongoing process of absorbing and making sense of new information, highlighting queries or thoughts that participants might have had early on and then forgotten or discounted as they read on.

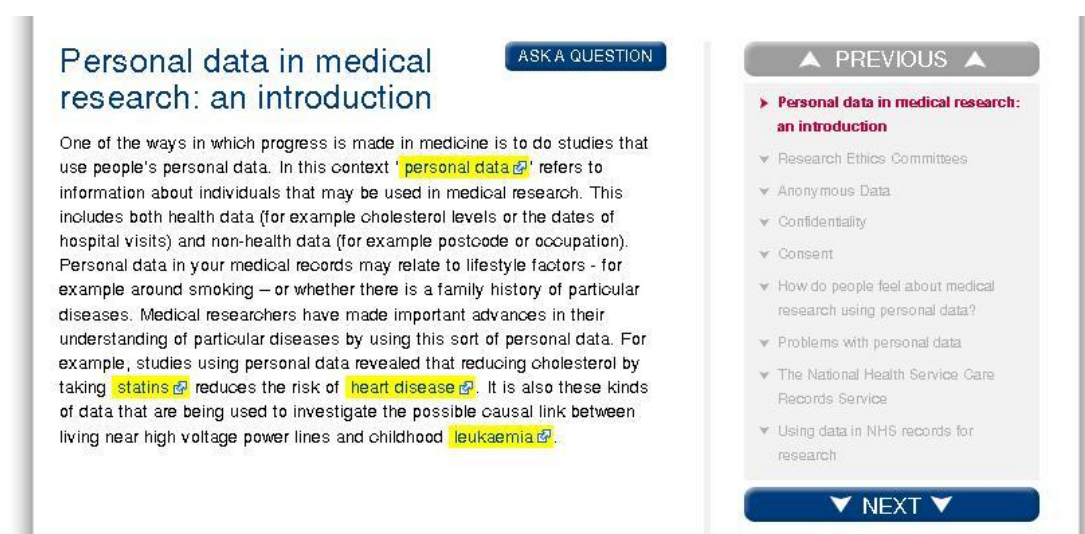


Figure 2 The main reading interface

Participants were therefore presented with text a short section at a time, as shown in the screenshot above. In the instructions, they were encouraged to ask a question or make a comment the moment one occurred to them, making use of the "Ask a Question" button at the top of each section. In addition, if they had not already made use of this button on a given section, participants were prompted to ask questions or make comments when they clicked on the "NEXT" button, as shown below. The times participants spent in the reading pane, providing voluntary questions and comments and providing prompted questions and comments were recorded.

As shown in Figure 3, the 'ask a question' dialogue box, whether opened spontaneously or at the end of a section, never obscured the main text.

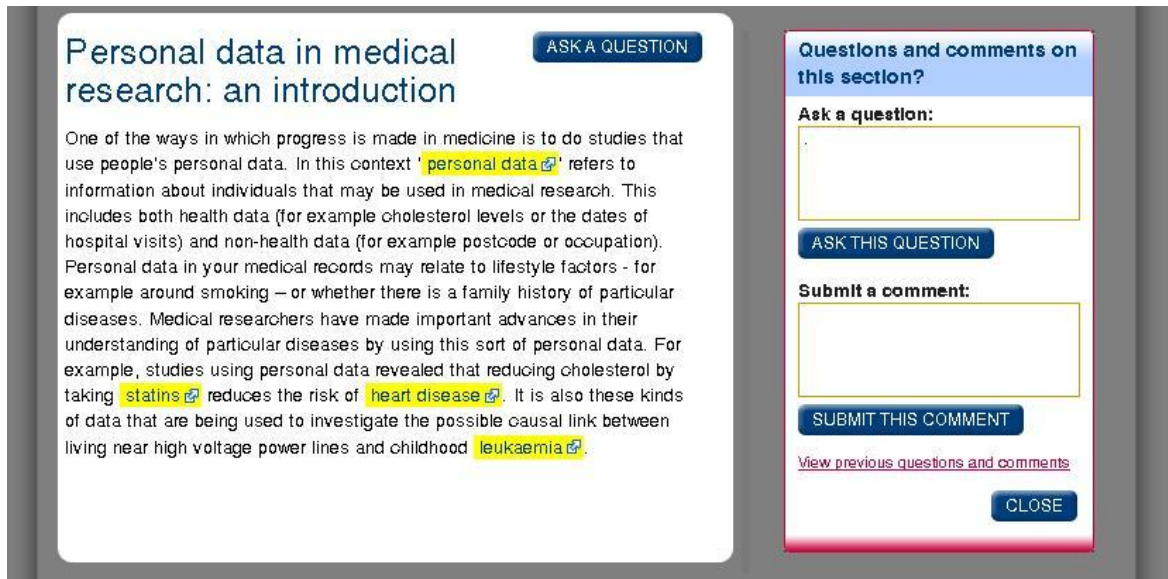


Figure 3 A prompted question dialogue box – note that the main text remains visible

If a respondent wanted to review a piece of text they had read previously at any time, they were able either to click "PREVIOUS" or to use the navigation on the right hand side (see below) to jump back to that point, although they were not permitted to jump forwards over unread sections.



Figure 4 The reading pane's navigation bar

Complex phrases in the text were also highlighted with further information available for participants if they clicked on these elements, as shown below. Which of these elements, referred to as 'glossary terms', were clicked on was then recorded, together with how long they kept the pop up open (the text underneath was faded and partially obscured until the pop up was closed).

Anonymous Data

Some information used in research is anonymous – this information does not identify the individual. The research example – that an unnamed person has a high level of cholesterol, or has tested positive for a sexually transmitted infection – will have been removed. (Another sort of anonymous information that might identify the person – such as name or postcode – will have been removed. (Another sort of anonymous information is **pseudonymised information**).

ASK A QUESTION

PREVIOUS

pseudonymised information [close]

This is like anonymised information in that the holder of the information cannot identify an individual. However, it is different in that the original provider of the information may retain a way of identifying individuals, for example by attaching a code so that the data will only be identifiable to those who have access to the code.

Figure 5 A glossary term pop up

Once all of the text had been read, participants were asked whether they had any additional questions before being presented with all of the comments they had made during the course of the exercise and asked which of them they still wanted responses to (see Figure 6). In this way, participants could filter out questions they felt had now been answered.

Review questions

In the next stage of the study we will be looking at what you have asked and a researcher at the University of Surrey will be working to put some material together that - at least in part - will help give some answers to your questions. We will then come back to this information in about a weeks time and you will have a chance to look at this material and again be able to feed back your views about this.

Please review your questions and comments and **unselect** any you **would not** like feedback on.

- What is personal data?
- "Identifiers" is overly complex here.
- I'm not sure I understand this entirely - how are these records held and where?

CONTINUE

Figure 6 Participants were asked to select which questions they wanted answered

3.3 Recording and reporting data

While participants were reading and responding to the text, information was being collected on how long they spent on each section, which glossary terms they read, whether they returned to any sections and so on. In the administration section of the website, the project team was able to view both each individual's responses and the aggregated data for an entire wave or for all participants (see below). This data could also be downloaded in the form of a Microsoft Excel-compatible CSV file.

In addition to the responses to the pre-text, post-text and post-task questions, and the questions and comments asked by participants, the system also logged the following details:

- Time spent completing pre-text questions, post-text questions and post-task questions;
- Time spent reading each section of the text;
- Number of times each section was visited;

- Total reading time;
- Number of glossary terms read;
- Time spent reading glossary items;
- Number of questions asked;
- Number of comments made;
- Time spent asking prompted questions/comments;
- Time spent asking spontaneous questions/comments;
- Time spent selecting which questions the respondent wants feedback on.



Figure 7 The administration interface showing some of the data for a single participant

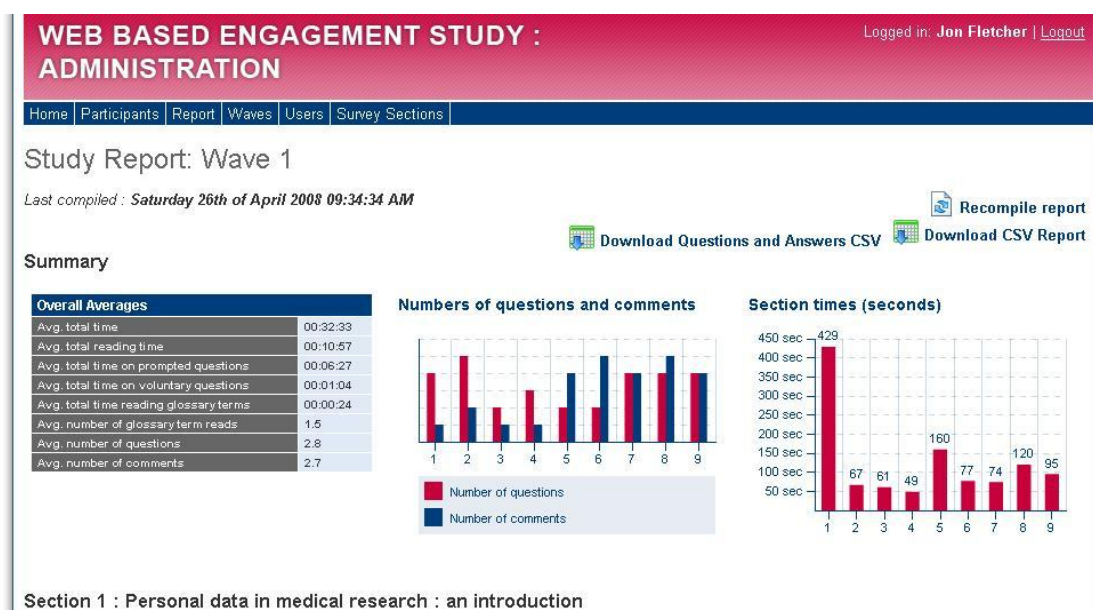


Figure 8 The administration interface showing some of the data for a whole wave

3.4 Testing

A considerable amount of testing was carried out during the development of the site to ensure that it was as bug-free as possible before launch. This was carried out in two waves, first by the project team and then, once we were reasonably satisfied that the basic functionality was in place, by colleagues, friends and family. The feedback received during the process was invaluable in ensuring that the text was readable and jargon free, that unexpected user actions did not 'break' the site and that navigation was as intuitive and natural as possible. It also gave us a reasonably good idea of how long it would take participants to complete the exercise.

3.5 Compiling the text

The challenge in compiling the text was to draw on official sources of information about the uses of personal data in biomedical research, to present this in a balanced way, and to write it in a tone that was accessible and understandable without being too simplistic. We drew on a range of sources (the main ones are detailed below³) and informally piloted the text extensively before entering the online pilot phase.

3.6 Providing feedback

After participants had completed the first phase of the exercise and submitted their comments and questions we contacted participants via the email address of the PI to let them know we had received their comments and questions. We said that we would be in touch again on a specified date with some material that we hoped would be useful in response to these. Two of the team then worked on gathering material for this feedback and each participant was then sent an email with this material. Providing replies that were personal and written in a reasonably conversational style was a lengthy and time consuming process. Copies of all replies sent are available on request and some examples are included in Appendix C. There was some duplication in the sorts of questions that were asked and thus material gathered early on was useful in compiling later replies. For waves 3 and 4 White October assisted in expediting the feedback process by adding a link to Outlook to the admin area of the site for

³ (1) Academy of Medical Sciences (AMS) (2006) Personal data for public good: using health information in medical research, AMS: London www.acmedsci.ac.uk/p48prid5.html

(2) AMS response to the House of Commons Health Committee Inquiry into the Electronic Patient Record and its use www.acmedsci.ac.uk/download.php?file=/images/publication/AMSsubmi.pdf

(3) Text of letter sent from the Department of Health to patients expressing concern over electronic care records www.connectingforhealth.nhs.uk/newsroom/all_images_and_docs/guardian-letter.pdf

(4) NHS Care Record Guarantee 2007 www.connectingforhealth.nhs.uk/nigb/crsguarantee

(5) Report of the Care Record Development Board Working Group on the Secondary Uses of Patient Information <http://www.connectingforhealth.nhs.uk/crdb/workstreams>

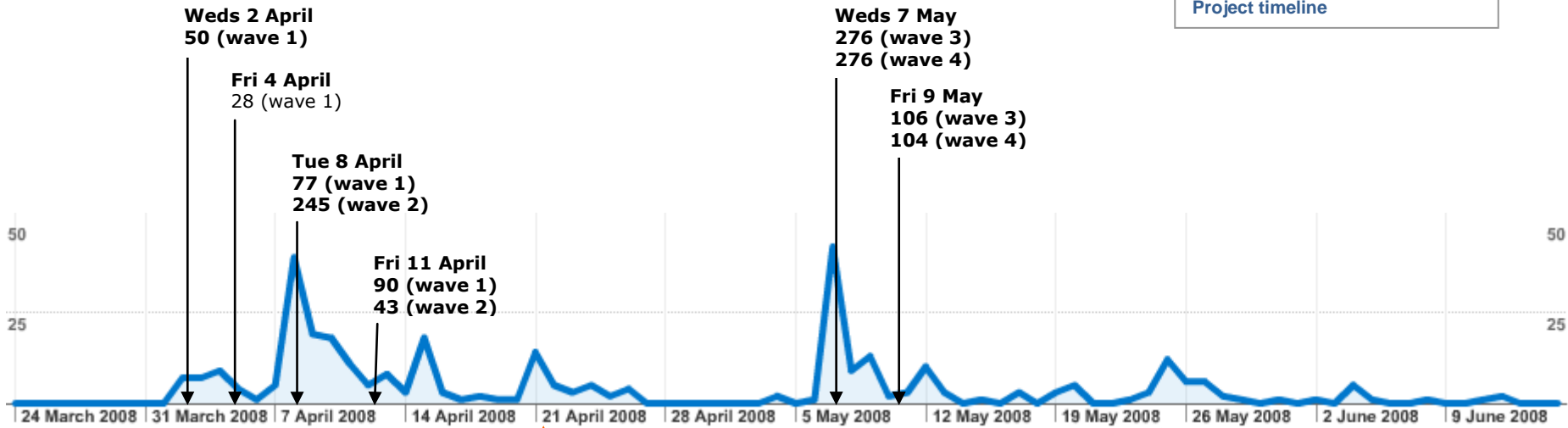
(6) Patient Information Advisory Group response to Report of the Care Record Development Board Working Group on the Secondary Uses of Patient Information <http://www.advisorybodies.doh.gov.uk/PIAG/piagresponse-CRDB-SUS.pdf>

each participant which contained the 'top and tail' of each reply. We then simply added the feedback that had been prepared to individual questions.


