SYSTEMATIC REVIEW

“Giving something back”: A systematic review and ethical enquiry of public opinions on the use of patient data for research in the United Kingdom and the Republic of Ireland

[version 1; referees: 1 approved with reservations]

Jessica Stockdale, Jackie Cassell, Elizabeth Ford

1Department of Primary Care and Public Health, Brighton and Sussex Medical School, Brighton, UK
2Department of Philosophy, School of History, Art History and Philosophy, University of Sussex, Brighton, UK

Abstract

Background: Use of medical data for secondary purposes such as health research, audit, and service planning is well established in the UK. However, the governance environment, as well as public opinion and understanding about this work, have lagged behind. We aimed to systematically review the literature on UK and Irish public opinions of medical data use in research, critically analysing such opinions though an established biomedical ethics framework, to draw out potential strategies for future good practice guidance and inform ethical and privacy debates.

Methods: We searched three databases using terms such as patient, public, opinion, and electronic health records. Empirical studies were eligible for inclusion if they surveyed healthcare users, patients or the wider public in UK and Ireland and examined attitudes, opinions or beliefs about the use of patient data for medical research. Results were synthesised into broad themes using a Framework Analysis.

Results: Out of 13,492 papers and reports screened, 20 papers or reports were eligible. While there was a widespread willingness to share EHRs for research for the common good, this very rarely led to unqualified support. The public expressed two generalised concerns through a variety of hypothetical examples. The first of these concerns related to a party’s competence in keeping data secure, while the second was associated with the motivation a party might have to use the data.

Conclusions: The public evaluates trustworthiness of research organisations by assessing their competence in data-handling and motivation for accessing the data. Public attitudes around data-sharing exemplified several principles which are also widely accepted in biomedical ethics. This provides a framework for understanding public attitudes, which should be considered in the development in any guidance for regulators and data custodians. We propose four salient questions which data guardians should address when evaluating proposals for the secondary use of data.

Open Peer Review

Referee Status: Invited Referees

version 1

Invited Referees

Sarah Cunningham-Burley, University of Edinburgh, UK
Corresponding author: Elizabeth Ford (e.m.ford@bsms.ac.uk)

Author roles: Stockdale J: Conceptualization, Data Curation, Formal Analysis, Writing – Original Draft Preparation, Writing – Review & Editing;
Cassell J: Conceptualization, Writing – Review & Editing; Ford E: Conceptualization, Formal Analysis, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

How to cite this article: Stockdale J, Cassell J and Ford E. “Giving something back”: A systematic review and ethical enquiry of public opinions on the use of patient data for research in the United Kingdom and the Republic of Ireland [version 1; referees: 1 approved with reservations] Wellcome Open Research 2018, 3:6 (doi: 10.12688/wellcomeopenres.13531.1)

Copyright: © 2018 Stockdale J et al. This is an open access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Grant information: This work was supported by the Wellcome Trust [202133]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Introduction
The use of medical data for secondary purposes such as health research, audit, and service planning is well established in the UK, and technological innovation in analytical methods for new discoveries using these data resources is developing quickly. Data scientists have developed, and are improving, many ways to extract and process information in medical records. This continues to lead to an exciting range of health related discoveries, improving population health and saving lives. Nevertheless, as the development of analytic technologies accelerates, the decision-making and governance environment as well as public opinion and understanding about this work, has been lagging behind\textsuperscript{1}.

Public opinion and data use
A range of small studies canvassing patient opinions, mainly in the USA, have found an overall positive orientation to the use of medical data for societal benefit\textsuperscript{2-7}. However, recent case studies, like NHS England’s ill-fated Care.data scheme, indicate that certain schemes for secondary data use can prove unpopular in the UK. Launched in 2013, Care.data aimed to extract and upload the whole population’s general practice patient records to a central database for prevalence studies and service planning\textsuperscript{8}. Despite the stated intention of Care.data to “make major advances in quality and patient safety\textsuperscript{9}”, this programme was met with widely reported public outcry leading to its suspension and eventual closure in 2016. Several factors may have been involved in this failure, from the poor public communication about the project, lack of social licence\textsuperscript{10}, or as pressure group MedConfidential suggests, dislike of selling data to profit-making companies\textsuperscript{11}. However, beyond these specific explanations for the project’s failure, what ignited public controversy was a concern with the impact that its aim to collect and share data on a large scale might have on patient privacy. The case of Care.data indicates a reluctance on behalf of the public to share their medical data, and it is still not wholly clear whether the public are willing to accept future attempts at extracting and linking large datasets of medical information. The picture of mixed opinion makes taking an evidence-based position, drawing on social consensus, difficult for legislators, regulators, and data custodians who often respond to personal or media generated perceptions of public opinion. However, despite differing results of studies, we hypothesise that there may be underlying ethical principles that could be extracted from the literature on public opinion, which may provide guidance to policymakers for future data-sharing.

Governance and legal framework of data use
The Data Protection Act (1998) is the main legislation governing the use of patients’ medical data, soon to be replaced by General Data Protection Regulation (2018). This law covers personal, or patient identifiable data in the UK. Personally Identifiable Information (PII) is defined as information that can be used on its own, or with other information, to identify, contact or locate a single person, or to identify an individual in context. Where patient data is being used at scale for indirect patient care (that is, service planning or research), without explicit consent and while still patient-identifiable, its use must be approved by the Confidentiality Advisory Group England (CAG) within the Health Research Authority.

Another potential route to the use of data for these purposes without individual patients’ consent, is the de-identification of data so that it is no longer classified as personal data. De-identification involves removal or replacement of personal identifiers so that it would be difficult to re-establish a link between the individual and their data\textsuperscript{11}: “The challenge is to balance the levels of de-identification that are acceptable to the patients, research participants, clinicians, researchers, institutions, and federal requirements”\textsuperscript{11}.

