

Privacy and the use of health data for research

Christine M O'Keefe and Chris J Connolly

Analysis of the vast archives of clinical and health system data can yield information that is vital to effective health policy development and evaluation. It can also lead to enhanced clinical care through evidence-based practice and safety and quality monitoring. However, analysis must be conducted in such a way that standards of privacy and confidentiality are not compromised for individual health care consumers. In recognition of Australia's international leadership in scope and extent of health-related data collected at the population level, the Population Health Research Network (PHRN) (<http://www.phrn.org.au>) has been established to provide Australian researchers with access to linkable, de-identified data from a wide range of health data sets, across jurisdictions and sectors.

The use of sophisticated data analysis and data-mining tools can increase the risks of privacy breaches occurring.¹⁻⁷ This topic is becoming more important in the context of the PHRN investments designed to improve accessibility to Australian health-related data for the research sector.⁸

In this article, we review privacy regulation and privacy perception in Australia, in the context of health data used in research. We also provide a summary of resources for researchers interested in privacy issues related to secondary use of health data.

Privacy regulation in Australia

General privacy laws

The general privacy legislation currently in place in Australia is shown in Box 1.

- *Definition of personal information.* There are some minor inconsistencies between the definitions of personal information in the different Acts. The definitions are regularly tested in privacy complaints.⁹
- *Definitions of use and disclosure.* Within the privacy regulation framework, there are different provisions for *use* and *disclosure*, although it can be difficult to determine whether a given scenario involves use, disclosure or both.
- *Consent for disclosure.* There are considerable inconsistencies in federal and state privacy regulation surrounding consent for disclosure.¹⁰⁻¹² Strict application of the consent provisions in some cases has forced researchers to seek alternative methods of access to data without triggering consent provisions.
- *De-identification.* Some privacy laws include specific provisions for de-identification and de-identified data, limited to certain types of research.

Health-specific privacy laws

The main health-specific privacy laws are shown in Box 2. Health information is represented differently in different laws: it is included either in personal information or sensitive information, or is defined separately.

Other legislative privacy requirements

Enforceable guidelines provide an additional layer of privacy regulation for health research without consent. The key federal guidelines are those approved under section 95 (for Medical

ABSTRACT

Objective: We reviewed resources for researchers interested in privacy issues surrounding secondary use of health data for research. These included applicable privacy regulations and available information on privacy perception in Australia. The review is timely because the current Australian Population Health Research Network infrastructure investments are likely to attract new researchers to the field.

Data sources: We used Australian federal, state and territory regulations and programs, polls and surveys, public speeches and academic literature, and some international resources.

Data synthesis: We identify four themes (de-identification, consent, bias and participation) emerging as areas of concern from the review, and discuss issues relevant to these themes. We provide arguments that excessive privacy regulation has a negative effect on public health research.

Conclusions: There is little evidence of privacy complaints or breaches in health research, but significant concerns about consent and de-identification appear to persist in the community. New researchers need to take account of privacy regulation and may wish to take account of privacy perception when designing study and consent processes.

MJA 2010; 193: 537-541

research) and section 95A (for national privacy principles about health information of the *Privacy Act* 1988).

An analysis conducted by the National Health and Medical Research Council (NHMRC) of the use of these guidelines in practice¹³ found:

- Among consumers, there was a low level of awareness of privacy legislation and people had difficulty distinguishing between "confidentiality" and "privacy". Consumers were uncertain about providing consent for the use of their data.
- Health professionals tended to equate confidentiality with privacy and always maintained patient confidentiality.
- Researchers reported difficulty in accessing registries and inconsistencies in decisions made by human research ethics committees regarding access and disclosure.
- Data custodians believed that there was no need for researchers to have access to identified data and felt that they obtained the same benefit from de-identified information.
- Ethics committees believed that interpreting privacy legislation was complex, and they were most strongly opposed to researchers having access to health information without consent.
- The research community has had some difficulty in using the guidelines, and the initial test for compliance rests with ethics committees that appear to have applied the test inconsistently.