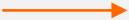
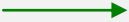
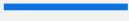
Project Timeline

The timeline on the following page tracks the course of the project, from 2 April 2008 when the first invitations were issued to Sunday 25 May 2008, when the final wave of responses to questions was sent out.

Project timeline



KEY

- 
 Invitations issued by ICM – date and number of invitations
- 
 Responses to participants' questions emailed
- 
 Participants emailed confirmation of their questions
- 
 Number of unique visitors

Tues 8 April Wave 1
Tues 15 April Wave 2
Mon 21 April Wave 2
Tues 15 April Wave 1
Mon 19 May Waves 3 & 4
Sat 24 May Wave 3
Sun 25 May Wave 4

4 Results

We have addressed the question of the feasibility of engaging with people in this way in several different ways. We will present the results of this study in relation to seven main questions:

1. How many participants completed the study and who were they?
2. Did participants produce questions and comments?
3. How much time did participants spend engaging with the exercise?
4. How did participants' attitudes relate to the extent and nature of their engagement?
5. Did the extent to which people were incentivised relate to the extent and nature of their engagement?
6. What did people think of the exercise?
7. What sort of questions did participants ask and what sort of comments did they make?

4.1 How many participants completed the study?

Eighty five people took part in the study. This consisted of 4 waves: 25 people took part in the first wave and then 20 people in each subsequent wave. Of the 85 participants that took part in the first part of the study (completing the pre- and the post-text questions), 74 (87%) of them completed the exercise by returning to complete the final set of questions.

Who were they?

In order to recruit participants into the study, invitations to take part were issued through 'It's Your View', an online panel of around 100,000 people administered by ICM. While this allowed us to ensure that a good mix of ages, genders and socio-economic grades were invited and gave us control over how many invitations were issued, the fact that the participants were used to taking part in online polling almost certainly had an influence on the way they interacted with the website. There was evidence, for instance, that one respondent was completing several surveys simultaneously, switching between windows in his browser. Although ICM applies strict rules about this sort of behaviour and removes offenders from their panel, it is extremely difficult to totally prevent it.

The site went live on 2 April 2008 when the first round of invitations was sent out by ICM. It was originally intended that four waves of invitations would be staggered to allow for changes to be made to the site in between waves. In the event, very little adjustment was required after the first two waves, so waves three and four were run simultaneously.

Since part of the aim of this project was to assess how an online tool might be used to engage the general public, two different incentive schemes were tested over the duration of the project. Invitations sent out in waves one and two offered participants £25 to complete the survey, while participants in the final two waves were rewarded in line with ICM's standard incentive scheme, which pays £1 for every five minutes spent completing a survey.

In total, 1,295 invitations to take part in the study were sent out on our behalf by ICM, in response to which 95 people registered on the site – a 7.6% conversion rate. Of these, 85 took part – 6.6% of the total number of people invited. These are broken down by wave in the table below.

Wave	Number of invitations	Number of registrations	% of invitations converted to registrations	Number of completions	% of invitations resulting in a completed survey
One	245	26	10.6	25	10.2
Two	288	24	8.3	20	6.9
Three	382	22	5.8	20	5.2
Four	380	23	6.1	20	5.2
TOTAL	1,295	95	7.6	85	6.6

Table 1 Invitations issued by ICM resulting in either registrations or completed surveys

While this response rate may seem low, it is not altogether surprising. ICM normally expects a 20% response rate for standard polls administered through the It's Your View system. Since our exercise asked potential participants to leave that system, incorporated unusual incentive payments and involved direct interaction with external organisations, we might reasonably have expected conversion rates to be considerably lower.

The figures in Table 1 seem to suggest that the higher (and fixed) incentive of £25 in waves one and two did act as a draw for potential participants. Later we consider the relationship between the different incentive schemes and the way in which people engaged with the exercise (see section 4.5).

Beyond the actual number of participants, we were also able to collect data on participants' age, gender, location and socio-economic status.

Age

Table 2 and Figure 9 show the ages of the participants in each wave.

Age	All waves		Wave 1		Wave 2		Wave 3		Wave 4	
	No. of participants	%	No. of participants	%	No. of participants	%	No. of participants	%	No. of participants	%
20-34	23	27	10	40	8	40	2	10	3	15
35-49	20	24	7	28	3	15	5	25	5	25
50+	42	49	8	32	9	45	13	65	12	60

Table 2 Age of participants

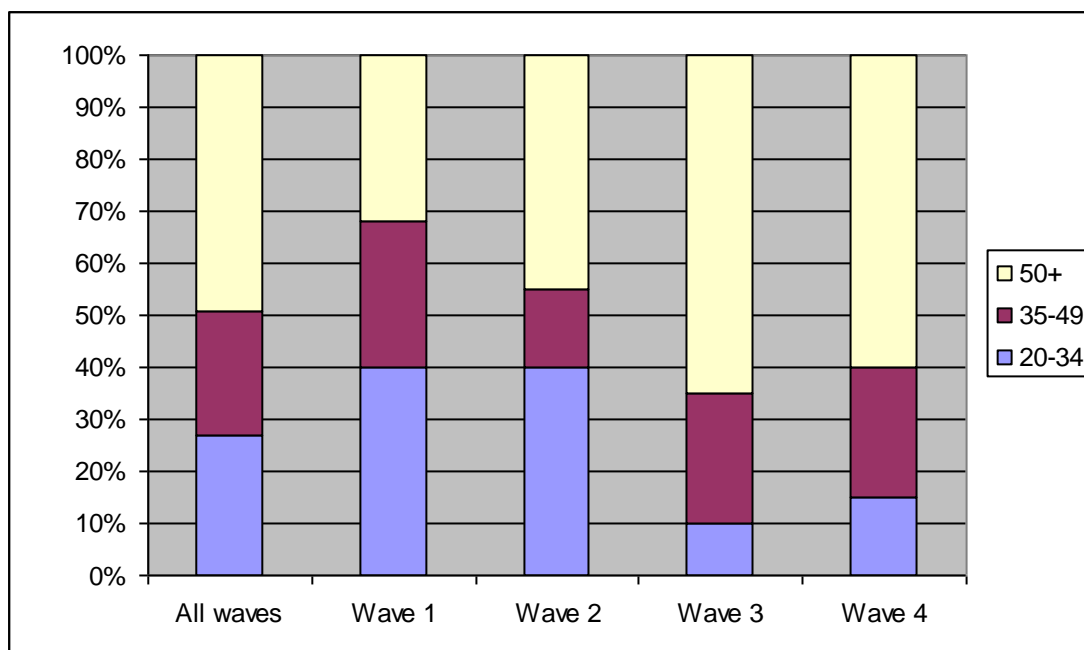


Figure 9 Age of participants in each wave (percentage of participants)

The older profile of participants in waves three and four (the mean age in wave 3 was 55 and 52 in wave 4 compared with 41 and 45 in waves one and two respectively) is due to different approaches to weighting the invitations issued by ICM. During waves one and two, when it was unclear what the response rates would be, ICM sent out invitations in relatively small batches. Relatively early on we noticed that those registering tended to be older and asked ICM to correct this in subsequent batches of invitations by weighting them more towards younger participants. As a result, 171 invitations were sent out to people aged between 18 and 30 in wave one, compared with 48 to people aged between 31 and 50 and 26 to people aged 51 or over. Wave two had a similar profile.

By the time we issued invitations for waves three and four, ICM had a rough idea what the response rates would be and sent out invitations according to a roughly nationally representative age profile: 135 invitations to those between 18 and 30 (135 in wave four), 155 to those between 31 and 50 and 92 to those aged 51 or over.

We asked ICM whether it was typical to see a greater number of older people taking part in an online survey. They responded as follows:

"Generally the older age groups tend to be more responsive and respond faster than the younger age groups. We find this on all of the surveys we do with our panel. It may also have been that as this was different to the usual type of survey we do, and needed more time, younger participants may have been put off."

Gender

As shown in Table 3, the gender profile of the participants was more equal, reflecting an equal distribution of invitations between men and women.

Age	All waves		Wave 1		Wave 2		Wave 3		Wave 4	
	No. of participants	%	No. of participants	%	No. of participants	%	No. of participants	%	No. of participants	%
Male	42	49	11	44	10	50	10	50	11	55
Female	43	51	14	56	10	50	10	50	9	45

Table 3 Gender of participants

Socio-economic Status

Since this project involved the development of a bespoke analytical tool, it was not possible to incorporate a sophisticated tool for recording – or calculating – socio-economic grades. Ideally this would have taken participants through a sequence of questions mirroring the broad classifications within the MRS *Occupational Groupings: A Job Dictionary* handbook, incorporating all of the subtleties and caveats that go with that – job titles with more than one meaning, or the former employment of retired respondents, for instance. In the event, we had to settle for a rudimentary, catch-all solution, but asking the following question:

What is the occupation of the chief income earner in your household? Please provide as much detail as possible (i.e. if they manage a department, roughly how many people work in the department?) (The chief income earner is the person whose annual income is largest)

We then cross referenced each response with the MRS 'Job Dictionary' by hand. As shown in Table 4, the SES profile of the participants was weighted towards B, C1 and C2 in all waves. We did not have SES information for 13 participants (e.g. they simply specified 'retired'). It is perhaps unsurprising that our survey tool, unlike other online polls that use socio-economic quotas to meet nationally representative profiles, failed to attract large numbers of grade A respondents (for whom the incentive may have had little appeal) and grade E respondents (many of whom may not have internet access).

Grade	All waves		Wave 1		Wave 2		Wave 3		Wave 4	
	No. of participants	%	No. of participants	%	No. of participants	%	No. of participants	%	No. of participants	%
A	5	7	2	9	1	6	0	0	2	12
B	19	26	9	41	3	17	3	20	4	24
C1	20	28	5	23	4	22	6	40	5	30
C2	18	25	2	9	6	33	5	33	5	29
D	7	10	3	14	2	11	1	7	1	6
E	3	4	1	2	2	11	0	0	0	0

Table 4 SES of participants

Location

The map in Figure 10 illustrates the distribution of participants across England, Scotland and Wales. This data was collected using web statistics software and suggests that the ICM invitations were successful in securing responses from a broad range of locations.



Figure 10 Location of participants

In addition to those who fully completed the survey first time, a number of people registered and then failed to proceed through the exercise. Some of these later returned, re-registered and completed the process normally; others did not. A third group of participants simply failed to return to the site after having their questions answered, or if they did, they did not complete the final set of questions. The numbers of people falling into each of these categories are outlined below:

- Failed to complete the first part of the exercise and did not return: 7 participants
- Initially failed to complete the first part of the exercise but returned and completed: 4 participants
- Failed to complete the final post-task questions: 11 participants

4.2 Did people ask questions and make comments?

One of the key questions of interest in this study was whether we could design a tool that would elicit questions or comments from the participants. The tool asked for comments and questions separately. Participants were encouraged to ask questions and make comments but they were free to move through each of the Sections without doing this.

57 participants (67%) produced at least one question or comment; 28 (33%) did not ask a question or make a comment. The number of questions or comments that were made ranged

from none to twenty and in total 348 questions and comments were submitted. The number of questions and comments submitted by participants is summarised in Table 5.

Number of questions and comments	Number of participants	Percentage
0	28	32.9
1 - 3	16	18.8
4 - 6	19	22.4
7 - 9	11	13
10+	11	13

Table 5 Total number of questions and comments submitted by participants.

In fact these figures do not give the whole picture insofar as sometimes more than one question was asked within a single box. Also of course, comments varied greatly in length. (See more about this in Section 4.3 below).

Figure 11 depicts the number of questions and comments that were contributed across each section of the text: Section 2 (Research Ethics Committees) prompted the greatest number of questions and Section 6 (How do people feel about medical research using personal data?) the fewest. In total there were 155 question entries. For comments Section 2 saw the fewest and Section 8 (the NHS Care Records Service) the greatest number. In total there were 193 comment entries.

It is interesting that for Sections 1 - 4 the number of 'questions' exceeded the number of 'comments' and that this pattern was reversed for Sections 5-9. There are several possible reasons for this. It could be that the nature of the material covered in the later sections lent itself to comment to a greater extent than earlier sections. Arguably for example, material in Sections 6 - 8 about how people feel about personal data in medical research and problems with personal data more generally are more likely elicit reflection and evaluation rather than questions.

