Using routinely collected patient data with and without consent: trust and professionalism

Simon de Lusignan MBBS MSc MD (Res) FRCGP FBCS CITP
Editor, Informatics in Primary Care and Reader in General Practice and Biomedical Informatics Division of Community Health Sciences St George’s – University of London, UK

Introduction

The founding fathers of academic primary care conducted research using routinely collected practice data. William Pickles’ description of infectious disease, Frans Huygen’s Families with their Illness, and John Fry’s Common Morbidity were produced in an era of paper data collection and provide examples of general practice research that changed the face of medicine.1

The development primary care informatics as disciplines has largely come from the development of general practice computer systems which provide access to routinely collected data.2

Patient data has always been used by health services and researchers for some purposes without consent. However, the boundaries for access and use are not clearly defined and this is an issue for the public, health services and researchers. The lack of clear limits to access and use of data within the health service may have contributed to the creation of ‘The Big Opt out’, a campaign to help people opt out of having summary health data about them shared across the NHS computer system.3 Notwithstanding the reluctance of some people to share data across the NHS, a recent survey suggests that a majority of people are happy for their personal health data to be used in research.4

In the UK, further consultations about how to use routinely collected data are currently underway. Connecting for Health has just closed a consultation on the use of routinely collected data. This consultation is due to report in February.5 Also, the Wellcome Trust has recently published a consensus statement which sets out to define appropriate boundaries for researchers using primary care data. The report provides an excellent statement of principles for researchers; though perhaps inevitably does not define or address some of the informatics principles which will be needed to underpin its implementation.6,7

This edition of the journal contains three papers which collectively describe and define our current dilemmas and suggest way that we might remedy the current situation.

Solution 1: Customised calculation of the risk of breaching privacy

Navarro argues that we must move from a situation where we assume that all data and risks of privacy breaches are equal to one where we assess the risk breach according to individual circumstances.8 He tells us that fundamentally two things affect the risk of a breach of privacy: the nature of the data – essentially its granularity; and how much we can trust the person who has access to it.

The nature of the data

Some data are so non-specific that minimal security is needed; while others are much more vulnerable. He sees ‘Inference attack’ – where information security is breached by piecing together pieces of information available at a low security level – as the main risk to personal privacy. This problem has been widely reported and accepted within bioinformatics for some time, yet this learning has not been assimilated into the wider informatics or research communities.9,10 For example, in the recent Connecting for Health public consultation a fictitious anonymised medical record is created for ‘Pam’ (Figure 1). This record gives the precise dates of six items of medical history and one operation. A major diagnosis is labelled by year of onset, and another surgical procedure by year. Precise height and weight are given, along with smoking habit, first part of postcode, county and town. A public figure might have one or more of these pieces of information in the public domain and knowledge of these (e.g. from local press reports) may provide access to others which...
that individual might not wish to disclose. In this fictitious case Pam may have not have disclosed her depression. It is hard to see why so many precise pieces of contextual information are needed?

**Trust**

Trust is hard to measure. However, the importance of trust resonates with many of us who work in primary care. It is self-evident that some individuals and groups or organisations are clearly more trustworthy than others; and this same principle apply in information security. There maybe less risk of a privacy breach from sharing real records with a senior researcher than from allowing a technician to download a complete data base using an ‘approved’ process. Emerging professionalism in informatics should help engender greater trust. The creation of a UK Council for Health Informatics Professionals (UK CHIP) and of improved career pathways in the Swindell’s report represents progress.

In summary, Navarro would have us move from a policy of blanket measure to one in which we calculate ‘privacy breach risk’. Risk assessments carried out as part of normal health and safety processes look at risk as a simple formula:

\[
\text{Risk} = \text{Hazard (the danger)} \times \text{Likelihood (the chance of it happening)}
\]

Navarro sees:

Privacy risk breach = Granularity (Nature) of the data \times The Trustworthiness of the Individual.

