Health Research Authority Response to the Nuffield Council on Bioethics Open Consultation on Biological and Health Data: The collection, linking, use and exploitation of biological and health data: ethical issues

1. Introduction

1.1 The Health Research Authority (HRA) was established in December 2011 in England to promote and protect the interests of patients and the public in health research. We strive, with partners, to make sure the UK is a great place for health research. Recognising that many members of the public want the opportunity to participate in research, we aim to ensure that health research involving them is ethically reviewed and approved, that they are provided with the information that they need to help them decide whether they wish to take part, and that their opportunity to do so is maximised by simplifying the processes by which high quality research is assessed. In doing this, we will help to build both public confidence and participation in health research, and so improve the nation's health.

1.2 In preparing this response the HRA has sought comments and advice from the National Research Ethics Advisors’ Panel (NREAP)\(^1\) the Confidentiality Advisory Group (CAG)\(^2\) and the Chairs of Research Ethics Committees flagged for the review of research databases.

1.3 The HRA, through its Research Ethics Committees (RECs) and the Confidentiality Advisory Group (CAG) have robust governance mechanisms in place to ensure that research involving biological and health data is consistent with applicable data legislation and ethical requirements. The Integrated Research Application System (IRAS) application form requires applicants to provide detailed information regarding, amongst other things: the use and storage of personal data (whether identifiable or anonymised) including how it will be stored; who has access and length of storage tailored to the type of research being undertaken.

2. HRA RESPONSE:

General Comments

2.1 Whilst the governance arrangements for the collection and use of biological and health data are relatively robust within the EU, the security and confidentiality of such data cannot be guaranteed once the data are made accessible internationally, e.g. in

---

1 NREAP is comprised of eight members including individuals with expertise in moral philosophy, research ethics committees, the pharmaceutical industry, patient and public involvement and clinical research. The panel is independent but hosted within the HRA and is a resource available to all RECs, funded by the UK Health Departments, within England and the devolved nations. The panel’s primary role is to help research ethics committees deliver robust, consistent and fair decisions through consultation with all stakeholders, including RECs.

2 CAG provides expert and independent advice to the HRA on access to confidential patient information for medical research purposes under Section 251 of the NHS Act 2006 and the Health Service (Control of Patient Information) Regulations 2002 in line with the Health Research Authority Directions 2013. This includes providing advice in relation to regulations 2, 3(4) and 5, in line with regulation 7. By agreement, the HRA’s CAG also provides advice to the Secretary of State for Health (SoS) via the Department of Health’s External Relations Directorate on all other non-research requests to access confidential patient information in relation to the SoS’s functions under the Health Service (Control of Patient Information) Regulations 2002.
response to the current drive for more trial data to be “open access”. The internationalisation of data presents many challenges for both research governance and informed consent especially in relation to how those data will be used and by whom.

2.2 It is important to bear in mind that the issues surrounding the use of biomedical data, and the acceptability of the uses to which such data may be put, cannot be adequately addressed by simply taking into account the majority view. Whilst it is likely that the majority of people are broadly in agreement with the optimisation and maximisation of the use of health data, some sections of society may feel more vulnerable to harm arising from wider access to linked data.

2.3 Likewise, whilst the linking of large quantities of data would appear to present a relatively benign opportunity to benefit health research and consequently the general public, it may be detrimental to specific research participants, such as those with mental health issues. Such individuals may need specific assurances and or enhanced protection concerning the use of their data.

2.4 As in any research, but more so in the context of large sets of linked data, the role of informed consent is crucial. Research participants’ understanding of key terms such as “anonymised”, “linked anonymised”, “unlinked” and “pseudonymised” is often poor and not assisted by the lack of commonly agreed definitions. Furthermore, truly informed, valid consent might be seen to be contingent upon a full understanding of all possible uses of an individual’s data. Clearly this will not always be possible when consent is given and it will therefore be important to be clear about whether broad consent is being asked for or more specific, limited consent. The very nature of ‘big data’ and open access to it will mean that it will be almost impossible to clearly state all of the future uses of the data. Thus, consideration might be given to the use of technology to invite the individual concerned to opt-in or out of specific proposed uses that arise where they are outside of the original consent. The use of technology to achieve this raises further important questions regarding access to such technologies that different social groups (class, age etc.) have, or will have in the future, and potential inequalities that might arise from unequal access.