Australian Law Reform Commission Review of the Privacy Act

The final report of the Australian Law Reform Commission (ALRC) Review of the Privacy Act was delivered on 30 May 2008,¹⁴ and the government released the first stage of its response on 14 October

1 General privacy legislation currently in place in Australia

Jurisdiction	Legislation	Regulator
Federal	<i>Privacy Act 1988 (Cwlth)</i>	Federal Privacy Commissioner
Australian Capital Territory	<i>Privacy Act 1988 (Cwlth)</i>	Federal Privacy Commissioner
New South Wales	<i>Privacy and Personal Information Protection Act 1998</i>	NSW Privacy Commissioner
Northern Territory	<i>Information Act 2002</i>	NT Information Commissioner
Queensland	<i>Information Privacy Act 2009</i>	QLD Information Commissioner
South Australia	<i>Cabinet Administrative Instruction 1/89 2009</i>	Privacy Committee of South Australia
Tasmania	<i>Personal Information Protection Act 2004</i>	Ombudsman Tasmania
Victoria	<i>Information Privacy Act 2000</i>	Victorian Privacy Commissioner
Western Australia	No laws	Not applicable

Cwlth = Commonwealth.

2009.¹⁵ Exposure draft legislation containing an important element of the first stage response, the proposed Australian privacy principles, was released on 24 June 2010.¹⁶

In response to the ALRC recommendations 6-1 to 6-7, there are likely to be revisions to definitions in the Privacy Act, including that in recommendation 6-1, to define “personal information” as “information or an opinion, whether true or not, and whether recorded in a material form or not, about an identified or reasonably identifiable individual”. The Office of the Privacy Commissioner is likely to publish guidance on the meaning of terms such as *identified* and *reasonably identifiable*.

Regarding health research, recommendation 65-1 is likely to give rise to the replacement of the various existing guidelines on privacy and research by a formal set of research rules issued primarily by the NHMRC.

Other key accepted recommendations can be summarised as follows:

- *Recommendations 65-2 and 65-3.* “Research” should be extended to include human research more generally and the compilation or analysis of statistics.
- *Recommendation 65-6.* When a research proposal seeks to rely on the research exceptions in the Privacy Act, it must be reviewed and approved by a human research ethics committee.
- *Recommendation 66-1.* The research rules issued by the NHMRC should address the question of the collection, use or disclosure of

personal information without consent for inclusion in a database or register for research purposes, and that approval to establish such a database does not extend to future unspecified uses.

- *Recommendation 66-3.* The research rules issued by the NHMRC should address the circumstances and conditions under which it is appropriate to collect, use or disclose personal information without consent in order to identify potential participants in research.

Privacy perception in Australia — the evidence

In this section we review available evidence of community attitudes and public perceptions regarding privacy in the context of using health data for research, focusing on de-identification, consent and participation.

De-identification

In reporting on Australian Government Department of Health and Ageing (DoHA) qualitative research, Taylor¹⁷ noted that “consumers are not familiar with the term ‘de-identified data’ and even when it’s explained to them, it’s a concept that they are not all that comfortable with”.

In a poll conducted by the Australian Medical Association (AMA) in 2005,¹⁸ 60% of respondents reported that they were slightly or very concerned about the de-identification process.

2 Health privacy legislation currently in place in Australia

Jurisdiction	Health privacy legislation	Regulator
Federal	<i>Privacy Act 1988 (Cwlth)</i>	Federal Privacy Commissioner
Australian Capital Territory	<i>Health Records (Privacy and Access) Act 1997</i>	Community and Health Services Complaints Commissioner
New South Wales	<i>Health Records and Information Privacy Act 2002</i>	Public sector: internal review Private sector: Privacy NSW
Northern Territory	None currently in place	Not applicable
Queensland	<i>Information Privacy Act 2009</i>	Health Quality and Complaints Commission
South Australia	<i>Code of Fair Information Practice</i>	Not applicable
Tasmania	None currently in place	Not applicable
Victoria	<i>Health Records Act 2001</i>	Health Services Commissioner
Western Australia	None currently in place	Not applicable

Cwlth = Commonwealth.