It should also be noted that privacy can be achieved through other means than de-identification of patient data such as controlling linkage with other sources of information, robust protection with computing security systems, and giving access only to trusted users\textsuperscript{12}. “De-identification should be considered a necessary but insufficient means of protecting health privacy. In accordance with this view, health information should be collected, maintained, disclosed, and used in the least identifiable form consistent with the purpose of the information”\textsuperscript{13}.

Striking a balance
While it is clearly important to make sure patient privacy is protected, it is also argued that the societal benefit of medical research using health big data, which may save lives, should be given ethical weight. This is argued on the basis that harms to patients may occur where these rich data sources are not used to improve our understanding of health conditions and treatments\textsuperscript{14}. While individual privacy and societal benefit are often portrayed as being in opposition, for the future of health big data research, a way to achieve both to the satisfaction of patients, legislators and researchers must be found. Recent work has sought to identify the key issues of patient responses to data-sharing\textsuperscript{15}. It still remains to be established whether these issues are connected by any system of ethical values. The Four Principles approach, established by Beauchamp and Childress, has become a canonical text and established approach to evaluating the ethical aspects of medical practice and decision-making\textsuperscript{16}. The Four Principles proposed by Beauchamp and Childress are respect for autonomy, nonmaleficence, beneficence, and justice. We aim in this study to systematically review and thematically analyse UK and Irish studies exploring patient and public opinions on medical data being used for the secondary purpose of research, using the lens of the Four Principles to understand and map the results we find onto an established ethical framework. We aim to draw out potential strategies for future good practice guidelines around data privacy, to guide data custodians, the health data research community and the public.

Methods
We followed the PRISMA guidelines for the conduct and reporting of this review\textsuperscript{17}.

Search strategy
We searched PubMed, Web of Science, and Scopus between 03/10/16 and 11/10/16 using the following search string: (Public OR Patient OR People) AND (Attitudes OR Knowledge OR Opinions OR Views OR Perceptions) AND (“Care.data” OR “Electronic Health Record” OR “Electronic Health Data” OR “Electronic Medical Record” OR “Electronic Medical Data” OR ...)
“Personal Health Information” OR “Personal Health Record” OR “Electronic Patient Information” OR “Electronic Patient Data” OR “Electronic Patient Record” OR “Data linkage” OR “Data sharing”) AND (Research). We restricted our search to publications from 2006–2016 inclusive. We also searched the grey literature using the search string: “public attitudes” AND “sharing” AND “health data” on Google (in June 2017). The first 20 results were selected and screened. The following inclusion criteria were then applied:

1. Empirical studies using any methods reported as a full length peer review manuscript or published report.
2. Healthcare users, patients or the wider public as participants
3. Examining attitudes, opinions or beliefs about the topic of use of patient data for medical research.
4. Studies using a UK or Irish sample, written in English. We chose to keep our review to these two countries because of similarities in their socialised healthcare systems, and because of the well-established use of patient data within these jurisdictions.

Studies were excluded if they were:

1. Studies focused more broadly on digital technologies in health care where the focus was on use of digital methods or records rather than public attitudes
2. Studies focused on patient and practitioner attitudes to analogous areas such as biorepositories, genetic testing and genomic research.
3. Non-empirical reviews of legislation, policy, ethical challenges etc.

Using these criteria, the articles extracted from the literature search were screened based on their title, then abstract (by author JS), then finally the choice of full text papers for the review was undertaken by two authors (JS and EF).

Quality assessment
Study quality was assessed using the Mixed Methods Appraisal Tool (MMAT)\(^9\). This tool was designed for the appraisal of studies in mixed methods systematic reviews and attempts to appraise the quality of methodology, rather than the quality of reporting. All studies meeting the inclusion criteria above were assessed using 6 criteria. The first two are the same for all studies: is there a clear research question or objective, and does the data collected address the research question or objective? A further 4 questions were specific to the study type. Studies were given a score out of 6 depending on how many of the 6 criteria they met, and were rejected if they did not meet at least the first two criteria. Two papers were excluded on the basis of scoring zero on all criteria.

Data extraction
We extracted author names, dates, location, type of study (qualitative or quantitative), methods used, number of participants, their backgrounds or roles, ages, genders, and the study findings which fitted into the themes relating to research questions reported below.

Synthesis of results
The full text of eligible articles were read iteratively by two authors (JS and EF) with the aim of extracting coherent themes. In the first iteration of reading and coding the results of the papers, authors focussed on nine questions.

1. Are patients/public aware of electronic health records (EHRs) and their secondary uses?
2. Are patients/public concerned about the privacy and security of their medical data?
3. Are patients/public willing to share their medical data for research, policy and planning?
4. What consent model do patients/public prefer?
5. How does data being identifiable or anonymised affect patient/public preferences?
6. Which organisations are most and least trusted with patients’/public data?
7. What are the reported perceptions of risks and benefits of sharing medical data for research?
8. Are there any other ways in which willingness to share could be increased?
9. Is there any one social group who are overly concerned with the sharing of EHRs?

A framework\(^9\) was created with a column for each of the 9 questions and data was extracted from each study where it fitted into these categories. Following this data extraction, the two authors (EF and JS) discussed refining and combining extracted data into as smaller number of themes. In a second iteration of data extraction, authors re-read articles and extracted data into 7 themes. For interpretation and synthesis, a data driven approach was taken, trying to make meaning from first order data reported in the papers (i.e. statistics or participant quotes). Where themes were populated mainly by summary of quantitative data, a straightforward reports of papers’ findings is given. Where contributing papers were mainly qualitative e.g. in the ‘Trust theme, we undertook a deeper analysis of meaning within findings guided by both meta-synthesis principles\(^9\) and established principles of bioethics\(^9\).