It may also be that as people worked through the exercise and became more familiar with the area that they became more confident to comment. A third possibility is that as the content of the Sections were of course linked, that new thoughts and possibilities for elaboration were exhausted before the end of the material. It is also worth noting that after all the sections of text had been viewed there was a 'mopping up' section where participants were invited to make additional comments and this option was taken up 27 times (23 different participants). If we consider this an indicator of engagement with the material this is particularly noteworthy insofar as participants were not told ahead of time that they would have this opportunity nor were they able to see any sections of text at this point.

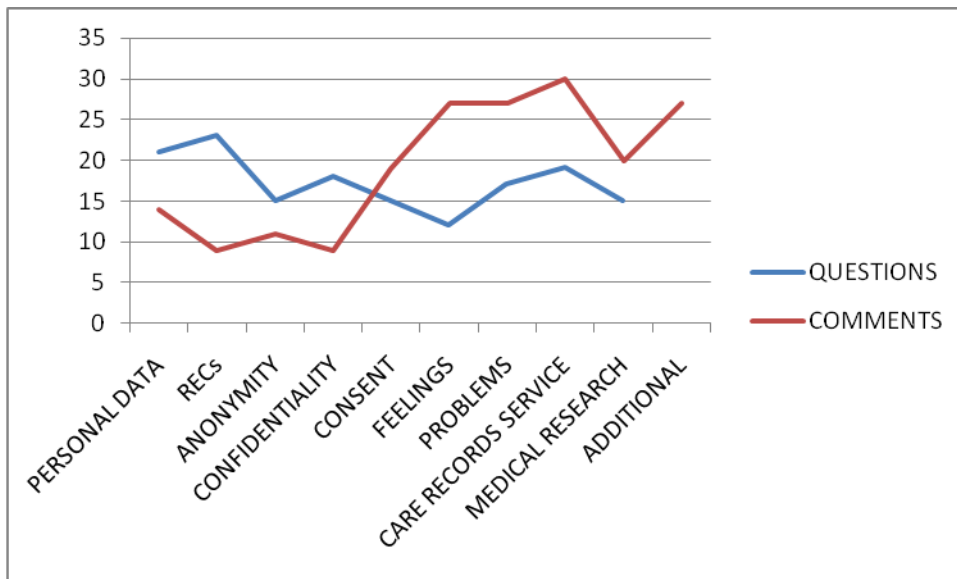


Figure 11 Total number of questions and comments recorded in each section

4.3 How much time did people spend on the exercise?

One of the capabilities of the tool we developed is to record behavioural measures that can serve as proxies for the extent to which people have engaged with it. One such measure is time. The software allowed us to measure the times that participants spent within different components of the tool. These measures are of interest in themselves but more importantly we are interested in the way in which they relate to:

- Our other main behavioural proxy of engagement: the numbers of questions asked and comments made
- Attitudinal measures
- Self reported knowledge
- Changes in attitudes and knowledge over the course of the online deliberation

Total time

The total time score at Time 1 (following completion of the post-text questions) is comprised of six components: the time spent (1) answering pre-text questions; (2) reading the text; (3) reading glossary terms; (4) producing questions and comments, (5) reviewing questions asked, and (6) answering the post-text questions. The time taken completing the post-task questions (at Time 2) was also recorded but is not being considered here.

Of course the time that participants took on different components of the exercise (e.g. reading the material, recording questions or comments, or answering the pre or post-text questions) may not be an accurate reflection of the time which the participant actually spent with engaging with the material. For example they may have taken a phone call whilst they were completing the exercise. In reporting the times that participants took for each of the components of the exercise, and in order to explore the relationship of time variables to other measures, it is

important to identify those that took a much greater time as statistically these 'outlier' scores will unduly bias means and variances.

The total time scores of three participants were classed as outliers in that the time they took was different from the mean time by more than two standard deviations. They took 1hr:43mins, 2hr:44mins and 3hr:7mins respectively. In all statistical analyses involving time these three participants were excluded. For the other 82 participants the time taken ranged between 6 and 55 minutes with a mean time of 24 minutes.

It is important to note however, that two of these outlier times were recorded by participants with the highest number of questions and comments – 18 and 20. In this sense these participants were taking part in the task as we intended; that is to say that greater time taken was related to a greater number of questions and comments being produced. Thus in analyses that do not involve measures of time (for example, when we look at the comments that participants made about the process (Section 4.6), or the nature of the questions that were asked (Section 4.7) these participants are included.

Time on each section

We can extend our analysis of the way in which participants asked questions and asked comments (in section 4.2 above) by considering the time spent asking questions or making comments in relation to each section of text. The mean length of time that was spent asking questions and making comments for those that made at least one question or comment is depicted in Figure 12 below. This does not suggest that participants were less willing to engage with the exercise in the later sections. Clearly there was considerable variation between sections, with the least time being taken in relation to questions or comments around anonymity (section 3 - 31secs) and the most in relation to the section on problems (section 7 - 1min:23secs).

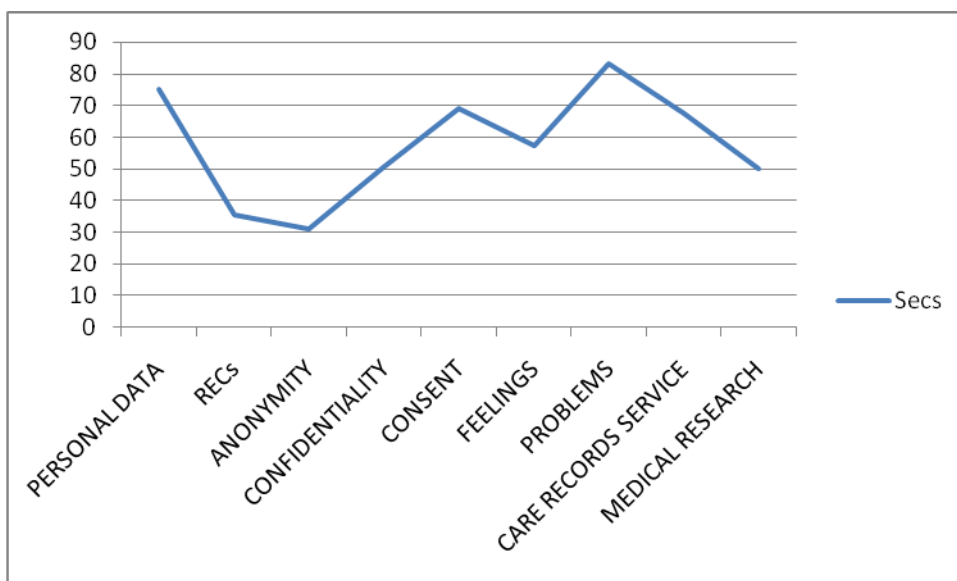


Figure 12 Mean length of time spent asking questions/making comments (for those that made at least one question or comment) by section

Reading time

Moving on from this consideration of total time, we can look more closely at one component of this: reading time. This refers to the sum of time spent reading the text in each of the sections and does not therefore include time spent asking questions or comments, or reading glossary terms. Figure 13 depicts the reading time taken in relation to each section. Of course, to some extent reading time reflects the number of words and indeed, a comparison of reading time per section with the number of words per section (depicted in Figure 14) illustrates this⁴.

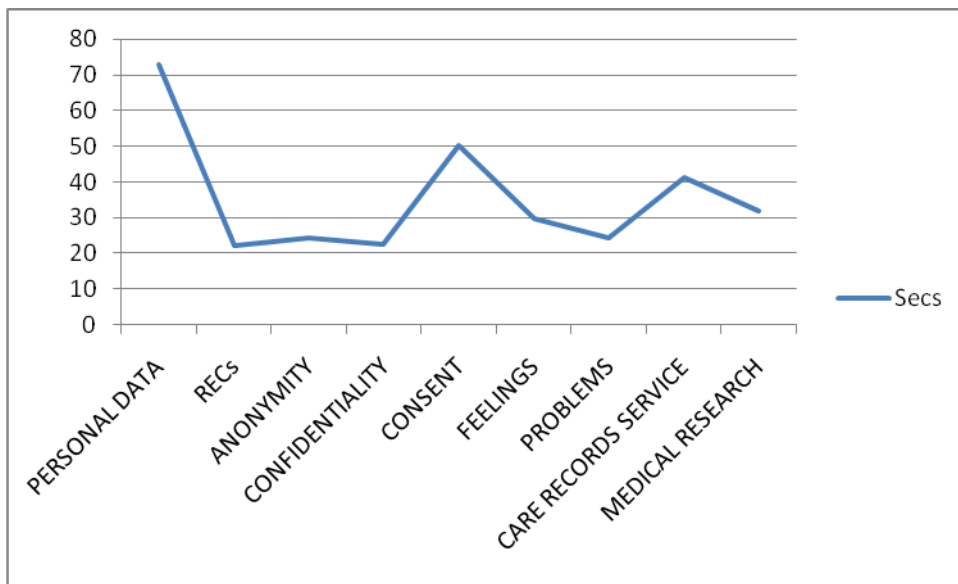
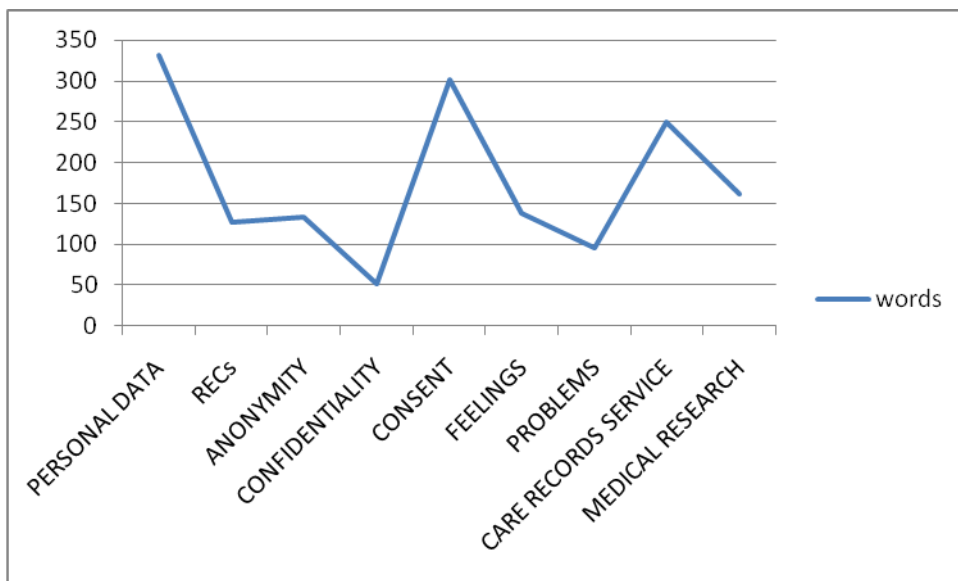


Figure 13 Reading time in relation to each section of text



⁴ We recognise that reading time will also be a function of a range of other factors relating to characteristics of both the material being read and of the reader. Resource constraints precluded us from formally considering these on this occasion.

Figure 14 Number of words per section of text

Glossary terms

Participants also had the opportunity to explore highlighted glossary terms. In total there were 10 glossary terms (see Appendix A). The facility to monitor participant exploration of glossary terms was designed as a further proxy for engagement with the material provided insofar as it is a choice that the participant makes to seek out the material underlying the term and doing this is not required to move further through the exercise.

Forty-five participants did not click into any glossary terms and 40 did. For those that did click into at least one glossary term the time they spent on the 10 terms in total ranged between 2 seconds and 162 seconds with a median time of 22 seconds.

Time-based measures of engagement - overview

Reading time, time producing questions and comments and time spent on looking at glossary material can be considered as three time based indicators of participant engagement with the material. As such we would expect measures of these activities should be correlated with one another. Table 6 below depicts the relationships between these time based engagement measures and the other measure of engagement: numbers of questions and comments produced. (We have also included a column where we have weighted actual reading time in relation to the length of the text being read).

	Reading time (weighted)	Reading time (unweighted)	Glossary time	Questions/ comments time	Number of questions/ comments
Reading time (weighted)					
Reading time (unweighted)	0.92**				
Glossary time	0.50**	0.41**			
Questions/ comments time	0.35**	0.30**	0.19		
Number of questions/ comments	0.40**	0.40**	0.25*	0.51**	

* ≤ .05; ** ≤ .01

Table 6 Relationships between time-based engagement measures and other measures of engagement

Overall there is a clear picture of a cohesive set of relationships between the different measures of behavioural engagement with the text. Time spent reading the text was significantly correlated with time spent looking at glossary terms, with time spent producing questions and comments and with the number of questions and comments produced. Time spent looking at glossary terms was significantly related to the number of questions and comments produced though not to the time spent producing them.

4.4 Engagement and attitudes

Another dimension of considering the way in which the tool worked is to examine the relationship between people's views about personal data and the nature and extent of their engagement with the study: What difference does the level of engagement with the task make to people's views?

Who produces most questions and comments?

The first question we asked was, "How confident are you that you understand the phrase 'personal data'?" No-one selected the option that they did not understand it at all. Two people said they were not very confident; 20 that they were quite confident; 31 said they were very confident and 32 that they were extremely confident.

There was a small but significant difference between those that did and did not go on to ask questions/make comments in that those who were less confident that they understood the phrase 'personal data' (mean = 3.96) were more likely to ask questions/make comments ($t(83) = 2.06, p \leq .01$) than those that were more confident (mean 4.36).

