8 June 2018

Brighton Citizens' Jury on Free-Text Patient Data

Commissioned by the Brighton and Sussex Medical School and Healtex

Citizens Juries c.i.c. & Jefferson Center

Brighton Citizens' Jury on Free-Text Patient Data

COMMISSIONED BY THE BRIGHTON AND SUSSEX MEDICAL SCHOOL AND HEALTEX

Introduction

This is an overview of proceedings of the citizens' jury on the use of free-text patient data convened from 6 to 8 June 2018 at Jury's Inn Waterside Hotel in Brighton, UK.

Overview of Proceedings

Jury Mission and Goals

A group of 18 jurors representing a broad cross-section of the public living in and around Brighton worked together to respond to a series of questions related to when free-text notes and letters in a patient record should be anonymised and used for health and healthcare research.

The Citizens' Jury was asked to:

- Learn about two scenarios involving a fictional patient - Tom - where information about Tom was recorded as part of providing healthcare to Tom, and then requested by researchers
  - One scenario about Tom's physical health (diabetes)
  - One scenario about Tom's mental health
- Vote on whether free-text and/or coded patient data recorded as part of Tom's healthcare should be anonymised and made available for research in these scenarios
- Report the reasons that inform the votes
- Generalise from the specific case of Tom to:
  - Vote on questions about anonymisation of free-text patient data, and on different kinds of processing and uses of anonymised free-text patient data for health and healthcare research
  - Report reasons to support, and be concerned about, using free-text patient data in this way
  - Suggest how concerns about using anonymised free-text patient data for research could be overcome.

Statement from participants

To our friends, neighbours, health professionals, the public

Through this process we learned from a wide range of experts, deliberated with one another, and developed a better understanding of how patient data is processed, anonymised, and used. We had the opportunity to carefully consider the potential benefits and risks of using free-text patient data in health and healthcare research.

Though we were a diverse cross-section of the public we were often on the same page about the topics we discussed, though we disagreed at times and were not always unanimous in our decisions. We were able to contribute to the discussion in a variety of ways and were allowed to make up our own mind about the topics we considered.

We hope this report helps in the development of future guidelines around the use of free-text patient data for research and that others will benefit from our ideas. While we are a cross-section of the public we hope that our views would reflect broader perceptions about this topic – that we expect research be conducted with the good of the general public in mind and that adequate steps be taken to safeguard our data and information. We ask that others take the opportunity to review their own records when available, to be thoughtful about how and when they consent to collecting and using their records, and be better informed about this topic as a result of our work.

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

I. 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

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.

Important considerations in reaching these conclusions

Over the course of the three days, we heard evidence from witnesses, and deliberated together to reach the answers to the jury questions set out above. We identified and discussed many different reasons and arguments over the three days. The things we considered most important in reaching our conclusions are set out in the tables below.

The Process of Anonymising and Coding Free-Text Patient Data

Things that made us confident about the process	Reservations we had about the process
• There are many checks and legal and non-legal controls built into the processes used to anonymise and use free-text • There are strong security protections and it would be very difficult to identify individuals from the data processed • Computers are now better than humans at anonymising free-text	• The error rate for "under-scrubbing" (i.e. not anonymising free-text data that should be anonymised) is too high • The checks and controls are done by staff involved in the process (self-policing) and differ across the NHS • It is time-consuming and costly to do

Opt in vs Opt Out

Reasons to support opt in	Reasons to support opt out
• Opt in would encourage more informed participation among patients	• Consent is beneficial but need to ensure that those with limited capacity are catered for e.g. mental health, religion, education, parents, carers

Drawing a distinction between physical and mental health data

Reasons to treat physical and mental health data differently	Reasons to treat physical and mental health data the same
	• If we treat physical and mental health differently it may safeguard those who are sensitive about their mental health, and may be easier for research purposes (coding).
	• If we treat physical and mental health differently, funding may be better protected.

Legal protections we considered important

- Patient data will be protected where there is a reasonable expectation of confidentiality but in certain circumstances your information may be disclosed eg. to a statutory body
- Data protection law protects how personal data are stored and used and needs to define what is meant by anonymisation and pseudonymisation
- Patient data is sensitive personal data and different from personal data
- You have a right to complain about a breach but by then the information will already be out

there
- The public should be informed about how to make a complaint
- We should know what breaches have happened and what harm was caused.

Free-text use	
Important reasons to use free text	**Important reasons to be cautious about using free text**
• If we allow use of free text we gain the potential to gather more information to improve research and create a healthier population.	This might lead to: • Misuse of our data • Production of inaccurate data • Unclear decisions of what is important, what is included/excluded, etc., • Misunderstanding and manipulation of data to influence people / make commercial profit.