Use of Free-text Health Data

A report of a citizens' jury designed to explore when and how free-text data in patient records should be used

Commissioned by

and

Developed by

and

June 2018

Table of Contents

Contents

Introduction

On 6 June2018, 18 people gathered at Jury's Inn Waterside Hotel in Brighton and began a three-day "citizens' jury". The task for these citizens was to tackle a set of jury questions. The questions were designed to test a number of scenarios where patient data, both coded and free-text data, could be shared for research. Participants were asked to judge how data should be processed and used for research in these specific scenarios, and in general.

Over the three days, the citizens heard from, and asked questions of, expert witnesses, and worked in groups on the jury questions. They reached conclusions together, and were polled on their individual views. They identified reasons for their answers.

This report explains why the jury was carried out, how it was designed, what the jurors did, the jury results, and a discussion of the findings.

Further information about the jury can be found at: www.healtex.org/jury

Why the citizens' jury was carried out

Much is known about public attitudes to the use of patient data in general. However, to date no research has asked whether the public feel differently about use of coded (structured) and free-text (unstructured) data. Despite this, custodians of patient data routinely remove free-text data from datasets provided for research. Addressing this gap in knowledge of public opinion is a crucial step towards informing custodians' data access policies. To this end, Brighton and Sussex Medical School, sponsored by Healtex, commissioned Citizens Juries CIC to design and run a three-day citizens' jury on 6-8 June 2018 to ask a cross-section of the public what they think about de-identified medical text data being used for research.

Jury and survey design

Citizens' jury design

The citizens' jury was planned, designed and refined over a period of eight months. There are many aspects to the jury design including:

- The jury questions;
- the jury demographics and recruitment approach;
- the brief and selection of individuals to act as expert witnesses;
- the brief and selection of individuals to act as members of the oversight panel;
- the programme of activities across the three days; and
- the design of the questionnaires completed at the start and end of the jury.

The design documentation is available at: www.healtex.org/jury

Bias, both conscious and unconscious, is an important criticism of citizens' juries.[1] For example, it is very difficult to know what constitutes "impartial information" or balanced argument, and almost every design choice, even down to a bullet point on a presenter's slide, could be challenged on grounds that it might manipulate the citizens' jury towards one outcome or another.

Bias can be monitored and minimised but not eliminated. To monitor and minimise bias on this project, an oversight panel was appointed to review the jury design and materials, and report potential bias. The panel members were fully satisfied that the jury was successfully designed to minimize bias. The end-of-jury questionnaires also asked the jurors about bias. In these, 15 jurors reported thought they were given a fair balance of information, and three jurors thought that there was some bias in favour of using free-text data for research.

Other design controls used to monitor and minimise bias included:

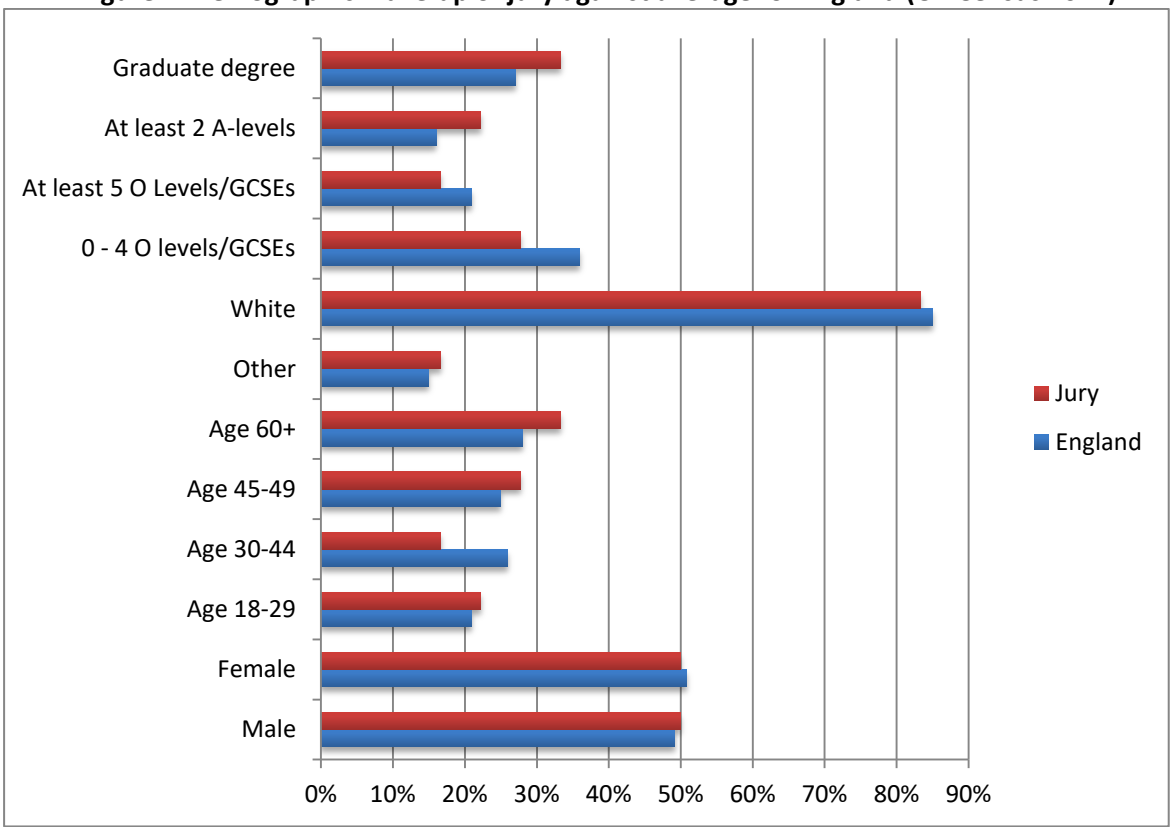
- The jury funders were involved in setting the jury questions but were independent from the jury process and outcomes;
- The jury worked with facilitators to construct their own report of their findings; and
- The detailed jury design and results documentation are being published.

