electronic Journal of Health Informatics
http://www.ejhi.net
2006; Vol 1 (1): e7

Research within the Privacy Regulations: Problems and Solutions for Database Custodians

Ea Mulligan[1], Wendy A. Rogers[2], Annette Braunack-Mayer[3]

[1]School of Law, Flinders University, Adelaide, SA, Australia
[2]School of Medicine, Flinders University, Adelaide, SA, Australia
[3]School of Public Health, Adelaide University, SA, Australia

Abstract

State and Federal legislation governing health information and privacy in Australia is complex and relatively untested, causing confusion amongst database custodians as to what conduct is required. Some database custodians believe that providing privacy will allay public anxiety and consequently support research. Others argue that data managers have become fearful of litigation and that this will restrict the access of researchers to data. Two of the significant ethical issues to be considered are the right to privacy, and whether using information poses a risk to data subjects. Data custodians have sought to address concerns about privacy in two main ways. The first is by seeking informed consent from those whose data is collected. There are significant, but not insurmountable, practical difficulties in seeking consent from large numbers of individuals. The second way of addressing privacy concerns has been through security measures which are designed to reduce risks to data subjects. These measures are often nominated as a response to privacy requirements; however, they do not necessarily offer the opportunity to consent to information disclosures. In this paper we present a review of the current literature on possible responses of database custodians to demands for increased privacy. We outline a variety of examples and responses, some of which have the effect of restricting research, while others are enabling research to proceed with greater privacy protection in place. We argue that finding ways to proceed with research while protecting privacy requires attention to a range of factors. There are challenges both in engaging populations about consent procedures, and in encouraging the use by researchers and health care professionals of technical solutions where these are available.

Keywords: Confidentiality, Privacy, Ethics, Research, Medical records

1 Introduction

Legislation expanding the control which individuals may exercise over the use and disclosure of personal (and especially health) information has placed new demands upon registries and other database custodians. New legislation includes amendment of the Commonwealth Privacy Act to apply to all health information in the private sector and introduction of health specific privacy legislation in a number of Australian states. These changes, in addition to pre existing common law protection for confidential information, have produced a complex legal environment in Australia which is relatively untested. Not surprisingly, health data custodians have been confused as to the most appropriate response. There has been considerable professional discussion concerning what conduct is required to comply with the various requirements of different statutes. The response of database custodians to demands for increased privacy protection has been mixed, while the ultimate impact on research productivity is not yet clear.

In this paper we acknowledge the problems which health data custodians confront in providing privacy protection, consider the ethical issues involved and review some of the responses of database custodians. The main responses have been to seek individual consent for the inclusion and use of data in registries and databases,

The *electronic Journal of Health Informatics* is an international journal committed to scholarly excellence and dedicated to the advancement of Health Informatics and information technology in healthcare. ISSN: 1446-4381

and to devise business rules which reduce the likelihood that clinical data can be traced back to the individuals who were its source.

2 Data protection may restrict or enable research

Some registry custodians support the practice of giving individuals the opportunity to consent, on the grounds that public confidence is necessary if research is to proceed. James Morrow has said of the UK Epilepsy and Pregnancy Register that while 'case ascertainment would probably be higher if informed consent was not necessary, this is an inevitable trade off to maintain patients' trust' (Morrow 2001). In their analysis of the statutory regulation of confidentiality and privacy affecting the National Cancer Institute Breast Cancer Surveillance Consortium, which has a centralised register of mammographic data in the United States, Carney et al point out that the public are increasingly aware of instances of misuse of health records. They argue that providing privacy guarantees is the antidote to public mistrust:

For research studies to gain the participation needed by the public, the confidentiality of research data must be honored and protected. Otherwise it will be impossible to conduct research (Carney, Geller et al. 2000 p 377).

Other health data custodians have voiced their concern at the impact which increasing privacy protection will have on research. A case report from custodians of a large data repository in the USA notes that

Recent deliberations over Health Insurance Portability and Accountability Act (HIPAA) requirements have cast a specter of concern over any secondary uses of data from electronic medical records systems (Moehr and Daniel 1998 p 394).