This relationship is both surprising and encouraging in relation to the possible future uses for this engagement tool. Certainly in a face to face situation one might reasonably expect that greater confidence was associated with a greater willingness to make comments and ask questions. Equally, a feeling of uncertainty about a subject might make people more reticent about revealing their lack of knowledge. In this instance it seemed that voluntarily asking questions/giving comments was undertaken more readily by those that were less sure that they understood the topic being considered. It may be that the online environment enables people to feel less concerned about showing any perceived ignorance.

Engagement and perceived knowledge

Two questions were asked about perceived knowledge at three stages: pre-text, post-text and post-task.

- How much would you say you know about the ways in which personal data can be used in medical research? (1 – nothing; 10 = a lot)
- How much would you say you know about the current limits on use of your personal data for medical research? (1 – nothing; 10 = a lot)

We wanted to look at participant responses to these questions at each of the three stages and, more importantly, to explore whether – and how – the extent of their engagement with the tool made a difference to their responses.

We constructed a single perceived knowledge measure and conducted a repeated measures ANOVA. Our engagement measure here was whether people had asked questions/made comments (yes/no) and this was included as a between subjects factor. There was a significant

main effect of self assessed knowledge – it increased across the 3 time points ($F(2) = 19.13, p \leq .01$). However there was also a significant interaction effect with the 'asking questions' variable ($F(2) = 6.29, p \leq .05$). Examination of the contrasts reveals that this result is due to the difference between the post-text and post-task ratings. This can be seen in Figure 15 below.

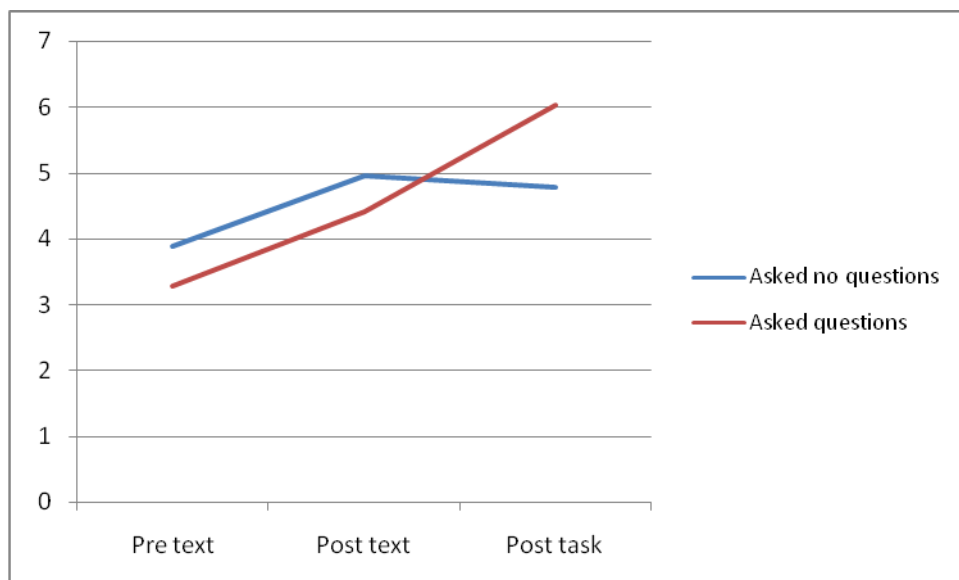


Figure 15 The relationship between self-assessed knowledge and propensity to ask questions at different stages of the research exercise

Ratings of self assessed knowledge of those participants that had produced questions or comments increased after they received relevant feedback to their questions. Of course, those that had not asked any questions did not receive any feedback and their self-assessed knowledge decreased slightly between post-text and post-task. This certainly indicates that at some level participants engaged with the feedback they received in a way that led them to assess what they knew about how personal data can be used and the limits of its use more positively.

We can address the same question using an alternative measure of engagement: reading time⁵. There was a significant main effect of self assessed knowledge – it increased across the 3 time points ($F(1.67, 78.43) = 22.45, p \leq .001^6$). However there was also a significant interaction effect with the 'reading time' variable ($F(1.67, 78.43) = 4.11, p \leq .05$). Examination of the contrasts reveals that this interaction is significant both in relation to the difference between the pre-text and the post-text ratings and the post-text and post-task ratings. This can be seen in Figure 16 below.

⁵ This is a single measure which was computed by (1) weighting the reading time of each section by taking account of the word length of each section (2) creating a mean reading time length (3) converting this to a categorical variable with 2 levels: the top third of scores were designated as high reading time and the bottom third as low reading time.

⁶ The F ratios have non-integer degrees of freedom as it was necessary to use the Greenhouse-Geisser correction as Mauchly's test indicated that the assumption of sphericity had been violated.

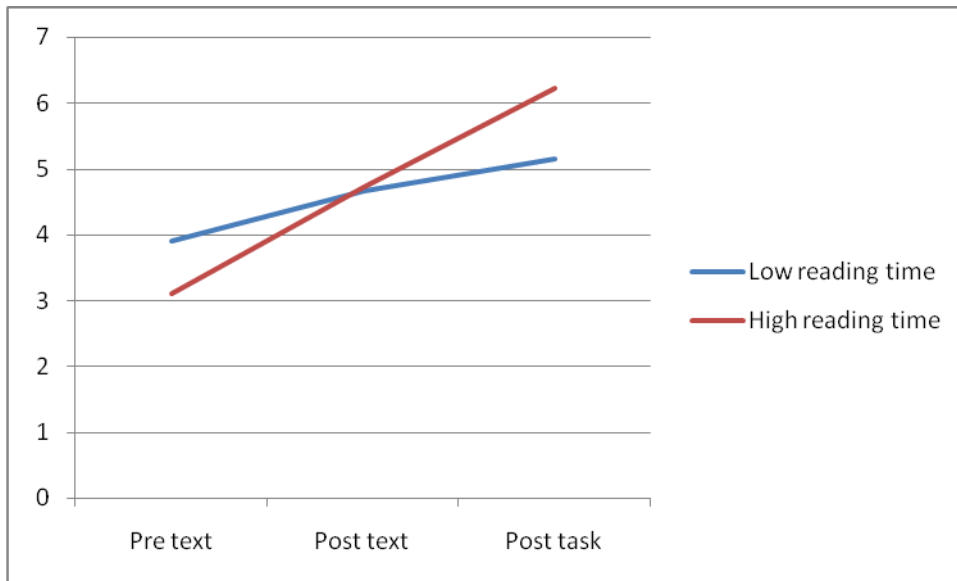


Figure 16 The relationship between self-assessed knowledge and reading time

It is noteworthy also that in both of these analyses the people that engaged more (by producing questions and by spending longer reading the text) were also those that assessed their knowledge as lower at the pre-text stage. This echoes the finding reported in relation to 'Who produces most questions and comments?' above.

How comfortable are you?

A single question asking how comfortable people were with the idea of their personal data being used for medical research was also asked at three stages: pre-text, post-text and post-task. Response options ranged from 1 (not at all comfortable) to 5 (extremely comfortable).

As above, a repeated measures ANOVA was conducted to explore whether how comfortable people were with the idea of their data being used for medical research at different stages was affected by the extent of their engagement with the task (as measured by whether or not they produced questions or comments).

The analysis here provides a different picture to that found in relation to perceived knowledge. Although there was a significant main effect of ratings of comfortableness over time ($F(2, 144) = 3.46, p \leq .05$), this was not affected by whether or not the participants had asked any questions. The mean ratings at the three time points are depicted in Figure 17 below.

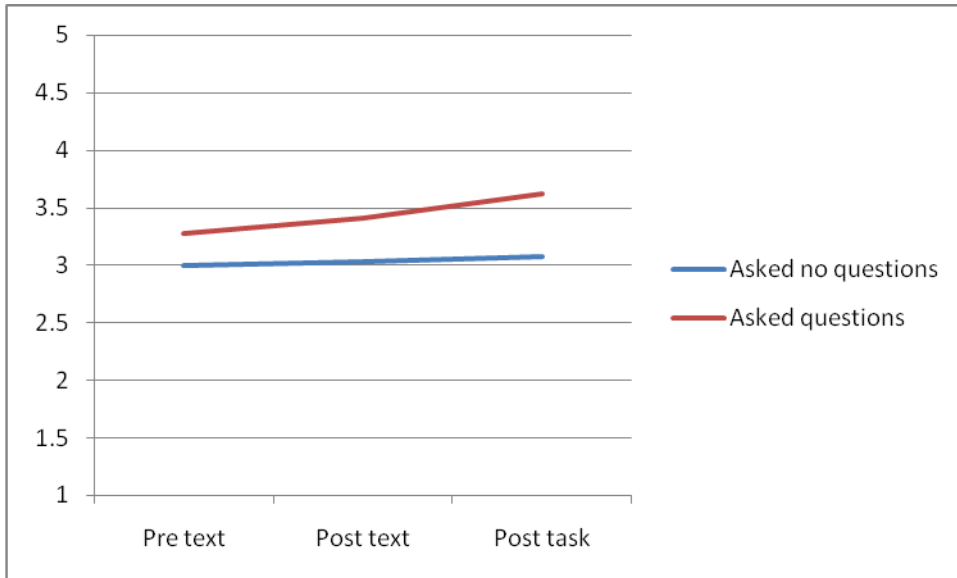


Figure 17 The relationship between people's feelings of comfort in relation to the use of their personal data for medical research and the propensity to ask questions

We then explored whether ratings of comfortableness changed in relation to the time that participants had spent reading the material. As with the 'producing questions and comments' measure of engagement, there was a significant – though small - main effect of ratings of comfortableness over time ($F(2, 144) = 3.46, p \leq .05$) but this was not affected by time spent reading the material. The mean ratings at the 3 time points are depicted in Figure 18 below.

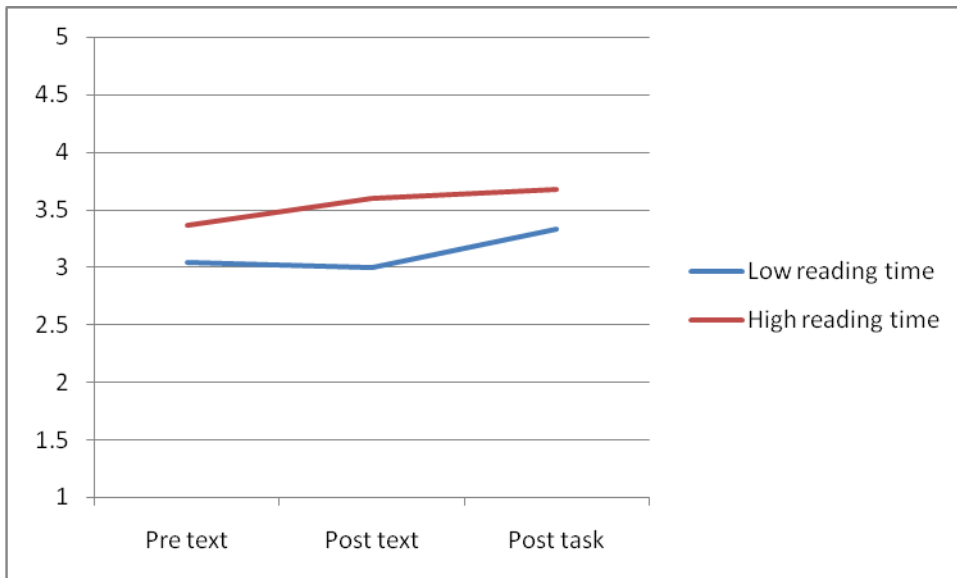


Figure 18 The relationship between people's feelings of comfort in relation to the use of their personal data for medical research and reading time

Other ratings were obtained at the post-text and the post-task stage. Here too we can explore whether any change in participant ratings was a function of the extent to which people engaged

with the task. We will use whether or not people produced questions or comments as the proxy for engagement here⁷.

The four items are:

1. How important is it that personal data used in medical research remain anonymous?
2. Should researchers ever be able to use medical data that identify the people they relate to without obtaining consent?
3. Do you think that a system such as the one currently being developed by the NHS which allows medical staff to access your medical records from anywhere in the country, is a good idea?
4. How comfortable are you with the idea of personal data in the NHS system being used for medical research?

There was only a significant interaction effect, for one of the four items: the degree of engagement was predictive of a significant change in participant ratings for item 1. There was no overall difference between the importance of anonymity ratings between post-text and post-task. However, these ratings were significantly affected by whether or not participants had engaged by asking questions/making comments ($F(1, 72) = 6.58, p \leq .05$). This interaction effect is depicted in Figure 19 below. Those that had produced questions or comments said that it was less important that data should remain anonymous than those that did not.