Jury recruitment

Jury recruitment

In total, 227 people applied to be a juror by completing an on-line survey. Shortlisted candidates had a brief telephone interview so that any ineligible candidates (e.g. healthcare professionals) could be identified and excluded. 18 people were recruited from in and around the Brighton area. The sample chosen was controlled for gender, age range, ethnicity (in terms of white/other), and educational attainment (see chart below). The percentage mix of these control categories matched closely the demographics of people in England (as recorded in the UK Census 2011). The table below shows the demographics of the 18 people who began and completed the three-day process.

Figure 1: Demographic make-up of jury against average for England (UK Census 2011)



Applicants also answered a question taken from a national survey to test their prior views on balancing privacy with health record sharing.[1, p. 59] The range of views represented on the jury was designed to match closely those reported in the national survey. However, one participant who withdrew just before the meeting had answered the survey question as "very unwilling" to share their medical records, and he could only be replaced at that stage by a person who was "fairly willing" to share. Thus overall the mix of people on the final jury was very slightly more in favour of sharing medical records for research than the views expressed in the national survey, although it was still within the target levels set in advance for this criterion.

Of the 18 jurors, 11 people were found an advertisement placed on the Indeed jobs website, 5 through Brighton's Community Base website, and one through word of mouth. 5 people were in full or part-time employment, 6 were self-employed, 3 were unemployed, 2 were retired and 2 self-classified as having an "other" employment status. Each juror was paid £300 for three days plus a £25 expense allowance. Three "reserve jurors" were also recruited. One participated until lunch on day 1 and was paid £75. The two other reserves became full members of the jury, after one person called in sick and another left for personal reasons early on day 1.

The jury process and jury report

The 3-day jury programme:

- Was facilitated by Kyle Bozentko, Executive Director of the Jefferson Center, supported by Lamiece Hassan, a research fellow from University of Manchester;
- Included evidence presented by seven expert witnesses;
- Engaged jurors in group exercises and deliberation;
- Ended with an end-of-jury questionnaire the end of day 3.

On day three of the jury proceedings, every member of the jury voted electronically on the jury questions, and identified reasons for their answers. Kyle Bozentko, the facilitator of the jury from the Jefferson Center, then constructed the jury's report from their votes and reasons. The jurors were led page-by-page through the jury's report, which was displayed on a large projector screen, to gain the jurors' acceptance that it fairly represented their views. That report is published at: www.healtex.org/jury. A direct extract from that report – jury questions and answers – follows below. A spreadsheet containing the full set of jury questions and results may be made available on request from: Dr Elizabeth Ford at E.M.Ford@bsms.ac.uk

Jury Questions and Answers

Introduction

Please read the following scenario about Tom and his health record. You will be shown the contents of his health record as it builds up over the course of the story. Your job is to decide what data from Tom's record should be used for health research, and what Tom should be told and/or asked when data from records about Tom are used.

Tom scenario part 1: Tom registers with a GP

Tom and his family emigrated from the UK soon after Tom was born. Aged 43, Tom returns to England and settles in Anytown. Soon after arriving, Tom registers with a GP. He has no health

records from the previous countries he has lived. He registers with a large local practice, Anytown Health Centre. He provides some basic details to the GP receptionist, including his full name, date of birth, and his new address. The receptionist suggests that Tom makes an appointment to see Dr. Jones, explaining that he can provide Dr. Jones with his medical history.