Custodians of medical data in the United Kingdom complain of confusion caused by the Data Protection Act (Stroble, Cave et al. 2000; Coleman, Evans et al. 2003). Database custodians are said to be "fearful of litigation if they allow any access for research or even audit without each patient's informed consent" (Peto and Fletcher 2004).

Researchers report that some custodians of health data have become "reluctant to continue supplying data to researchers" with the result that epidemiological research "is threatened by too rigid and inconsistent interpretation of the Data Protection Act" (Evans, McNaughton et al. 2001 p 672). There are reports of increasing difficulty in identifying suitable research subjects for recruitment (Redsell and Cheater 2001) and complaints that compliance with the Data Protection Act is expensive and time consuming, drains resources (Evans, McNaughton et al. 2001) and acts as an obstacle to research (Peto and Fletcher 2004).

3 The ethical issues

From an ethical perspective, there are three main issues at stake. One is the important benefit to society gained from medical research. Preservation of this public good may justify forgoing some other public or private benefits. This is the position taken by the National Health and Medical Research Council in its ethical guidelines on research (NHMRC 1999).

A second is the right to privacy, supported by legislation, which emphasises the importance of people being able to decide freely who should have access to their personal information. The right to privacy is recognised by researchers and custodians who accept that most patients wish to exercise control over whether their health information is used in research (Fletcher, Marriott et al. 2004; Billings, Kohn et al. 1992; Sankar, Mora et al. 2003).

The third ethical issue revolves around whether or not using information poses a risk of harm to those whose information is used. Some researchers doubt that there are real risks to data subjects from which they should be protected (Evans, McNaughton et al. 2001 p 672), while others note that there have been misuses of health information (Marwick 1996). In 1997 the *British Medical Journal* published two articles which defended the need for epidemiological research data gathering without patients' consent (Smith 1997; Doyal 1997). This stimulated many letters to the editor. Researchers argued that consent should not be dispensed with, but that instead research should be better designed (Pfeffer and Alderson 1997). Journalists pointed out that consent is not obtained in many research projects and presented findings that the US Food and Drug Administration had identified over 2000 instances of research proceeding in the absence of consent during its inspections of clinical trials (Epstein and Sloat 1997).

4 Responses to the demands of privacy protection

At present there are two main ways in which data custodians and researchers are responding to the increasing demands for privacy protection. The first is to seek informed consent from those whose data is subject to collection in a registry or other health database. Data subjects are given information about any intended uses and offered the opportunity to refuse inclusion of their data. This response meets ethical requirements as well as complying with legal requirements in many jurisdictions. The second response is to develop mechanisms for ensuring protection of privacy, on the grounds that privacy is not breached if the data are sufficiently de-identified and protected such that they can not be linked back to individuals. This response does not allow data subjects to control the uses of their data, but it does minimise the potential for individuals to be harmed.

5 Consent

Custodians of both research databases and clinical information systems have commented on the practical problems associated with seeking consent from large numbers of individuals. Database custodians have argued

that seeking consent from individuals for inclusion of health data in population research would result in unrepresentative sampling (AlShahi and Warlow 2001), and that it is impractical (Holman 2001), or undesirable (Evans, McNaughton et al. 2001) or both (Finkelstein 1999).

Despite the difficulties, some data managers have set themselves to the task of communicating with large numbers of people. For example, leaflets were used in the UK to inform local residents of a local record linkage project involving the records of three general practices, two healthcare trusts, an ambulance service and a social service. The leaflets were distributed to 88,000 households (approx 225,000 residents). Only 82 calls were received by the advertised information line, while the information website received 1306 hits in six months. Participants were not asked to consent, but were offered the opportunity to withdraw their data from the project. A total of 10 individuals asked for their records to be excluded (Adams, Budden et al. 2004).

