INFORMATICS IN PRIMARY CARE

Leading article

Effective pseudonymisation and explicit statements of public interest to ensure the benefits of sharing health data for research, quality improvement and health service management outweigh the risks

Simon de Lusignan

Editor in Chief – Informatics in Primary Care
Professor of Primary Care and Clinical Informatics, Chair and Head of Department of Health Care Management and Policy, University of Surrey.

Cite this article: de Lusignan S. Effective pseudonymisation and explicit statements of public interest to ensure the benefits of sharing health data for research, quality improvement and health service management outweigh the risks. *Inform Prim Care*. 2014;21(2):61–63.

http://dx.doi.org/10.14236/jhi.v21i2.68

Author address for correspondence:
Simon de Lusignan
Editor in Chief – Informatics in Primary Care
Professor of Primary Care and Clinical Informatics, Chair and Head of Department of Health Care Management and Policy, University of Surrey.

Accepted May 2014

ABSTRACT

This journal strongly supports the sharing of data to support research and quality improvement. However, this needs to be done in a way that ensures the benefits vastly outweigh the risks, and vitally using methods which are inspire both public and professional confidences – robust pseudonymisation is needed to achieve this. The case for using routine data for research has already been well made and probably also for quality improvement; however, clearer mechanisms are needed of how we test that the public interest is served. Ensuring that the public interest is served is essential if we are to maintain patients' and public's trust, especially in the English National Health Service where the *realpolitik* is that patients can opt out of data sharing.

SHARING HEALTH INFORMATION HAS BENEFITS AND THERE ARE INTERNATIONAL STANDARDS FOR PSEUDONYMISATION

Sharing personal health information, in a way that keeps it private, has long been a tradition in the United Kingdom, with these data supporting a wide research and quality improvement agenda. The use of data in this way underpins our understanding of health and disease, particularly in primary care.

> *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.*

Use of these data has provided opportunities and at the same time challenges.[1] Whilst there is always scope to improve the way these data are used, we also need to improve the way they are kept private, as the scope for inference attack and other ways of potentially reidentifying individuals becomes easier with advances in technology.[2,3]

Pseudonymisation, the process by which data have personal details removed and substituted by other flags, provides a mechanism by which privacy can be maintained, at the same time as allowing linkage of data to other key information. Pseudonymisation can be conducted using an International Standards Organisation (ISO) defined process. This process was defined by its Technical Committee (TC) on Health Informatics (ISO/TC 215). This TC has a number

of working groups (WGs), including WG4, which looks at security. WG4 produced a number of technical specifications (TSs), including ISO/TS 25237:2008 on pseudonymisation.[4] This standard

- sets out the basic concept of pseudonymisation;
- provides different use cases for pseudonymisation that can be both reversible and irreversible;
- describes a method for pseudonymisation services both organisational and technical;
- gives a guide of how to conduct a risk assessment for reidentification;
- specifies a policy framework and minimal requirements for
  - the trustworthy practice of a pseudonymisation service;
  - controlled re-identification;
- specifies interfaces for the interoperability of services.[5]

More recent developments in pseudonymisation have included how to give patients control[6] and enhanced methods to underpin translational research.[7] The technical requirements for providing such a solution are well described.[8] There have been recent efforts to create an open pseudonymisation standard for the National Health Service (NHS) that would enable pseudonymisation at source.[9]

MAINTAINING PUBLIC TRUST AND OPTING OUT

Trust and public faith in the professionalism of health service managers are essential and cannot be taken for granted.[10] The provision of the Data Protection Act (DPA), which in turn is based on a European treaty, provides principles that ensure that there is only proper use of data.[11] The key principles of the DPA classify those who handle personal data as

- 'data controllers' – who hold data, such as general practices and hospitals;
- 'data processors' – who process personal data on behalf of the data controller;
- neither – some organisations involved with data may neither be a data controller or processor.

All organisations that hold personal data must be registered under the provisions of the DPA, and they must also adhere to its principles. These include responding to 'subject access requests' and ensuring that processing of data is fair, lawful, purposeful, not excessive, kept secure and not sent overseas. Health organisations are not exempt from these provisions and compliance helps to ensure public and patient trust.

Concerns, primarily in England, about the use of data within the NHS may have the potential to result in many individuals opting out of allowing their data to be shared. There appears to be a *realpolitik* that individuals will have the right to opt out of having their data uploaded. There have been number of objections raised about the use of health data by the English NHS and a lack of awareness of any information being provided to the public.[12] Should significant numbers opt out then this would seriously undermine the representativeness of these data?

Currently, most data sources have a review process prior to data being made available for research or other quality improvement processes. The same standards of review need to be applied when data are to be used outside the health care system. This should include a clear statement of the benefits, purpose and any risks.

Overall, the impression of the current care.data scheme, in the English NHS, is one that lacks the confidence of much of the public and the profession. This article suggests that the way to restore this is to need to be explicit about the purpose for which data are going to be used, be clear about the public interest and move to pseudonymisation of data at source.[13]

SUMMARY – WHAT SHOULD BE DONE?

