

DIGITAL TECHNOLOGY Big data and the NHS – we have the technology, but we need patient and professional engagement

Authors: Neil R Lawrence^A and Stephen H Bradley^B

ABSTRACT

The potential promise of ‘big data’ in the NHS has not been overhyped. However, cyber security and linkage attacks remain ongoing concerns, and previously damaging projects, such as care.data and the Royal Free’s collaboration with google DeepMind, have raised concern among patient groups. We must use technology itself to minimise these risks, while publicising the good news stories and the positive case for using patient data in research. ‘#Datasaveslives’ is a national campaign, launched by the Farr Institute in 2014, that aims to spread the importance of our duty to share patient data for the benefits of health outcomes. 2018 is an incredibly important year for the future of data sharing. We urge our colleagues to join the campaign by sharing stories online with #Datasaveslives to promote how learning from patient data helps improve healthcare for us all.

KEYWORDS: Data, research, Caldicott, #datasaveslives, information

Introduction

Artificial Intelligence and ‘big data’ analytics of healthcare data have been placed at the centre of the UK government’s life sciences industrial strategy.¹ For once, the potential promised by these technologies does not appear to be overhyped.² The UK has an extraordinary competitive advantage, with the NHS, a global leader, holding vast repositories of healthcare data. Yet, although wonderful things are possible with the technology we already have, huge hurdles remain around governance frameworks, establishing procedures for gaining access to data, and the fundamental acquiescence of society to trade privacy risks for improved medical knowhow. Accessing healthcare data for researchers can be time consuming, often requiring duplicative permission processes from multiple organisations. Combining data about the same patient from two or more discrete collections (‘linking’ data sets) promises some of the biggest rewards, but remains a laborious process.

Barriers to progress

If the rewards are so promising, why is regulation so restrictive? Cybersecurity is a common patient worry, unsurprisingly given the

widespread coverage and ongoing discussion of the WannaCry ransomware attack. Although cybersecurity will always require large defensive resources, the more unwieldy threat is from ‘linkage attacks’.³ Innovative techniques have the potential to piece together linked data sets with varying levels of accuracy. Although the separate data sets might have had obvious identifying demographic data removed (pseudonymised; Fig 1), the bespoke nature of a patient’s attributes might still make them identifiable should those data be combined with data from another source. A linkage attack attempts to reidentify individuals from ‘pseudonymised’ data sets and, in some cases, might even be done by mistake. Aggregation of data makes linkage attacks more difficult, but reduces data utility.

Technology itself could be the answer to linkage attacks. If we can hold data within a central secure repository, with an intelligent program that can directly question the data set and provide an aggregated answer, or reveal risk factors or advantageous treatments, we might find a solution. A similar system could even work with a distributed network of systems provided they interoperate, negating the need for a central repository. Patient confidence might increase with the knowledge that researchers simply have to consider the research questions and apply the answers, without ever having access to patient-identifiable data. However, the processes and safeguarding that would have to be in place for such a black box methodology to be accepted by the research community are difficult to conceptualise.

It is important to recognise that patient surveys and feedback consistently show that patients are broadly in favour of their data being used for research to improve care. This is among a series of data protection controversies, to include learning from care.data⁵ through to DeepMind’s work on acute kidney injury.⁶ This approval always comes with various riders, with reassurance required over data sharing with private companies and cybersecurity concerns as recurrent themes.⁷

Winning hearts and minds

To make the case for data sharing, we need to start publicising the good news stories and the positive case for using patient data in research. We know that the perspective of individuals on the use of their data shifts when they become patients, those with chronic diseases finding it easier to understand the advantages of data sharing.⁶ A greater understanding of the rationale for using data for improving patient outcomes is associated with greater acceptance and support for

Authors: ^Aclinical fellow, NHS Digital, UK; ^BGP and clinical research fellow, University of Leeds, Leeds, UK