There is some doubt, however, that household leafleting is an effective public education method. In a similar project in South Staffordshire involving a population of 60,000 households, follow up surveys were conducted with general practice patients and shoppers in the area. While there were no requests to opt out of this project, and a large majority (80%) felt comfortable with it, less than half were aware that their data had been networked and only 15% understood that they could opt out, despite the public education campaign. People were relatively uninterested in health data linkage, making it a challenge to ensure that they were made aware of these data linkage projects (Adams, Budden et al. 2004).

The Mayo Clinic's collection of over 5 million medical records have been used to provide data for tens of thousands of studies, and the Rochester Epidemiology Project which links these records with those from other treatment centers has supported more than a thousand publications describing the natural history of various diseases (Melton 1997). The advent of federal privacy legislation (the Health Insurance Portability and Accountability Act 2000) in the United States brought with it the requirement for individual patient consent and the Mayo Clinic sought this. At that time 94% of current Mayo Clinic patients (and 97% at a partner institution) agreed to provide written consent for the use of their records. This suggests two things; one, that a large proportion of patients will consent if given the opportunity; and two, that the logistical problems of obtaining consent from a large number of potential research participants can be overcome.

From these examples, it would appear that a large proportion of people in the general population (at least in the UK and the USA) will not object to health data linkage for clinical care and for research purposes.

The practical problems identified by database custodians are how to provide information to large numbers of people, how to obtain and record consent from them and how to provide effective opt out options for the few individuals who do refuse consent. Model protocols for transferring health and consent information between heath facilities which automatically protect transferred health information according to the patients' consent conditions are in development. An example is the eConsent model described by O'Keefe et al. This model captures patients' expressed wishes regarding who can access their information, attaches these to existing or new electronic records (or parts of the record), and applies the consent conditions to grant or deny access requests (O'Keefe, Greenfield et al. 2005).

Reducing the participation rate in large population based studies by omitting those who decline to consent or those who cannot provide consent (for example because they cannot be traced, or have died) can undermine the meaningfulness of epidemiological research. Researchers faced with the prospect of being required to seek individual consent to health data linkage have put forward examples of research designs which would be frustrated (Bruinsma, Venn et al. 2000). This finding is supported by other studies indicating that those who seek to opt out have different demographic profiles from those who consent to have their health data linked for research purposes (Woolf, Rothemich et al. 2000; Yawn, Yawn et al. 1998). For example, voluntary recruitment into the registry of the Canadian Stroke Network resulted in a seriously biased group and the assessment that this registry will never be able to provide a representative data set (Tu, Willison et al. 2004).

Some large population studies have, however, obtained consent and found few differences between those who consented and those who did not. The Norwegian Women and Cancer Study invited 179,388 women to complete their questionnaire. Response rates were only 45 to 60% (depending on age group) but there were few demographic differences between those who did and those who did not participate when they were compared using a national population register. This research group concluded that there was no major selection bias introduced by voluntary participation.

6 Obscuring links between individuals and data

An alternative to seeking informed consent is to develop systems that protect privacy during data linkage. Security measures which are designed to reduce risks to data subjects who have not been offered the opportunity to consent are often nominated as privacy protecting measures (Berman 2002; Evans, McNaughton et al. 2001).

Administrative procedures or business rules for record linkage processes offer a high level of security to the stored data, although reverse engineering can be employed to reassemble information into a form which can be identified with individuals (Sweeney 2002). One methodology involves the separation of personal identifiers from clinical information and their separate encryption by the reporting clinician, submission of these data to a "trusted third party" who allocates an identifier specific to the paired data items

and forwards personal identifiers and clinical data to separate repositories. Clinical data from disparate databases can only be linked by the "trusted third party" using the unique identifier which it allocated. Neither researchers nor the specific disease registers are provided with personal identifiers (Churches 2003).

This system requires legal protection and financial support from government and reflects many of the elements of the work of the Western Australian Data Linkage Unit (Kelman, Bass et al. 2002). In Western Australia the custodians of disease registries and health databases that contain both personal identifiers and clinical information sign memoranda of understanding which authorise a trusted third party (the Linkage Unit) to identify data concerning the same individual across different databases. The Linkage Unit allocates unique anonymous identifiers for each individual which are specific to the particular research project. The researcher then uses the unique identifier to obtain clinical data directly from each database and is the only person to hold the complete file of linked clinical data. Researchers hold clinical but not identifying data, while the linkage unit accesses personally identifying information but not clinical data.