Tom sees the GP

At his appointment, Tom tells Dr. Jones what he knows about his medical history. He is not aware of his parents and other close family suffering from particular illnesses, except that his mother had type 2 diabetes. Tom explains he is overweight, and that he gets thirsty, and urinates often. Dr. Jones pricks his finger, and tests Tom's blood with a special strip. Dr. Jones notes down these symptoms, the high blood glucose level result, and the suspected diagnosis of type 2 diabetes. She asks Tom to attend the next diabetes clinic at the practice, and to avoid sugary foods in the meantime.

A diabetes research team approaches the practice asking for data

Anytown Health Centre takes an active interest in health research, and has close ties with Anytown University's Department of General Practice. The university has a signed data sharing agreement with the practice that has been approved by Anytown University's research ethics committee. The agreement states how the university will protect the data properly, and states that any proposed new use of the data must be approved in advance by Anytown Health Centre and the university's research ethics committee.

Sue Stark, Anytown Health Centre's manager, receives a letter from Prof. Smith, the lead researcher for a university project approved by the research ethics committee. Prof Smith wants to use anonymised general practice data – with patient identifiers removed - to identify characteristics of patients with both suspected and confirmed type 2 diabetes. This could enable the researchers to identify early signs of the disease, helping GPs and patients to spot type 2 diabetes earlier, and reduce complications.

Sue emails Prof. Smith to say she will check this with the GPs. She asks the researcher to send a list of the items he is seeking. Prof. Smith replies saying he wants data about all patients with suspected and confirmed type 2 diabetes. Some of the data items are recorded by GPs as codes (e.g. "1234" for type 2 diabetes), and some in free text boxes (e.g. "gets very thirsty at night"). The codes for all of the patients' diagnoses are also required. The GPs review the list of data items and agree that there would only be a very small risk of the researchers discovering the identity of a patient from the coded data items.

Questions and jury answers about the Tom scenario part 1

Q1 Should Anytown Health Centre agree to release the coded data items about Tom and all the other patients in the practice with suspected or confirmed type 2 diabetes?

If you chose "other", please explain. [50 words maximum]

Of the 18 jurors who responded:

- 8 said: a) "Yes"
- 7 said: b) only if Tom and the other patients can opt out
- 2 said: c) only if Tom and the other patients can opt in
- 0 said: d) No
- 1 said: e) Other

The person who answered "e) other" said they supported option b) above but with the caveat that the opt out system be much easier and transparent.

The GPs are less sure about the free-text data because they do not know what sensitive or revealing information it might contain. Sue goes back to check this with Prof. Smith. He says that the university has software that automatically removes text that could be used to identify a person. Prof. Smith also says that the data counts as anonymised in law as only two researchers will have access to the coded and free-text data, and that the risk of identifying a patient is very small.

Q2. Should Anytown Health Centre also agree to release the free-text data items about Tom and all the other patients in the practice with suspected or confirmed type 2 diabetes?

If you chose "other", please explain. [50 words maximum]

Of the 18 jurors who responded:

- 4 said: a) "Yes"
- 10 said: b) only if Tom and the other patients can opt out
- 1 said: c) only if Tom and the other patients can opt in
- 0 said: d) No
- 3 said: e) Other

Of the three people who answered "e) other":

- One said they supported "option b) above but with the caveat that the opt out system be much easier and transparent"
- One said "Only if Tom and the other patients can opt out but they must be given full information in order to make an informed choice"
- One said they supported "option c) but no reason why those with a physical health condition might not be able to help with the research".

Q3. If your answer to Q2 is different to your answer to Q1, please give reasons for your answers. [50 words maximum]

There were four responses:

- "There may be more sensitive information including about other people in free text and some identifiers may slip through"
- "Because in the first case it's coded data but free text in the second"
- "Further information to Tom as to how data will be used i.e. more people with diabetes will gain better healthcare or a cure may be found"
- "Because free-text data is far more sensitive than coded information is and can be easily readable and leaked more easily".

Tom scenario part 2: Tom hears voices

When speaking to the nurse at the diabetes clinic, Tom mentions that he is feeling low. The nurse recommends he discusses this with his GP, and so Tom makes an appointment with Dr. Jones. At the appointment, Dr Jones hears from Tom that he has been feeling low but also that he has been hearing voices: imagining he is having conversations with people he once knew. Dr. Jones suggests that it would be useful for Tom to talk to a specialist mental health practitioner at Anytown Mental Health Trust. Dr. Jones refers Tom to the mental health trust to have an assessment and discuss what might be done to address his symptoms.