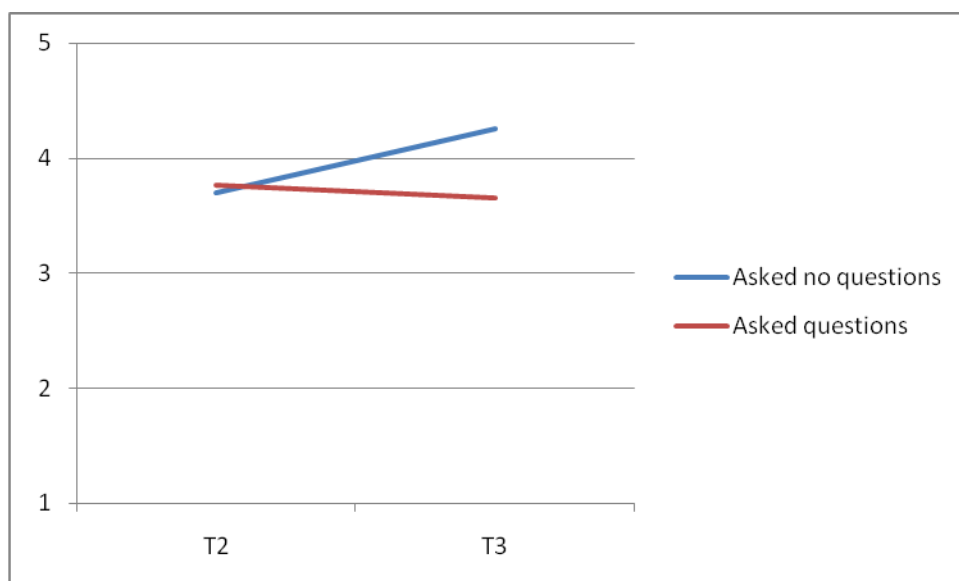


Figure 19 The relationship between ratings of the importance of anonymity and reading time

The report by the Academy of Medical Sciences notes the value of a less conservative approach to anonymity: there are occasions where it is vital that researchers have access to identifiable data. This was referred to in the Confidentiality section: 'Confidentiality is an important concept

⁷ The reading time engagement measure is much less relevant as we are focusing upon the change occurring between post-text and post-task ratings (both of which are after the reading of the text has taken place). In contrast the producing questions/comments engagement measure is very relevant at this stage as one group receives feedback and the other does not.

because personal data that has had its identifiers removed – i.e. that is anonymous - is often not very useful to researchers'. This text was viewed by all participants prior to the post-text stage. It was however, after at the post-text stage (where those that had asked questions received feedback) that this difference between those that had and had not asked questions became apparent. Of course not all feedback related to this point in particular although this was one of the key themes of the questions that were asked.

4.5 Engagement and incentives

One of the issues that we took the opportunity to explore in this study was the effect of the different methods of incentivising participants on the way in which people engaged with the tool. Waves 1 and 2 were recruited on the basis that they would receive a flat rate incentive of £25 – this was a higher rate than normally available through the recruitment company. Waves 3 and 4 were incentivised at the normal rate of £1 per 5 minutes. The mean amount paid to Wave 3 and 4 participants was £7 (we will therefore refer to this group as the Low Incentive group and to those receiving the £25 as the High Incentive group).

When we look at how the completion rate varied in relation to the size of the incentive offered we see that in the high incentive group 43 (95.5%) returned to complete the post-task section and 2 did not. In the low incentive group 31 (77.5%) returned and 9 did not. (It is worth noting in passing that the three participants that spent the greatest amount of time in the exercise (see section 4.3) were all from Wave 1 and thus not incentivised in relation to time spent).

When we consider the differences between the two incentive level groups, there are no significant differences between them in terms of reading time, time producing questions or comments or in time looking at glossary terms.

There is however a difference between the two groups in what we might consider to be the measure most indicative of 'active engagement': the number of questions and comments that they produced. The low incentive group asked significantly fewer questions (mean = 2.68) than the high incentive group (mean = 4.69). Because the low incentive group had a higher percentage of older people we included age in the analysis also to check whether the greater representation of older participants in the lower incentive group might explain this. An ANOVA showed that there was no effect of age and no interaction effect on number of questions or comments produced, simply a main effect of incentive ($F=7.42$, $df = 1$, $p \leq .01$). Those promised the higher incentive produced significantly more comments/questions.

Participants in the two incentive groups did not vary in relation to any of the pre-text measures.

4.6 What did people think of the exercise?

At the end of the post-text section, participants were asked a series of questions about the exercise that they had taken part in. They were asked how easy it was to understand (a) the initial instructions, (b) what the research was for, (c) what they were expected to do at each point, (d) to read the text, (e) to understand the text, and (f) to make comments and questions.

There were six response options ranging from extremely difficult (1) to extremely easy (2). Participant ratings can be seen in Table 7. There were no differences between the three age groups of 18-30, 31-50 and 51+ or between SES groups (A/B, C1/C2 and D/E) on any of these variables.

Question (n= 85 for all questions)	Extremely/very/quite easy	Extremely/very /quite difficult	Mean
Understand initial instructions	79 (93%)	6 (7.1%)	4.9
what the research was for	81 (94%)	4 (4.7%)	4.9
what you were expected to do at each point	82 (96%)	3 (4%)	4.9
to read the text	85 (100%)	---	5.2
to understand the text	85 (100%)	---	5.1
to make comments and questions	79 (93%)	6 (7.1%)	5

Table 7 Participants' assessments of the ease of use of various aspects of the online tool

We also asked participants to rate the length and detail of the text. 79% rated the length of the text as 'just right' and 82% rated the level of detail as 'just right'.

In the light of this feedback after Waves 1 and 2 we decided it was not necessary to conduct any further adjustments to the tool and to proceed with the final two waves with no changes.

We also asked people how easy it felt to ask questions about the information (where 1 meant questions came easily to mind and 10 that it was very difficult to think of anything). The results were quite evenly spread here though more inclined towards the view that it was easy rather than difficult, as depicted in Table 8. It is noteworthy too that there is a significant negative correlation between this 'ease of asking questions' measure and the number of questions ($r = -.3, p \leq .01$): those who thought it was more difficult to think of questions actually produced fewer.

Rating	Number of participants	Percentage
1 - 2	11	22.4
3 - 5	37	34.1
6 - 8	24	27.0
9 - 10	13	16.4

Table 8 Participants' assessments of the ease of asking questions

As part of the post-task survey we asked participants for their views about participating in this exercise. Everyone was asked about how enjoyable the exercise was; how well the exercise had held their interest and the objectivity of the information provided (where 1 was low/poor and 10 was high/good). Table 9 depicts the mean ratings:

Question (n= 74 for all questions)	Mean	Standard Deviation
How enjoyable the exercise was	8.08	1.79
How well the exercise held your interest	8.35	1.62
The objectivity of the information provided	8.47	1.38

Table 9 Participants' assessments of the exercise

Clearly people were very positive about the exercise overall: they enjoyed it; felt that it held their interest and that the information provided was objective. It is noteworthy that there were significant relationships between these subjective evaluations and the actual extent of participant engagement as measured in the number of questions and comments that they had contributed. (Correlation between rating of how enjoyable exercise was and number of questions and comments, $r=.30$, $p \leq .01$; rating of how well exercise kept interest, $r=.30$, $p \leq .01$; rating of objectivity of information $r=.29$, $p \leq .05$). Although the ratings by all participants were generally positive, people were more positive about the exercise when they had engaged with it to a greater extent by asking questions.

Finally, we asked those that asked questions and made comments how useful the feedback that they had received was and how satisfied they were with it. Both ratings were on a five point scale from (1) not at all useful/satisfied at all to (5) extremely useful/satisfied. All participants were asked how effective the exercise had been as a way of engaging people about personal data. Ratings ran from (1) not at all effective to (5) extremely effective. Table 10 depicts the mean ratings.

Question	Mean	Standard Deviation
How useful was the feedback you received? (n= 51)	4.31	.65
How well the exercise held your interest (n = 51)	4.49	.61
How effective was the exercise as a way of engaging people about personal data?	4.11	.77

Table 10 Participants' assessments of feedback and the effectiveness of the exercise

Sixty four percent of participants requested a report of the study and 87% said that they would be happy to be contacted in the future to take part in a further study in this area. We are preparing a short report to send to those participants who requested it.

4.7 What were people's comments and questions?

A key aim of developing this on line engagement tool was to facilitate people's reflection on a topic which they knew little about. The essence of the tool, as it stands, is a structure that facilitates an exchange between two people about a topic. This is done within a 'thin' context of social interaction: there is no face to face exchange, and no interaction with a wider social group. In this thinner context however, it does aim to capture the key aspects of deliberation⁸.

- a. Informational: to challenge people with new information and in relation to beliefs that they already hold
- b. Argumentative: to draw attention to the way in which issues link up, to their coherence or lack of it and to make hidden assumptions visible
- c. Reflective: to encourage people to reflect on their preferences in a context where these have to be justified
- d. Social interaction: to enable people to talk and to listen and to do this in a way that makes the place of this within in a social group salient.

One of the key outputs of this tool is the qualitative data: the nature of the comments and questions people made as part of their engagement with the exercise. Of course we might expect that when this tool is used in future one of the key points of interest for those who use it is the insight that the questions and comments afford about a particular topic. At this stage though our focus has been on the evidence that we have that the tool is capable of eliciting and facilitating participant engagement with a topic. We will do this by looking at the extent to which there is evidence that different aspects of deliberation occurred.

Evidence for information seeking

There was clear evidence that the information that was provided about personal data and its use in medical research stimulated a search for new information. It is likely that this search was particularly facilitated by the encouragement the exercise provided to ask questions. Examples of information that was sought in relation to each section of the text can be seen in Table 11 below. To give some indication of the breadth of questioning, each question in this Table was asked by a different participant.

Evidence for argumentation

Many comments and questions were contributed that indicate the use of argumentation. This material (1) indicated links between issues, (2) sought to establish, or challenge, coherence and (3) to make hidden assumptions visible.

⁸ Dryzek J & List C (2003) Social Choice Theory and Deliberative Democracy: A Reconciliation, *British Journal of Political Science* 33(1): 1-28

Examples are provided in Table 12. Again this is done in relation to each section of text in order to acknowledge the way in which this material drew on the context(s) that had preceded it. It is noteworthy that some sections seemed to elicit more 'argumentation' comments and questions than others. There was evidence of variation between individuals in the number of such comments/questions they came up with. Overall many of the questions and comments that were produced showed that participants were concerned to interrogate the text for assumptions and implications and to probe the value of the concepts that were being explored in relation to the wider social and political context.

Evidence for reflection

In a true deliberative exercise, insofar as participants state their preferences, this is done in a context where they may be called upon to justify their views and to be accountable for them. Of course for this reason, in a face-to-face situation, people may choose not to explicitly express their preferences. In this online conversation however, there were many examples of participants stating their preferences and using detail (or lack of it) from the text to justify or to provide caveats to the position taken. As the conversational 'turns' of the exercise were limited in this pilot study, it was not possible (on this occasion) to explore how further exchanges would facilitate the development of reflection. Some preferences are expressed in relation to social value judgements. Others are expressed with reference to the personal and these are often linked to the expression of affect.

Examples of participant reflection about their preferences can be seen in Table 13.

Evidence for social interaction

As noted at the outset, the deliberative context of social interaction for this tool is 'thin': there is no face to face exchange, and no interaction with a wider social group. It is noteworthy however, that participants drew on their own social contexts in constructing questions and comments. For example, they related their own family, health or occupational experiences to validate or challenge the information being provided in the text and also referred to the media or to other imagined publics of whom they had no direct experience.

More importantly, however, it is important to note that participants' responses were often written in a conversational way and were predicated on the assumption that there was a dialogue being conducted with the research team. The tone of many of the replies suggested this: 'are you saying that..?'; 'I see your point but..'; 'Exactly!'; 'Sorry I took so long on this bit - had a phone call!'; 'Hmmm.....we will see'; 'No thank you to the 2nd paragraph.'; 'No comment on this one I am afraid'.

SECTION	INFORMATION REQUESTED
SECTION 1 PERSONAL DATA IN MEDICAL RESEARCH – AN INTRODUCTION	How are things like medical history and lifestyle factors submitted from medical records to researchers...how is info gathered?(P83); How intricate does the non-health data get before being classified as 'really personal'? (P125); Who decides how when and what personal data is collected (P174)
SECTION 2 RESEARCH ETHICS COMMITTEES	Who makes up these research ethics committees and what criteria do they use? (P80); Who regulates the Research Ethics Committee and its decisions? (P88); Of which bodies / individuals are these committees formed? Govt. Dept? Watchdog? Medical Board? Private Sector? (P94)
SECTION 3 ANONYMOUS DATA	Can the researcher access the code to identify the person? (P108); It says 'some information is anonymous' - in what instances does it remain anonymous and what kind of cases is it not anonymous? (P112); Some information used in research is anonymous but which information is not? (P171)
SECTION 4 CONFIDENTIALITY	Why is anonymous data often not very useful? (P78); Is consent only given once or is it asked for each different study? (90); If consent is given for example for an external body to access your medical report is all this 'secret' non-relevant information also given out? (P127)
SECTION 5 CONSENT	In what situations is consent required for the use of anonymous personal data and why is it necessary in some cases and not in others? (75); What is stopping them using the data without consent
SECTION 6 HOW DO PEOPLE FEEL ABOUT MEDICAL RESEARCH USING PERSONAL DATA	Can you ask not to have your data used for specific research that you may not agree with? (P180); How is consent obtained from an individual? (P91); How can individuals be confident that personal information is secure and will not leak to other parties? (P73)
SECTION 7 PROBLEMS WITH PERSONAL DATA	What measures are in place to prevent this happening again? (P96); What protection against this would be put in place? (128); Could the same thing occur with Medical Records as the Child Benefit CD's (P123)
SECTION 8 THE NHS CARE RECORDS SERVICE	Is there any chance that a person's file can be 'accidentally' deleted? (P77); would the individual be able to see their records? (P108); who decides to whom this data needs to be released? (P168)
SECTION 9 USING DATA IN NHS RECORDS FOR RESEARCH	Is the use of 'England' in the second paragraph the usual synonym for the UK, or is it a system confined to the English NHS? (P113); Who will be held responsible if something goes wrong? (p122); What lessons can we learn from other countries about the use and centralisation of medical data?