Other protocols have been proposed which utilise public key cryptography to obscure identifying information during data linkage between two or more data bases (Agrawal, Evfimievski et al. 2003). While these protocols cannot protect against collusion between a researcher and a database to disclose identifying information, they do offer a high level of security. Protocols have also been developed which allow data linkage between databases without disclosure of identifying information to any party outside the originating data source, or which allow extraction of a cohort of data from a database without revealing membership of the cohort to the source database (O'Keefe, Gu et al. 2004).

All of these methodologies to obscure data minimise the risk that health information will be disclosed. Reducing risks to data subjects may enhance their confidence and address their concerns, but it does not extend control to them over the use and distribution of their data.

7 Conclusion

Database custodians confront problems as they comply with privacy and data protection requirements. They may respond in a variety of ways. Some responses have the effect of restricting research, other responses enable research to proceed with privacy protection in place. Responses revolve around either seeking consent, or using data management systems which make it difficult to trace individuals from data.

The technical solutions for protecting privacy are complex, which may be one reason why they are not always utilised. Information system developers have observed that functions such as user authentication and data transfer audit may be ignored by health care providers, even when these are available (Moehr and Daniel 1998).

In summary, there is unlikely to be a single solution to the challenges posed by performing data-based research with adequate privacy protections. Data custodians and researchers may need to tailor their responses to types of population or register, or to specific research needs. There are difficulties in engaging the population in debate about data uses, and in getting researchers and health care professionals to use technical solutions where these are available. Meeting these challenges to ensure high ethical standards in research involving databases and registers requires time and effort, but is necessary to maintain public support for research.

References

Adams, T., M. Budden, et al. (2004). "Lessons from the New Hampshire electronic health record pilot project: Issues of data protection and consent." *British Medical Journal* 328(7444): 871-4.

Agrawal, R., A. Evfimievski, et al. (2003). "Information Sharing Across Private Databases." *SIGMOD*: 86-97.

AlShahi, R. and C. P. Warlow (2001). "'No Consent Please-we're British' Will Observational Research in Neurology Survive Changes in Data protection Law and Confidentiality Guidance?" *Neurology, Neurosurgery & Psychiatry* 71(3): 424.

Berman, J. J. (2002). "Confidentiality Issues for Medical Data Miners." *Artif Intell Med* Sept- Oct 26(1-2): 25-36.

Billings, P., M. Kohn, et al. (1992). "Discrimination as a Consequence of Genetic Testing." *American Journal of Human Genetics* 50: 467 - 482.

Bruinsma, F., A. Venn, et al. (2000). "Accessing Patients' Records Without Individual Consent for Epidemiological Research." *Journal of Law and Medicine* 8: 76 - 80.

Carney, P., B. Geller, et al. (2000). "Current medicolegal and confidentiality issues in large mulitcenter research programs." *American Journal of Epidemiology* 152(4): 371-78.

Churches, T. (2003). "A proposed architecture and method of operation for improving the protection of privacy and confidentiality in disease registers." *BMC Medical Research Methodology* 3(1): 1-13.

Coleman, M., B. Evans, et al. (2003). "Confidentiality and the public interest in medical research - will we ever get it right?" *Clinical Medicine* 3(3): 219-28.

Doyal, L. (1997). "Journals Should Not Publish Research to Which Patients Have Not Given Fully Informed Consent - With Three Exceptions." *British Medical Journal* 314: 1107 - 11.

Epstein, K. and B. Sloat (1997). "Informed Consent is not Always Obtained in the United States." *British Medical Journal* 315(7102): 247.

Evans, J., D. McNaughton, et al. (2001). "Pharmacoepidemiological research at the Medicines Monitoring Unit, Scotland." *Pharmacoepidemiology and Drug Safety* 10: 669-73.