A few weeks later Tom is assessed by Ahmed Hussein, a psychiatric nurse at the Anytown Mental Health Trust. They talk about Tom hearing voices. Ahmed begins to build a picture of when this happens and how it affects Tom's life, making notes in Tom's record. Tom says the imaginary conversations are not usually distressing, and that the main problem is that they interfere with his concentration. He is particularly concerned not to lose his new job. Ahmed explains that medication and cognitive-based therapy are options that may help him control the voices. They agree that Ahmed will arrange for Tom to see a psychiatrist within the mental health trust. After Tom has left, Ahmed types up a summary of what Tom has told him, saves it within a new record for Tom within the mental health trust's patient record system. Ahmed also creates a referral letter for Tom to be seen by the psychiatrist.

Researchers request data to investigate how hallucinations affect daily life

Prof. Brown, one of Prof. Smith's colleagues at Anytown University, is doing a research study about how hearing voices and having hallucinations affects people's lives. She has been looking at messages on internet forums for people who have times when they hear voices and have hallucinations. Some people posting online say that these episodes interfere with their work, and that this is often overlooked, or not addressed within the care decisions made by healthcare professionals.

Prof Brown needs some basic data about patients who have experienced hallucinations and heard voices, like age and gender, and some coded data including all their previous medical diagnoses. She also wants the free-text notes that are recorded on the mental health trust patient record system. This is because the coded data does not have all the details needed for her research. The free-text will be analysed by computer software to strip out identifying information such as names and dates of birth. The usual protections in place at Anytown University will also apply. Only the research team will be given access to this sensitive data. Prof. Brown suggests that with all the protections in place, there is only a very small chance of identifying a patient from the data, so the data set she requires counts as anonymised in law.

All this is explained in a letter to the research department at Anytown Mental Health Trust. The research lead in the trust brings it to the next senior management meeting where the issue is discussed thoroughly.

Questions and jury answers about the Tom scenario part 2

Q4. Should Anytown Mental Health Trust agree to release the free-text data items about Tom and all the other patients in the trust who hear voices or have hallucinations?

Of the 18 jurors who responded:

- 4 said: a) "Yes"
- 10 said: b) only if Tom and the other patients can opt out
- 2 said: c) only if Tom and the other patients can opt in
- 0 said: d) No
- 2 said: e) Other

Of the two people who answered "e) other":

- One said they supported "option b) above but with the caveat that the opt out system be much easier and transparent"
- One said "only if Tom can opt out because here we are dealing with sensitive mental health

data, after being properly informed".

Q5. If your answer to Q4 is different to your answer to Q2, please give reasons for your answers. [50 words maximum]

There were no responses to this question.

General questions and jury answers about the use of free-text data

Q6. How comfortable are you with anonymisation of free-text patient data:

I. Where done by a person (researcher or healthcare professional)?
- 5 said: a) Comfortable
- 11 said: b) Somewhat comfortable
- 0 said: c) Neither comfortable nor uncomfortable
- 2 said: d) Somewhat uncomfortable
- 0 said: e) Uncomfortable

II. Where done by a computer?
- 5 said: a) Comfortable
- 12 said: b) Somewhat comfortable
- 0 said: c) Neither comfortable nor uncomfortable
- 1 said: d) Somewhat uncomfortable
- 0 said: e) Uncomfortable

III. Where done by a combination of a person and a computer?
- 8 said: a) Comfortable
- 10 said: b) Somewhat comfortable
- 0 said: c) Neither comfortable nor uncomfortable
- 0 said: d) Somewhat uncomfortable
- 0 said: e) Uncomfortable

Q7. You have heard reasons to support the process of anonymising, coding and using free-text data for health research, and reasons to be concerned about the process. Given these, to what degree do you support the use of free-text data from patients' records for health research?
- 6 said: a) Strongly supportive
- 12 said: b) Fairly supportive
- 0 said: c) Neither supportive not unsupportive
- 0 said: d) Fairly unsupportive
- 0 said: e) Strongly unsupportive

Q8 You have heard about several different ways in which free-text data can be anonymised, coded and used for health research. How supportive are you of each of these processes?