Results
A total of 13,472 peer-reviewed papers were found through the systematic search, as well as 20 reports found through the grey literature search. Of these, 20 UK and Ireland based papers met the inclusion criteria and were included in the review\(^\text{21–39}\) (Supplementary File 2). Studies which reported time periods indicated that data was collected from 2004 to 2016, although seven studies published between 2011 and 2016 did not report the data collection period. Research participants included patients,
service-users, lay persons, those living with chronic conditions, and the general public ranging from 16 years of age to over 75. Five of the studies included the views of health researchers, health professionals, industry experts, NHS managers and other key stakeholders. Seven of the papers were quantitative, using surveys or structured questionnaires. Ten of the studies were qualitative, using focus groups and one-to-one interviews, and there were three mixed methods studies. Details of studies are reported in Table 1.

Quality assessment
Studies’ quality scores ranged from 3 to 6 out of a possible 6, scores of individual studies are shown in Table 1. Two studies which otherwise met inclusion criteria were rejected on the basis of quality and do not appear further in the results40-41.

Themes elicited from the studies
The seven themes identified in and elicited from the studies were: Knowledge and Awareness of Electronic Records; Willingness to Share; Privacy; Trust; De-identification and Consent Preferences; Increasing Trust; and Demographic Differences. The contribution of each study to each theme is shown in Table 2.

Knowledge and awareness of electronic records
Generally, knowledge of the content and electronic collection of GP records among respondents was high. One quantitative study reported that a moderately high proportion of respondents at 59% had prior awareness of EHRs39. Another quantitative study reported that levels of understanding of the information recorded by GPs were high without giving exact numbers35. One qualitative study reported that across participants, there was a high awareness of the kind of information that usually held in general practice records1. Nevertheless, participant awareness of specific uses of routinely collected patient data was low. For instance, two quantitative studies reported that 82%33 and 80%33 of the general public had not heard of the National Cancer Registry, while another study reported that patients were not only inadequately informed about Care.data, but were also unaware of the project39. Two studies indicated that understanding of medical research using patient data was low33,31, while another suggested that participants were unaware of how their data was currently used36. Another demonstrated limited public grasp of a range of concepts related to patient information use, such as de-identification, data science, the benefits of aggregate data, and the role of private companies in the healthcare system. People with lower understanding of these issues were more likely to have concerns about commercial access to health data33.

Willingness to share
In many of the studies, participants expressed willingness to share their EHRs for secondary purposes like research, policy and planning, despite the range of concerns discussed below. Among the quantitative studies, public support for a national EHR system was reported at 62.5%36, 62.47%3, and 81%32, while support for sharing information in general was reported 73%3. Specifically, support for sharing data with researchers was reported at 68.7%34, 83%36, 77%34, and 81.4%39. In the qualitative studies, participants identified willingness to share their EHRs for secondary purposes with the “common”, “greater” or “public good”30,21,22,26,31; “social responsibility”38,39; “altruistic attitudes”92; and “giving something back”97 to “other people” or “future generations”32,26-30,38. For example, in one study it was stated:

“I’m saying yes because I think there is a greater good. (Participant 1, Group 2.)35

Such reasoning was largely predicated on the understanding that medical research using EHRs could lead to benefits such as the improvement of healthcare services, or innovations in the diagnosis and treatment of disease. For example:

I think if you are going to do something, eczema, allergies, something that affects one in five people you need the huge samples in order to do it. (Patient Interview L2,39)

And:

. . . because you never know where research is going to go. You don’t know where some brilliant young scientist’s mind linking up different things, you know. And you cannot put a halt on that, a break on that. (Di, Focus Group 1,26)

Moreover, it was also understood that using EHRs might be a better way of doing and facilitating research:

. . . I mean it’s a better system than it is at present, because you are going to get 100% response that way or near enough and the present system is that the GPs put out things on spec to people that may want to join this thing and they may get a very low return. (Male, Patient Focus Group 3,21)

This suggests that, here the “common good” consists of the collective public health benefits brought about by the improvement of the services, practices and methods of healthcare through secondary uses of data. Willingness to share appears connected to idea of an individual having a personal responsibility, obligation or duty to help bring about this common good:

Once you have been in receipt of the excellent kind of care and treatment that I’ve had, I think you have a social responsibility that if you can help the next generation by having your information provided to the researchers to [do] some good. (Focus Group 3,39)