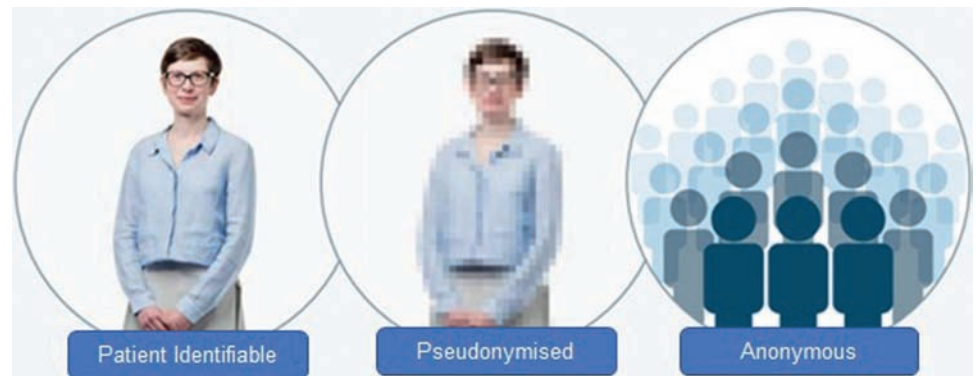


Fig 1. The spectrum of identifiability.⁴

its use.⁸ Therefore, patients themselves are our best ally. '#Datasaveslives' is a national campaign, launched by the Farr Institute in 2014, that aims to spread the importance of our duty to share patient data for the benefits of health outcomes. Patient support has materialised through groups such as 'Use My Data', who are progressively advocating for the use of patient data by commercial companies with the appropriate controls in place. Indeed, companies such as Optum have revealed amazing insights using large American health data sets, and make their services available to the NHS. We can learn much from the National Institute for Health and Care Excellence (NICE), which is among the world leaders in patient involvement when producing guidelines. The more we routinely involve patients in research ethics applications and study design, the more the acceptability of using patient data is likely to permeate the population.

Following Dame Caldicott's review and the national consultation that revealed public and professional approval of the recommendations,⁹ the opt-out process is being streamlined by NHS Digital. Rather than having to write to a GP to opt out of secondary uses of their data, patients will be able to use a simple website from May 2018.¹⁰ Yet both professional and patient levels of understanding around this remain low, resulting in the risk that patients will default to the risk-averse behaviour of opting out of sharing their data, without realising the opportunity cost.

Conclusions

2018 is an incredibly important year for the future of data sharing. Join the campaign by sharing stories online with #Datasaveslives to promote how learning from patient data helps improve healthcare for us all. We urge readers to view the inspiring videos of health data sharing at the Farr Institute website.¹¹ ■

References

- 1 *Life Sciences Industrial Strategy – A report to the Government from the life sciences sector*. www.gov.uk/government/uploads/system/uploads/attachment_data/file/650447/LifeSciencesIndustrialStrategy_acc2.pdf [Accessed 23 July 2018].
- 2 Armstrong S. Data, data everywhere: the challenges of personalised medicine. *BMJ* 2017;359:j4546.
- 3 *Protecting against linkage attacks that use 'anonymous data'*. www.marklogic.com/blog/protecting-linkage-attacks-use-anonymous-data/ [Accessed 23 July 2018].
- 4 *New words and pictures to explain anonymisation*. <https://understandingpatientdata.org.uk/news/new-words-and-pictures-explain-anonymisation> [Accessed 23 July 2018].
- 5 Trigg N. *Care data: how did it go so wrong?* www.bbc.co.uk/news/health-26259101 [Accessed 23 July 2018].
- 6 *NHS digital strategy consultation*. www.youngpeopleshealth.org.uk/wp-content/uploads/2016/05/NHS-digital-consultation-April-16-final.pdf [Accessed 23 July 2018].
- 7 *Public attitudes to commercial access to health data*. <https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-summary-wellcome-mar16.pdf> [Accessed 23 July 2018].
- 8 *Citizens' jury 2017*. www.connectedhealthcities.org/get-involved/citizens-juries/ [Accessed 23 July 2018].
- 9 *Your data: better security, better choice, better care*. www.gov.uk/government/uploads/system/uploads/attachment_data/file/627493/Your_data_better_security_better_choice_better_care_government_response.pdf [Accessed 23 July 2018].
- 10 *Government confirms national opt-out for patient data*. www.digitalhealth.net/2017/07/government-confirms-national-opt-out-for-patient-data/ [Accessed 23 July 2018].
- 11 www.farrinstitute.org/public-engagement-involvement/datasaveslives [Accessed 23 July 2018].

Address for correspondence: Dr Neil Lawrence, NHS Digital, Data and Integration, 1 Trevelyan Square, Boar Lane, Leeds LS1 6AE, UK.

Email: neil.lawrence2@nhs.net