I. Where it is coded by the healthcare professional (e.g. GP or nurse) who provides care and records the free-text
- 7 said: a) Strongly supportive
- 8 said: b) Fairly supportive
- 1 said: c) Neither supportive not unsupportive
- 2 said: d) Fairly unsupportive
- 0 said: e) Strongly unsupportive

II. Where it is first anonymised by computer and/or person, then provided to a research team who will read the free text in order to gain a deep understanding of a specific thing (qualitative analysis)

7 said: a) Strongly supportive
10 said: b) Fairly supportive
0 said: c) Neither supportive not unsupportive
1 said: d) Fairly unsupportive
0 said: e) Strongly unsupportive

III. Where it is first anonymised by computer and/or person, then coded by a medical student and checked by a healthcare professional from the research team

11 said: a) Strongly supportive
6 said: b) Fairly supportive
0 said: c) Neither supportive not unsupportive
1 said: d) Fairly unsupportive
0 said: e) Strongly unsupportive

IV. Where it is first anonymised by computer and/or person, then coded by a medical student and checked by a healthcare professional, and then used to develop a computer program which would automatically code other patient records for research

10 said: a) Strongly supportive
8 said: b) Fairly supportive
0 said: c) Neither supportive not unsupportive
0 said: d) Fairly unsupportive
0 said: e) Strongly unsupportive

V. Where it is first anonymised by computer and/or person, then automatically coded by a computer program and checked by a healthcare professional,

7 said: a) Strongly supportive
10 said: b) Fairly supportive
1 said: c) Neither supportive not unsupportive
0 said: d) Fairly unsupportive
0 said: e) Strongly unsupportive

Note that in all the cases above apart from II., codes are created from free-text data and used for research.

Q9. What are the main reasons to support the process of anonymising, coding and using free-text data for health research?

- There is a large amount of free text data in patient records, particularly for mental health cases. This free text can be richer than coded data, adding "flesh" to the coded data within a patient record.
- This richer data can enable better research that could lead to better treatments, improve care, and may save lives.
- There is a low risk of re-identification when processing free text if proper procedures are followed.
- An opt out system gives a larger, more representative sample of the population for research than an opt in system which can lead to more accurate research and better results.

- When millions of records need to be processed by computer and there may be too many for humans to process effectively these processes can support better research.

Q10. What are the main concerns about the process of anonymising, coding and using free-text data for health research?

- If people believe their data are unsafe, they may withhold important information when seeing their doctor.
- The law requires "fair processing" – patients must be informed of the uses of their data but sometimes they are not.
- There is a lack of awareness about how patient data is used, or by whom, and that patients can opt out.
- People who might otherwise be willing to share information may be less willing to do so if they are unable to either give permission or be informed and able to opt out.
- Data processed to remove identifiers does not always mean it is completely anonymous
- Free text data is sensitive and inherently more identifying than coded data.
- Computer programs are currently unable to remove identifiers to an acceptable level with 100% accuracy.
- Free text patient data could contain information about other patients, judgements, offhand comments and other data requiring interpretation, and could be misinterpreted by researchers.
- There is a procedure in place (section 251) to ask for legal approval to process free text data without requiring consent in specific scenarios.
- Despite safeguards that might be in place, IT and data protection systems may be at-risk of being accessed by third-parties who seek unauthorized access to records and data.

Q11. Can you suggest how these concerns could be overcome?

Patients should be comprehensively informed at the outset about how, when, and under what conditions their free text might be processed, anonymised, coded, and analysed for research purposes. This should include:

- information that communicates their rights (to file complaints, to access their own information, etc.);
- a Privacy Statement;
- how data will be protected from breach during processing, analysis, and once research is completed;
- whether or not and how their information will be anonymised;
- who would access it and for what purpose; and plans for long-term storage or management of their data.

Researchers should communicate how decisions are made about who, why, and under what circumstances patients' data and records are being used in language that is accessible and easy to understand.

Efforts could be undertaken to involve patients in various elements of research and ethics decision-making (such as patients sitting on ethics boards) so that these processes are more open and transparent.

There has to be continuous improvement in the methods used for coding, anonymising, and processing free text, as well as in systems for safeguarding IT systems that secure access to data to improve performance, data protection, and public confidence.

Providing an option for people to access (published) research which utilises their records or data might be useful in maintaining trust.

End-of-jury questionnaire results

All 18 jury members of the jury completed the end-of-jury questionnaire at the end of day 3. This section shows results for the 18 jurors who completed both questionnaires. The questionnaire design and the full results are available at: www.healtex.org/jury.

One question, taken from a 2016 Ipsos MORI poll of the public commissioned by the Wellcome Trust [2], was asked in order to select a broadly representative sample of jury members in terms of balancing information sharing for public benefit and protecting privacy. This question was asked when jurors applied to participate in the jury, and then again in the end-of-jury questionnaire so as to gauge whether, and if so how, their views had changed by the end of the jury process. The question, and the answers given by the 18 jury members, are shown in the table below.