Consent

Surveys by the Office of the Privacy Commissioner in 2001 and 2004 found that about 64% of respondents said consent should be sought for the use of de-identified data for research, while 33% said that use without consent was fine.^{19,20} In the 2007 survey, 51% said that consent should be sought, while 46% said that consent should not be sought.²¹

Similarly, the DoHA research¹⁷ found that consumers supported the use of data in research and registers, provided the data were de-identified and the purpose was legitimate and worthwhile. If identified data were to be used, consumers expected to be informed and their consent to be sought.

In the contrasting AMA poll,¹⁸ about 80% of respondents thought that their doctor should ask permission before allowing their de-identified data to be used for medical research, government purposes or commercial purposes. The comments provided suggest that some respondents may have overlooked the fact that the survey was only about de-identified data.

Comparative results can be seen in the United States National Consumer Health Privacy Survey 2005²² and an Australian perspective is provided by the Australian Consumer Association.²³

It is interesting to compare individuals' preferences for consent to the use of de-identified data for research with the legislation which permits the use of de-identified and sometimes even identifiable health data for research without consent, under certain conditions.

Participation

The use of an individual's health data for research can be viewed as participation by that individual in the research. An individual may have an objection to the purpose of the research on moral grounds even when there is no risk of identification or personal consequences.²⁴

The AMA poll¹⁸ found that 67% of respondents would give permission for their de-identified data to be used for research, 45% would give permission for government purposes and 32% would give permission for commercial purposes, showing that some participation concerns existed for a significant number of respondents.²⁵

De-identification, consent and bias

De-identification

De-identification is a complex issue surrounded by a lack of standard terminology and clarity. However, it is important because it underpins many health information privacy regulations.

First, it is often not clear what is meant when the term "de-identified" is used to refer to data. Sometimes it appears to mean simply that nominated identifiers such as name, address, date of birth and Medicare number have been removed from the data. At other times its use appears to imply that individuals represented in a dataset cannot be identified from the data, although it can also be unclear what this means. Simply removing nominated identifiers is often insufficient to ensure that individuals represented in a dataset cannot be identified. It can be straightforward to match some of the available data fields with the corresponding fields from external datasets, and thereby obtain enough information to determine individuals' names either uniquely or with a low degree

of uncertainty. This is particularly true of health information or of information which contains times and/or dates of events.

In Australia, the *National statement on ethical conduct in human research*²⁶ avoided the term "de-identified data" because its meaning is unclear. Instead, it proposed that data may be collected, stored or disclosed in three mutually exclusive forms: *individually identifiable*, *re-identifiable*, and *non-identifiable*. One problem with this approach is the datasets that do not fit into any of the defined categories.

In contrast, the *US Health Insurance Portability and Accountability Act 1996* (HIPAA) (<http://www.hipaa.org>) provides a useful legislative test for de-identification that provides certainty for the research community and for ethics committees.

Considering the issues surrounding the concept of de-identification in the Australian context would help to address some of the concerns highlighted by the NHMRC analysis,¹³ outlined under "Other legislative privacy requirements" above. The de-identification test contained in the HIPAA is a useful example of a legislative test that provides certainty for the research community. On the other hand, there may be a significant burden of compliance: if an organisation has many datasets then it would take a great deal of time for a person to perform the tasks outlined.

Consent and bias

Bias refers to the distortion of study results due to flaws in design or analysis. There is concern and some evidence that selection effects from consent processes lead to bias in research results.

Some investigations have been done on the possibility that consent processes may lead to bias in the makeup of study groups, and that this in turn may jeopardise the quality and applicability of the results. Woolf and colleagues²⁷ concluded that:

Patients who release personal information for health services research differ in important characteristics from those who do not ... older patients and those in poorer health were more likely to grant consent. Quality and health services research restricted to patients who give consent may misrepresent outcomes for the general population.

With regard to population health, Stanley^{28,29} has stated that:

The advantage of population record linkage [without consent], from an epidemiological perspective, is that it is not biased and no-one is excluded. This relates to human rights because generally the people who are excluded from studies are the most marginalised. The results are useful for the whole population.