It is important that the provisions of the DPA are strictly adhered to. Whilst there is no suggestion that these are being breached, the development of a clearer way of developing the public interest case would help to ensure that key DPA principles such as purpose, adequacy and not being excessive have been carefully thought through. The review process that most data sources go through prior to releasing data should explicitly include this function.

The technical approach to data extraction needs to have the confidence of informed professionals. Several groups are arguing that the way forward should be one of pseudonymisation at source, as this provides the best possible mechanism for ensuring privacy. Research and demonstration projects to establish or refute this notion need to be urgently put in place.

Finally, we should not miss out on the potential benefits of research and quality improvement that can be derived from routine data. There is potential to improve patient safety: and address major quality problems using routine data. The rogue GP Howard Shipman could have been flagged,[14] and the Mid-Staffordshire Trust can was identified sooner from routine data; had we had screening mechanisms in place.[15]

The benefits for research, quality improvement and patient safety from the processing of routine data should not be lost as a result of overstatement of the risks. However, we need to be much more explicit about the benefits to the public of processing such data and implement methods of pseudonymisation that minimise any risk of privacy breeches, and have the confidence of experts in our domain.

REFERENCES

1. de Lusignan S and van Weel C. The use of routinely collected computer data for research in primary care: opportunities and challenges. *Family Practice* 2006;23(2):253–63. http://dx.doi.org/10.1093/fampra/cmi106. PMid:16368704.
2. Navarro R. An ethical framework for sharing patient data without consent. *Informatics in Primary Care* 2008;16(4):257–62. PMid:19192326.
3. Claerhout B and DeMoor GJ. Privacy protection for clinical and genomic data. The use of privacy-enhancing techniques in medicine. *International Journal of Medical Informatics* 2005;74(2–4):257–65. http://dx.doi.org/10.1016/j.ijmedinf.2004.03.008. PMid:15694632.
4. de Meyer F, De Moor G and Reed-Fourquet L. Privacy protection through pseudonymisation in eHealth. *Studies in Health Technology and Informatics* 2008;141:111–18. PMid:18953131.
5. International Standards Organization (ISO). Technical Committee: ISO/TC 215 Health Informatics. Pseudonymization ISO/TS 25237:2008. URL: http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=42807&commid=54960
6. Neubauer T and Heurix J. A methodology for the pseudonymization of medical data. *International Journal of Medical Informatics* 2011;80(3):190–204. doi: 10.1016/j.ijmedinf.2010.10.016. http://dx.doi.org/10.1016/j.ijmedinf.2010.10.016. PMid:21075676.
7. Aamot H, Kohl CD, Richter D and Knaup-Gregori P. Pseudonymization of patient identifiers for translational research. *BMC Medical Informatics Decision Making* 2013;13:75. doi: 10.1186/1472-6947-13-75. http://dx.doi.org/10.1186/1472-6947-13-75. PMid:23883409; PMCid:PMC3733629.
8. Ganslandt T, Mate S, Helbing K, Sax U and Prokosch HU. Unlocking data for clinical research - the German i2b2 experience. *Applied Clinical Informatics* 2011;2(1):116–27. doi: 10.4338/ACI-2010-09-CR-0051. http://dx.doi.org/10.4338/ACI-2010-09-CR-0051. PMid:23616864; PMCid:PMC3631913.
9. University of Nottingham, QResearch. Open pseudonymiser. URL: http://www.openpseudonymisation.org/
10. de Lusignan S. Using routinely collected patient data with and without consent: trust and professionalism. *Informatics in Primary Care* 2008;16(4):251–4. PMid:19192325.
11. de Lusignan S, Chan T, Theadom A and Dhoul N. The roles of policy and professionalism in the protection of processed clinical data: a literature review. *International Journal of Medical Informatics* 2007;76(4):261–8. http://dx.doi.org/10.1016/j.ijmedinf.2005.11.003. PMid:16406791.
12. Hoeksma J. The NHS's care.data scheme: what are the risks to privacy? *BMJ* 2014;348:g1547. doi: 10.1136/bmj.g1547. http://dx.doi.org/10.1136/bmj.g1547
13. Hagger-Johnson GE, Harron K, Goldstein H, Parslow R, Dattani N, Borja MC et al. Making a hash of data: what risks to privacy does the NHS's care.data scheme pose? *BMJ* 2014;348:g2264. doi: 10.1136/bmj.g2264. http://dx.doi.org/10.1136/bmj.g2264. PMid:24667239.
14. Aylin P, Best N, Bottle A and Marshall C. Following Shipman: a pilot system for monitoring mortality rates in primary care. *Lancet* 2003;362:485–91. http://dx.doi.org/10.1016/S0140-6736(03)14077-9; http://dx.doi.org/10.1016/S0140-6736(03)14518-7.
15. Bottle A and Aylin P. Intelligent information: a national system for monitoring clinical performance. *Health Services Research* 2008;43(1 Pt 1):10–31. PubMed PMID: 18300370; PubMed Central PMCID: PMC2323144.