Ipsos MORI survey question: *"How willing or unwilling would you be to allow your medical records to be used in a medical research study? The information given to researchers would not include your name, date of birth, address or any contact details."*

Figure 5: Summary of juror answers

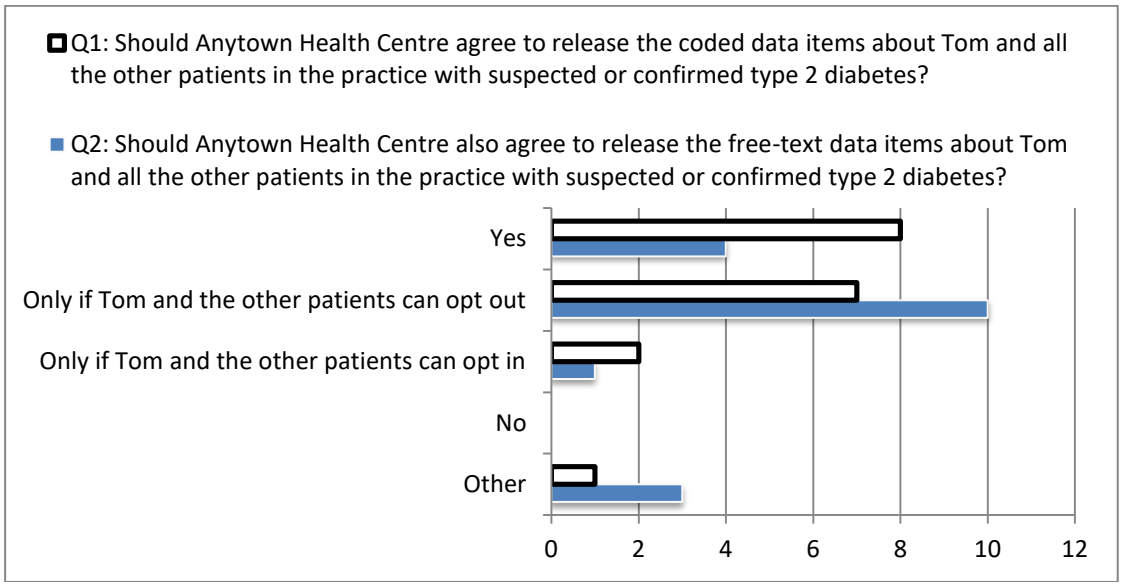
How willing or unwilling?	Pre-jury questionnaire	End-of-jury questionnaire
Very willing	9	9
Fairly willing	6	9
Don't know	0	0
Fairly unwilling	2	0
Very unwilling	1	0

The figures above suggest a general movement towards greater willingness to allow use of medical records for research. However, over the three days, 12 out of 18 people changed their minds and gave a different answer to the question to the one they originally provided during jury recruitment, with 7 people becoming more willing and 5 becoming less willing.

Main conclusions from citizens' jury

The results from the citizens' jury show broad support for the use of free-text health data being used for health-related research. This was shown in voting on the Tom scenario, and for the general questions that followed; the same pattern of results appeared across the first 8 jury questions. The jury was slightly more cautious about using free-text than coded health data for research but were nevertheless broadly supportive as long as there was a means of opting out. This is exemplified in the two graphs below with jury answers to question 1 and 2.

Comparison of Jury Answers to Q1 (coded data) and Q2 (free-text)