Privacy
Despite the general willingness to share EHRs for secondary purposes, many qualifying concerns were raised by participants32,26-33,36,39,40. This suggests that although the sharing of EHRs is largely seen as being for the overall common good, participants believe that it also has the potential to create new risks, and increase existing ones. The various perceived risks involved in sharing EHRs were well described with participants frequently citing such things as hacking39,56, unintentional data leakage or loss39, identity theft30, unauthorised access39, errors in medical records39, unnecessary stigmatising judgements in clinical settings39, consequences for employment, pension
<table>
<thead>
<tr>
<th>Study Reference</th>
<th>Setting</th>
<th>Method</th>
<th>Data Collection Period</th>
<th>Sample</th>
<th>Quality Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>Audrey, S., et al., BMC Medical Ethics. 2016. 17(1): p. 53.</td>
<td>Qualitative: focus groups with participants from ALSPAC.</td>
<td>Bristol, England</td>
<td>February and July 2006</td>
<td>Total n=55, 56.4% female 43.6% male Ages: 17–19</td>
</tr>
<tr>
<td>25</td>
<td>Campbell, B., et al., Quality and Safety in Health Care, 2007. 16(6): p. 404–408.</td>
<td>Qualitative: postal questionnaire</td>
<td>South-West England</td>
<td>October to December 2004</td>
<td>Total n=19, female (n=12), male (n=6), Unstated (n=1) Ages: &lt;60 (n=1); 60–74 (n=15); +75 (n=3).</td>
</tr>
<tr>
<td>26</td>
<td>CM Insight and Wellcome Trust, Summary report of qualitative research into public attitudes to personal data and linking personal data. 2013.</td>
<td>Qualitative: focus groups and one-to-one telephone interviews.</td>
<td>London, England</td>
<td>29 April to 12 May 2013</td>
<td>Total n=50, Ages: 18–70 Focus group respondents were recruited as owners of store loyalty cards, smart phones and social media users. Telephone interviewees were recruited as especially pro-privacy or cautious about sharing personal data.</td>
</tr>
<tr>
<td>28</td>
<td>Grant, A., et al., BMC Health Services Research, 2013. 13: p. 422.</td>
<td>Qualitative: focus groups and semi-structured interviews.</td>
<td>Tayside and Lothian, Scotland</td>
<td>Between February and June 2011</td>
<td>Total n=64, Focus Groups: Patients (n=37), Health services researchers (n=10) Interviews: GPs and Practice managers (n=17).</td>
</tr>
<tr>
<td>29</td>
<td>Haddow, G., et al., Journal of Evaluation in Clinical Practice, 2011. 17(6): p. 1140–1146.</td>
<td>Qualitative: focus groups.</td>
<td>North East Scotland</td>
<td>May and June 2009</td>
<td>Total n=19, female (n=12), male (n=6), Unstated (n=1) Ages: &lt;60 (n=1); 60–74 (n=15); +75 (n=3).</td>
</tr>
<tr>
<td>30</td>
<td>Hays, R. and DakeWhite, B, BMC Public Health. 2015. 15: p. 838.</td>
<td>Qualitative: using tweets.</td>
<td>Twitter</td>
<td>Over 18 days during February and March 2014</td>
<td>Total n=3577 tweets containing the hashtag #caredata. 30% contributors</td>
</tr>
<tr>
<td>Study no.</td>
<td>Reference</td>
<td>Method</td>
<td>Setting</td>
<td>Data Collection Period</td>
<td>Sample</td>
</tr>
<tr>
<td>----------</td>
<td>-----------</td>
<td>--------</td>
<td>---------</td>
<td>------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>32</td>
<td>Ipsos Mori, Macmillan Cancer Support, Cancer Research UK, Perceptions of the Cancer Registry: Attitudes towards and awareness of cancer data collection. 2016.</td>
<td>Quantitative: online survey.</td>
<td>England</td>
<td>13 June to 4 July 2016</td>
<td>Total n=2,033 Adults who have, or have had cancer (PLWC) (n=1,033) Adults from the general public (n=1,000). All 18+</td>
</tr>
<tr>
<td>33</td>
<td>Ipsos Mori, Wellcome Trust, The one-way mirror: Public attitudes to commercial access to health data. 2016.</td>
<td>Mixed methods study. Quantitative: face-to-face survey Qualitative workshops</td>
<td>England, Wales, Scotland and Northern Ireland</td>
<td>September to December 2015</td>
<td>Quant: n = 2017, age 16+; Qual n= 247; Members of the general public, patients, and ALSPAC cohort members (n=212) GPs and hospital doctors (n=35)</td>
</tr>
<tr>
<td>39</td>
<td>Stevenson, F., et al., Family Practice, 2013. 30(2): p. 227–232.</td>
<td>Qualitative: focus groups and interviews</td>
<td>Not stated</td>
<td>Not stated</td>
<td>Total n=57, Patients (n=50), Staff members (n=7).</td>
</tr>
</tbody>
</table>
Participants frequently identified two key elements that could be determined in relation to their information. The first was whether information is revealed to, or accessed by another party:

My concern is exactly that: who has access to my files and how can we make sure that only those I want to have access would have access? (Focus Group 12, 36)

A second element concerned how this information should be used, or analysed after it being revealed to another party:

At the end of the day, it’s not who has access to it all, it’s how they use it, I think is the main concern for us, for everybody, . . how they use it. (Person with MS, Focus Group 7, 22)

### Table 2. Contribution of Studies to Themes.

<table>
<thead>
<tr>
<th>Study</th>
<th>Knowledge and Awareness of Electronic Records</th>
<th>Willingness to Share Data</th>
<th>Privacy</th>
<th>Trust</th>
<th>De-identification and Consent</th>
<th>Increasing Trust</th>
<th>Demographic Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audrey et al. 2016</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baird et al. 2009</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barrett et al. 2006</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buckley et al. 2011</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Campbell et al. 2007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>CM Insight and Wellcome Trust 2013</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Clerkin et al. 2013</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grant et al. 2013</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haddow et al. 2011</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hays &amp; Daker-White 2015</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Hill et al. 2013</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ipsos Mori, MRC 2007</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ipsos Mori, MacMillan, CRUK 2016</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ipsos Mori, Wellcome Trust 2016 One Way Mirror</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ipsos Mori, Wellcome Trust 2016 Monitor Report 3</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Luchenski et al. 2013</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Papoutsi et al. 2015</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Riordan et al. 2015</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spencer et al. 2016</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stevenson et al. 2013</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

eligibility, or insurance costs, social discomfort and community embarrassment, re-identification, access without explicit consent, the use of EHRs for financial gain, aggregating data to a group’s disadvantage, access, use and governance of data by the government. The breadth of this list demonstrates the structural complexities of the particular, concrete situations which study participants imagine may arise from the misuse of their data. Several studies connected these risks and the concept of privacy. Privacy was widely conceptualised as a process whereby an individual determines for themselves what happens with the information relating to them:

Seemingly radical idea: let PATIENTS control who can access their personal medical data! #caredata. (Twitter user)
These two factors were necessary components in identifying what was and was not acceptable when it came to unlocking the potential of health data.

**Trust**

Views on storing and using EHRs were linked to the kind of trust or distrust the public had in an organisation or individual using or accessing the data.