Balancing privacy and research

There is no fundamental disagreement in the literature that the rights of the individual with respect to privacy need to be balanced against the public interest in the outcomes of health research. However, there is a range of views on where the appropriate balance lies.³⁰⁻³² A mutually satisfactory balance for consumers and the community is likely to be achieved by a combination of policy-centric³³ and technology-centric³⁴ measures.

There are at least four arguments that excessive privacy regulation has a negative effect on public health research. These include arguments about efficiency and quality of health research, the risk of avoidable harm to research subjects and the interests of the community versus those of the individual.

Efficiency of health research

The perception is that overheads resulting from privacy regulation hamper research efficiency in Australia.^{35,36} ALRC recommendations 65-3, 65-6 and 66-2 potentially exacerbate this situation.

Quality of health research

The fear is that selection effects from privacy-related processes including consent will lead to results bias. ALRC recommendations 66-1 and 66-3 (that organisations developing systems “to allow the linkage of personal information for research purposes should conduct a Privacy Impact Assessment”) were accepted in principle and potentially exacerbate this situation.

Risk of avoidable harm to research subjects

Avoidable harm may be caused to research subjects if they are exposed to sensitive medical information during overt data collection. For example, a request for consent to link treatment records with cancer registries could cause anxiety.³⁷

Interests of the community versus those of the individual

There is a perception that excessive privacy regulation denies the community the full potential benefits of health research based on more complete data. The moral dimension of this work has been addressed directly by Australian researchers as follows:

The examples provided demonstrate that *only* complete population data obtained by such linkage is inclusive of *all* those often underrepresented or excluded in many studies ...²⁸

This relates to human rights because generally the people who are excluded from studies are the most marginalised.²⁹

How does the ethics committee, or privacy officer in an organisation interpret [the Privacy Act's public interest exceptions to consent gathering]? You might expect that the ethical considerations would determine the outcome. However, it is more likely that the overriding consideration will be legal liability.³⁵

Discussion and conclusion

There is little evidence of privacy complaints or breaches in health research. However, privacy regulation and privacy perception are both key factors in the health research context, acting as potential restraints on some types of research that could deliver considerable public benefit. Further, significant concerns regarding consent and de-identification remain in the community. In particular, the proportion of individuals who believe that consent should be required even where information is de-identified is likely to remain at significant levels (perhaps somewhere between one-quarter and one-third of the population) for some time to come.

Will these community concerns impact on health research? Ultimately, decisions on research are made by ethics committees applying guidelines that allow some balance between privacy and research. The decision is therefore taken out of the hands of individual consumers, but these community concerns help to shape privacy regulation and will have an indirect influence on the decisions of ethics committees.

Under the changes proposed by the ALRC, a single set of formal research rules issued by the Privacy Commissioner will guide all decisions by ethics committees. This may lead to improved

consistency in outcomes that attempt to balance privacy rights with the public interest.

The ALRC recommendations also leave room for technical solutions to play an increased role in allowing personal information to be de-identified for research purposes. Recent advances in the techniques for de-identifying personal information³⁴ provide some hope that de-identification can occur without a negative impact on data quality.

Acknowledgements

Most of this work was conducted as a project for the Preventative Health National Research Flagship, and we thank the other project stakeholder representatives who gave valuable feedback. Additional research and guidance was provided by Galexia consultants Peter van Dijk and Francis Vierboom.

Author details

Christine M O'Keefe, PhD, MBA, BSc(Hons), Operations Director and Science Leader (Privacy and Confidentiality)¹