Table 11 Information requested in relation to each text section (The numbers after each quote refers to the participant number)

SECTION	ARGUMENTATION
SECTION 1 PERSONAL DATA IN MEDICAL RESEARCH – AN INTRODUCTION	Is using personal data on paper really a valid way of making causal links. From experience I know that many patient may be on prescription drugs, but don't actually take them. (P78) Not clear at this juncture exactly how the progression from data capture / analysis to study occurs (P94)
SECTION 2 RESEARCH ETHICS COMMITTEES	What qualifications do these members have, or are they just pen pushers [&] quangos? (P154); I have read that consent has not always been obtained in the past and where it has often been ambiguous. (P75)
SECTION 3 ANONYMOUS DATA	Should you reveal the result as high cholesterol or positive for sexually transmitted disease to the individual? This way they can change their lives accordingly. (P74) This is fine but what happens if the sexually transmitted disease is rife in a certain area. Would it not be wise for the area to know what is going on and stop the spread of the disease? (P152)
SECTION 4 CONFIDENTIALITY	This is fine in theory it is the non-compliance and failure to hold information securely that is the problem (P76) Would it not be better to only withhold records if people do not give consent (assume consent is given unless told otherwise)? (P180) Are you saying that in most cases the doctor would ask permission for disclosure? This seems to contradict earlier statements. (P120)
SECTION 5 CONSENT	You could get patients to sign a consent form to use their data, say on every 1st visit of a year...or something like that. Or postal. Without a true centralised data base system, it would be a nightmare. (P78) Given that most people (myself included) distrust scientists and their motives for research, I can only say that I'm pleased there are such stringent restrictions on gaining consent. (P139) From what I've read it sounds like the consent process is convoluted and very restricted. It's a wonder any research based on personal data can be conducted! (P94)
SECTION 6 HOW DO PEOPLE FEEL ABOUT MEDICAL RESEARCH USING PERSONAL DATA	I realise that medical research HAS to be done but my argument is that I would like to be able to give my permission if I agree with what they are researching (P148) Personal details which can breach confidentiality should not be allowed!! (P154)
SECTION 7 PROBLEMS WITH PERSONAL DATA	In the case of Alder Hay etc, I think many people would have given their consent if the procedures and possible benefits had been explained to them in the correct way. People need to be treated properly and as individuals by medical staff, scientists and bureaucrats and not patronised (P75) And so there should be - this is a real problem that all the accountable parties, including the Medical field, should address - security and privacy of data and individuals is a sensitive subject (P94)
SECTION 8 THE NHS CARE RECORDS SERVICE	In the past relatives of mine have worked for the computer records department and they have had access to records. There is no way of stopping this happening. (P88) Confidentiality is the biggest concern, and recent errors of lost data show that no matter how high levels of security are, there is still space for human error. Patients would need to be able to trust the system. (P112)
SECTION 9 USING DATA IN NHS RECORDS FOR RESEARCH	How can we be assured that the information will only be used for the purposes of the research or for patient/doctor purposes? (P128) We are not interested in political debate but responsibility and accountability. A single person should be accountable and heads should roll if something goes wrong. (P122)

Table 12 Argumentation in relation to each text section

SECTION	REFLECTION ON PREFERENCES
SECTION 1 PERSONAL DATA IN MEDICAL RESEARCH – AN INTRODUCTION	I am happy to volunteer information or to give permission for data which is already available to be used in medical research (P113); I am concerned it would get into the wrong hands. I think I should be the only one apart from my doctor who has access to my records. (148) I believe this is an excellent way to use information and should be encouraged. (P166)
SECTION 2 RESEARCH ETHICS COMMITTEES	Why withhold data if the research is important for the future of mankind? (P152)
SECTION 3 ANONYMOUS DATA	So long as i myself get the help. I don't care if the researcher knows who I am. (P108) I would be happy for my information to be use anonymously but not so happy about pseudo-anonymised information. (P139); Personally i do not feel anonymity is of great importance in this matter (P166)
SECTION 4 CONFIDENTIALITY	I would not have an objection to personal data being disclosed by my GP as long as I knew what it was being used for and had agreed to the disclosure. (P139) I agree with this entirely - if the doctor wishes to pass on my details then he should ask me first. (P148)
SECTION 5 CONSENT	I have no problem with this provided the information is held securely. (P76) I do NOT agree with the government giving permission about my details being used (P178)
SECTION 6 HOW DO PEOPLE FEEL ABOUT MEDICAL RESEARCH USING PERSONAL DATA	I agree with the last sentence. I am sceptical. People are careless. (P76); I also think that if your details were somehow used for illegal purposes that the research responsible would take full liability, if i had these reassurances I would be happy for mine and my family's details to be used. (P81)
SECTION 7 PROBLEMS WITH PERSONAL DATA	This is correct, and I am personally very anxious [not] to lose any of my data. It is shameful that these things are happening thro' negligence. (P122) I think most people accept that there can never be a guaranteed 100% confidentiality. However, loss of sensitive personal data is something that would deter me from sharing information about myself. (P139)
SECTION 8 THE NHS CARE RECORDS SERVICE	This worries me if personal medical data is lost. (P134) When i delivered my second baby my hospital was closed, I had to travel 30 miles to the next hospital, my records were not transferred in time and I had problems with the birth. If my records were computerised I would not have had the problems i did, so i think national records are a good idea from personal experience. (P81) Even though the NHS has these guarantees, I am still sceptical about the security of the data. (P85)
SECTION 9 USING DATA IN NHS RECORDS FOR RESEARCH	no problem about using the data for research or any other use in the health service again provided the data is available on a need to know basis and excludes government and law enforcement agencies (P169) No thank you to the 2nd paragraph. I suspect there is already more than enough data about me on record and that is far more than is truly needed. (P85) Colin Blakemore is brilliant - to think every person in England would have a better chance of surviving an illness. (P108)

Table 13 Reflection on preferences in relation to each text section

5 Discussion and implications

A small pilot study was conducted to develop a web based tool suitable for encouraging citizen deliberation around science and technology. The way in which people engaged with it was monitored. Uniquely, behavioural and attitudinal measures were captured alongside extensive qualitative material. Analyses have been conducted that consider some of the relationships between these data. This chapter contains some final reflections on the results of the pilot study and some suggestions as to possible ways of taking this work forward.

5.1 Measuring behaviour

One of the unique facets of this online tool is its capacity to objectively capture the ways in which people are behaving as they move through the exercise. These behavioural measures related to one another in coherent and consistent ways. Perhaps the most passive measure of engagement (reading time) was positively related to the more active time based measures (glossary reading time and time spent producing questions and comments) as well as to our key indicator of active engagement: the number of questions and comments produced. The behavioural measures also related to self-reported claims about behaviour. For example, those who said they found asking questions difficult were those that asked fewer questions.

5.2 Asking questions

We are aware of little research in relation to public engagement that has systematically considered the questions that people ask⁹. This is perhaps surprising given the focus on mechanisms of engagement that purport to facilitate the expression of participant perspectives. Arguably those few methods of engagement that do focus on eliciting participants' questions tend to be small in scale (e.g. citizens' juries) and certainly have limited capacity to monitor and record the questions that people ask or to relate them to the way in which people have actually engaged with the deliberative process¹⁰.

One aim in developing and piloting this tool was to facilitate participants' question asking and, in response to this, to provide material that was tailored to the particular issues that they had raised. The results of this pilot study certainly suggest that this focus on stimulating participant questions may be a particularly productive one. There were a range of important differences between those who chose to ask questions and make comments and those who did not. For example the data suggested that it was those who considered that they knew less that showed a greater propensity to engage in this way. This runs counter to our experience of face-to-face

⁹ See the following for an exception: Falchetti E, Caravita S & Sperduit A (2007) What do laypersons want to know from scientists? An analysis of a dialogue between scientists and laypersons on the web site Scienzaonline *Public Understanding of Science* 2007; 16; 489

¹⁰ Of course this is usually outside the interest, as well as the capacity, of those convening such engagement exercises

methods of engagement where greater knowledge of the topic area tends to be associated with greater confidence to participate. In such fora, those who purport to know less, initially at least tend to sit back (and of course the differential ease of recruiting such participants tends not to be monitored or reported). Thus, those who are less confident that they have any knowledge of the subject matter may, in engagement terms, constitute the 'hard to reach'¹¹. The greater willingness of these participants to engage by asking questions and making comments suggests that the 'privacy' afforded by our online tool and the 'faceless' nature of the researcher with whom they are interacting makes engaging and seeking out information easier for them. The further development of this tool would thus seem to hold particular promise for use by those who are less confident in what they know.

5.3 Changing assessments of knowledge

The results of this pilot study also showed both in relation to the more passive reading time measure of engagement *and* the more active 'producing comments and questions' measure that more active engagement with the tool led to a significantly greater increase in self-assessed knowledge. This is important as the change in this measure seems to represent peoples' genuine sense that they have learned something.

5.4 Engagement and feeling comfortable

It is also worth highlighting the lack of impact of engagement (either in terms of time or in relation to questions and comments produced) on the single attitudinal measure that was included at pre-text, post-text and post-task: assessments of how comfortable people were with the idea of their personal data being used for medical research. We acknowledge that any larger study would have to explore this more thoroughly, for example by including more attitudinal measures or perhaps by considering text material that was specifically designed to address comfort issues. At this stage though what we have found does link with the literature in this area and serves to highlight the importance of not considering increasing knowledge instrumentally as a simple route to acceptance.

5.5 Asking questions as deliberation

Examination of the nature of the questions and comments produced by participants suggested that they embodied some of the key indicators of deliberation. The information that was provided about personal data stimulated a search for new information; provided a context for the deployment of argumentation and allowed people to reflect on, and state, their preferences. Further developments of the tool could certainly consider ways in which the social context could

¹¹ In line with this, rather than the 'hard to reach' we prefer the expression, the 'hardly reached', first coined, we believe by London Sustainability Exchange. It reflects the assessment that effective ways of engaging with these 'groups' have not been developed rather than inferring that the cause of being hard to reach is located in the characteristics of the group itself.

be further enhanced although this may result in some of the possible benefits from a 'thin' social context being lost (e.g. willingness to admit lack of knowledge and to actively engage even when this is the case).

5.6 The topic

Where the tool is used in the future it is likely that those using it will value the insights that the questions and comments bring about the topic itself. The insights provided about participant understanding of using personal data in biomedical research have not been covered in this report as, in line with the aims of the study, our focus has rather been upon assessing the feasibility of the tool. However, to give an example, one of the key questions that a number of participants asked concerned the issue of why it might be important for researchers to use data that were identifiable rather than anonymous. This pattern of questioning would be the sort of insight that could easily feed into information campaigns or communication strategies in this area.

It is also quite possible that patterns of engagement may vary across topic areas. This too is something that could be explored in future use of the tool.

5.7 Incentives

Although this was only a small study we took the opportunity to vary incentive rates across waves in order to make some assessment of the level of incentive that would be required to encourage good levels of participation. As it turned out, the higher rate was a flat rate and the lower rate was incentivised in relation to time spent doing the exercise. It seemed that the flat rate of £25 was linked to slightly higher levels of participation in the exercise itself. Additionally, those participants receiving the lower rate, whilst not differing significantly on the behavioural measures of time, did produce fewer questions and comments.

In many ways this raises more questions than it answers. One might wonder, for example, about the extent to which this difference was a product of the rate being higher than existing expectations of those taking part in the ICM panel – there is no guarantee that non-ICM respondents would register the same level of engagement for the same incentive. Indeed, those less used to taking part in online surveys may engage *more*. It may also be that for many purposes a lower level of incentive would be 'good enough' – it certainly was not the case that participants on the lower rate did not engage at all. Given that levels of satisfaction were high it may also have been that participants would happily have engaged in another exchange with the research team for the same money, for example, giving fuller reflections on what they thought of the feedback material and what other questions this raised for them.

5.8 Further development of the online deliberation tool

There are a wide range of possible uses for the tool both in terms of topic areas that could be explored and in relation to the sorts of media that could be incorporated within it. For example, with further development work, in addition to text, participants could view images, listen to audio material or watch video footage. Indeed part of the exercise could involve them in selecting the media that they prefer to use (and from a research perspective one could assess whether and how the media selected differently affected behavioural measures of engagement or self report measures).

It would also be interesting to consider how best to extend the qualitative analysis that has been done. For example it would be possible to consider whether particular types of deliberation (e.g. information seeking, argumentation, reflection) are linked to different sets of beliefs about the use of personal data or to different profiles of attitude change (for example use of argumentation may be linked to more lasting attitude change). It may also be that different 'deliberation profiles' (e.g. exclusive use of information seeking; no evidence of argumentation) may be linked to the time people spent reading the text, looking at glossary terms etc.

The tool could be used to provide a snapshot of the way in which people's questions about a topic do or don't change over time, to see what questions media coverage of a particular topic stimulates or for assessing the effects of subtle changes in the language used in the text. Equally, it could be developed as a platform for a 'conversation' with the same participants over a longer period of time. It could provide detailed social intelligence to inform communications or, more substantively, to elicit the insights *of* people rather than simply providing insights *about* people. Different levels of incentives are likely to be needed for different purposes and will also of course reflect the value the tool user places upon the information they will gain.

The Trust may also wish to consider the possibility of returning to the participants of the current study. For example it would be possible to see whether the differences in self assessed knowledge of those who engaged more or less actively are maintained over a longer period of time.