2.5 We should not assume that the political and technological circumstances pertaining at the time of data collection and consent will persist throughout the lifetime of the data being held. We should also not underestimate the possibility that currently held anonymised data might be de-anonymised, through technological advances, in the future. Similarly, the current political landscape may undergo changes such that data

---

3 This diversity of terms was highlighted in the National Information Governance Board (NIGB) (now the Confidentiality Advisory Group) response to EU consultation on the legal framework for the fundamental right to protection of personal data (2009) (http://ec.europa.eu/justice/news/consulting_public/0003/contributions/publicAuthorities/national_information_governance_board_en.pdf):

“2) De-identification and handling linked anonymised data

Whilst having a broad definition of personal data is helpful in bringing more data within the remit of oversight bodies, the lack of agreed definitions of de-identified and linked anonymised or pseudonymised data, where data are de-identified to the recipient but not the original source, presents challenges.

A major challenge with de-identification techniques relate to maintaining the quality and utility of data whilst also protecting personal identities of the individuals to whom the original data relate. Clarifying standards related to de-identification and how pseudonymised/linked anonymised data should be handled would therefore be helpful. Similarly providing incentives to organisations to invest in privacy enhancing technologies and other technological solutions would also be helpful.”
not accessed by Government or other authorities currently, or data accessed solely for currently understood notions of the “common good”, may be used in the future for purposes we would not anticipate nor condone currently. For example, the current moratorium restricting insurance company access to the results of genetic testing could potentially change in the future. Thus, it may be considered necessary to forewarn participants of the possibility of such future risks.

2.6 The possibility of future unforeseen and unquantifiable risks resulting from sharing large data sets will need careful consideration and explanation to the individuals being asked to consent to the use of their data. Researchers need to be trained to explain and discuss these putative risks, and the risks themselves need to be widely debated in public fora.

2.7 The term 'personal data' has not been defined or included in the glossary included in the Nuffield Council's consultation document. There are a number of changes to definitions being made at present following the publication of 'The Information Governance Review Report' which includes a new definition of 'personal confidential data':

"... personal information about identified or identifiable individuals, which should be kept private or secret. For the purposes of this review ‘personal’ includes the Data Protection Act definition of personal data, but it is adapted to include dead as well as living people and ‘confidential’ includes both information ‘given in confidence’ and ‘that which is owed a duty of confidence’ and is adapted to include ‘sensitive’ as defined in the Data Protection Act."

'Personal data' is currently defined in the Data Protection Act (1998) as:

"...data which relate to a living individual who can be identified –
(a) from those data, or
(b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller,

and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual."

It would be helpful to clearly define the Council's understanding of this term as it applies to this consultation and the final report.

2.8 The current consultation document glossary defines “Confidentiality” as: “the restriction of access to certain data”. Two examples are given: "commercial confidentiality" and the "doctor patient relationship". The Council might consider that this should include reference to the common law duty of confidentiality as set out in the Department of Health document 'Confidentiality: NHS Code of Practice' (November 2003):

"Common Law of Confidentiality

---

6 http://www.ico.org.uk/for_organisations/data_protection/the_guide/key_definitions#personal-data
30. This is not codified in an Act of Parliament but built up from case law where practice has been established by individual judgements. The key principle is that information confided should not be used or disclosed further, except as originally understood by the confider, or with their subsequent permission.Whilst judgements have established that confidentiality can be breached ‘in the public interest’, these have centred on case-by-case consideration of exceptional circumstances. Confidentiality can also be overridden or set aside by legislation.

Consultation question 1: Do biomedical data have special significance?

2.9 Biomedical data do have special significance particularly when they relate to an identifiable individual. The sensitivity of that data will vary depending upon the context in which the data is collected, stored and used (and the view taken by the individual that the data refer to).

2.10 Biomedical data can be seen as distinct from other sources of data as they explicitly lay bare our vulnerabilities and weaknesses, potentially leaving us open to exploitation. Biomedical data can be used to predict the future health status of individuals and therefore has potentially serious implications for other areas of life, including access to financial services such as insurance.

2.11 What might be considered to be “biomedical data” may become difficult to define, as more and more data are collected in non-clinical settings. For example, it is possible that purchases logged by supermarket loyalty cards can be used to infer “biomedical” information such as whether an individual is pregnant or diabetic. The use to which commercial organisations will put such data and the confidentiality that might be expected by the individual to which these data apply (notwithstanding the ‘small print’ of any agreement entered into) may be very different to the expectations and governance that apply within the NHS.

2.12 The extent to which genomic data sets should be regarded as belonging to one individual, and to what extent other interests should be recognised is a question that is routinely encountered and handled in the clinical setting where diagnostic genetic testing is performed. In this setting, genetic testing will clearly have implications for family members but there is a clearer presumption of benefit for other family members in being diagnosed than there is in the context of research. The idea that a genomic data set does not belong to a single individual is reasonably well established within the clinical sphere; but what is more important is how those data are used. Such use, clinically and in the research context, is subject to regulation. However, where data are shared in large data sets, the subsequent use of those data is difficult to manage and may encompass diseases beyond those that were the subject of the initial investigation or research question.