Chris J Connolly, LLB, Director²

1 Preventative National Health Research Flagship, CSIRO Mathematics, Informatics and Statistics, Canberra, ACT.

2 Galexia, Sydney, NSW.

Correspondence: Christine.OKeefe@csiro.au

References

- Gomatam S, Karr AF, Reiter JP, et al. Data dissemination and disclosure limitation in a world without microdata: a risk-utility framework for remote access servers. *Stat Sci* 2005; 20: 163-177.
- O'Keefe CM, Good NM. Regression output from a remote analysis server. *Data Knowl Eng* 2009; 68: 1175-1186.
- Reiter JP. Model diagnostics for remote-access regression servers. *Stat Comput* 2003; 13: 371-380.
- Reiter JP, Kohnen CN. Categorical data regression diagnostics for remote servers. *J Stat Comput Sim* 2005; 75: 889-903.
- Reznek AP. Recent confidentiality research related to access to enterprise microdata. Washington, DC: US Census Bureau, Center for Economic Studies, 2006. <http://www.oecd.org/dataoecd/6/30/37503027.pdf> (accessed Sep 2010).
- Sparks R, Carter C, Donnelly J, et al. A framework for performing statistical analyses of unit record health data without violating either privacy or confidentiality of individuals. Proceedings of the 55th session of the International Statistical Institute, Sydney, 2005 (CD-ROM).
- Sparks R, Carter C, Donnelly J, et al. Remote access methods for exploratory data analysis and statistical modelling: privacy-preserving analytics. *Comput Methods Programs Biomed* 2008; 91: 208-222.
- Australian Government Department of Innovation, Industry, Science and Research. Strategic roadmap for Australian research infrastructure. Canberra: Commonwealth of Australia, 2008. http://ncris.innovation.gov.au/Documents/2008_Roadmap.pdf (accessed Sep 2010).
- Office of the Privacy Commissioner. Privacy Commissioner concludes investigation into CAMM Pacific and Health Communications Network Limited [media release]. http://www.privacy.gov.au/news/media/05_02.html (accessed Sep 2010).
- New South Wales Administrative Decisions Tribunal Appeal Panel. Vice Chancellor, Macquarie University v FM [2003] NSWADTAP 43. http://www.lawlink.nsw.gov.au/lawlink/privacynsw/ll_pnsw.nsf/pages/PNSW_07_cnadtap43 (accessed Sep 2010).
- New South Wales Administrative Decisions Tribunal. KJ v Wentworth Area Health Service [2004] NSWADT 84. http://www.lawlink.nsw.gov.au/lawlink/privacynsw/ll_pnsw.nsf/pages/PNSW_07_cnadt84 (accessed Sep 2010).
- New South Wales Administrative Decisions Tribunal. MT v Director General, NSW Department of Education and Training [2004] NSWADT 194. <http://www.austlii.edu.au/au/cases/nsw/NSWADT/2004/194.html> (accessed Sep 2010).