5.9 Problems

It is worth re-iterating that if the tool is to be used as a way of engaging large numbers of people, considerable demands are made upon researcher time when providing information in response to the questions participant questions ask. Of course it would be possible, though to our mind, not ethical, to do this in a more cursory manner than was the case in this pilot study. The feedback participants got was clearly appreciated.

We recognise that all the participants in this study had at least some experience of participating in online surveys. Clearly, by definition, the use of this tool is geared to people that have access to a PC in the home environment and thus it is likely to be systematically excluding those of lower SES. In future studies it may be worth using recruitment routes that capture people that, though computer users, are not part of an online panel. This of course does have cost implications.

In conclusion it would seem that this pilot study has tested a tool that has considerable potential for future use both for online deliberation and learning and simultaneously to allow us to address research questions about the relationships between behavioural measures of engagement, self reported knowledge and attitudes, and the questions that people ask.

Appendix A - The text

NOTE: Items shown in red were clickable glossary terms. The information provided when these were clicked upon is listed under the main text in each section.

Section 1

Personal data in medical research: an introduction

One of the ways in which progress is made in medicine is to do studies that use people's personal data. In this context 'personal data' refers to information about individuals that may be used in medical research. This includes both health data (for example cholesterol levels or the dates of hospital visits) and non-health data (for example postcode or occupation). Personal data in your medical records may relate to lifestyle factors - for example around smoking - or whether there is a family history of particular diseases. Medical researchers have made important advances in their understanding of particular diseases by using this sort of personal data. For example, studies using personal data revealed that reducing cholesterol by taking statins reduces the risk of heart disease. It is also these kinds of data that are being used to investigate the possible causal link between living near high voltage power lines and childhood leukaemia.

Personal data: The Data Protection Act 1998 defines personal data as: 'Data which relate to a living individual who can be identified from those data, or from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller.'

Statins: Statins are a class of drugs that are used to lower cholesterol levels in people with, or at risk of, cardiovascular disease.

Heart disease: The Heart Protection study identified 130,000 people through hospital records without their prior consent. After informed written consent was given over 20,000 took part in a study that clearly showed the importance of statins in reducing the risk of heart attacks and strokes in high risk people.

Leukaemia: One recent study found there was relatively greater risk of leukaemia for children living within 200 metres of a power line compared to children matched for sex and age who lived more than 600 metres away. It is not clear whether this result is due to chance or some as yet unknown biological mechanism. Further research is needed.

Section 2

Research Ethics Committees

Groups known as research ethics committees keep an eye on what it is that medical researchers want to do with personal data and can grant - or withhold - permission for the research. There are several important concepts that ethics committees use in deciding whether research can go ahead. Three of the key ones are anonymity, confidentiality and consent.

Research ethics committees: NHS Research Ethics Committees (RECs) have been established throughout the UK for many years with the purpose of safeguarding the rights, dignity and welfare of people participating in research in the NHS. Potential research participants at NHS organisations in the UK will come under the protection of a REC. The REC is entirely independent of the researcher and the organisations funding and hosting the research.

Section 3

Anonymous Data

Some information used in research is anonymous - this means that the information does not identify the individual. The researcher only knows - for example - that an unnamed person has a high level of cholesterol in his or her blood, or has tested positive for a sexually transmitted disease. Any information that might identify the person - such as name, address and postcode - will have been removed. (Another sort of anonymity is provided by pseudonymised information).

Pseudonymised information: This is like anonymised information in that the holder of the information cannot identify an individual. However, it is different in that the original provider of the information may retain a way of identifying individuals, for example by attaching a code so that the data will only be identifiable to those who have access to the code.

Section 4

Confidentiality

Confidentiality means that a doctor or other health professional will hold secret all information relating to a patient, unless the patient gives consent permitting disclosure. Confidentiality is an important concept because personal data that has had its identifiers removed – i.e. that is anonymous – is often not very useful to researchers.

Section 5

Consent

The regulations around getting consent for medical research using personal data are complicated. Sometimes consent is needed even when people cannot be identified from their data. On the other hand the use of identifiable data may be allowed without consent being obtained but this can mean a long and complicated process of deciding whether **the laws governing medical research** allow this. The **Academy of Medical Sciences** explain that it is often not practical to seek consent, for example, where the data were collected a long time ago – it would take too long to do this and cost too much. Asking for consent may cause distress to people – for example, they may not wish to be reminded of a previous diagnosis. It may also be the case that some sorts of people say yes and other sorts say no and that this can lead to research being biased. Organisations like the Academy of Medical Sciences have argued that the rules around using personal data in medical research are so strict that it is often too difficult to get permission to do studies using personal data that has already been collected and that this is stopping vital medical research being done.

The laws governing medical research: The most important laws governing medical research using personal data include: Data Protection Act 1998; Common Law of confidentiality; Human Rights Act 1998 and Section 60 of the Health and Social Care Act 2001.

Academy of Medical Sciences: One of the objectives of the Academy of Medical Sciences is to encourage the pursuit of internationally competitive medical science and the translation of that knowledge, and its associated technologies, from the laboratory bench to the delivery of healthcare. Another is to engage with the public to build confidence in the practice of medical research and to address public concerns.

Section 6

How do people feel about medical research using personal data?

Even though there is generally no direct medical benefit to the individuals taking part in such studies people are well aware of the benefits of their personal data being used for medical research. They know that they are doing something for the 'greater good', helping other people and even future generations of their own family. Some recent research for the Medical Research Council concluded that if people are informed about what it means for their data to be used in medical research they are generally positive about this. Research for the Wellcome Trust found that, although there was strong support for the value of medical research that used personal data, people were sometimes sceptical about promises of confidentiality because of experiences in their everyday lives where such promises were broken through error, deliberate intrusion or lack of care.

Section 7

Problems with personal data

In the past there has been a great deal of media attention on failures in medical practices in this area, for example when organ samples were saved and stored without consent at Alder Hey hospital. Recently there has been a lot of publicity about personal data more generally. First the Government admitted that a copy of HM Revenue and Customs child benefit data – including millions of bank details – had been lost. Later they announced that a copy of records from over three million candidates for the driving theory test had also been lost.

Section 8

The National Health Service Care Records Service

The debate about the use of personal data in medical research has recently focused on a new system currently being developed within the NHS – a data base called the **NHS Care Records Service** that holds all our health records electronically. One part of this system, the Summary Care Record, aims to ensure that information about each individual is available wherever it is needed – at a GP's surgery, a hospital appointment or in Accident and Emergency for example.

Already some people have expressed concerns about various aspects of this system. One concern is about whether or not medical records will remain confidential. In response to this those developing the system say that the NHS Care Records Service will use the strongest national and international security measures for handling information. These will ensure that information is stored safely, stays private and is accessed appropriately. Details of the NHS commitment to protect our privacy are outlined in the [Care Record Guarantee](#).

NHS Car Records Service: The aim of this is to "improve the safety and quality of your care. Over time, the NHS Care Records Service will begin to provide healthcare staff with quicker access to reliable information about you to help with your treatment, including in an emergency"

Care Record Guarantee: This gives various guarantees, for example, you can choose not to have information in your electronic care records shared and a record will be kept of everyone who looks at the information the NHS Care Records Service holds about you.

Section 9

Using data in NHS records for research

A potentially important secondary use of the data in the NHS Care Records Service is for research. The government and the Academy of Medical Sciences believe that national electronic care records give the UK the potential to become a world leader in medical research that uses health information. Colin Blakemore, former chief executive of the Medical Research Council, stresses that medical research using the health records of large numbers of people can provide important information that would be difficult or impossible to gain in any other way.

In 2007 the Summary Care Records service was rolled out in six Primary Care Trusts. An evaluation of how well this has worked will inform how quickly this system will be implemented across the rest of England. The target is for an electronic record to be created for every person in England by 2010. The debate around the uses of identifiable personal data for purposes other than direct patient care is still going on.

Appendix B – The questionnaire

TIME ONE – Pre-text

Q1 How confident are you that you understand the phrase "personal data"?	
Response	Routing
Extremely confident	Q2
Very confident	Q2
Quite confident	Q2
Not very confident	Q2
I do not understand the phrase "personal data" at all	Begin exercise

Q2 On a scale of one to ten, how much would you say you know about who currently holds data about you? (where one equals nothing and ten equals everything you feel you need to know)	
Response	Routing
1 - nothing	Q3
2	
3	
4	
5	
6	
7	
8	
9	
10 – everything I need to know	

Q3 How comfortable are you with the idea of some organisations holding your personal data?	
Response	Routing
Extremely comfortable	Q4
Very comfortable	
Quite comfortable	
Not very comfortable	
Not at all comfortable	

Q4 Sometimes, personal data can help scientists in carrying out medical research. On a scale of one to ten, how much would you say you know about the ways in which personal data can be used in medical research?
(where one equals nothing and ten equals everything you feel you need to know)

Response	Routing
1 - nothing	Q6
2	
3	
4	
5	
6	
7	
8	
9	
10 - everything I need to know	

Q5 How comfortable are you with the idea of your personal data being used for medical research?

Response	Routing
Extremely comfortable	Q6
Very comfortable	
Quite comfortable	
Not very comfortable	
Not at all comfortable	

Q6 On a scale of one to ten, how much would you say you know about the current limits on use of your personal data for medical research?
(where one equals nothing and ten equals everything you feel you need to know)

Response	Routing
1 - nothing	Begin exercise
2	
3	
4	
5	
6	
7	
8	
9	
10 - everything I need to know	

TIME ONE – Post-text

Repeat of relevant baseline questions

Q7 On a scale of one to ten, how much would you say you know about the ways in which personal data can be used in medical research?
(where one equals 'nothing' and ten equals 'a lot')

Response	Routing
1 - nothing	Q8
2	
3	
4	
5	
6	
7	
8	
9	
10 – a lot	

Q8 How comfortable are you with the idea of your personal data being used for medical research?

Response	Routing
Extremely comfortable	Q9
Very comfortable	
Quite comfortable	
Not very comfortable	
Not at all comfortable	

Q9 On a scale of one to ten, how much would you say you know about the current *limits* on use of your personal data for medical research?
(where one equals 'nothing' and ten equals 'a lot')

Response	Routing
1 – nothing	Q10
2	
3	
4	
5	
6	
7	
8	
9	
10 – a lot	

More detailed questions on the topic

Q10 More specifically, how comfortable are you with the idea of the following types of personal data being used in medical research?

Response	Extremely comfortable	Very comfortable	Quite comfortable	Not very comfortable	Not at all comfortable
Your postcode (e.g. to track which diseases occur most frequently in which areas)	Q12	Q12	Q12	Q12	Q11
How often you visit your GP (e.g. to monitor demand for health services)	Q12	Q12	Q12	Q12	Q11
Your cholesterol levels (e.g. to enable researchers to better understand the link between cholesterol and heart disease)	Q12	Q12	Q12	Q12	Q11
Your occupation (e.g. to identify whether or not some occupations are more likely to lead to ill health than others)	Q12	Q12	Q12	Q12	Q11
Your previous medical history (e.g. to help establish whether particular diseases are linked)	Q12	Q12	Q12	Q12	Q11

Q11 In the previous question, you said that you would not be comfortable with your XXXXXX being used in medical research. Please say why.

Q12

Q12 Which of the following kinds of organisation, if any, regulate the use of personal data for medical research? (select any that apply)

Response	Routing
General Medical Council	Q13
Universities UK	
Research Ethics Committees	
House of Commons Committee on Standards & Privileges	
County Councils	
The Academy of Medical Science	

Q13 How important is it that personal data used in medical research remain 'anonymous'? (i.e. that people cannot be identified through their data)

Response	Routing
Extremely important	Q14
Very important	
Quite important	

Not very important	
Not at all important	
Don't know	

Q14 Should researchers ever be able to use medical data that identify the people they relate to without obtaining consent?
(please explain your answer)

Response	Routing
Yes, always	Q15
Yes, under certain circumstances	
No, never	
Don't know	

Q15 To what extent would you trust the following people or organisations to store your personal data?

Response	Completely	A great deal	Quite a lot	Not a lot	Not at all
Your local hospital					
Your local GP (doctor) surgery					
The Department of Health					
Any other government department					
A university					
A pharmaceutical company					
Another private company					
The National Health Service (NHS)					

All responses route to Q16

Q16 To what extent would you trust the following people or organisations to use your personal data responsibly for medical research?

Response	Completely	A great deal	Quite a lot	Not a lot	Not at all
Your local hospital					
Your local GP (doctor) surgery					
The Department of Health					
Any other government department					
A university					
A pharmaceutical company					
Another private company					
The National Health Service (NHS)					
All responses route to Q17					

Q17 Do you think a system – such as the one currently being developed by the National Health Service, which allows medical staff to access your medical records from anywhere in the country - is a good idea?

Response	Routing
Yes, it's essential (please explain why)	Q18
I think the advantages outweigh the disadvantages (please explain why)	
I have no feelings either way (please explain why)	
I think the disadvantages outweigh the advantages (please explain why)	
I think it's a terrible idea (please explain why)	
It depends (please explain why)	
Don't know (please explain why)	

Q18 How comfortable are you with the idea of personal data in the NHS system being used for medical research?

Response	Routing
Extremely comfortable (please explain why)	Q19
Very comfortable (please explain why)	
Quite comfortable (please explain why)	
Not very comfortable (please explain why)	
Not at all comfortable (please explain why)	
It depends (please explain why)	

The task itself

Q19 Thinking about the exercise you have just completed, how easy was it to...