You have to trust people. (Fiona, Focus Group 2, 28)

Where participants distrusted organisations who would handle their data, this generally occurred along two lines:

1. Distrust of a party’s ability, or competence, to ensure data security.
2. Distrust of a party’s motivations.

In terms of a party’s competence, participants may agree that a particular party can store and use their EHR in principle, but are concerned that they are not able to guarantee the level of security required by such personal data due to institutional incompetence. One such party was “the NHS”33,36. For example, in one study a large majority, 71.3% of respondents, voiced doubts about the ability of the NHS to guarantee the security of EHRs, yet 53.5% of those respondents would nevertheless support the development of a national EHR46. On the incompetence and inefficiency of the NHS, participants stated the following things:

I just have very little faith in the way that the NHS handles databases. I don’t think it’s got a very good record... (Focus Group 3, 29)

Always thought that [the NHS] would mess it up (Focus Group 11, 36)

#NHSPatientdata scheme handling a ‘masterclass in incompetence’ #CareData #NHS [link] [link], (Twitter user, 29)

However, in the qualitative studies, participants expressed a generalised trust towards the NHS, especially when concerning GPs:

I mean I can trust the doctors and all... but other people, no. Once it leaves the NHS, I’d be wondering where it’s going and who’s looking at it. (Participant 1929)

... once it goes out of the NHS, the NHS have no control over it whatsoever. (Di, Focus Group 1, 26)

Participants tended to say that the data would be safer in the hands of the NHS or a public sector or independent organisation, and that private companies were less likely to be as diligent in their handling of it33.

When it came to an organisation’s motivation, there was a strong sense that any access and use of the data must be for the good of the individual patient or the common good of the public. Many studies indicated that any kind of data handing for private interests would be unacceptable36–30,36,39. In terms of the possible consequences, a recurring theme was that if a party had the wrong competences or motivations, this could lead to substantial harm on both an individual or collective societal level. For instance, as the following quote illustrates, it was identified that the private profit motivations of insurance and marketing companies could lead to harms on an individual level:

One of my fears was if it somehow goes astray from there and somebody, for instance, like insurance companies, get a hold of it they could use it to their advantage and the patient’s disadvantage. (P2, Focus Group, 27)

However, direct harm to individuals is not a necessary factor in determining the wrongness of certain motivations. It was also indicated that even if no particular individual is disadvantaged, allowing those with private interests to access public data can constitute a collective harm. This is because there is a strong sense that data should only be used to benefit either individuals:

Financial gain comes into it then so why should you then let them look at your records? They are going to gain out of it and you’re not... (Participant 2, Group 2, 30)

Or, the public at large:

If there was a large commercial company... [that] had free and easy access to people’s medical records I don’t think that would be right. It would further their research into the particular drug or treatment, but it’d also further their profits that would be wrong. But if it was for medical research for everybody then that would be different. (Participant 6, Group 3, 36)

Despite this firm belief, several of the studies indicated a tension in the status of pharmaceutical companies whose products are indispensable to medicine and the health of populations, but which ultimately operate in a profit driven capacity22,27,30,31,36. As Grant et al.27 write, this leads some participants to see the involvement of pharmaceutical companies as a “necessary evil”.

This dimension was further discussed in the grey literature which revealed a more nuanced picture regarding public opinion towards the commercial uses of data. Support for commercial access to health data raised from 54% to 61% when taking into account the possibility of new treatments being discovered28,29,31,36, and participants were indifferent to who conducts research so long as the objective is to increase knowledge around the causes and cures of ill health30. This suggests that participants recognise that not all commercial uses of data are done from purely privately interested motivations, but that at least in part can involve public motivations too. In explaining the apparent reluctance of the public to accept certain private interests so as to ensure public benefits, one study identified that participants did not currently feel that they could evaluate the motivations of commercial organisations using data, which created an unclear conception of what the public could stand to gain through these uses of data. As a result, participants tended to fall back into wider assumptions, personal beliefs and prejudices regarding private companies31.
De-identification and consent preferences

In the quantitative studies, 67.5%20 of respondents in 2011 and 91%31 of respondents in 2015 were clear that although it was fine for researchers to access their EHRs, they still expected to be asked for consent when their identifiable data was accessed for secondary purposes. However, there was less consensus over de-identified data, with 83.7%24, 51%33, and 49.3%37 of respondents reporting willingness to share or extracted without consent. Reasons for concern around de-identification also emerged in the qualitative studies where participants questioned what would qualify as identifying information48, whether de-identification could be achieved effectively31,36, whether it was sufficient for the elimination of consent21,30 and highlighted the risks of re-identifying individuals26,29.

Several studies also indicated substantial concerns about the opt-out rather than opt-in consent which was proposed in schemes such as Care.data20,39, while others noted that participants generally thought about consent along opt-in lines when asked for their opinions21. Participants expressed worries about whether people would really understand the concept of opting-out9. They also criticised opt-out on the basis that it was unethical and illegal22. However, in one quantitative study 52% of the general public supported the opt-out method of collection for the National Cancer Registry21, while a minority of participants in another study acknowledged that opt-out might be a better option given the impracticalities of opting-in37.

The problem of selection bias and its connection with consent arrangements was explored in three studies21,30,36. In two studies, some participants identified the potential for bias if the information which was gained was neither accurate nor balanced21,30:

If they’ve got mental health illness then . . . that might affect their willingness, so it might be hard to . . . gather enough information. I think that might be biased. . .
(Male ID47, 21)

Participants also recognised that larger, more representative samples could be gained by an opt-out process:

You are going to get 100% response that way, or near enough and the present system is that the GPs put out things on spec to people that may want to join this thing and they may get a very low return. (Male, Patient focus group 3, 27)

This prompted discussion in one study about the importance of mitigating the requirement of consent by de-identifying information:

There is certain situations where you might be able to, it might be acceptable to ask or it might be acceptable just to go ahead and get it—as long as it wasn’t directly linked back to you as a person, it would be alright. . .
(Female, ID6, 21)

In another study38, after receiving presentation about selection bias, participants recognised the difficulties faced by researchers. Interestingly, when asked if this information had changed their opinion about using health data without consent, several participants out of the group who at first indicated reluctance, reported that they had indeed changed their minds. A quantitative study showed that a substantial minority of respondents at 20%33 believe that consent may not be needed if it is not practical to obtain.