Finkelstein, M. (1999). "Does the CMA's Privacy Code go too far? Or far enough?" *Canadian Medical Association Journal* 160(6): 781.

Fletcher, J., J. Marriott, et al. (2004). "Data protection, informed consent, and research: Interpretation of legislation should reflect patients' views." *British Medical Journal* 328(7453): 1437.

Holman, C. D. (2001). "The impractical

nature of consent for research use of linked administrative health records." *Australian and New Zealand Journal of Public Health* 25(5): 421-22.

Kelman, C. W., A. J. Bass, et al. (2002). "Research Use of Linked Health Data - A Best Practice Protocol." *Australian and New Zealand Journal of Public Health* 26(3): 251-255.

Marwick, C. (1996). "Increasing Use of Computerised Record Keeping Leads to Legislative Proposals for Medical Privacy." *JAMA* 276: 270-72.

Melton, J. (1997). "The Threat to Medical - Records Research." *New England Journal of Medicine* 337(20): 1466 - 1470.

Moehr, J. R. and J. G. Daniel (1998). "Adoption of security and confidentiality features in an operational community health information network: the Comox Valley experience - case example." *International Journal of Medical Informatics* 49: 81-87.

Morrow, J. (2001). "Data Protection and Patient's Consent: Informed Consent Should be Sought Before Data are Used by Registries." *British Medical Journal* 322(7285): 549-550.

National Health and Medical Research Council. (1999). *National Statement on Ethical Conduct in Research Involving Humans*. Commonwealth of Australia, Canberra.

O'Keefe, C., P. Greenfield, et al. (2005). "A Decentralised Approach to Electronic Consent and Health Information Access Control." *Journal of Research and Practice in Information Technology* 37(2): 161-78.

O'Keefe, C., L. Gu, et al. (2004). *Privacy-Preserving Data Linkage Protocols*. WPES ACM Workshop on Privacy, Washington DC.

Peto, J. and O. Fletcher (2004). "Data Protection, Informed Consent, and Research." *British Medical Journal* 328(7447): 1029-30.

Pfeffer, N. and P. Alderson (1997). "The Central Problem is Often Poor Design and Conduct of Trials." *British Medical Journal* 315: 247.

Redsell, S. and F. Cheater (2001). "The data Protection Act (1998): implications for health researchers." *Journal of Advanced Nursing* 35(4): 508-13.

Sankar, P., S. Mora, et al. (2003). "Patient Perspectives on medical Confidentiality; a review of the literature." *Journal of Internal Medicine* 18(8): 659.

Smith, R. (1997). "Informed Consent; the Intricacies." *British Medical Journal* 314: 1059-60.

Stroble, J., E. Cave, et al. (2000). "Data protection legislation: Interpretation and barriers to research." *British Medical Journal* 321(7265): 890-92.

Sweeney, L. (2002). "k-Anonymity: A Model for Protecting Privacy." *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10(5): 557-70.

Tu, J., D. Willison, et al. (2004). "Impracticality of Informed Consent in the Registry of the Canadian Stroke Network." *New England Journal of Medicine* 350(14): 1414-21.

Woolf, S., S. Rothemich, et al. (2000). "Selection Bias From Requiring Patients to Give Consent to Examine Data for Health Services Research." *Archives of Family Medicine* 9(10): 1111-8.

Yawn, B., R. Yawn, et al. (1998). "The Impact of Requiring Patient Authorization for Use of Data in Medical Research." *Journal of Family Practice* 47(5): 361-5.

Correspondence

Ea Mulligan
Research Associate, School of Law, Flinders University
Adelaide SA
Australia

ea.mulligan@flinders.edu.au

Wendy A Rogers
Associate Professor Medical Ethics and Health Law,
Department of Medical Education
Flinders University
GPO Box 2100
Adelaide, SA 5001
Australia

wendy.rogers@flinders.edu.au

A.Braunack-Mayer
Senior Lecturer
School of Public Health
Adelaide University
Adelaide SA
Australia

annette.braunackmayer@adelaide.edu.au