- 13 Hill D. Consumer attitudes to privacy in an e-health environment [speech]. Proceedings of the National Health Information Summit: transferring Australia's health system; 2004 Dec 6-7; Melbourne. Canberra, Australian Health Information Council, 2005.
- 14 Australian Law Reform Commission. For your information: Australian privacy law and practice (ALRC Report 108). Sydney: Commonwealth of Australia, 2008. <http://www.alrc.gov.au/publications/report-108> (accessed Oct 2010).
- 15 Australian Government Department of the Prime Minister and Cabinet. Enhancing national privacy protection: Australian Government first stage response to the Australian Law Reform Commission report 108. Canberra: DPMC, 2009. http://www.pmc.gov.au/privacy/alrc_docs/stage1_aus_govt_response.pdf (accessed Sep 2010).
- 16 Parliament of Australia, Senate. Exposure drafts of Australian privacy amendment legislation. Canberra: Senate Finance and Public Administration Committee, 2010. http://www.aph.gov.au/Senate/committee/fapa_Ctte/priv_exp_drafts/index.htm (accessed Sep 2010).
- 17 Taylor J. Consumer attitudes to privacy in an e-health environment [speech]. Proceedings of the National Health Information Summit: transferring Australia's health system; 2004 Dec 6-7; Melbourne. Canberra, Australian Health Information Council, 2005.
- 18 Australian Medical Association. AMA poll shows patients are concerned about the privacy and security of their medical records. Canberra: AMA, 2005. <http://www.ama.com.au/web.nsf/doc/WEEN-6EG7LY> (accessed Sep 2010).
- 19 Office of the Federal Privacy Commissioner. Privacy and the community. Sydney: OFPC, 2001. <http://www.privacy.gov.au/materials/types/research/view/6614> (accessed Oct 2010).
- 20 Office of the Federal Privacy Commissioner. Community attitudes towards privacy 2004. Sydney: OFPC, 2004. <http://www.privacy.gov.au/materials/types/research/view/6615> (accessed Oct 2010).
- 21 Office of the Privacy Commissioner, Australia. Community attitudes to privacy 2007. Sydney: OFPC, 2007. <http://www.privacy.gov.au/materials/types/download/8820/6616> (accessed Oct 2010).
- 22 Forrester Research. National Consumer Health Privacy Survey 2005. Oakland, CA: California HealthCare Foundation, 2005. <http://www.chcf.org/publications/2005/11/national-consumer-health-privacy-survey-2005> (accessed Sep 2010).
- 23 Ballenden N. Transforming Australia's health system [speech]. Proceedings of the National Health Information Summit: transferring Australia's health system; 2004 Dec 6-7; Melbourne. Canberra, Australian Health Information Council, 2005.
- 24 Privacy in Research Ethics and Law. Recommendations from PRIVIREAL to the European Commission. Sheffield, UK: PRIVIREAL, 2003. <http://www.privireal.org/content/recommendations/#Recc> (accessed Sep 2010).
- 25 Newell C. Health information management and information communications technology for public health and population research: a consumer perspective [speech]. Proceedings of the National Health Information Summit: transferring Australia's health system; 2004 Dec 6-7; Melbourne. Canberra, Australian Health Information Council, 2005.
- 26 National Health and Medical Research Council. National statement on ethical conduct in human research (2007). Canberra: NHMRC, 2007. <http://www.nhmrc.gov.au/publications/synopses%20/e72syn.htm> (accessed Sep 2010).
- 27 Woolf SH, Rothemich SF, Johnson RE, Marsland DW. Selection bias from requiring patients to give consent to examine data for health services research. *Arch Fam Med* 2000; 9: 1111-1118.
- 28 Stanley F. Record linkage — public good or invasion of privacy? Proceedings of the 25th International Conference of Data Protection and Privacy Commissioners; 2003 Sep 10-12; Sydney, Australia. <http://www.privacyconference2003.org/program.asp#fiona> (accessed Sep 2010).
- 29 Stanley F. Australian Health Information Council conference 2004 [speech]. Proceedings of the National Health Information Summit: transferring Australia's health system; 2004 Dec 6-7; Melbourne. Canberra, Australian Health Information Council, 2005.
- 30 Kelman CW, Pearson S-A, Day R, et al. Evaluating medicines: let's use *all* the evidence. *Med J Aust* 2007; 186: 249-252.
- 31 Peto J, Fletcher O, Gilham C. Data protection, informed consent, and research. *BMJ* 2004; 328: 1029-1030.
- 32 Wanless D. Securing good health for the whole population. London: HM Treasury, 2004. http://www.hm-treasury.gov.uk/consultations_and_legislation/wanless/consult_wanless04_final.cfm (accessed Sep 2010).
- 33 Room S. Data protection, informed consent, and research: Data Protection Act does not bar medical research [letter]. *BMJ* 2004; 328: 1437.
- 34 O'Keefe CM. Privacy and the use of health data — reducing disclosure risk. *Electronic J Hlth Informatics* 2008; 3: e5.
- 35 Gun R. Privacy law is kneecapping epidemiological research. *Australas Epidemiologist* 2005; 12.1: 2-4.
- 36 Skene L. Proliferating ethics committees and privacy legislation: new fetters on scientific research. *Australas Epidemiologist* 2005; 12.1: 16-18.
- 37 Hill D. Testimony for the National Health and Medical Research Council, Senate Legal and Constitutional References Committee, inquiry into the Privacy Act 1988, 20 May 2005. Canberra: Commonwealth of Australia, 2005. <http://www.aph.gov.au/hansard/senate/commtee/S8383.pdf> (accessed Sep 2010).

(Received 8 Oct 2009, accepted 3 May 2010)

□