Increasing trust

Across studies, participants identified several different infrastructure arrangements which could increase willingness to share EHRs for secondary purposes. Participants indicated that no single organisation should be responsible for deciding who could access and use their EHRs, rather a committee of stakeholders was called for, including Caldicott Guardians, research consultants, members of the public, GPs, social services staff, charities, funders, patients etc.22,28. It was also felt that greater transparency was needed in regards to safeguarding processes and data sharing arrangements20,38, including stiff penalties or fines for misuses of data28,33; the publication of results31; clear guidelines and laws to regulate access and use of data29; and, regulators and parties accessing data to be held to high standards31. Several studies also indicated that participants wanted a better understanding about the nature of EHR initiatives, medical research31, the purposes and benefits of using data22,33, de-identification and aggregation31, and also why in some situations consent might not be practical31. More generally, participants wanted the security of records to be ensured23,33; for private profit to be capped33; and denial of third party access33. In several studies, participants also indicated their preference to retain granular control over the data in their EHR using an explicit opt-in consent scheme, the right to withdraw at any time and ability to tailor sharing preferences22,27,29,38.

Despite the breadth and diversity of participant suggestions to increase willingness, it might be that no one, or any specific combination of strategies will amount to a gold standard of acceptability or social licence. One study found that no particular safeguard made sharing data with commercial companies any more acceptable than any other31. However, in the same study, participants were significantly less likely to endorse sharing data without any safeguards (49% agreed) compared to with safeguards (56–64% agreed, depending on the safeguard). This suggests that the precise nature of the safeguard is less important to improving willingness to share than knowing there are safeguards in place.

Demographic differences

We aimed to ascertain whether the included studies indicated a level of heightened concern, worry or fear among one or more specific social groups and we restricted this analysis to quantitative studies which are designed to enable such contrasts. Although participants were asked a variety of different questions across each survey, we evaluated responses on the basis of whether they indicated an overall negative or positive attitude towards the sharing of EHRs for secondary purposes such as research. For example, in Papoutsi et al.36, participants were asked if they would be more worried about the security of their information if it were part of a national EHR register, while Buckley et al.34 asked if they would allow their EHRs to be...
provided to researchers without their explicit consent. Despite the differing approaches of these questions, we concluded that a response indicating more worry about security, and one indicating less likelihood of granting researchers access without explicit consent, were comparative insofar as they represented a negative attitude towards sharing of EHRs.

Within quantitative studies, findings were reported across a whole range of demographic differences. Between studies, comparison could only be made between age range, levels of education, and ethnicity. We found conflicting findings in all three of these categories. We found evidence that both younger people and older people would favour sharing their data, that people with lower levels of education were both more and less likely to agree to sharing without consent, and that people of non-white ethnicity were both more and less likely to support EHRs and think of them as secure. For a full break down of the demographic results, see Table 3.

**Discussion**

We found that knowledge of the content and collection of electronic patient data was reasonably high, but knowledge about the secondary uses, such as data sharing for research, was low. Nevertheless, when asked, participants were generally willing to share their data for the “common good”, subject to safeguards. Willingness was qualified with concerns about privacy, and loss of control over personal information. Participants feared adverse outcomes less when they trusted both the motivation of research organisations to conduct research for the common good, and the competence of organisations to handle the data safely and without compromise. When evaluating opinions on consent mechanisms, findings suggested that educational and deliberative research into public opinion may provide different answers from snapshot surveys. Results suggested a range of mechanisms to increase public trust, and the overarching theme here was transparency of motivation, data handling and data flow.

Beyond this, however, further analysis reveals how closely this public rationale maps to, and can be interpreted by, some of the foundational moral principles which Beauchamp and Childress identify as paramount to governing biomedical practice. For instance, the included studies indicate that there is a widespread willingness to share EHRs for secondary purposes, in principle. This belief was held on the basis that, using and accessing data in such a way can bring about benefits which are in the interests of all individuals, or in other words, the “common good”. This strongly suggests that what forms the basis of this belief is a general expectation that if members of the public can contribute to the welfare of each other by sharing data, then they feel a moral obligation to do so. This can be interpreted as a more specific formulation of the principle of beneficence, which urges us to act, where we can, to promote good. Interestingly, this shows that the public feel that they, and not just health professionals and researchers, have a moral duty to uphold within medical practice, and give something back for the good of others where possible.

Nevertheless, willingness rarely led to unqualified support of the schemes designed to enable secondary use. This was withheld because, in practice, it was felt that key values would not, or could not, be ensured, thus bringing with it the risk of individual and collective harm. Two values in particular emerged as pertinent to conferring support: the first related to an organisation having the right expertise, or competencies, to ensure data security, and the second pertained to an organisation having the right motivations, i.e. those which serve the public, rather than private interests. A party’s failure to show that they will meet either or both of these values led to the erosion of trust in that party, and ultimately, in support for sharing health data in general. The public might feel justified in objecting to irresponsible, or insecure use of data because it is likely to cause individual harm; a direct violation of the principle of nonmaleficence. Similarly, the use of data for private gain may be said to be in violation of the principle of justice because it is generally unfair to exploit something for reasons other than what it was intended for. Finally, the use of patient data without consent may be seen to violate the principle of respect for autonomy. This is only a brief sketch of how such acceptable ethical principles are reflected in public opinion about sharing health data, however, what it demonstrates is the cogency of public ethical reasoning and its relationship to the Four Principles framework, which appears to be a constituent part of non-specialist thinking about the ethical practicalities of healthcare and medicine.