Response	Extremely easy	Very easy	Quite easy	Quite difficult	Very difficult	Extremely difficult
Understand the initial instructions	Q21	Q21	Q21	Q21	Q20	Q20
Understand what the research is for	Q21	Q21	Q21	Q21	Q20	Q20
Understand what you were expected to do at each point	Q21	Q21	Q21	Q21	Q20	Q20
Read the text	Q21	Q21	Q21	Q21	Q20	Q20
Understand the text	Q21	Q21	Q21	Q21	Q20	Q20
Make comments and ask questions	Q21	Q21	Q21	Q21	Q20	Q20

Q20 In the previous question, you said that it was difficult to XXXXXX. Please tell us a bit more about why this was the case.

Q21

Q21 Do you think the text was...

Response	Routing
Too long?	Q22
Too short?	
Just the right length?	

Q22 Do you think the text was...

Response	Routing
Too detailed?	Q23
Not detailed enough?	
Just right?	

Q23 On a scale of one to ten, overall how easy did it feel to think of questions about the personal data information?
(where one means that questions easily came to mind and where ten means that it was very difficult to think of anything to ask)

Response	Routing
1 – questions came easily to mind	Q24
2	
3	
4	
5	
6	
7	
8	
9	
10 – it was very difficult to think of anything to ask	

Q24 As a way of informing the general public about a complicated issue and finding out about their questions and concerns, are there any ways you can think of in which this exercise could be improved?

Q25 It may be that answering our questions has prompted some other questions or comments you have about personal data. If so, please note these in the space below.

Thank and close

TIME THREE

Repeat of relevant baseline questions

Q26 On a scale of one to ten, how much would you say you know about the ways in which personal data can be used in medical research?
(where one equals nothing and ten equals everything you feel you need to know)

Response	Routing
1 - nothing	Q27
2	
3	
4	
5	
6	
7	
8	
9	
10 – everything I need to know	

Q27 How comfortable are you with the idea of your personal data being used for medical research?

Response	Routing
Extremely comfortable	Q28
Very comfortable	
Quite comfortable	
Not very comfortable	
Not at all comfortable	

Q28 On a scale of one to ten, how much would you say you know about the current limits on use of your personal data for medical research?
(where one equals nothing and ten equals everything you feel you need to know)

Response	Routing
1 - nothing	Q29
2	
3	
4	
5	
6	
7	
8	
9	
10 – everything I need to know	

More detailed questions on the topic

Q29 More specifically, how comfortable are you with the idea of the following types of personal data being used in medical research?

Response	Extremely comfortable	Very comfortable	Quite comfortable	Not very comfortable	Not at all comfortable
Your postcode (e.g. to track which diseases occur most frequently in which areas)	Q31	Q31	Q31	Q31	Q30
How often you visit your GP (e.g. to monitor demand for health services)	Q31	Q31	Q31	Q31	Q30
Your cholesterol levels (e.g. to enable researchers to better understand the link between cholesterol and heart disease)	Q31	Q31	Q31	Q31	Q30
Your occupation (e.g. to identify whether or not some occupations are more likely to lead to ill health than others)	Q31	Q31	Q31	Q31	Q30
Your previous medical history (e.g. to help establish whether particular diseases are linked)	Q31	Q31	Q31	Q31	Q30

Q30 In the previous question, you said that you would not be comfortable with your XXXXXX being used in medical research. Please say why.

Q31

Q31 Which of the following kinds of organisation, if any, regulate the use of personal data for medical research?

(select any that apply)

Response	Routing
General Medical Council	Q32
Universities UK	
Research Ethics Committees	
House of Commons Committee on Standards & Privileges	
County Councils	
The Academy of Medical Science	

Q32 How important is it that personal data used in medical research remains 'anonymous'?
(i.e. that people cannot be identified through their data)

Response	Routing
Extremely important	Q33
Very important	
Quite important	
Not very important	
Not at all important	
Don't know	

Q33 Should researchers ever be able to use medical data that identify the people they relate to without obtaining consent?
(please explain your answer in each case)

Response	Routing
Yes, always	Q33
Yes, under certain circumstances	
No, never	
Don't know	

Q34 To what extent would you trust the following people or organisations to store your personal data?

Response	Completely	A great deal	Quite a lot	Not a lot	Not at all
Your local hospital					
Your local GP (doctor) surgery					
The Department of Health					
Any other government department					
A university					
A pharmaceutical company					
Another private company					
The National Health Service (NHS)					

All responses route to Q35

Q35 To what extent would you trust the following people or organisations to use your personal data responsibly for medical research?

Response	Completely	A great deal	Quite a lot	Not a lot	Not at all
Your local hospital					
Your local GP (doctor) surgery					
The Department of Health					
Any other government department					
A university					
A pharmaceutical company					
Another private company					
The National Health Service (NHS)					

All responses route to Q36

Q36 Do you think a system – such as the one currently being developed by the National Health Service - that allows medical staff to access your medical records from anywhere in the country is a good idea?

Response	Routing
Yes, it's essential (please explain why)	Q37
I think the advantages outweigh the disadvantages (please explain why)	
I have no feelings either way (please explain why)	
I think the disadvantages outweigh the advantages (please explain why)	
I think it's a terrible idea (please explain why)	
It depends (please explain why)	
Don't know (please explain why)	

Q37 How comfortable are you with the idea of personal data in the NHS system being used for medical research?

Response	Routing
Extremely comfortable (please explain why)	Q38
Very comfortable (please explain why)	

Quite comfortable (please explain why)	
Not very comfortable (please explain why)	
Not at all comfortable (please explain why)	
It depends (please explain why)	

Questions on the exercise as a whole

Q38 This study has provided you with the opportunity to ask questions that we have then tried to answer. How useful was the feedback you received?

Response	Routing
Extremely useful	Q39 (unless asked no questions, in which case, Q41)
Very useful	
Quite useful	
Not very useful	
Not at all useful	

Q39 To what extent were you satisfied with the feedback we provided?

Response	Routing
Extremely satisfied	Q41
Very satisfied	Q41
Quite satisfied	Q41
Not very satisfied	Q40
Not at all satisfied	Q40

Q40 Why were you unsatisfied with the feedback we provided? Is there any way we could improve it?

	Q41Q31

Q41 Overall, as a way of engaging people in a short discussion about the use of personal data in medical research, do you think this exercise has been...

Response	Routing
Extremely effective?	Q42
Very effective?	
Quite effective?	
Not very effective?	
Not at all effective?	

Q42 Please rate the study between one and ten according to the following criteria, where one is low/poor and 10 is high/good.

Criteria

How enjoyable the exercise was

How well the exercise kept your interest

The objectivity of the information provided

How useful the feedback was

Q43

Q43 Since you completed the first part of this exercise, you may have discussed the project with friends or family, or looked for additional information on personal data or the project itself. We're interested to know how often this has happened, so if you did follow up the exercise in any way, please give details.

All responses route to Q44

Q44 What, if anything, do you think is the most important thing you have learned through taking part in this exercise?

All responses route to Q45

Q45 As a way of informing the general public about a complicated issue and finding out about their questions and concerns, are there any ways you can think of in which this exercise could be improved?

Thank and close

Appendix C - Example replies to participants

Example 1

Dear xxxx

Our team at the University of Surrey have been looking at the questions that you made in the first part of the research and have tried to bring together some up to date material from official sources that are relevant in addressing the points you made. We may not have addressed everything you said - mostly this is the case where your point was more of a comment than a question.

There is quite a bit of information out on the web about the personal data and its use in medical research and where we can we have pointed you in the direction of useful web pages in case you would like to follow any of this up at your leisure.

how can individuals be confident that personal information is secure and will not leak to other parties

The interesting question you raise here is a concern of many people. The NHS recognises the need to provide strong safeguards of personalised medical information and communicates the approach it takes to this in the 'Care Record Guarantee'. The answer we give to your second question below is also relevant to this question of how individuals can be confident that their personal information is secure.

How can the information be kept current with as little human handling as possible.

We think that your question here picks up on one of the key challenges in this area. To take the new NHS Care Records Service as the example, the NHS promises to use the strongest national and international security measures for handling your information - more secure even than internet banking. <http://www.nhscarerecords.nhs.uk/your-confidentiality> . NHS Staff will have to pass three tests to access your records:

- They will have to be registered as NHS staff and be issued with a Smartcard and pass code which works like a chip and PIN bankcard
- They have to be recognised by the system as providing care or treatment to you
- They will only be able to see the sorts of information they need to give you that care or treatment

Anyone accessing NHS records will automatically have their details recorded – who they are and what they did as part of an electronic record audit trail. Patients will be entitled to see this information.

<http://www.nhscarerecords.nhs.uk/frequently-asked-questions/confidentiality-security-and-access/confidentiality-security-and-access#11>

Because of the potential value of the NHS Care Records Service for 'secondary' purposes (such as medical research) thought is being given to the best way of managing this. One suggestion

(made by the Care Record Development Board see <http://www.connectingforhealth.nhs.uk/crdb/workstreams/secusesreport.pdf> if you are interested to read more) is that an "honest broker" takes on responsibility for ensuring the data is appropriately managed and made available to researchers in the right way. Others have pointed out that this idea is actually more problematic than it sounds – for example one report questioned exactly what the legal basis is for transferring patient identifiable information to 'honest brokers'.

I do subscribe to the belief that personal information may be useful to create potent strategies for health and well being to monitor patient care and early detection.

This is certainly the view of many leaders in the NHS . The CRDB report (mentioned above) outlines some of the ways in which using people's data for medical research has led to huge medical advances – they particularly mention a greater understanding of HIV prevention, the relationship between smoking and lung cancer and the ill effects of the use of aspirin for children.

Please do give these points some thought and when you have done this please log back in to complete the final part of the study. You need to double click the following link and then log in with your email address:

<http://wbe.whiteoctober.co.uk/time3/73/265981e7185e298c1d9c35e5b559dd0a/>

You will need to do this by Monday 21st April at the latest. The final part of the study will take between 5 and 10 minutes to complete.

Once again – thank you very much for taking part in this research.

With our best wishes

Example 2

Dear xxxx

Our team at the University of Surrey have been looking at the questions that you made in the first part of the research and have tried to bring together some up to date material from official sources that are relevant in addressing the points you made. We may not have addressed everything you said - mostly this is the case where your point was more of a comment than a question.

There is quite a bit of information out on the web about the personal data and its use in medical research and where we can we have pointed you in the direction of useful web pages in case you would like to follow any of this up at your leisure.

Is using personal data on paper really a valid way of making causal links? From experience i know that many patient may be on prescription drugs, but don't actually take them.

This is a very valid point. There is quite a bit of research in relation to 'adherence' and researchers using personal data for research will be very aware that having a particular sort of prescription for taking drug X is not the same as the person *taking* drug X. Of course there are many other sorts of research using personal information that would be less open to this sort of criticism.

Why is anonymous data often not very useful

A report by the Academy of Medical Sciences gives 4 key reasons why anonymised data may not be very useful and why access may be required to identifiable information.

- ***To assess/avoid double counting.*** For example, congenital anomaly registers were set up in response to the thalidomide tragedy. Anomalies can be reported by more than one source (e.g. paediatricians, midwives, genetic counselling services etc) and matching reliable personal information is the only way to identify duplicates and avoid double counting.
- ***For longitudinal research.*** Understanding how exposure to a risk factor influences later health requires that information on an individual be updated over time. This is not possible if data are irreversibly anonymised.
- ***For validation.*** It is important to ensure that that data subject x really is data subject x. This can only be achieved using identifiers to match the records.
- ***Identifiers contain useful information.*** Many of the identifiers that might be stripped from data during anonymisation are useful to research. For instance, where you live (postcode), how old you are (date of birth), and what your job is (occupation) are all routinely used as important factors in analysing data.

You could get patients to sign a consent form to use their data, say on every 1st visit of a year...or something like that. Or postal. Without a true centralised data base system, it would be a nightmare.....and no....the system one system is not working!

This is a very interesting point and finding a really workable system of providing consent for (e.g.) using personal data for research is a real challenge. One of the key documents about the using personal data in medical research suggested that as the NHS Care Records System becomes more established, that this could be used to record people's preferences for participating in research. Then the consent choices of an individual could be automatically adhered to when providing data from their record. (If you are interested you can see some details about this in a short publication in 2007 by the Care Record Development Board. You can find this at <http://www.connectingforhealth.nhs.uk/crdb/workstreams/secusesreport.pdf> - the section on consent is page 15) Potentially one problem with providing consent ahead of time as it were is that people are most willing to consent if they have a clear indication of the type of research that their data will be used for. It is however often difficult to anticipate in advance what these detailed uses might be.

I can understand that the NHS needs to update its systems. Unfortunately a national system is a long time away, as so many GP practices run on a different system. Also there is an increase in patients having private insurance, private clinics are excluded from national systems.

Work on the NHS Care Records system is well under way. Six 'Early Adopter' Primary Care Trusts started work to create Summary Care Records in Spring 2007 and the roll out will continue in 2008. The target is for an electronic record to be created for every person in England by 2010. Within private health care a record will be sent to the GP about (for example) what operation has been conducted.

Please do give these issues some thought and when you have done this please log back in to complete the final part of the study. You need to double click the following link and then log in with your email address

<http://wbe.whiteoctober.co.uk/time3/78/2c7947f49631d7244d4508e15ad61343/>

You will need to do this by Monday 21st April at the latest. The final part of the study will take between 5 and 10 minutes to complete.

Once again – thank you very much for taking part in this research.

With our best wishes