**Solution 2: An ethically sustainable integrated approach to privacy using health cards**

Neame, in the second paper in this edition paints a rather different picture. He speaks from the experience of someone who has worked extensively implementing healthcare IT information security in New Zealand. The paper is in two parts: (1) the legal framework and principles; (2) using health cards.

**Legal framework**

The first part of his paper deals with the European legal framework which defines privacy and the potential legal risks if it is breached. Neame explains that research and audit are secondary uses of data and therefore if data are identified they require the informed consent of the patient. Where de-identification has occurred in a way that re-identification is impossible then consent is not required. Neame lists four principles which need to be met:

- patients should have control of access to their data
- only de-identified data are needed for most secondary purposes
- where data are needed: seek consent and make minimal disclosure
- consent and permission should be simple for patients and professionals.

He also explains how the context and content of the record can be separated and that most of the risks
associated with context are already known. Among the contextual items he feels it is essential to remove some which are included in the example anonymised record for Pam Smith (Figure 1):

- biometric identifiers (possibly including Pam’s exact height and weight)
- all dates (there are seven precise dates recorded on Pam’s example anonymised record and three items listed by year)
- geographical indicators which embrace less than 20,000 people (we don’t know the size of Pam’s town).

This policy would systematically remove many of the risks identified by Navarro; though both papers identify the same types of data being vulnerable to inference attack.

**Using health cards**

The health card is presented as the logical way of providing patients with control. Patients can even be allowed more than one ‘identity’ if they wish to keep certain items confidential from one provider.

In summary, using identifiable data for secondary purposes without consent is highly risky and likely to be illegal. Clear principles give patients much greater control of access to their records and separate context from content and remove much of the context which lend records vulnerable to inference attack.

**Consensus statement on using routinely collected data for research**

The final paper dealing with privacy issues is a workshop report from the UK’s Society for Academic Primary Care (SAPC) annual scientific meeting. Routinely collected primary care data have been described as ‘Goldmines for research’, yet many of us involved in research find it time consuming to get through the approval process.

This paper juxtaposes a summary of the Wellcome consensus report described in the introduction and the outputs of a workshop held at the conference. It contains practical suggestions for using technology, and developing the right contractual framework with practices. Its recommendations include the importance of keeping the GP as the (trusted) advocate for their patients’ advocate.

**Summary**

We are all committed to keeping health data private; and whilst Navarro and Neame have completely different models their very different approaches both seek to mitigate the same risks. Both authors make a valuable contribution to this current debate. Navarro suggest we should measure ‘privacy risk breach’ and to present a simple theoretical framework for its assessment. Neame urges us to think carefully and separately about the context and content of data and to adopt strict rules for sharing context. He also presents a much more patient-centred framework for sharing health data, and health card controlled data access.

Hinds presents a model whereby researchers can continue to utilise these data goldmines – something the public currently support. However, this is not support we should take for granted and active engagement of the public in this research and the past and current consultations are to be welcomed in this context.

Emerging professionalism in informatics should help develop better definitions and core generalisable theory about risks to patients’ privacy about how to mitigate them; but more importantly informaticians who can be trusted. Further research is needed to explore how to measure the risks and consequences of privacy breach. However, if sharing of health data is blocked then patients may continue to suffer because information known about them in one part of the health system is not made available in another; and research which might improve patient care will not be done. Informatics groups should not allow themselves the luxury of adopting an unconstructive critical approach – but instead should either support current consensus statements or come up with specific proposals of how these might be improved.

**REFERENCES**

3 The Big Opt Out. www.thebigoptout.com/
5 NHS Connecting for Health. *Consultation on Public, Patients, and Other Interested Parties Views on Additional Uses of Patient Data*. Report to be published at www.connectingforhealth.nhs.uk/susconsultation/
6 Wellcome Trust. *Towards Consensus for Best Practice: the use of patient records for research in general practice*. www.wellcome.ac.uk/GPrecords

ADDRESS FOR CORRESPONDENCE
Simon de Lusignan
Reader in General Practice and Biomedical Informatics
St George’s – University of London
London SW17 ORE
UK
Tel: +44 (0)20 8725 5661
Email: slusigna@sgul.ac.uk