2.13 Researchers do not always understand how to effectively de-identify biomedical data. It would be helpful to provide greater clarity around practical and effective de-identification options for biomedical data.

2.14 There will often be uncertainty surrounding the significance that identifiable data may have for individuals, both currently and in the future, and this will have implications for the information-giving and consent process. For example, where someone gives consent for their tissue sample to be analysed for inclusion within a genome database, consideration should be given regarding the implications for consent and what should happen in the event that the (linked) data:
• show that they possess a gene related to a potentially fatal or other serious condition that is untreatable?; or
• show that they possess a gene related to a potentially fatal or other serious condition that is potentially treatable? (e.g. gene for familial breast cancer treatable by double mastectomy or intensive surveillance); or
• reveal genetic information, that, with current scientific knowledge, suggests potential but uncertain significance for their health.

2.15 Paternity issues can also arise in genomic studies which involve family members sharing genomic sequences. This raises challenging ethical issues for the researchers in terms of containing this information.

2.16 It is important to discuss such issues more widely (and initiatives such as this Nuffield Council consultation make a valuable contribution to that), with a view to establishing a consensus. This will help avoid problems that might arise from the proliferation of such data sets whilst the ethical issues are being debated and decided locally and independently.

Consultation question 2: What are the new privacy issues?

2.17 The question of whether the privacy issues raised by ‘big data’ are new in kind or scale is very difficult, if not impossible, to address. Our ability to predict advances in technology or changes in political landscape that will affect the implications and the subsequent future risks/harms involved in the setting up and sharing of large data sets will always be limited.

2.18 There is a danger that our current state of knowledge may lead us to underestimate the possibility that data presumed to be anonymous now may no longer be anonymous in the future (through advancements in technology allowing the data to be de- anonymised)\(^7\).

2.19 Public confidence and trust in research and its governance is extremely important. If the increasing use of large data sets for research undermines that trust, this might lead to a lack of confidence and a subsequent reduction in the number of people willing to participate in research. This might also affect the public’s willingness to provide data for clinical purposes.

2.20 The growth of social networking websites and an increasing commercial market in the sharing of personal data has resulted in a marked divide between younger and older people in their willingness to share personal data (other than biomedical data) more generally. A recent survey\(^8\) showed that younger UK consumers (18 to 34-year-olds) are more comfortable providing personal information that will be used by commercial organisations to tailor future personal offers and communication. Younger people appear willing to accept a reduction in privacy and to share personal data where this

\(^7\) For example, future technological advances may be able to take MRI or CT data providing detailed facial bone structure and use this to digitally reconstruct the overlying soft tissue resulting in a recognisable face. This image might further be linked to a specific individual using face recognition software able to search the internet, including social media sites, for matches amongst uploaded photographs.

would result in personal gain. Such attitudes might facilitate a market for personal data involving a public with a greater awareness of the value of their personal data as a commodity to be bought and exchanged. This commodification of personal data, involving an increased appreciation of its commercial value, may have implications for future non-commercial use of biomedical and health data.

Consultation question 3: What is the impact of developments in data science and information technology?

2.21 Modern technology has made access to, and analysis of, very large amounts of patient data for research much easier than before, and is likely to increasingly facilitate cross-disciplinary research. Such research, conducted in accordance within established legal and governance frameworks, is essential for continued advances and improvements in patient care.

2.22 Paradoxically, advancements in technology have made security both easier and more difficult. The use of paper-based data storage systems do not allow for robust security and confidentiality, however, it is extremely difficult to extract meaningful data from paper patient records due to the large amount of time and effort required. In contrast, modern information systems make it possible to access and analyse data far more effectively.

Consultation question 4: What are the opportunities for, and the impacts of, use of linked biomedical data in research?

2.23 We know from the work recently undertaken by the Wellcome Trust that the public have a high level of trust in the way that their personal data is handled within the NHS and they can appreciate the benefits of linking aggregated data. They are fearful of data ‘falling into the wrong hands’ and suspicious of commercial gain. Recent public engagement work conducted by the HRA also indicates that the general public are suspicious of research funded by the pharmaceutical industry and usually unaware of the links between the pharmaceutical industry and the NHS. The Wellcome Trust report concludes that provided that the objective of linking health data is to increase knowledge about the causes and cures of ill-health, there is less public concern as to who carries this out.