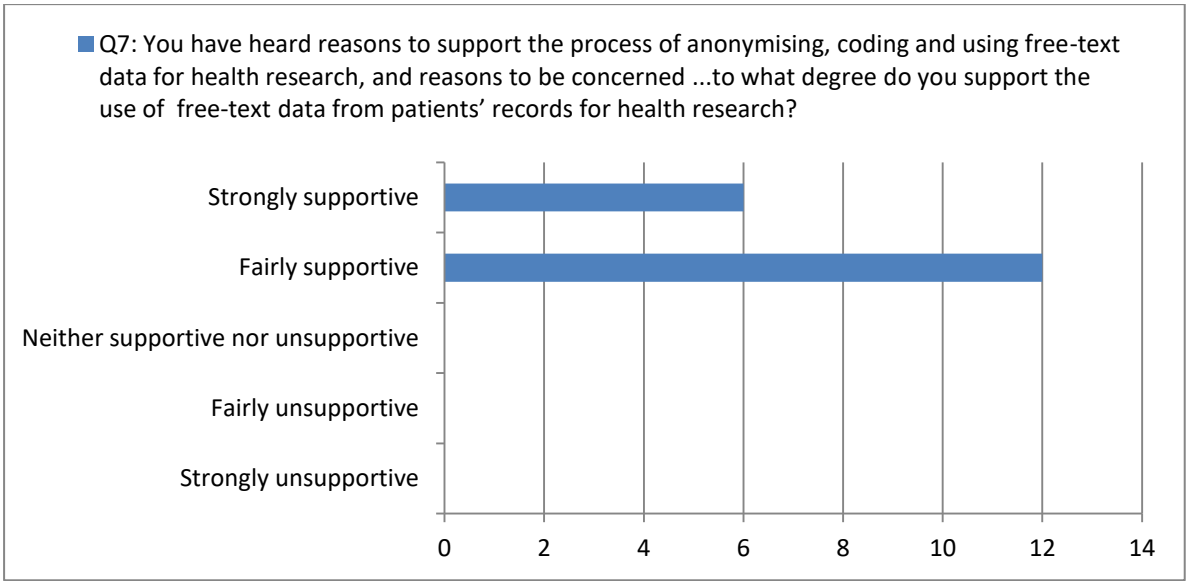


The jury heard evidence about the process used for anonymising and using free-text data for research, and potential arguments for its use and the reasons to be cautious. Their concerns were expressed in their answers to jury question 10 (see above for full list), and these included:

- Data processed to remove identifiers does not always mean it is completely anonymous
- Free text data is sensitive and inherently more identifying than coded data
- Computer programs are currently unable to remove identifiers to an acceptable level with 100% accuracy
- Free text patient data could contain information about other patients, judgements, offhand comments and other data requiring interpretation, and could be misinterpreted by researchers.

Nevertheless, the jury's answers to question 7 suggest that people supported the processing and use of free-text health data for health-related research. Note that the jurors were guided to consider academic and NHS research and asked not to consider research by private companies.

Jury Answers to Q7



Appendix 1: further information about the jury

The Citizens' Jury Method

Like much public policy, balancing privacy and sharing free-text data from patient records is a complex area with a lot of information and many arguments to consider. Surveys and focus groups provide useful information about what the public thinks, but they are not mechanisms to inform people. A citizens' jury can tell policymakers what members of the public think once they become more informed about a policy problem. In a citizens' jury, a broadly representative sample of citizens are selected to come together for a period of days, hear expert evidence, deliberate together, and reach conclusions about questions they have been set.

They are a form of "deliberative democracy", based on the idea that individuals from different backgrounds and with no special prior knowledge or expertise can come together and tackle a public policy question. A citizens' jury is a particularly relevant method for informing public bodies making value judgements. Some organisations have used citizens' juries to *make* policy decisions, even though members of juries are not elected and cannot be made accountable for decisions. For example, Melbourne City Council has appointed a citizens' jury to determine how to allocate its A$5 billion budget, and the council is implementing virtually all of the jury's recommendations.[3] The Citizens' Council in Ireland is currently considering many important questions. Its first topic was whether to change the Irish Constitution on abortion, where its advice to a parliamentary committee led to the May 2018 national referendum.

Expert witnesses

Seven expert witnesses were chosen to provide relevant information to the members of the jury to enable them to answer the jury questions. Each witness answered questions posed by the jurors. They all presented slides which were reviewed for bias in advance by the oversight panel (see below). One witness (Dr. Jon Fistein) was asked to be a "balancing witness", engaging in dialogue with John Carroll and Elizabeth Ford so as to help the jury to consider to understand more about the risks as well as the benefits of processing patient records to extract free-text data for research.

The expert witnesses were issued with a brief prior to preparing their presentations. It is published at: www.healtex.org/jury

The following table was provided (in ring binders) to jurors about each witness.

Day	Expert Witness		Topic	Section for Slides
Day 1 PM	Prof. Jackie Cassell	Trained as a medical doctor, now Department of Primary Care and Public Health at Brighton and Sussex Medical School	Patient records	14
Day 1 PM	Prof. John Carroll	Professor of Computational Linguistics (Informatics) at the University of Sussex	Computers anonymising free-text	15
Day 1 PM	Dr. Elizabeth Ford	Lecturer in Research Methodology, Brighton and Sussex Medical School	Processes to extract information from free text documents	16

Day 2 AM	Prof. Bobbie Farsides	Professor of Clinical and Biomedical Ethics	Ethics	17
Day 1 PM (no slides) & Day 2 PM	Dr Jon Fistein	Trained as a medical doctor and barrister, now Associate Professor in Clinical Informatics at the University of Leeds.	Balancing witness (day 1) Law (day 2)	18
Day 2 PM	Dr. Angus Roberts	Senior Lecturer in Health Informatics, King's College London	Case for using free-text	19
Day 2 PM	Mr. Phil Booth	Co-ordinator, medConfidential	Case for being cautious about using free-text	20

The oversight panel

The oversight panel was appointed to help monitor and minimise bias. The panel reviewed the citizens' jury questions and design, and much of the detailed jury documentation, including the jury questionnaires and the slides from the presentations by the impartial expert witnesses, resulting in some changes to these materials. The oversight panel members, chosen for their knowledge of the topic and lack of conflict of interest in any particular jury outcome, were:

- Ms Jenny Westaway, Head of Office of the National Data Guardian
- Ms. Dawn Monaghan, Head of Data Sharing and Privacy, NHS England;
- Dr. Mary Tully, Director of Public Engagement, Connected Health Cities.