In concrete terms, this means that the public will not outright reject any research initiative, or project, on the sole basis that it requires the use of health data for purposes other than direct care. However, what will be demanded is that projects of this nature not only maintain a secure environment for health data, but also set research objectives which are primarily concerned with contributing to the common good. So long as these values are maintained, it is possible that the public may not object to companies outside the NHS accessing and using their data. In light of these findings, any researcher, regulator or company must ensure that the use of health data neither brings about risk of personal harm nor is conducted with the dominant aim of creating private profit. Consequently, any governmental safeguards that are instantiated to protect data must seek to preserve either of these principles as a matter of course.

**Strengths and limitations**

We conducted a wide search and sifted a huge number of papers, including grey literature reports. The search was challenging due to wide range of terms used within the literature for secondary data usage, and for expressing the concept of public opinion and attitudes. We cast a wide net and spending time excluding papers, and we therefore believe this review encompasses all available research meeting our criteria up until the search was conducted. Our findings were deliberately limited to UK and the Republic of Ireland to create a manageable, relevant and comparable body of literature. This enabled us to look for underlining principles for publics exposed to a particular type of healthcare system, but are obviously only applicable within these contexts. Synthesis of results was also challenging as there was a wide range of study types, using different methods. The small samples and low response rates in the majority of studies is also likely to have introduced bias into the findings, as it is
### Table 3. Study findings on Demographic Differences

<table>
<thead>
<tr>
<th>Group</th>
<th>Indicative of Negative Attitude</th>
<th>Indicative of Positive Attitude</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td>Compared to those aged 25–34, respondents between the ages of 35–64 were more likely to report they would be worried about the security of their records as part of a national EHR(^{36}).</td>
<td>Increase in age by each 10 year increment was significantly associated with an increased likelihood of reporting that any info can be provided to researchers without asking for consent(^{24}).</td>
</tr>
<tr>
<td></td>
<td>Compared to those aged 25–34, respondents over 35 years old were more likely to report less confidence in the ability of NHS security and were less likely to report that EHRs were equally or more secure than paper records(^{16}).</td>
<td>Older people (55–64, 65+) were more likely to find a drug company conducting research into the unwanted side effects of a drug using deidentified data to be more acceptable than younger people (16–24, 25–34, 35–44, 45–54)(^{33}).</td>
</tr>
<tr>
<td></td>
<td>Older people were increasingly more likely to report that they would not be in favour of a national EHR compared with 25–34 year olds(^{36}).</td>
<td>Those aged 55–64 tended to agree that research should be conducted by commercial organisations if there is a possibility of new treatments being discovered in comparison to 16–24s and 35–44s(^{33}).</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td>Respondents with lower educational qualifications were more likely to expect to be asked for explicit consent before their deidentified records were accessed(^{37}).</td>
<td>Compared with participants with higher degrees, individuals with no academic qualifications were less likely to say that they would worry about security if their record was part of a national EHR(^{36}).</td>
</tr>
<tr>
<td></td>
<td>Compared with completion of third level education, completion of only primary level education was associated with increased likelihood of reporting that any info can be provided to researchers without asking for consent(^{16}).</td>
<td>Compared with completion of third level education, completion of only primary level education was associated with increased likelihood of reporting that any info can be provided to researchers without asking for consent(^{24}).</td>
</tr>
<tr>
<td><strong>Socioeconomic Status</strong></td>
<td>Those of a lower socioeconomic status were more likely to be concerned about privacy(^{23}).</td>
<td>Those in the lower socioeconomic group DE (43%) were more likely to support companies using health data collected in the NHS to help target health products at different groups of people(^{33}).</td>
</tr>
<tr>
<td></td>
<td>Those in socioeconomic groups C2 and DE were less likely than those in AB and C1 to view the use of health data as having a potential benefit to society(^{39}).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Those in the lower socioeconomic group DE were less likely to say they trusted a variety of people with their health data; say that the advantages outweigh the disadvantages of using health data in research; and say that researcher can use data without prior consent than ABs(^{37}).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Those in socioeconomic groups C1 and C2 were less likely than ABs to allow their health data to be used(^{31}).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Those in socioeconomic groups DE (46%) were less likely to support commercial organisations to undertaking health research with health data than AB (62%)(^{38}). Those in socioeconomic groups DE (26%) were less likely to support commercial organisations to undertaking health research with health data than AB (30%)(^{35}).</td>
<td></td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td>Black British respondents were more likely to say they would not support the development of a national EHR system compared with White British respondents(^{26}).</td>
<td>Compared with White British groups White non-British, Asian, British Asian, Black-African, Caribbean, and British Black groups were more likely to say that EHRs are as secure, or more secure that paper records(^{36}).</td>
</tr>
<tr>
<td></td>
<td>Respondents identifying as belonging to an ethnic group other than White British were more likely to expect to be asked for explicit consent before their deidentified records were accessed(^{37}).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Those whose ethnicity was not White British were more likely to be concerned about the invasion of privacy(^{23}).</td>
<td></td>
</tr>
</tbody>
</table>
probable that only members of the public most interested in the issues consented to take part in the research. This means that each study likely represents a narrow range of views. It is not clear how this might have affected results across the whole range of studies, but it is likely that the themes and views represented here are not a complete picture of the public’s opinions. Additionally, certain research questions of particular interest were not asked of participants and therefore our understanding of public opinion is still limited. One example of this is whether the use of medical text (in contrast to structured data in medical records) elicits specific privacy concerns for the public.

Our analysis was informed and influenced by our respective backgrounds in philosophy, psychology and epidemiology. While attempting to be data-led, we must acknowledge that we may not have been wholly neutral in approach. However, our review highlights similar themes to Aitken et al., suggesting a cross-over with other syntheses in this area.