2.24 The linking of NHS patient data, combined with the Office for National Statistics (ONS) death registration data, offers an important and unique opportunity to further scientific knowledge of the natural history of diseases, assessing the efficacy of treatment and care, undertaking audit etc. The joint NHS England and the Health & Social Care Information Centre's consultation on 'NHS Hospital Data and Datasets' identifies that an enhanced patient data set will allow researchers to "analyse patterns and trends across the country, develop more sophisticated analytical and predictive tools, and conduct more rigorous evaluations... Overall, we envision a virtuous cycle where richer

---

datasets and greater transparency will lead to greater participation and better care for all".

2.25 There may be potential public concern that if commercial, rather than academic or health interests, increasingly drive the linking and use of biomedical data this may lead to a culture of ‘profit’ rather than ‘cooperation’ in this area. The commodification of health data, however, may allow some patient groups to obtain more leverage with pharmaceutical companies when negotiating access to those groups’ health data, particularly where small patient groups are able to form alliances.

2.26 The more that data sharing transcends national boundaries the more difficult, if not impossible, it will be to effectively monitor and control.

**Consultation question 5: What are the opportunities for, and the impacts of, data linking in medical practice?**

2.27 The advancement of the argument that there may be a “moral obligation” to allow one’s data to be used in research could raise concerns that the importance of ‘altruism’ as a current driver of research participation\(^{12}\) may be diminished. In addition, there could be public concern that any unwillingness to share data in this way, in the face of an implied “obligation”, might adversely affect an individual’s treatment options. The promotion of any such argument, and mitigation of its attendant concerns, would need to be addressed through an effective communications strategy.

2.28 Increased data linking across the NHS may raise concerns around the quality control of data input and database validity.

**Consultation question 6: What are the opportunities for, and the impacts of, using biomedical data outside biomedical research and health care?**

2.29 One of the main challenges will be the sharing of health data outside of the standard clinical controls and governance arrangements of the NHS and public concern regarding the commercial use of their data.

**Consultation question 7: What legal and governance mechanisms might support the ethical linking and use of biomedical data?**

2.30 The HRA would like to draw the Council’s attention to the existing legal mechanism provided for by Section 60 of the Health and Social Care Act 2001 as re-enacted by Section 251 of the NHS Act 2006 which allows the Secretary of State for Health to make regulations to set aside the common law duty of confidentiality for defined medical purposes (including data linkages). The HRA took on responsibility for Section 251 in April 2013, establishing the Confidentiality Advisory Group (CAG). CAG provides expert and independent advice to the HRA on access to confidential patient information for medical research purposes as well as advice to the Secretary of State.

---

for Health via the Department of Health’s External Relations Directorate on all other non-research requests to access confidential patient information in relation to the SoS’s functions under the Health Service (Control of Patient Information) Regulations 2002.

2.31 Whilst the principle of “solidarity” will have some purchase when applied to public health systems such as the NHS, it has less relevance to the commercial sector. Thus, references to the “public good” as an incentive and justification for the collection, sharing and use of data cannot be relied upon where those data will be shared beyond the public health system with commercial companies.

2.32 The current principle of consent which allows a person to withdraw from a study, such as a clinical trial, also allows for data retention following their withdrawal (in the absence of a specific request for the removal of their data). Without data retention, there is a possibility that study results will be biased e.g. where participant drop out is due to side effects of the treatment. In addition, data will often be anonymised after it is linked, making it difficult or impossible for researchers to remove a specific individual’s data.

2.33 One of the best methods for ensuring that biomedical data are used ethically is to educate and inform both the public and researchers on the ethical and legal issues involved. This will facilitate discussion, negotiation and the use of biomedical data from an informed position. There is also an opportunity for greater patient and public involvement in this area. Working more closely with patients and the public at an early stage will ensure that their concerns are taken in account in the development of large biomedical datasets.

2.34 In relation to the use of opt-in or opt-out systems we believe there is a strong presumption in favour of an opt-in system when linking identifiable biomedical data. Increasing use of, and access to, technology will facilitate the seeking of opt-in consent for future use of an individual’s data more easily (where existing consent does not cover the intended use). Such consent respects the rights and autonomy of the individual to control their data and engenders confidence that data will not be used inappropriately leading to greater public support for the use of their biomedical data.

---

13 The HRA provide a number of training days that are open to Research Ethics Committee members, researchers and R&D managers including the one day training workshop: “Personal Data in Research”.

14 The HRA are committed to public involvement and have recently published the “HRA Strategy for Public Involvement” (16/09/13) which details three specific objectives: 1. Develop the HRA into an effective “involving” organisation; 2. Embed public involvement into the core business of the HRA and 3. Develop the role of the HRA with its partners to support the spread of public involvement in health research. http://www.hra.nhs.uk/documents/2013/10/hra-public-involvement-strategy-circulation-september-2013.pdf