The brief for the oversight panel is available at: www.healtex.org/jury. Each member of the panel completed a questionnaire about bias, which are published at the same site. The three panel members were "completely satisfied" that the jury was designed to minimise bias.

Citizens' jury project team and funders

The citizens' jury was funded by Brighton and Sussex Medical School, which received the funds following a researcher grant application to Healtex, the UK Healthcare Text Analytics Research Network, which is funded by the Engineering and Physical Sciences Research Council. The lead commissioner of the work from Brighton and Sussex Medical School was Dr. Elizabeth Ford.

The project manager of the citizens' jury was Dr. Malcolm Oswald, Director of Citizens Juries CIC and an Honorary Research Fellow in Law at The University of Manchester. Chris Barnes and Amanda Stevens from Citizens Juries CIC recruited and supported the jurors, and jury process. The lead jury facilitator was Kyle Bozentko, Executive Director of the Jefferson Center in the USA. Kyle, with support from his colleague Larry Pennings, worked closely with Malcolm to design the jury, and in particular the three-day jury activity programme. Kyle then facilitated the jury with Dr Lamiece Hassan, a researcher in health informatics and public engagement at the University of Manchester.

The citizens' jury programme of activities

The activities were designed primarily by the Jefferson Center in line with their citizens' jury method [4] and managed by the two facilitators. The jury ran Wednesday 6 June to Friday 8 June, 9:30 am to 5:00 pm. There was lunch, plus a tea/coffee break in the morning and afternoon.

When	Main content	Expert Witnesses involved
Day 1, AM	• Consent forms • Introductions • Why are we here? • Jury simulation exercise • Witness on introduction to Health Records with jury creating a health record (part 1)	Prof. Jackie Cassells, Head of the Department of Primary Care and Public Health, Brighton and Sussex Medical School (BSMS)
Day 1, PM	• Witness on introduction to Health Records with jury creating a health record (part 2) • Jury exercise: anonymising a health record • Two witnesses on free-text anonymisation and processing, with balancing witness and jury deliberation	Prof. Jackie Cassells Supported by Malcolm Oswald Prof. John Carroll, Professor of Computational Linguistics, University of Sussex and Dr Liz Ford, Lecturer in Research Methodology in Primary Care and Public Health at BSMS; Dr Jon Fistein, University of Leeds (balancing witness)
Day 2, AM	• More jury deliberation on anonymisation and processing of free-text • Witness on ethical considerations (e.g. Opt in vs. opt out, Mental/Physical Health distinction, Benefits/Risks of using free text) and jury deliberation	Prof. Bobbie Farsides (on ethics)
Day 2, PM	• Introductory presentation on the law and how patient data is protected • Jurors record initial answers to jury questions (on paper, or electronic questionnaires and individual's answers are printed then given to jurors next day) • Two partial witnesses: one argues for using free-text, and one for caution about using free-text health data, then deliberation	Dr Jon Fistein, University of Leeds Dr. Angus Roberts, Senior Lecturer in Health Informatics, King's College London Phil Booth, MedConfidential
Day 3, AM	• Jury deliberation on jury questions • Prioritising reasons • Jury voting on jury questions re Tom • Deliberation on how to generalise from specific case of Tom	None
Day 3, PM	• Voting about general jury questions • Jury report preparation • End of jury questionnaire	None

Appendix 2: Bibliography

1. Armour, A., *The citizens' jury model of public participation: a critical evaluation*, in *Fairness and competence in citizen participation*. 1995, Springer. p. 175-187.
2. Ipsos MORI. *The One-Way Mirror: Public attitudes to commercial access to health data*. 2016 [cited 17 Jun 2016]; Available from: https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf.
3. Reece, N. *Experiment pays off: Melbourne People's Panel produces quality policy*. 2015 [cited 15 Feb 2015]; Available from: http://www.theage.com.au/comment/experiment-pays-off-melbourne-peoples-panel-produces-robust-policy-20150628-ghzoz4.html.
4. Jefferson Center. *The Citizens' Jury Handbook*. 2004 [cited 09 Feb 2016]; Available from: http://www.epfound.ge/files/citizens_jury_handbook.pdf.
5. Wellcome Trust. *Wellcome Trust Monitor Report - Wave 3*. 2016 [cited 02 Feb 2018]; Available from: https://wellcome.ac.uk/sites/default/files/monitor-wave3-full-wellcome-apr16.pdf.