Future directions
Our research clearly demonstrates the relevance of the widely accepted Four Principles approach to public deliberation on EHR research. We recommend that frameworks for regulators and data custodians deciding on access to data for secondary purposes encourage consideration of the following questions:

1. Do the methods of data collection in the proposal respect individual patient autonomy? (Respect for Autonomy)

Given that public concern about data sharing is currently widespread, it is essential that individuals have the option to opt-out of any data collecting schemes like, for instance, Care.data. This is of particular importance while public understanding and transparency of data sharing is low, since these factors often contribute towards the overall feeling that there is not enough information to base decisions on. Giving individuals the right to opt-out, and ensuring such a process is as transparent and streamlined as possible is vital for maximising public trust in an initiative.

2. Could granting access to the data, or granting a particular use of the data, lead to individual harm? (Nonmaleficence)

Regarding the access and use of data for research purposes, data custodians and regulators must develop sophisticated methods to identify any underlying risk to individuals in submitted proposals. Possible risks to consider are articulated earlier in the results section of this paper. Mitigations of risk may include high standards for data storage security, restrictions on data linkage where necessary, and evaluation of analytical methods. Depending on the dataset in question, and those involved, parties may need to submit further applications if they wish to use data they have already been granted access to for new purposes.

3. Are the objectives and the intended outputs primarily concerned with contributing to the public good? Do they have clear scientific value? (Beneficence)

In order to maintain a minimum standard of justice, ethical bodies and regulators must also evaluate proposals on the basis of their intended aims and whether they significantly contribute towards the common good. As part of this process, such bodies must identify what private value may be unlocked for the party involved, and assess whether any reported objectives and intended outputs are reasonably achievable in the proposal.

4. Related to 3; is any agreement between the NHS and organisations providing analytics (private or public) fair and just? Will it lead to an outcome which will provide benefit to the key stakeholders of the data, i.e. the patients? (Justice)

The engagement of industry and private companies to provide data analytics will be crucial to maximise benefits from patient data in the future. However, agreements with industry must be transparent and critically, must be fair, representing benefit or gain for both parties, and for the public. Failure to achieve fairness or transparency in data-sharing agreements may result in a collective harm, that is, a loss of public trust in the endeavours of research.

Conclusions
We have shown that public thinking about the privacy issues around sharing patient data for research maps onto established biomedical ethical principles. These can be used to frame guidance for data custodians, regulators and researchers when planning or approving research using patient data.

Competing interests
The authors declare no competing interests.

Grant information
This work was supported by the Wellcome Trust [202133].

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.
References

26. Summary report of qualitative research into public attitudes to personal data and linking personal data. CM Insight, Welcome Trust. 2013. Reference Source
31. The Use of Personal Health Information in Medical Research General Public Consultation. Ipsos Mori, Medical Research Council. 2007. Reference Source
33. The one-way mirror: Public attitudes to commercial access to health data. Ipsos Mori, Welcome Trust. 2016. Reference Source
34. Wellcome Trust monitor report wave 3: Tracking public views on science and biomedical research. Ipsos Mori, Welcome Trust. 2016. Reference Source
On the whole this is a clearly presented systematic review (a copy edit is required as there are a few typos) and it reinforces the findings of a similar systematic review that I am co-author on, as the authors note in their conclusion. However, this review included quantitative studies and focused only on UK and Ireland, so the articles included do not fully overlap - the reviews were different in scope. So this is an additional contribution to the literature on public attitudes to data linkage and sharing for health research.

The process of the systematic review is delineated well and there is sufficient information on each included article for the reader to be able to access these and also to relate the findings of the review to those articles. The authors also cite some other relevant literature not included in the review. The authors are appropriately cautious in their interpretation of the findings from various studies, as these are often small scale, limited response rates etc.

I have a few concerns about the paper. The authors do not seem to be aware of existing governance structures, carefully developed alongside research on public attitudes and legal and ethical analyses. Health is a devolved matter in the UK. They need to read and make reference to the Scottish Government’s Data Linkage Framework (http://www.gov.scot/Topics/Statistics/datalinkageframework), the Guiding Principles for Data Linkage, and the terms of reference for the Public Benefits and Privacy Panel for Health and Social Care. Perhaps also look at the FARR Institute website to see how this major initiative is promoting safe use of health data for research purposes. There are some key reports that have not been identified by their search that are highly relevant: Public Acceptability of Cross-Sectoral Data Linkage (http://www.gov.scot/Publications/2012/08/9455); Public acceptability of data sharing between public, private and third sectors for research purposes (http://www.gov.scot/resource/0043/00435458.pdf); Aitken et al (2011)¹.

These would all help the authors craft more apposite recommendations.

A few other points – while supportive of an approach that identifies core principles, I’m not sure that Beauchamp and Childress’ four principles for biomedical research translate as easily as they suggest. I think some reference to emergent frameworks that speak to a social licence might be more compelling and the core principles that might underpin such a license. Public health ethics might help here.

A more minor point – the authors start by referring the medical data but really they are focussed on health data – a broader term. I also wonder why they use the term public opinion instead of attitudes. It may be that these terms are used differently in quantitative and qualitative research perhaps, but some
justification would be helpful. During the presentation of the findings of the review, they also refer to GP records, the Electronic Health Record, Cancer Registries – maybe clarify what type of records the studies they are referring to – or use the overarching term of EHR. The authors touch on the differences in the way in which public’s views are accessed and I think that point bears further elaboration.

References

Are the rationale for, and objectives of, the Systematic Review clearly stated?
Yes

Are sufficient details of the methods and analysis provided to allow replication by others?
Yes

Is the statistical analysis and its interpretation appropriate?
Not applicable

Are the conclusions drawn adequately supported by the results presented in the review?
Yes

Competing Interests: No competing interests were disclosed.

Referee Expertise: Sociology of health and illness, public engagement in health research and medical technologies

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.